GESELLSCHAFT
FÜR INFORMATIK



Sandra Schulz, Natalie Kiesler (Hrsg.)

DELFI 2024
Die 22. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.

09.09.-11.09.2024
Fulda

Gesellschaft für Informatik e.V. (GI)

Lecture Notes in Informatics (LNI) - Proceedings
Series of the Gesellschaft für Informatik (GI)

Volume P-356

ISBN 978-3-88579-255-0
ISSN 1617-5468

Volume Editors
Prof. Dr. Sandra Schulz
Universität Hamburg
Mittelweg 177
20148 Hamburg
sandra.schulz@uni-hamburg.de

Prof. Dr. Natalie Kiesler
Technische Hochschule Nürnberg Georg Simon Ohm
Keßlerplatz 12
90489 Nürnberg
natalie.kiesler@th-nuernberg.de

Series Editorial Board
Andreas Oberweis, KIT Karlsruhe,
(Chairman, andreas.oberweis@kit.edu)
Torsten Brinda, Universität Duisburg-Essen, Germany
Dieter Fellner, Technische Universität Darmstadt, Germany
Ulrich Flegel, Infineon, Germany
Ulrich Frank, Universität Duisburg-Essen, Germany
Michael Goedicke, Universität Duisburg-Essen, Germany
Ralf Hofestädt, Universität Bielefeld, Germany
Wolfgang Karl, KIT Karlsruhe, Germany
Michael Koch, Universität der Bundeswehr München, Germany
Peter Sanders, Karlsruher Institut für Technologie (KIT), Germany
Andreas Thor, HFT Leipzig, Germany
Ingo Timm, Universität Trier, Germany
Karin Vosseberg, Hochschule Bremerhaven, Germany
Maria Wimmer, Universität Koblenz-Landau, Germany

Dissertations
Steffen Hölldobler, Technische Universität Dresden, Germany
Thematics
Agnes Koschmider, Universität Kiel, Germany
Seminars
Judith Michael, RWTH Aachen, Germany

© Gesellschaft für Informatik, Bonn 2024
printed by Köllen Druck+Verlag GmbH, Bonn



This book is licensed under a Creative Commons BY-SA 4.0 licence.

Vorwort

Die 22. Fachtagung Bildungstechnologien (DELFI 2024) an der Hochschule Fulda steht unter dem Motto “Offene Bildung: Durch Technologie, Transparenz und Nachvollziehbarkeit die Zukunft gestalten”. Die diesjährige Fachtagung widmet sich damit vor allem den Bildungstechnologien im Kontext innovativer, offener technologischer Lösungen, und möchte dabei besonders auf das Spannungsfeld zwischen offenen Forschungsdaten, offenen Technologien und offener Bildung eingehen. In einer immer stärker vernetzten, und scheinbar intelligenten Bildungswelt, ist die Offenheit von Bildung und Bildungstechnologien zentral für deren Unabhängigkeit. Daher sind auch Qualitätsmerkmale wie Nachvollziehbarkeit, Transparenz und Nachhaltigkeit immer mehr von Bedeutung in der Konzeption, Vernetzung und Nutzung von Technologien in Bildungskontexten.

KI und insbesondere Generative KI (genKI) ist auch im Jahr 2024 in aller Munde und begegnet uns nahezu täglich in Lehre und Forschung. Viel mehr noch, es wird immer schwieriger zu identifizieren, in welchen Technologien KI verwendet wird. Bereits in der Schule sind Lehrkräfte damit konfrontiert, Aufgaben so zu gestalten, dass die eigene Leistung der Schüler:innen auf bestimmter Schöpfungshöhe erkennbar ist. In der beruflichen Bildung, Weiterbildung, und Hochschule erleben Lehrende die gleiche Herausforderung. Unterricht und Lehre müssen zunehmend so entwickelt werden, dass sie genKI konstruktiv einsetzen und den kritisch-reflektierten Umgang damit fördern. Weiterhin stehen wir vor der Herausforderung, Lehre und Prüfungen zu gestalten, um die notwendigen Kompetenzen der Lernenden zu fördern. GenKI veranlasst uns insbesondere an den Hochschulen dazu, neu darüber nachzudenken, was wir vermitteln wollen und sollten, und welche Kompetenzen tatsächlich zukuntfsorientiert sind. Gleichzeitig bietet genKI vielseitige Potentiale für individualisiertes Feedback und die Unterstützung von Lernenden in selbstgesteuerten Lernszenarien. Auch Lehrende können von der Generierung von Inhalten, Übungsaufgaben, und Feedback profitieren. Jedoch ist auch im Kontext von genKI ein hohes Maß an Transparenz bzgl. Forschungs- und Trainingsdaten sowie deren Methodik und Limitationen unabdingbar. Auch aus diesem Grund widmen wir das diesjährige Tagungsmotto der DELFI den offenen Bildungstechnologien.

In Vorbereitung auf die Tagung haben wir uns als PC-Chairs gemeinsam mit dem Tagungsteam bewusst für einige organisatorische Veränderungen entschieden. Dazu gehört u.a. die Anpassung der Seitenlimits für die Beitragskategorien Demo und Poster auf 4 Seiten. Zudem haben wir für die Bewertung der Seitenlimits die Literatur ausgeschlossen. Ein weiteres Novum ist der Verzicht auf einen gedruckten Tagungsband, um finanzielle und materielle Ressourcen zu schonen. Die digitale Verfügbarkeit erscheint uns als zeitgemäße Lösung. Zudem haben wir auf die Degradierung von Beiträgen verzichtet, und Beitragende dazu ermutigt, sich im Vorfeld bewusst für ihre Beitragskategorie zu entscheiden. Für alle Beitragsformate wurde nicht zuletzt eine transparente, nachvollziehbare und offene Zugänglichkeit der dazugehörigen Artefakte gewünscht. Beiträge mit offen verfügbaren Forschungsdaten (inkl. Software) werden im vorliegenden Tagungsband daher mit einem Badge ausgezeichnet.



Die 125 eingereichten Beiträge wurden in einem doppelt blinden Verfahren von Mitgliedern des Programmkomitees begutachtet. Nach einem Paper-Bidding-Verfahren und der Abgabe von 2 bzw. 3 Gutachten (je nach Kategorie) schloss sich eine Diskussionsphase zwischen den zugewiesenen Komiteemitgliedern an, sodass zahlreiche divergierende Einschätzungen zusammengeführt werden konnten. Nach der finalen Diskussion der verbleibenden Einreichungen in der Sitzung des Programmkomitees Ende Mai konnten 55 Beiträge angenommen werden (Annahmerate 44%). Unser Dank gilt vor allem den Gutachter:innen, die Ihre Einschätzungen in überwältigendem Maße zeitgemäß eingereicht haben. Auch allen Autor:innen gilt an dieser Stelle unser herzlicher Dank. Ohne Beiträge aus Forschung und Praxis kann es keine DELFI geben.

Die angenommenen Beiträge sind im vorliegenden Tagungsband zu finden. Es handelt sich konkret um 11 Langbeiträge, 17 Kurzbeiträge, 11 Praxisbeiträge, 1 Positionspapier, 11 Demos, und 4 Posterbeiträge. Hinzu kommen die Kurzdarstellungen der eingeladenen Keynote-Vortragenden Dr. Hieke Keuning (Utrecht University) und Dr. Daniel Schiffner (DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation).

Während der Begutachtung wurden außerdem Beiträge aus den Kategorien Lang- und Kurzbeiträge als Best Paper nominiert. Die Beiträge mit den besten Bewertungen sind im vorliegenden Tagungsband entsprechend gekennzeichnet. Wir beglückwünschen die Autor:innen zur Nominierung und wünschen viel Erfolg bei der Präsentation.

Dank der großen Zahl von 125 Einreichungen, und der Auswahl der besten Beiträge haben wir ein umfassendes, vielseitiges Programm konzipiert, welches nicht nur das Tagungsmotto abbildet, sondern auch aktuelle Trends wie etwa generative KI, und die Nutzung und Evaluation von Bildungstechnologien in verschiedenen Kontexten, beispielsweise im Lehramt, Schule, und in der Hochschulbildung.

Wir freuen uns darauf, Sie im September auf der 22. DELFI in Fulda begrüßen zu dürfen, und wünschen allen Teilnehmenden eine interessante, konstruktive, und schöne Tagung! Wir hoffen, Sie erhalten zahlreiche Denkanstöße und Inspiration für die (Weiter-)Entwicklung offener Bildungstechnologien und entsprechender Forschungs- und Lehrprojekte.

Im August 2024,

Sandra Schulz und Natalie Kiesler (die Herausgeberinnen)

PS: Nicht zuletzt gilt unser ganz herzlicher Dank dem gesamten Organisationsteam:

Jan-Torsten Milde (Demo- und Poster Chair),
Hans-Martin Pohl (Lokaler Organisator) und
Johannes Konert (General Chair und Lokaler Organisator).

Tagungsleitung

Gesamtleitung:	Johannes Konert, Hochschule Fulda
Leitung des Programmkomitees:	Sandra Schulz, Universität Hamburg
	Natalie Kiesler, Technische Hochschule Nürnberg
Workshop-Leitung:	Natalie Kiesler, Technische Hochschule Nürnberg
	Sandra Schulz, Universität Hamburg
Poster und Demos:	Jan-Torsten Milde, Hochschule Fulda
Lokale Leitung:	Johannes Konert, Hochschule Fulda
	Hans-Martin Pohl, Hochschule Fulda

Programmkomitee

Marc Berges	Friedrich-Alexander-Universität
Nadine Bergner	RWTH Aachen
Steven Beyer	Universität Duisburg-Essen
Rainer Blum	Hochschule Fulda
Andreas Breiter	Universität Bremen
Josef Buchner	St. Gallen University of Teacher Education
Julian Dehne	Universität Göttingen
Ira Diethelm	Carl von Ossietzky Universität Oldenburg
Ralf Doerner	RheinMain University of Applied Sciences
Matthias Ehlenz	RWTH Aachen
Albrecht Fortenbacher	Hochschule für Technik und Wirtschaft Berlin
Paul Grimm	Hochschule Darmstadt
Josef Guggemos	University of Education Schwäbisch Gmünd
Stefan Göbel	TU Darmstadt
Joerg Haake	FernUniversität in Hagen
Andreas Harrer	University of Applied Sciences and Arts Dortmund
Anja Hawlitschek	Otto-von-Guericke-Universität Magdeburg
Peter Henning	Karlsruhe University of Applied Sciences
Claudia Hildebrandt	Heidelberg University of Education
Sebastian Hobert	TH Lübeck
Ioana Jivet	FernUniversität in Hagen
Marco Kalz	Pädagogische Hochschule Heidelberg
Paul-Thomas Kandzia	Duale Hochschule Baden-Württemberg Lörrach
Michael Kerres	University Duisburg-Essen
Natalie Kiesler	Technische Hochschule Nürnberg
Ralph Koelle	University of Hildesheim
Johannes Konert	Hochschule Fulda
Bastian Küppers	RWTH Aachen
Roland Küstermann	Duale Hochschule Baden-Württemberg Karlsruhe
Carsten Lecon	Hochschule Aalen
Thiemo Leonhardt	RWTH Aachen
Andreas Lingnau	Deutsche Hochschule für angewandte Wissenschaften
Joern Loviscach	Fachhochschule Bielefeld

Ulrike Lucke	Universität Potsdam
Ulrike Mascher	Goethe-Universität Frankfurt
Jan-Torsten Milde	Hochschule Fulda
Tobias Moebert	Universität Potsdam
Andreas Mühling	Christian-Albrechts-Universität zu Kiel
Miriam Mulders	Universität Duisburg Essen
Svenja Noichl	RWTH Aachen
Niels Pinkwart	Humboldt-Universität zu Berlin
Hans-Martin Pohl	Hochschule Fulda
René Röpke	TU Wien
Guido Rößling	Technische Universität Darmstadt
Michael Rücker	Friedrich-Schiller-Universität Jena
Daniel Schiffner	DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Jan Schneider	DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Sandra Schulz	Universität Hamburg
Clara Schumacher	Humboldt Universität zu Berlin
Andreas Schwill	Universität Potsdam
Niels Seidel	FernUniversität in Hagen
Sabine Seufert	Universität St. Gallen
Katharina Simbeck	Hochschule für Technik und Wirtschaft Berlin
Volodymyr Sokol	RWTH Aachen
Patrick Stalljohann	Fachhochschule Münster
Sven Strickroth	Ludwig-Maximilians-Universität München
Michael Striewe	Hochschule Trier
Heinrich Söbke	Bauhaus-Universität Weimar
Veronika Thurner	University of Applied Sciences Munich
Alexander Tillmann	Goethe-Universität Frankfurt
Sarah Voß-Nakkour	Goethe-Universität Frankfurt
Martin Wessner	Hochschule Darmstadt
Arno Wilhelm-Weidner	VDI/VDE-IT
Raphael Zender	Humboldt-Universität zu Berlin

Inhaltsverzeichnis

Eingeladene Vorträge

Hieke Keuning
The interplay between rich and big data in programming education research 19

Daniel Schiffner
(Re)-Using the Data: Address Challenges in Research Data via FAIR and APIs 23

Wissenschaftliches Programm

Best Paper Kandidaten Langbeiträge

Frank Wehrmann, Raphael Zender
Towards Nonverbal Communication in Multi-User Virtual Reality Learning Environments: A Systematic Review 31

David Mack, Ludger Schmidt
Unterstützung einer komplexen Montageaufgabe durch eine Augmented-Reality-Anleitung mit Durchsichtdatenbrille 49

Andreas Scholl, Daniel Schiffner, Natalie Kiesler
Analyzing Chat Protocols of Novice Programmers Solving Introductory Programming Tasks with ChatGPT 63

Sven Judel, Christopher Zeise, Ulrik Schroeder
Informiertes Zustimmungsmanagement für Learning Analytics in Lernmanagementsystemen 81

Best Paper Kandidaten Kurzbeiträge

Jesper Dannath, Alina Deriyeva, Benjamin Paaßen
Evaluating Task-Level Struggle Detection Methods in Intelligent Tutoring Systems for Programming 97

Maurice Herwig, Norbert Hundeshagen, John Hundhausen, Stefan Kablowski, Martin Lange
Problem-Specific Visual Feedback in Discrete Modelling 107

David Baberowski, Thiemo Leonhardt, Lanea Lilienthal, Nadine Bergner
Beyond Realism: Rethinking Presence in Virtual Environments for Abstract Concept Learning 115

KI Einsatz in der Hochschule I

Peter Kießling, Florian Funke, Sven Hofmann
Entwicklung und Evaluation einer KI-Assistenz zur didaktisch-pädagogischen Unterstützung des Lernprozesses mit Programmieraufgaben 125

Thomas Ranzenberger, Carolin Freier, Luca Reinold, Korbinian Riedhammer, Fabian Schneider, Christopher Simic, Claudia Simon, Steffen Freisinger, Munir Georges, Tobias Bocklet
Multidisciplinary Approach to AI-based self-motivated Learning and Teaching with Large Language Models 133

Andre Cordes
Evaluation von LLM- und Intent-basierten Ansätzen zur Umsetzung eines Chatbots für die Unterstützung bei der Studienorganisation 141

Rebecca Finster, Linda Grogorick, Susanne Robra-Bissantz
ChatGPT erzähl mir eine Geschichte: Die Verwandlung von Lernwelten durch KI-gestützte Erzählungen 153

Bildungstechnologien in Lehramt und Schule

Moritz Kreinsen, Maimon Thiems, Sandra Sprenger, Sandra Schulz
A project-based AI and Data Education Concept for STEM Teacher Education in Teaching-Learning-Labs 163

Hassan Soliman, Milos Kravcik, Alexander Tobias Neumann, Yue Yin, Norbert Pengel, Maike Haag, Heinz-Werner Wollersheim
Generative KI zur Lernenbegleitung in den Bildungswissenschaften: Implementierung eines LLM-basierten Chatbots im Lehramtsstudium . . . 171

Nadine Nicole Koch, Ann-Kathrin Kapfenstein, Niklas Meißner, Maria Wirzberger
Enhanced Program Comprehension: Individualized Learning of Code Tracing with the Feedback Buddy 179

Alexander Aumann, Wolfgang Müller, Jörg Stratmann, Holger Weitzel
Entwicklung Medien(fach)didaktischer Kompetenzen bei Lehramtstudierenden auf Basis curricular verankerter Medienprojekte . . 201

VR/AR und immersives Lernen I

Axel Wiepke, Alexander Hacke, Ulrike Lucke, Christian Manz
Phasen der Problemlösekompetenz bei der Korrektur elektrischer Schaltkreise in einer VR-Lernumgebung 211

Florian Nowotny, Chantal Berkhan, Ulrike Lucke
Integration von Virtual Reality in der technischen Berufsbildung: Ein praxisorientierter Ansatz zur Konzeption immersiver Lernszenarien 225

Rene Wilzeck, Raphael Zender
Virtual Reality zur Professionalisierung von Sportlehrkräften 233

Guido Klees, Kirsten Santelmann, Markus Weber
*Immersive 360° Laborreise für Schüler*innen* 241

Open Science und OER

Natalie Kiesler, René Röpke, Daniel Schiffner, Sandra Schulz, Sven Strickroth, Matthias Ehlenz, Birte Heinemann, Arno Wilhelm-Weidner
Towards Open Science at the DELFI Conference 251

Frederic Salmen, Martin Breuer, Sergej Görzen, Malte Persike, Ulrik Schroeder
FAIR Learning Technologies with Web Components and Packages 267

Sebastian Speiser, Melanie Baur, Johanna Sedlmair
Implementation of flexible and open course material for computer programming courses 275

Jan Bernoth, Christian Riedel, Axel Wiepke, Firas Al Laban
From Receiving to Characterizing: Improving Training Strategies for Research Data/Software Management by another domain 281

VR/AR und immersives Lernen II

Andreas Fuchs, Sven Appel, Paul Grimm
Immersive Räume zur Kreativitätsunterstützung: Ein intelligenter Lehr- und Lernraum 291

Ulrike Mascher, David Fernes
360°-Perspektiven: von der physischen zur virtuellen Ausstellung im Hochschulkontext 305

Dennis Wohlers, Roland Schröder-Kroll, Svetlana Golovanev, Michael Geis, Maximilian Mushoff, Valerie Varney, Joachim Metternich, Rüdiger Weißbach, Axel Wagenitz
Konzeption eines adaptiven Systems für Augmented Reality Lehr-Lernsysteme mit individuellen Lernszenarien 313

Marc Schlicker, Raphael Zender
Schleif-R: Handwerkliches Schleifen in VR 321

KI Einsatz in der Hochschule II

Paul Christ, Jörg Haake, Torsten Munkelt
Automatische Generierung von Aufgaben für die konzeptuelle Modellierung 331

Alexander Gantikow, Sara Durski, Andreas Isking, Paul Libbrecht, Wolfgang Müller, Simon Ostermann, Sandra Rebholz
KI-basierte Analyse von E-Portfolios 347

Teresa Sedlmeier, Claudia Schmidt, Volker Sänger, Katrin Bauer, Michael Canz, Gisela Hillenbrand, Prabin Dahal, Saptadi Nugroho
Learning Experience durch Content Curation und KI-basierte Lernempfehlungen 361

Theresa Kruse, Dominic Lohr, Marc Berges, Michael Kohlhase, Halimeh Moghbeli, Marcel Schütz
Term Extraction for Domain Modeling 369

Lernunterstützung durch Bildungstechnologien

Imen Azaiz, Sven Strickroth
Supporting Constructive Alignment: Visualization of the Alignment of Teaching Goals and Developed Assignments 381

Lea Reinartz, Sarah Schneeweiß, Andreas Harrer
Design, Realisierung und Evaluation von Scaffolds zur Unterstützung von Lernprozessen im Hochschulkontext 389

Armin Egetenmeier, Sven Strickroth
Challenges of Computer Science Tutors in Their Teaching: How can Educational Technology help? 397

Miriam Naß, Tabea Dobbrunz, Sina Warmer, Andreas Harrer, Norbert Pohlmann
Entwicklung, Evaluation und Distribution einer Selbstlernakademie zum Thema IT-Sicherheit 407

Michael Zohner, Philipp Brettschneider
Design und Implementierung einer interaktiven Lernumgebung im Bereich der IT-Sicherheit 415

Programmieren Lehren und Lernen in der Hochschule

Melanie Baur, Johanna Sedlmair, Sebastian Speiser
A novel project assignment fostering self-competence and reflection using robotics for first-year CS students 425

Annabell Brocker, Ulrik Schroeder
Correlation of Error Metrics in Python CS1 Courses 441

Marc Stelzner, Sarah Scheer, Max Kaussow, Stefan Fischer
Hackathon als Pflichtmodul im ersten Semester - Eine Fallstudie 449

Anna-Lena Roth, David James, Michael Kuhn, Johannes Konert
Enhancing Parallel Programming Education with High-Performance Clusters Utilizing Performance Analysis 457

Poster-Beiträge

Kensuke Akao
KI-basierte Unterstützung zum Abbau sprachlicher Barrieren für Kinder mit nichtdeutscher Herkunftssprache 467

Michael Striewe
Forschungsdaten in der Bildungstechnologie: Worüber sprechen wir und was ist wann relevant? 473

Maurice Herwig, Norbert Hundeshagen, Marit Kastaun, Cedric Kollenberg
Towards Computer-Aided Teaching of Reductions in TCS 479

Alina Deriyeva, Jesper Dannath, Benjamin Paaßen
Relation between struggle and learning personality in programming exercises 485

Demo-Beiträge

Felix Meinhardt, Sergej Görzen, Richard Werkes, Philipp Pelzer, Daniel Krause, Michael Schroeder, Ulrik Schroeder
Neugestaltung eines Mixed Reality Prototypen für das Schülerlabor Chemie 493

Christian Koch, Frederic Salmen
Web-based prototype of a visual and interactive deep learning simulation 499

Michael Striewe
Paare finden: Ein konfigurierbares Lernspiel mit Analyse-Dashboard . . . 505

Benjamin Ledel, Tabea Schwarz
Sicherer Einsatz von xAI in der Bildung: Erkennung von LLM-Halluzinationen bei der Generierung von Lehr- und Lernmaterialien 511

Zoe Jebing, Armin Egetenmeier, Sven Strickroth
easyProtocol: Ein digitales Unterstützungstool für mündliche Prüfungen . 517

Marcella Haller, Alexander Ast
Augmented Reality an der Schnittstelle von Hochschuldidaktik und Technologie: Integration von realen Objekten für medialgestütztes Lernen 523

Sergej Görzen, René Röpke, Ulrik Schroeder
BuddyAnalytics: A dashboard and reporting tool for study program analysis and student cohort monitoring 527

Matthias Ehlenz, Frederic Maquet
19squared - Interaktive XR-Inhalte gemeinsam selbst erstellen 533

Tabea Dobbrunz, Miriam Naß, Sina Warmer, Norbert Pohlmann, Andreas Harrer
Produktionstechnisches Vorgehen und Demonstration der Erstellung einer Selbstlernakademie für IT-Sicherheit 539

Sam Sabah, Alexander Tillmann, David Weiß, Hendrik Drachsler
Enhancing Geographical Learning Through GeoVR: Immersive Exploration and Topographic Analysis Educational Application 545

Dominic Lohr, Marc Berges, Abhishek Chugh, Michael Striewe
Adaptive Learning Systems in Programming Education: A Prototype for Enhanced Formative Feedback 549

Autorenverzeichnis

Eingeladene Vorträge

The interplay between rich and big data in programming education research

Hieke Keuning [1]

Abstract: To conduct solid research on how students learn programming, we need both 'rich data' and 'big data'. In the past decades, researchers have been collecting both types of data, such as large datasets of programs written by students, containing numerous mistakes, but also more fine-grained data, such as verbalizations of what students were thinking when solving a challenging programming problem. While there is an interplay between these two types of data, they are typically used to answer different questions. There are also several existing datasets available for conducting programming education research, however, these are more often 'big' rather than 'rich', and it is not trivial to find and use them. In this talk I will show several examples of my research, in which I have (re)used datasets to study aspects of how students learn to program, discussing the need for collecting, analyzing and sharing big as well as rich data.

Keywords: Computing education research, programming datasets

The need for rich and big data

Since computing education research took off in the 1960s, researchers have been collecting data of students learning to program [FR19]. Researchers have gathered large datasets of programs written by students [Ih15], containing a vast array of mistakes, but also more fine-grained data such as verbalizations of what students were thinking when solving a challenging programming problem. The analysis of this data has informed numerous new theories, interventions, and tools [Lu18].

If we want to conduct solid research on how our students learn programming, we need both 'rich data' and 'big data'. Brown and Guzdial [BG24] argue the different roles of these types of data, and the different questions they answer. Big data involves a large collection of data that is typically automatically collected, with little demographical details. Researchers can use this data to answer questions such as 'how often do students make mistake X.' Rich data is associated with a known context, and often collected and analyzed by a human. Questions such as 'why does a student ask for a hint' can be answered by studying rich data. The point Brown and Guzdial make in their paper originated from being annoyed by reviewers making unsuitable comments. In studies with big data, reviewers asked for contextual information. In studies with rich data, they complained about the small sample size. However, there is an *interplay* between these two types of data; they inform each other, for example by deriving hypotheses from rich data, to be confirmed by large datasets. According to Brown and Guzdial, there are very few studies that actually complement each other in such a way.

1 Utrecht University, Department of Information and Computing Sciences, Utrecht, The Netherlands, h.w.keuning@uu.nl, https://orcid.org/0000-0001-5778-7519

doi:10.18420/delfi2024_01

As an example from my own research, we have conducted several studies to investigate the style and quality of the code students write. We examined quality issues in 2.6 million Java code snapshots written by novice programmers from the Blackbox dataset [KHJ17]. We also conducted a study using an existing dataset of logged student actions and code snapshots in a tutoring system for code refactoring [OKJ23]. In this study we aimed to identify misconceptions that students have, by analyzing transitions from a correct to an incorrect code snapshot. This dataset can be considered as fairly 'big'. But how do we really know these 'misconceptions' are rooted in incorrect mental models students have, or if they are merely minor slips? To find out, we conducted a think-aloud study, collecting verbalizations of 11 students while they were refactoring code [OKJ24]. We recognized some of the misconceptions, which we could confirm by analyzing the accompanying thoughts of students. To further establish and expand the misconceptions, we would need to return to the 'big' datasets.

Availability of programming education data

Another aspect of data in programming education research is its availability and accessibility. Sharing data has not always been the priority of researchers; it is common to collect new data specifically for a study. However, there are also several existing datasets available for conducting research on learning programming, but it is not always trivial to find and use these datasets [Ki23], and they are more often 'big' rather than 'rich'. Although researchers have started to share their data more and more, there are several issues to address when re-using this existing data.

In the last two years many studies have appeared on GenAI's influence on computing education. In these studies, the AI model produces the data (e.g. programming solutions or feedback), or the data is from students interacting with an AI model. What does this mean for these datasets, since AI models are being updated all the time? These volatile datasets pose several challenges for re-use and replication [Pr23]. Not the data itself, but high-quality materials to generate this data might become more important.

The interplay

In this talk I will show several examples of my research, in which I have (re)used datasets to study aspects of how students learn to program. I will highlight the challenges that came with using these, and discuss the need for collecting, sharing, and analyzing big as well as rich data.

References

[BG24] Brown, N. C. C.; Guzdial, M.: Confidence vs Insight: Big and Rich Data in Computing Education Research. In: Proceedings of the 55th ACM Technical Symposium on Computer Science Education. S. 158–164, 2024, DOI: 10.1145/3626252.3630813.

[FR19] Fincher, S. A.; Robins, A. V.: The Cambridge handbook of computing education research. Cambridge University Press, 2019.

[Ih15] Ihantola, P.; Vihavainen, A.; Ahadi, A.; Butler, M.; Börstler, J.; Edwards, S. H.; Isohanni, E.; Korhonen, A.; Petersen, A.; Rivers, K.; Rubio, M. Á.; Sheard, J.; Skupas, B.; Spacco, J.; Szabo, C.; Toll, D.: Educational Data Mining and Learning Analytics in Programming: Literature Review and Case Studies. In: Proceedings of the 2015 ITiCSE on Working Group Reports. S. 41–63, 2015, DOI: 10.1145/2858796.2858798.

[KHJ17] Keuning, H.; Heeren, B.; Jeuring, J.: Code quality issues in student programs. In: Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education. S. 110–115, 2017, DOI: 10.1145/3059009.3059061.

[Ki23] Kiesler, N.; Impagliazzo, J.; Biernacka, K.; Kapoor, A.; Kazmi, Z.; Ramagoni, S. G.; Sane, A.; Tran, K.; Taneja, S.; Wu, Z.: Where's the Data? Exploring Datasets in Computing Education. In: Proceedings of the ACM Conference on Global Computing Education. S. 209–210, 2023, DOI: 10.1145/3617650.3624951.

[Lu18] Luxton-Reilly, A.; Simon; Albluwi, I.; Becker, B. A.; Giannakos, M.; Kumar, A. N.; Ott, L.; Paterson, J.; Scott, M. J.; Sheard, J. et al.: Introductory programming: a systematic literature review. In: Proceedings companion of the 23rd annual ACM conference on innovation and technology in computer science education. S. 55–106, 2018, DOI: 10.1145/3293881.3295779.

[OKJ23] Oliveira, E.; Keuning, H.; Jeuring, J.: Student Code Refactoring Misconceptions. In: Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education. S. 19–25, 2023, DOI: 10.1145/3587102.3588840.

[OKJ24] Oliveira, E.; Keuning, H.; Jeuring, J.: Investigating Student Reasoning in Method-Level Code Refactoring: A Think-Aloud Study. In: In submission. 2024.

[Pr23] Prather, J.; Denny, P.; Leinonen, J.; Becker, B. A.; Albluwi, I.; Craig, M.; Keuning, H.; Kiesler, N.; Kohn, T.; Luxton-Reilly, A. et al.: The robots are here: Navigating the generative ai revolution in computing education. In: Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education. S. 108–159, 2023, DOI: 10.1145/3623762.3633499.

(Re)-Using the Data

Address Challenges in Research Data via FAIR and APIs

Daniel Schiffner[1]

Abstract: With the increasing interest in research data, research data centers (RDCs) face several challenges with the collection, storage and dissemination of data. This keynote highlights some of the intricacies of managing and maintaining research data. It does so from a computer science lens to focus on storage infrastructures, API integration, and usability aspects. Moreover, the FAIR principles are a cornerstone guiding recent developments. This presentation will explore how adhering to these principles fosters open science and accelerates scientific processes. By aligning infrastructure development with the FAIR principles, RDCs and the National Research Data Infrastructure (NFDI) project aim to harmonize data management practices across disciplines, institutions, and borders. Their goal is to promote collaboration and maximizing the utility of research data in the long run. In this context, APIs play a pivotal role to facilitate seamless interaction with diverse data sources and platforms. By leveraging innovative approaches and emerging technologies, researchers can navigate and report research data more effectively, uncovering valuable insights and driving scientific progress.

Keywords: FAIR, Research Data, Challenges

1 Introduction

Research is a process. And research has evolved significantly with the advancement of technology, enabling the generation of vast amounts of data. However, managing this data effectively comes at a cost. The growth of volume and velocity requires new mechanisms to ensure long-term availability. Research Data Centers (RDCs) are at the forefront of addressing these challenges. They develop and maintain infrastructure for data donation, storage, and dissemination. Their task is to ensure that data is curated and preserved for future use, supporting the broader goals of open science.

In Germany, RDCs have been formalized in 1999 [Cl04], when the necessity for such infrastructures had been identified. By providing data to researchers, RDCs enable researchers to validate findings, build upon existing work, and avoid duplication of studies. To further increase quality, data is accompanied by metadata, which are supposed to make it easier for researchers to discover and reuse data across several data providers. RDCs also adhere to legal regulations and develop technical solutions to ensure security to both, the data and the researcher alike [HL22].

[1] DIPF | Leibniz Institute for Research and Information in Education,
d.schiffner@dipf.de, https://orcid.org/0000-0002-0794-0359

doi:10.18420/delfi2024_02

The FAIR principles[Wi16] are considered a cornerstone of modern data management practices. These principles guide the development of infrastructures and technologies to ensure data is Findable, Accessible, Interoperable, and Reusable. By promoting standardization, the FAIR principles enable researchers to share and reuse data across disciplines and institutions, enhancing collaboration and maximizing the utility of research data.

2 Challenges and Solutions

Despite all the efforts, managing research data still comes with several challenges, such as ensuring data quality, findability, and interoperability. Data curation involves the ongoing management of data to support its long-term viability. This includes tasks such as data cleaning, annotation, and organization. These steps are supposed to secure that data will be provided under agreed legal terms. However, due to changing technologies, RDCs are not able to effectively address the intricacies of data from specific contexts, e.g., computer science education or learning analytics, using discipline-specific methods. While ideas like Research Data Management Tools [En17] exist, important details are often lost in translation.

The diversity of data generated in our community not only challenges ourselves but troubles the long-term availability as well. There is no gold standard to structure, e.g., logdata collected from an assessment or store software artifacts, like dashboards and the data it visualizes. RDCs also need to be able to store and provide these in a way that researchers can extract necessary information in the future.

Another challenge is finding data. The current solution involves the development of comprehensive metadata and using them in search mechanisms. This enables methods like knowledge graphs [Oe22], but does not substitute the actual data.

To address these challenges, we must develop and implement innovative solutions and need to define what we, as a community require, to continue our research process. This involves adopting standards, developing and extending metadata practices, creating software tools, and leveraging APIs [KS22; KS23a].

Standards for describing our research artifacts need to include the environment we are working in. Software tools, databases and artifacts are an essential part of the research data. Currently there is no technical support to ensure the archiving of such artifacts, but this is not due to limitations of the technology, as shown by [Al23].

APIs have and will play a pivotal role in facilitating seamless interaction with diverse data sources and platforms, allowing to derive new insights. While RDCs are one backbone to provide scientific data, we need to provide different methods to utilize their potential [Li22].

3 Future

The future potential of research data management lies in the combination of efforts of RDCs and research communities. This involves fostering best practices, promoting data and software sharing and reuse [KS23b], and developing innovative solutions. We all play a crucial role in shaping the future of research data management. By actively participating in the development of data infrastructures, new systems will meet research needs and support workflows.

We also need to consider how others will use and interpret our data. Provide detailed and accurate metadata, use standard data formats, and share insights. This includes the reporting of mistakes and errors, which are an inevitable part of the research process (as we are all human). By openly discussing and addressing these, the quality and reliability of research data will improve.

No individual will overcome these limitations, so let us create a more transparent and open research community together.

References

[Al23] Al Laban, F.; Bernoth, J.; Goedicke, M.; Lucke, U.; Striewe, M.; Wieder, P.; Yahyapour, R.: Establishing the Research Data Management Container in NFDIxCS. Proceedings of the Conference on Research Data Infrastructure 1, 2023, ISSN: 2941-296X, DOI: 10.52825/cordi.v1i.395, URL: http://dx.doi.org/10.52825/CoRDI.v1i.395.

[Cl04] Clemens, W.: Rezension zu: Kommission zur Verbesserung der informellen Infrastruktur zwischen Wissenschaft und Statistik (Hrsg.): Wege zu einer besseren informationellen Infrastruktur. 2001. Sozialwissenschaften und Berufspraxis 27 (2), pp. 221–223, 2004.

[En17] Engelhardt, C.; Enke, H.; Klar, J.; Ludwig, J.; Neuroth, H.: Research data management organiser. In: Proceedings of the 14th International Conference on Digital Preservation. Pp. 25–29, 2017.

[HL22] Hoffstätter, U.; Linne, M.: Datenzugang. Einführung in das Thema Zugang zu Daten der Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften in Forschungsdatenzentren. de, 2022, DOI: 10.5281/ZENODO.7347063, URL: https://zenodo.org/record/7347063.

[KS22] Kiesler, N.; Schiffner, D.: On the Lack of Recognition of Software Artifacts and IT Infrastructure in Educational Technology Research. In. Gesellschaft für Informatik e.V., 2022, DOI: 10.18420/DELFI2022-034, URL: http://dl.gi.de/handle/20.500.12116/38835.

[KS23a] Kiesler, N.; Schiffner, D.: Open Science in den Bildungstechnologien: Zur Publikation und Begutachtung von Forschungsdaten inklusive Software im Rahmen der DELFI. In. Gesellschaft für Informatik e.V., 2023, DOI: 10.18420/WSDELFI2023-40, URL: https://dl.gi.de/handle/20.500.12116/43311.

[KS23b] Kiesler, N.; Schiffner, D.: Why We Need Open Data in Computer Science Education Research. In: Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1. ITiCSE 2023, ACM, 2023, DOI: 10.1145/3587102.3588860, URL: http://dx.doi.org/10.1145/3587102.3588860.

[Li22] Limani, F.; Johannes, R.; Zhang, Y.; Schiffner, D.: KonsortSWD Task Area 5 Measure 3: Milestone 1 & 2 Report. en, 2022, DOI: 10.5281/ZENODO.6497189, URL: https://zenodo.org/record/6497189.

[Oe22] Oelen, A.: Open Research Knowledge Graph - Using Human And Machine Intelligence To Generate A Scholarly Knowledge Graph. 2022, DOI: 10.5281/ZENODO.6501979, URL: https://zenodo.org/record/6501979.

[Wi16] Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; 't Hoen, P. A.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1), 2016, ISSN: 2052-4463, DOI: 10.1038/sdata.2016.18, URL: http://dx.doi.org/10.1038/sdata.2016.18.

Wissenschaftliches Programm

Best Paper Kandidaten Langbeiträge

Towards Nonverbal Communication in Multi-User Virtual Reality Learning Environments: A Systematic Review

Frank Wehrmann [1] and Raphael Zender [2]

Abstract: Virtual Reality is uniquely positioned to enable multi-user interactions through embodiment in virtual environments. Beyond verbal communication, this embodiment allows for various modes of nonverbal communication, such as body language, facial expressions, gestures, and avatar design. Nonverbal communication plays a significant role in face-to-face communication and is especially relevant in inclusive education, where both the diversity of expression of the students and their various communicative needs can be better catered to with a wider range of available communicative means. To explore the current state of practices and perceived benefits of nonverbal communication in multi-user virtual reality learning environments, a systematic review was conducted. The review identifies 14 relevant papers but concludes that this area represents a research gap. A discussion of potentials for nonverbal learning highlights future directions to address this gap.

Keywords: Virtual Reality, Nonverbal Communication, Multi-User, E-Learning, Embodied Learning

1 Motivation

Diverse communication plays a crucial role in achieving inclusive education [RS19a]. To foster an inclusive learning environment, it is essential to embrace diverse and multimodal communication styles, including verbal, written, and nonverbal channels. The Universal Design for Learning Guidelines advocate for offering multiple means for the engagement with learning content and the expression of ideas [CA24]. Communication is also of paramount importance in collaborative e-learning, where it serves as the foundation for peer interaction among learners engaged in synchronous cooperation or collaboration within learning environments [Di99].

Immersive Virtual Reality (VR), especially when based on Head-Mounted Displays (HMDs), provides a method to engage the physical dimension of communication through embodied learning [FQ23]. By leveraging HMDs and controllers, VR captures users' head and hand movements within applications [Dö19], and with the aid of technologies like data gloves and camera-based tracking for more accurate hand tracking [Li19] and algorithms for interpolating realistic body movements from limited tracking points (inverse kinematics) [Ol22], a more authentic simulation of body language is achievable.

A review of design for inclusive VR learning applications highlighted nonverbal communication (NVC) as a benefit of VR, emphasizing the significance of avatars with detailed body

1 Humboldt-Universität zu Berlin, Institut für Informatik, Unter den Linden 6, 10099 Berlin, Deutschland, frank.wehrmann@hu-berlin.de, https://orcid.org/0009-0007-5774-5045
2 Zeppelin Universität, Am Seemooser Horn 20, 88045 Friedrichshafen, Deutschland, raphael.zender@zu.de, https://orcid.org/0000-0001-9866-9455

doi:10.18420/delfi2024_03

language and facial expressions [WZ24]. However, the variety of embodied communication forms presents a research gap in multi-user VR learning contexts [Ma23], which motivates this review.

This paper presents a systematic review aimed at addressing the research question: What is the current state of practice, and what are the potential benefits of NVC for peer interaction in multi-user VR environments for learning and inclusion? The review begins with a discussion of the literature to establish key concepts, followed by a brief presentation of the review methodology and a narrative synthesis of the findings. A significant research gap was identified, as very few studies address educational contexts. These limited findings will be compiled and placed in the context of literature on NVC in learning environments to highlight unexplored potential in education. The conclusion will summarize the most critical insights and potentials and emphasize the need to address the identified research gap.

2 Background

2.1 Nonverbal Communication for Education and Inclusion

This review focuses on NVC for peer interaction, but past research has shown that nonverbal cues also provide teachers with insights into students' attitudes and emotions [Mi88; OR94], suggesting that teachers would also benefit from being able to observe NVC by students in Multi-User Virtual Environments (MUVE). NVC plays a critical role in collaborative learning. Despite being an under-researched area [ZSC24], insights from studies on collaboration in digital [SP13] and face-to-face settings [SWM21] suggest a correlation between the quality of collaboration and gaze behavior. Another study highlights an increase in creativity associated with nonverbal synchrony between collaborators [Wo14b]. Moreover, Won et al. [Wo14b] provide evidence that body language is correlated with mutual positive attitudes among interacting individuals, which, in turn, is linked to success in interpersonal interactions. A study found that in cooperative tasks requiring spatial coordination, dyads developed spontaneous nonverbal coordination without prior verbal agreement [PES14]. Although this example pertains to spatially coordinated tasks, it indicates an intuitive aspect of NVC that could potentially facilitate various forms of cooperation and collaboration.

Inclusion is interpreted as a comprehensive approach that addresses various inequalities, with a strong emphasis on providing equitable education for all. This approach focuses on processes such as interdependent cooperation, learner-driven communication, active and democratic participation, and reflection concerning biases [Fr19]. While accessibility is often the first aspect associated with inclusion, it is encompassed within the broader concept of participation. The notion of inclusive education is thus expanded to promote a positive attitude towards diversity and to productively utilize existing diversity within learning contexts.

NVC interacts with multiple dimensions of heterogeneity among learners. Culture and gender influence nonverbal behaviors [CM24; Co16; OR94], and Anderson [An23] argue that NVC enhances intercultural communication competency by offering a rich source of cultural and emotional information. Dey; Puntambekar [DP23] note that verbal and NVC are not directly linked; a student with low verbal participation might exhibit high nonverbal participation, indicating a preference for this communication style. Key functions of NVC include expressing emotions, conveying interpersonal attitudes, and presenting one's personality [OR94]. To include the real diversity of students in educational processes, it is essential to provide individual, multimodal, and diverse communication channels that allow them to fully participate and express themselves in meaningful ways.

The link between communication and inclusion extends beyond the dimension of diversity. Inclusive education aims to develop students' communicative competence within the domain of the subject being taught [RS19a; RS19b]. The acquisition of subject-specific language (e.g., the language of physics, biology, chemistry) is compared to and contextualized within foreign language acquisition [Le05; RS19b]. While this language is often conceptualized as written language, inclusive education emphasizes the importance of verbal communication [Fo17; Rö]. Additionally, NVC is pivotal in transitioning from everyday speech to subject-specific language. Students move from context-dependent language use in everyday speech and nonverbal registers (e.g., pointing at an experiment and saying "that right there") to increasingly abstract and context-independent descriptions as they acquire the subject's vernacular [Fo17; Le05; Le24]. Gesture-based collaboration has been shown to enhance perceptual learning [RZ22], and children prevented from gesturing rely more on non-present information, reducing their reasoning based on perceptually present information [GA13]. These findings suggest the potential of NVC as scaffolding for the early stages of development, when students still benefit most from using perceptually present information for heavily context-dependent communication.

2.2 Nonverbal Communication and VR

A review of communication in VR underscores the importance of avatars and nonverbal cues in influencing user interactions in social VR settings. It also reveals that VR communication bears more resemblance to face-to-face interaction than other communication media [WJF22]. However, this review did not specifically focus on Multi-User Virtual Environments nor discuss educational applications, underscoring the need for the present review.

The immersive nature of VR, particularly with advanced body-tracking technologies, facilitates a high degree of embodiment [Jo14]. Embodiment is the process through which a user assumes a simulated body within a virtual environment, referred to as an avatar [Mo14]. Embodiment allows for the incorporation of sensorimotor activities into the learning process [Jo14] and enhances processes of implicit learning through a sense of body ownership [Sl17]. Beyond its value for individual learners, embodiment in VR can introduce new opportunities for communication during collaboration. For example, individuals tend to

gesture more when establishing common ground in their exchanges, and gestures play a role in regulating turn-taking in discussions and directing attention towards objects [Na21].



Fig. 1: Mapping of user's real-world eye and mouth movements onto an avatar. [Ki23]

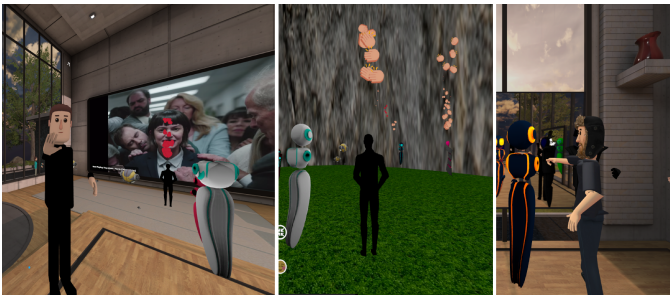


Fig. 2: Nonverbal cues in VR: extended hand, non-isomorphic emoji use, and pointing. [MFW20]

Embodied learning spans a spectrum from no immersion to full immersive motion-captured embodiment [Jo14]. Depending on the extent of motion capture used, VR can achieve the highest level of embodiment within this spectrum. In the context of MUVEs, not only is the perception of the environment as real (presence) important, but also the perception of other users' avatars as real to create the sensation of sharing the embodied space with others (social presence) [Ki23]. Factors influencing the degree of social presence include avatar representation [Ro16] and nonverbal cues such as facial expressions [Ki23] (refer to Fig. 1) and body language [SW21] (refer to Fig. 2). Moreover, avatars themselves can be viewed as a form of NVC [Mo14], serving as a powerful expression of self [Fr20] and identity [MF20]. Notably, users can adapt to more abstract avatars [Ro16; SW21] and can even embrace symbolic or non-isomorphic nonverbal expressions [Ba21; Id21]. While full avatar realism is not mandatory, it is preferred for its superior representation of realistic body language [AI21; Fr22; Ro16; Ro22].

Gesture interaction in VR, studied as a means of directly interacting with the virtual environment or applications [Li19], is not the primary focus of this review. Instead, this review emphasizes nonverbal cues used for peer interaction, such as facial expressions, gaze, gestures, proximity, emojis, drawing, and more [WJF22]. Communication is inherently

multimodal, integrating various sensory modalities and body-based capacities for expression [Na21]. Consequently, the implementation of nonverbal cues in virtual environments is critical. For instance, emojis can be effectively used if a convenient selection menu is implemented [THB20], and while tracking facial expressions is challenging with HMDs, it is a highly desirable feature [Ki23]. The simplest gestures in VR environments involve proximity and basic hand gestures, requiring only the implementation of avatars that can be seen and interact with others in virtual space. Limited gesture tracking is available with most VR systems through tracking hand and head position, but increasingly more systems, like the Apple Vision Pro and the Meta Quest series of headsets, feature camera-based hand tracking, which is able to track hand gestures without relying on controllers. The focus of this paper is on how NVC is implemented in MUVEs utilizing a variety of these technologies, but it is conceivable that as hand-tracking technology develops it may become the dominant technology.

3 Literature Review

The literature review involved a systematic review adapted from the PRISMA methodology [Pa21], see Fig. 3. The search strategy was a modified Population, Phenomena of Interest, Context (PICo) framework [Mu18]: **Population** (multi-user OR "multi user" OR cooper* OR collab* OR kooper* OR kollab* OR team* OR mehrspieler OR multiplayer OR multi-player), **Phenomena of Interest** (nonverbal OR non-verbal OR "non verbal" OR

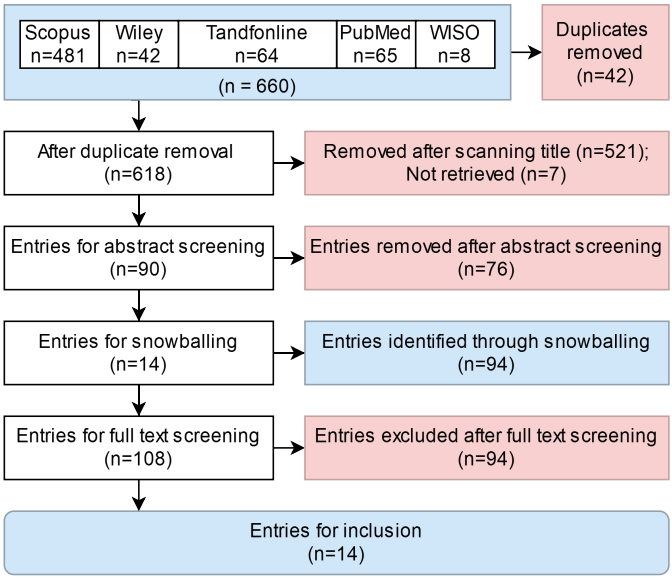


Fig. 3: Flowchart of search process

"body language" OR gestur* OR "facial expression" OR "eye movement" OR gaze OR "nicht verbal" OR körpersprache OR gest* OR mimik OR auge*), **Context** ("virtual reality" OR virtual-reality OR "virtuelle realität").

The search date was the 3rd of April 2024, for databases searched see Fig. 3. This selection was informed by recommendations [GH20] and prior experiences with databases. Only works published from 2013 onward were considered, because the widespread availability of relevant HMD hardware started at that time. Searches were limited to peer-reviewed journal articles and conference papers, with further restrictions in Scopus to entries tagged with the "virtual reality"keyword. The ACM Digital Library was not searched directly due to a very high volume of irrelevant results with this search strategy, but other databases yielded numerous ACM articles, indicating that relevant literature from there was still captured. Forward and backward snowballing techniques were employed [Wo14a] to mitigate any limitations in the search strategy.

After a review of the full texts, a set of 14 papers was identified adhering to these inclusion criteria: (1) Utilization of HMD-based VR. (2) Multi-user social interaction among peers. (3) Documentation of the presence or benefits of NVC. (4) A context involving meaningful multi-user tasks, not strictly confined to educational or learning contexts, as preliminary searches indicated scarce literature focusing on these areas.

4 Findings

Analysis of the identified records was done through a narrative synthesis [Ro09]. This approach was chosen because of the heterogeneous and mostly non-systematic nature of the data. The findings are divided into real-world applications, where actual case studies were conducted with actually implemented MUVEs, and experimental settings, where MUVEs were designed to gather data on a specific aspects of NVC in MUVEs. Applications are described in detail, to highlight the emergent usage of NVC they enable. Experimental results are synthesized to give an overview of the state of research. The synthesis of all records focused on identifying the implementation or emergence of NVC in peer interactions within these papers.

Notably, even among the identified records, there is a minimal focus on learning contexts, indicating a significant research gap. However, some findings highlight the potential of MUVEs with NVC for education and inclusion, which will be discussed following the presentation of the findings.

4.1 Applications

Krinkle Cube [LNA16] is a multiplayer game that incorporates gesture interaction. It utilizes cartoon-style avatars that lack facial expressions but are tracked through HMD

position and Leap Motion controllers for hand movements. Despite the avatars' positive reception, a desire for more facial expressiveness was noted. This is not an educational application, but players are able to see each other and spontaneously communicated through gestures within the game, suggesting that coexistence in a shared space for task collaboration fosters emergent communication forms which were not part of the games rules or task instructions. This paper presents its insights qualitatively while not providing any quotations of what the players actually said, instead summarizing them.

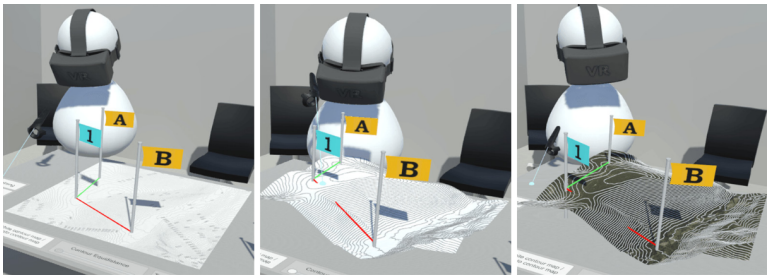


Fig. 4: Use of an abstract avatar and laser pointer in a geography education application [Ša18].

A **collaborative geography education application** [Ša18] employs abstract avatars to interact with authentic geospatial data. It features a shared virtual room for multiple users, tracking head and hand movements. Hands are represented as floating controllers with a laser pointer functionality (Fig. 4). The paper utilizes qualitative methodologies. Users reported that the lack of facial features hindered emotion perception, yet they adapted their communication strategies to operate based on posture and proximity. The implementation of objects within the virtual environment led to instances where users would obstruct each other's view by standing inside objects. The study suggests that despite these challenges, user adaptability helped them overcome some communication barriers and the authors suggest that initial adaptation problems would disappear during long-term educational interventions as users get used to the application.

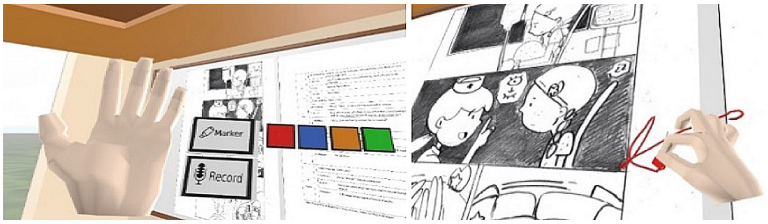


Fig. 5: Collaborative manga creation in SyncMeet with an annotation tool [TYT20].

SyncMeet [TYT20] is a virtual work environment designed for collaborative manga creation (Fig. 5). This is not an educational application, but educational usage for collaborative creative tasks are possible. It features simplistic avatars with head tracking and detailed hand models for gestures tracked through Leap Motion controllers. This application allows users to collaboratively work on manga, annotating it with a drawing tool. Data was collected

through a mixture of qualitative and quantitative data. Users found the VR meetings to be almost as engaging as face-to-face interactions. The lack of facial expressions was noted again. The process of discussing and annotating comics, while pointing at specific image elements, presents an interesting case for task-specific NVC, though the paper does not elaborate on these interactions. Quantitative insights of the paper focused on mental workload comparing different forms of interaction, noticing no significant findings.

A **Construction Sector Occupational Health and Safety VR experience** [Ta22] supports up to twenty users in a shared VR space with basic avatars that feature floating heads and hands tracked through the HMD and controllers. The hands are able to use laser pointers. Some hand gestures are tracked using Meta Quest controllers, which enable simple gestures like thumbs-up and pointing. Avatars can be customized with a nametag on their helmets, and movement within the VR environment is facilitated through walking (short distances) and teleportation (long distances). This application is not educational in nature, but its emphasis on collaborative virtual design processes could be valuable for educational contexts. The methodology of this paper is qualitative, presenting this application as a case study and conducting a thematic analysis. Users found that shared spatial visualization enhanced communication by offering nonverbal cues such as head direction and hand gestures. They suggested that avatar improvements could lead to a stronger sense of presence and richer NVC, such as blinking, gaze tracking, and lip syncing.



Fig. 6: Theatre rehearsal space and avatar appearance in VR Chat [Le23b].

The use of social VR platforms Altspace VR and VR Chat for **theatre rehearsals** [Le23b] showed that users preferred full-body avatars (refer to Fig. 6) employing inverse kinematics over floating torsos with hands. Because of the avatars' limited capability for NVC, particularly facial expressions, users felt the experience was akin to performing a radio drama rather than a stage play. Presence issues arose from avatars behaving unrealistically, e.g., through glitches, physics violations, and teleportation. However, VR rehearsals were beneficial for learning blocking (actor positioning in a scene), embodying characters, and fostering social bonds among participants. Users were able to memorize blocking positions and replicate them on a real stage, suggesting that MUVEs can effectively represent the nonverbal dimension of proximity. The study concludes that VR is beneficial for theatre rehearsals, even with limited gesturing and positioning capabilities, suggesting that advanced technologies like haptics and facial tracking might not be essential for creating valuable

MUVE experiences. Most of the data gathered was qualitatively, but some quantitative data was gathered as well. A high degree of engagement, presence, co-presence and a low degree of motion sickness were measured through questionnaires. This application is not strictly educational, because it was aimed at professional actors, but educational use for actors in training and theatre classes is possible.

4.2 Experiments

The papers covered in this section are experimental MUVEs, designed and used only for gathering data about peer interaction in MUVEs. As such they are not strictly educational, but their findings can highlight educational benefits. All of the papers gather some quantitative data, and 5 out of 9 papers additionally provide qualitative data extracted from interviews with the users. This section presents a narrative synthesis of the findings.

Research consistently demonstrates the efficacy of MUVEs for communication, highlighting their resemblance to face-to-face interactions [CSF18; Gr17; Le23a; SN18] and how embodiment positively affects cooperative tasks [CSF18; PS17; Wu21]. On the flip side, a lack of embodiment can diminish this effect and the resemblance to face-to-face interaction [SN18].

In terms of gaze behavior, findings are mixed. One study reported minimal impact from the realism of gaze [Se17], yet another indicated that more expressive avatars can encourage active gaze engagement [Le23a]. The expressiveness of avatars significantly influences the usage of nonverbal expressions, especially in direct social interactions [HOB18], with limited facial expressiveness reducing perceived social presence [SN18], while expressive facial features enhance it [Le23a] (refer to Fig. 7). This suggests that a lower degree to realism that can exaggerate the emotional expressiveness of avatars might be beneficial.

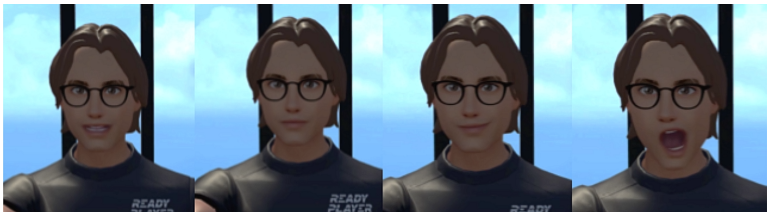


Fig. 7: Facial expressions on an avatar [Le23a].

Avatar embodiment can play a role in reducing social anxiety [Gr17], fostering positive attitudes towards collaboration [HOB18; PS17; Wu21], and effectively transmitting emotional cues [Ha18]. Enhancing NVC through technological means (exaggerating tracked facial expressions) is suggested to further improve these interactions [Ha18]. Interestingly, a lower degree of avatar realism may be beneficial here, as it could lead to a greater tolerance for tracking inaccuracies [HOB18] and does not significantly impact social presence negatively [Wu21].

Drawing and sketching activities particularly benefit from VR's interactive capabilities [CSF18]. Although the advantages of full-bodied avatars in single-user VR environments may not fully apply to MUVEs [HOB18], expressive avatars are generally linked to an enhanced sense of social presence [Le23a; Wu21]. For tasks that lack social interaction, facial expressions may not significantly affect performance [Le23a], indicating the nuanced role of nonverbal cues in different contexts. The need for nonverbal cues seems to vary across applications; some necessitate realistic cues for effective communication, while others do not [Gr17; Ha18; Le23a].

5 Potentials for Education and Inclusion

The findings presented permit several observations regarding the potential of MUVEs with NVC for education and inclusion. The existing studies do not enable a comprehensive empirical evaluation of these observations, and as such, they are subject to future assessment. Nonetheless, they provide a means to address the research gap identified by this review.

Approximating Face-to-Face Interactions in Virtual Environments appears feasible based on numerous accounts previously described. While Virtual Reality (VR) offers significant value across various learning scenarios, its effectiveness in creating sufficiently communicative learning environments largely depends on the design of the applications. The findings presented suggest that VR learning applications ought to be designed as multi-user platforms. Moreover, these applications should incorporate features such as detailed avatars capable of displaying facial expressions and supporting gestural communication, to foster social presence and immerse students in communicative learning contexts. This aspect is particularly crucial from the perspective of inclusive education, where the importance of student communication through diverse means is central.

The **emergence of spontaneous NVC strategies**, as observed with Krinkle Cube [LNA16], mirrors the findings detailed by Parikh et al. [PES14]. In instances where NVC supplements verbal communication, users have also positively acknowledged this additional value [Ta22]. The provision for employing diverse communication methods suggests that students who could benefit from this style of communication might increasingly utilize it, thereby enabling a type of interaction otherwise inaccessible in VR. It is conceivable that, due to the scant descriptions provided regarding the precise nature of nonverbal interactions in many of these studies, more such emergent interactions occurred but were not documented.

Communication About Perceptually Present Information in VR Visualizations represents a domain where one of VR's primary advantages—innovative ways of visualizing information—converges with the educational benefits of learning through perceptually present information, which facilitates gesture interactions [GA13; RZ22]. This benefit has been observed in interactions with geospatial data [Ša18], manga creation [TYT20], and construction safety [Ta22], where users were provided with tools such as laser pointers or

drawing capabilities. These tools enhance communication by enabling users to directly point to or interact with objects that are perceptually present.

The **utilization of avatars provides benefits for self-expression**, potentially even surpassing the capabilities of face-to-face interactions. One study highlighted that the "masking" effect facilitated by an avatar led to a reduction in social anxiety during collaborative tasks [Gr17]. Moreover, Maloney; Freeman [MF20] observed that social VR with avatars fosters experimentation with self-expression. They noted that users of social VR platforms experimented with avatars of different genders, potentially enhancing inclusion for queer individuals [Fr20]. All of this suggests that the ability to hide ones body behind an avatar, and to communicate not verbally but through NVC, can provide a feeling of safety to the users and helps them overcome social anxiety [MFW20; MZF20]. Another study discovered that marginalized users of a social VR platform utilized NVC as a means to shield themselves from undesired behaviors and to overcome biases and communication barriers associated with not being native English speakers [MFW20]. This indicates advantages for intercultural communication skills, aligning with research on NVC in face-to-face language learning [An23]. Beyond self-expression through avatars, Slater [Sl17] discussed how VR embodiment could reduce racial bias when users assume perspectives of bodies outside their in-group. These findings, primarily within the realm of social VR, suggest potential benefits that could extend to learning applications, provided avatar customization is adequately implemented. Such an approach could facilitate aspects of participation (through self-expression) and reflection (towards bias reduction), which are essential to inclusive education [Fr19].

An aspect related to the capacity for self-expression is **the ability to read emotions and foster social bonding** through NVC in VR. As noted by many discussed articles, the absence of facial expressions was a common critique. The availability of facial expressions would facilitate the perception of emotions and enhance social presence, a feature users desired. During VR theatre rehearsals, users reported a high degree of social bonding among participants [Le23b]. Other studies demonstrated that avatar embodiment could promote positive feelings towards collaborators [HOB18; PS17; Wu21] and assist in conveying emotional cues [Ha18]. An additional potential benefit for inclusion is the enhancement of facial expressions by some studies [Ha18; Le23a], which could enable learners who struggle to express emotions in face-to-face interactions to rely on technological support to communicate their feelings to collaborators.

Higher quality of collaboration and cooperation through nonverbal VR interaction has been observed in multiple studies [CSF18; PS17; Wu21] and appears to align with the positive effects NVC has on face-to-face interactions, as discussed in prior research [PES14; SP13; SWM21; Wo14b; ZSC24]. This may stem from the cumulative effects of all previously mentioned benefits, such as positive social attitudes towards collaborators, enhanced trust, a greater degree of self-expression, communication about perceptually present information, and spontaneous NVC strategies. There appears to be an interdependence between cooperation and communication, a theme also explored in the inclusive education

literature, which emphasizes the significance of cooperation and communication among other processes [Fr19].

6 Summary and Conclusion

This review underscores the capability of MUVEs to foster a level of social presence that rivals that of face-to-face communication by implementing NVC features. It has been observed that users are adaptable to varying levels of avatar fidelity, including aspects such as gaze, gestures, and bodily realism, tailored to the context of the task at hand. Tasks that demand a higher degree of social attentiveness benefit from greater levels of embodiment. Nonetheless, there is a prevalent call for avatars with more expressive facial features, highlighting an area ripe for technological improvement and a desire of users to experience higher social presence with their collaborators. Future work could cover the question, if less realistic avatars might provide benefits that outweigh the noted value of realistic representation.

Furthermore, it has been noted that specific tasks stand to gain from the unique effects of nonverbal interactions, such as the use of blocking in theater rehearsals, laser pointers in 3D visualisations, and drawing in collaborative creative projects. However, the spontaneous emergence of NVC strategies in collaborative scenarios hints at potential widespread benefits beyond specialized solutions.

The potential advantages for education and inclusion stemming from these observations were explored. MUVEs with nonverbal interaction capabilities offer a semblance to face-to-face interactions, enable the spontaneous emergence of NVC strategies, facilitate communication regarding perceptually present information in VR visualizations, allow for creative self-expression through avatars, enhance the ability to perceive and express emotions, strengthen social bonds, and foster higher quality collaboration and cooperation through nonverbal interactions. These insights suggest that NVC in MUVEs represents a promising area for future research, which should examine these aspects in greater detail. The scarcity of empirical studies specifically within educational contexts and the lack of many applications specifically designed for educational contexts limit this review, underscoring a significant research gap.

Initially aimed at spotlighting the importance of NVC in MUVEs for learning, this review instead unveiled a considerable gap in research on this topic. While most identified studies focused on non-educational tasks or social presence, their core findings are nonetheless pertinent to the exploration of NVC for peer interaction in educational settings. Future research should give precedence to educational applications, incorporating these insights into design guidelines for NVC and avatar creation in educational MUVEs. Moreover, the development of applications tailored to specific educational scenarios across various disciplines could utilize these findings for more comprehensive evaluations.

References

[AI21] Aseeri, S.; Interrante, V.: The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. eng, IEEE transactions on visualization and computer graphics 27 (5), pp. 2608–2617, 2021, DOI: 10.1109/TVCG.2021.3067783.

[An23] Anderson, K. A.: Nonverbal communication as a mediator of intercultural communication in English as a second language classrooms. en, Journal of Multilingual and Multicultural Development, pp. 1–13, 2023, DOI: 10.1080/01434632.2023.2262422.

[Ba21] Baloup, M.; Pietrzak, T.; Hachet, M.; Casiez, G.: Non-isomorphic Interaction Techniques for Controlling Avatar Facial Expressions in VR. In: Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology. ACM, Osaka Japan, pp. 1–10, 2021, DOI: 10.1145/3489849.3489867.

[CA24] CAST: Universal Design for Learning Guidelines version 2.2, 2024, URL: https://udlguidelines.cast.org/, visited on: 04/02/2024.

[CM24] Correa-Chávez, M.; Mejía-Arauz, R.: Cultural patterns in nonverbal collaboration between pairs of Indigenous children and pairs of urban middle-class children in México. en, Journal of Applied Developmental Psychology 90, p. 101606, 2024, DOI: 10.1016/j.appdev.2023.101606.

[Co16] Correa-Chávez, M.: Cultural patterns of collaboration and communication while working together among U.S. Mexican heritage children. en, Learning, Culture and Social Interaction 11, pp. 130–141, 2016, DOI: 10.1016/j.lcsi.2016.08.001.

[CSF18] Coburn, J.; Salmon, J.; Freeman, I.: Effectiveness of an Immersive Virtual Environment for Collaboration with Gesture Support Using Low-Cost Hardware. Journal of Mechanical Design 140 (4), 2018, DOI: 10.1115/1.4039006.

[Di99] Dillenbourg, P.: What do you mean by 'collaborative learning'? In (Dillenbourg, P., ed.): Collaborative learning: cognitive and computational approaches. Advances in learning and instruction series, Elsevier, Oxford, pp. 1–19, 1999.

[Dö19] Dörner, R.; Broll, W.; Grimm, P.; Jung, B., eds.: Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität. Springer, Berlin, Heidelberg, 2019.

[DP23] Dey, I.; Puntambekar, S.: Examining Nonverbal Interactions to Better Understand Collaborative Learning. In. Pp. 273–276, 2023, DOI: 10.22318/cscl2023.218179.

[Fo17] Fornol, S. L.: Bildungssprache – mehr als konzeptionelle Schriftlichkeit? In (Heinzel, F.; Koch, K., eds.): Individualisierung im Grundschulunterricht: Anspruch, Realisierung und Risiken. Springer Fachmedien, Wiesbaden, pp. 178–182, 2017, DOI: 10.1007/978-3-658-15565-0_30.

[FQ23] Fortman, J.; Quintana, R.: Fostering collaborative and embodied learning with extended reality: Special issue introduction. en, International Journal of Computer-Supported Collaborative Learning 18 (2), pp. 145–152, 2023, DOI: 10.1007/s11412-023-09404-1.

[Fr19] Frohn, J.; Brodesser, E.; Moser, V.; Pech, D., eds.: Inklusives Lehren und Lernen. Allgemein- und fachdidaktische Grundlagen. Julius Klinkhardt, Bad Heilbrunn, 2019.

[Fr20] Freeman, G.; Zamanifard, S.; Maloney, D.; Adkins, A.: My Body, My Avatar: How People Perceive Their Avatars in Social Virtual Reality. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. CHI EA '20, Association for Computing Machinery, New York, NY, USA, pp. 1–8, 2020, DOI: 10.1145/3334480.3382923.

[Fr22] Fraser, A. D.; Branson, I.; Hollett, R. C.; Speelman, C. P.; Rogers, S. L.: Expressiveness of real-time motion captured avatars influences perceived animation realism and perceived quality of social interaction in virtual reality. Frontiers in Virtual Reality 3, p. 981400, 2022, DOI: 10.3389/frvir.2022.981400.

[GA13] Goldin-Meadow, S.; Alibali, M. W.: Gesture's Role in Learning and Development. In (Zelazo, P. D., ed.): The Oxford Handbook of Developmental Psychology, Vol. 1. Oxford University Press, pp. 952–973, 2013, DOI: 10.1093/oxfordhb/9780199958450.013.0033.

[GH20] Gusenbauer, M.; Haddaway, N. R.: Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. en, Research Synthesis Methods 11 (2), pp. 181–217, 2020, DOI: 10.1002/jrsm.1378.

[Gr17] Greenwald, S. W.; Wang, Z.; Funk, M.; Maes, P.: Investigating Social Presence and Communication with Embodied Avatars in Room-Scale Virtual Reality. In (Beck, D.; Allison, C.; Morgado, L.; Pirker, J.; Khosmood, F.; Richter, J.; Gütl, C., eds.): Immersive Learning Research Network. Springer International Publishing, Cham, pp. 75–90, 2017, DOI: 10.1007/978-3-319-60633-0_7.

[Ha18] Hart, J. D.; Piumsomboon, T.; Lawrence, L.; Lee, G. A.; Smith, R. T.; Billinghurst, M.: Emotion Sharing and Augmentation in Cooperative Virtual Reality Games. In: Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts. CHI PLAY '18 Extended Abstracts, Association for Computing Machinery, New York, NY, USA, pp. 453–460, 2018, DOI: 10.1145/3270316.3271543.

[HOB18] Herrera, F.; Oh, S. Y.; Bailenson, J. N.: Effect of Behavioral Realism on Social Interactions Inside Collaborative Virtual Environments. Presence: Teleoperators and Virtual Environments 27 (2), pp. 163–182, 2018, DOI: 10.1162/pres_a_00324.

[Id21] Ide, M.; Oshima, S.; Mori, S.; Yoshimi, M.; Ichino, J.; Tano, S.: Effects of Avatar's Symbolic Gesture in Virtual Reality Brainstorming. In: Proceedings of the 32nd Australian Conference on Human-Computer Interaction. OzCHI '20, Association for Computing Machinery, New York, NY, USA, pp. 170–177, 2021, DOI: 10.1145/3441000.3441081.

[Jo14] Johnson-Glenberg, M.; Birchfield, D.; Tolentino, L.; Koziupa, T.: Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. Journal of Educational Psychology 106 (1), pp. 86–104, 2014, DOI: 10.1037/a0034008.

[Ki23] Kimmel, S.; Jung, F.; Matviienko, A.; Heuten, W.; Boll, S.: Let's Face It: Influence of Facial Expressions on Social Presence in Collaborative Virtual Reality. In. 2023, DOI: 10.1145/3544548.3580707.

[Le05] Leisen, J.: Wechsel der Darstellungsformen. Ein Unterrichtsprinzip für alle Fächer. Der fremdsprachliche Unterricht. Englisch 39 (78), pp. 9–11, 2005.

[Le23a] Le Tarnec, H.; Bevacqua, E.; Augereau, O.; De Loor, P.: Effect of Avatar Facial Expressiveness on Team Collaboration in Virtual Reality. In. 2023, DOI: 10.1145/3570945.3607330.

[Le23b] Levordashka, A.; Eastman, J.; Skoulikari, E. A.; Salagean, A.; Cosker, D.; Stanton Fraser, D.: An Exploration of Theatre Rehearsals in Social Virtual Reality. In: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. CHI EA '23, Association for Computing Machinery, New York, NY, USA, pp. 1–7, 2023, DOI: 10.1145/3544549.3585685.

[Le24] Leisen, J.: Sprachsensibler Fachunterricht | Sprachbildung und Bildungssprache, 2024, URL: http://www.sprachsensiblerfachunterricht.de/sprachbildung, visited on: 04/10/2024.

[Li19] Li, Y.; Huang, J.; Tian, F.; Wang, H.-A.; Dai, G.-Z.: Gesture interaction in virtual reality. Virtual Reality & Intelligent Hardware 1 (1), Number: 1, pp. 84–112, 2019, DOI: 10.3724/SP.J.2096-5796.2018.0006.

[LNA16] Lages, W. S.; Nabiyouni, M.; Arantes, L.: Krinkle Cube: A Collaborative VR Game Using Natural Interaction. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts. CHI PLAY Companion '16, Association for Computing Machinery, New York, NY, USA, pp. 189–196, 2016, DOI: 10.1145/2968120.2987746.

[Ma23] Mayer, A.; Kilinc, I.; Sprügel, K.; Häfner, P.: Towards Research Gaps in Collaborative Virtual Reality Environments for Education: A Literature Review. English, Lecture Notes in Networks and Systems 763 LNNS, ed. by M.E, A.; R, L.; T, T., pp. 513–525, 2023, DOI: 10.1007/978-3-031-42467-0_47.

[MF20] Maloney, D.; Freeman, G.: Falling Asleep Together: What Makes Activities in Social Virtual Reality Meaningful to Users. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play. CHI PLAY '20, Association for Computing Machinery, New York, NY, USA, pp. 510–521, 2020, DOI: 10.1145/3410404.3414266.

[MFW20] Maloney, D.; Freeman, G.; Wohn, D. Y.: Talking without a Voice: Understanding Non-verbal Communication in Social Virtual Reality. en, Proceedings of the ACM on Human-Computer Interaction 4 (CSCW2), Number: CSCW2, pp. 1–25, 2020, DOI: 10.1145/3415246.

[Mi88] Miller, P. W.: Nonverbal Communication. What Research Says to the Teacher. https://eric.ed.gov/?id=ED293190, ERIC, 1988.

[Mo14] Morie, J. F.: Avatar appearance as prima facie non-verbal communication. In: Nonverbal Communication in Virtual Worlds: Understanding and Designing Expressive Characters. ETC Press, Pittsburgh, PA, USA, pp. 77–102, 2014.

[Mu18] Munn, Z.; Stern, C.; Aromataris, E.; Lockwood, C.; Jordan, Z.: What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. en, BMC Medical Research Methodology 18 (1), p. 5, 2018, DOI: 10.1186/s12874-017-0468-4.

[MZF20] Maloney, D.; Zamanifard, S.; Freeman, G.: Anonymity vs. Familiarity: Self-Disclosure and Privacy in Social Virtual Reality. In: Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology. VRST '20, Association for Computing Machinery, New York, NY, USA, pp. 1–9, 2020, ISBN: 978-1-4503-7619-8, DOI: 10.1145/3385956.3418967.

[Na21] Nathan, M. J.: Foundations of Embodied Learning: A Paradigm for Education. Routledge, New York, 2021.

[Ol22] Oliva, R.; Beacco, A.; Navarro, X.; Slater, M.: QuickVR: A standard library for virtual embodiment in unity. Frontiers in Virtual Reality 3, 2022, DOI: 10.3389/frvir.2022.937191.

[OR94] O'Hair, M. J.; Ropo, E.: Unspoken messages: Understanding diversity in education requires emphasis on nonverbal communication. Teacher Education Quarterly, Publisher: JSTOR, pp. 91–112, 1994.

[Pa21] Page, M. J.; McKenzie, J. E.; Bossuyt, P. M.; Boutron, I.; Hoffmann, T. C.; Mulrow, C. D.; Shamseer, L.; Tetzlaff, J. M.; Akl, E. A.; Brennan, S. E.; Chou, R.; Glanville, J.; Grimshaw, J. M.; Hróbjartsson, A.; Lalu, M. M.; Li, T.; Loder, E. W.; Mayo-Wilson, E.; McDonald, S.; McGuinness, L. A.; Stewart, L. A.; Thomas, J.; Tricco, A. C.; Welch, V. A.; Whiting, P.; Moher, D.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. en, Systematic Reviews 10 (1), p. 89, 2021, DOI: 10.1186/s13643-021-01626-4.

[PES14] Parikh, S. P.; Esposito, J. M.; Searock, J.: The Role of Verbal and Nonverbal Communication in a Two-Person, Cooperative Manipulation Task. en, Advances in Human-Computer Interaction 2014, pp. 1–10, 2014, DOI: 10.1155/2014/375105.

[PS17] Pan, Y.; Steed, A.: The impact of self-avatars on trust and collaboration in shared virtual environments. en, PLOS ONE 12 (12), Publisher: Public Library of Science, e0189078, 2017, DOI: 10.1371/journal.pone.0189078.

[Rö] Rödel, L.: Konzeptionelle Mündlichkeit/ Schriftlichkeit, de, Seite, URL: https://pse.hu-berlin.de/de/forschung-und-lehre/projekte/fdqi-hu/projektergebnisse/Inklusionsglossar/konzeptionelle-muendlichkeit, visited on: 04/10/2024.

[Ro09] Rodgers, M.; Sowden, A.; Petticrew, M.; Arai, L.; Roberts, H.; Britten, N.; Popay, J.: Testing Methodological Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: Effectiveness of Interventions to Promote Smoke Alarm Ownership and Function. en, Evaluation 15 (1), pp. 49–73, 2009, ISSN: 1356-3890, DOI: 10.1177/1356389008097871.

[Ro16] Roth, D.; Lugrin, J.-L.; Galakhov, D.; Hofmann, A.; Bente, G.; Latoschik, M. E.; Fuhrmann, A.: Avatar realism and social interaction quality in virtual reality. In: 2016 IEEE Virtual Reality (VR). Pp. 277–278, 2016, DOI: 10.1109/VR.2016.7504761.

[Ro22] Rogers, S. L.; Broadbent, R.; Brown, J.; Fraser, A.; Speelman, C. P.: Realistic Motion Avatars are the Future for Social Interaction in Virtual Reality. Frontiers in Virtual Reality 2, p. 750729, 2022, DOI: 10.3389/frvir.2021.750729.

[RS19a] Rödel, L.; Simon, T.: 1.4.2 Kommunikation. In (Frohn, J.; Brodesser, E.; Moser, V.; Pech, D., eds.): Inklusives Lehren und Lernen. Allgemein- und fachdidaktische Grundlagen. Julius Klinkhardt, Bad Heilbrunn, pp. 43–45, 2019, DOI: 10.25656/01:16759.

[RS19b] Inklusive Sprach(en)bildung. Ein interdisziplinärer Blick auf das Verhältnis von Inklusion und Sprachbildung. de, ed. by Rödel, L.; Simon, T., Publisher: [object Object], 2019, DOI: 10.25656/01:17403.

[RZ22] Rau, M. A.; Zahn, M.: Nonverbal Collaboration on Perceptual Learning Activities with Chemistry Visualizations. In (Rodrigo, M. M.; Matsuda, N.; Cristea, A. I.; Dimitrova, V., eds.): Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium. Vol. 13356, Series Title: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 231–235, 2022, DOI: 10.1007/978-3-031-11647-6_41.

[Ša18] Šašinka, Č.; Stachoň, Z.; Sedlák, M.; Chmelík, J.; Herman, L.; Kubíček, P.; Šašinková, A.; Doležal, M.; Tejkl, H.; Urbánek, T.; Svatoňová, H.; Ugwitz, P.; Juřík, V.: Collaborative Immersive Virtual Environments for Education in Geography. en, ISPRS International Journal of Geo-Information 8 (1), p. 3, 2018, DOI: 10.3390/ijgi8010003.

[Se17] Seele, S.; Misztal, S.; Buhler, H.; Herpers, R.; Schild, J.: Here's looking at you anyway! How important is realistic gaze behavior in co-located social virtual reality games? In. Pp. 531–540, 2017, DOI: 10.1145/3116595.3116619.

[Sl17] Slater, M.: Implicit Learning Through Embodiment in Immersive Virtual Reality. In (Liu, D.; Dede, C.; Huang, R.; Richards, J., eds.): Virtual, Augmented, and Mixed Realities in Education. Series Title: Smart Computing and Intelligence, Springer Singapore, Singapore, pp. 19–33, 2017, DOI: 10.1007/978-981-10-5490-7_2.

[SN18] Smith, H. J.; Neff, M.: Communication Behavior in Embodied Virtual Reality. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, Montreal QC Canada, pp. 1–12, 2018, DOI: 10.1145/3173574.3173863.

[SP13] Schneider, B.; Pea, R.: Real-time mutual gaze perception enhances collaborative learning and collaboration quality. en, International Journal of Computer-Supported Collaborative Learning 8 (4), pp. 375–397, 2013, DOI: 10.1007/s11412-013-9181-4.

[SW21] Sun, Y.; Won, A. S.: Despite Appearances: Comparing Emotion Recognition in Abstract and Humanoid Avatars Using Nonverbal Behavior in Social Virtual Reality. Frontiers in Virtual Reality 2, p. 694453, 2021, DOI: 10.3389/frvir.2021.694453.

[SWM21] Schneider, B.; Worsley, M.; Martinez-Maldonado, R.: Gesture and Gaze: Multimodal Data in Dyadic Interactions. In (Cress, U.; Rosé, C.; Wise, A. F.; Oshima, J., eds.): International Handbook of Computer-Supported Collaborative Learning. Springer International Publishing, Cham, pp. 625–641, 2021, DOI: 10.1007/978-3-030-65291-3_34.

[Ta22] Taylor, S.; Hoang, T.; Aranda, G.; Mulvany, G. T.; Greuter, S.: Immersive Collaborative VR Application Design: A Case Study of Agile Virtual Design Over Distance. ng, International Journal of Gaming and Computer-Mediated Simulations 13 (4), Number: 4, pp. 1–14, 2022, DOI: 10.4018/IJGCMS.291538.

[THB20] Tanenbaum, T. J.; Hartoonian, N.; Bryan, J.: "How do I make this thing smile?": An Inventory of Expressive Nonverbal Communication in Commercial Social Virtual Reality Platforms. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. ACM, Honolulu HI USA, pp. 1–13, 2020, DOI: 10.1145/3313831.3376606.

[TYT20] Tenorio Morales, M.; Yamamoto, K.; Tsujino, Y.: SyncMeet: Virtual Work Environment for Collaborative Manga Creation. In. Vol. 12202 LNCS, pp. 518–532, 2020, DOI: 10.1007/978-3-030-49757-6_38.

[WJF22] Wei, X.; Jin, X.; Fan, M.: Communication in Immersive Social Virtual Reality: A Systematic Review of 10 Years' Studies. In: Proceedings of the Tenth International Symposium of Chinese CHI. ACM, Guangzhou, China and Online China, pp. 27–37, 2022, DOI: 10.1145/3565698.3565767.

[Wo14a] Wohlin, C.: Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. ACM, London England United Kingdom, pp. 1–10, 2014, DOI: 10.1145/2601248.2601268.

[Wo14b] Won, A. S.; Bailenson, J. N.; Stathatos, S. C.; Dai, W.: Automatically Detected Nonverbal Behavior Predicts Creativity in Collaborating Dyads. en, Journal of Nonverbal Behavior 38 (3), pp. 389–408, 2014, DOI: 10.1007/s10919-014-0186-0.

[Wu21] Wu, Y.; Wang, Y.; Jung, S.; Hoermann, S.; Lindeman, R. W.: Using a Fully Expressive Avatar to Collaborate in Virtual Reality: Evaluation of Task Performance, Presence, and Attraction. Frontiers in Virtual Reality 2, 2021, DOI: 10.3389/frvir.2021.641296.

[WZ24] Wehrmann, F.; Zender, R.: Inclusive Virtual Reality Learning: Review and 'Best-Fit' Framework for Universal Learning. en, Electronic Journal of e-Learning 22 (3), pp. 74–89, 2024, DOI: 10.1016/j.ridd.2006.09.002.

[ZSC24] Zhou, Q.; Suraworachet, W.; Cukurova, M.: Detecting non-verbal speech and gaze behaviours with multimodal data and computer vision to interpret effective collaborative learning interactions. en, Education and Information Technologies 29 (1), pp. 1071–1098, 2024, DOI: 10.1007/s10639-023-12315-1.

Unterstützung einer komplexen Montageaufgabe durch eine Augmented-Reality-Anleitung mit Durchsichtdatenbrille

Vergleich mit Smartphone- und Papieranleitung

David Mack [1] und Ludger Schmidt[1]

Abstract: Die hier vorgestellte Studie betrachtet die Montage eines Notgriffs zur Türentriegelung von Schienenfahrzeugen an einem originalgetreuen Arbeitsplatz mit unterschiedlichen Anleitungen. Dazu wurde eine AR-Anleitung mit einer Durchsichtdatenbrille neu entwickelt und mit einer Smartphone-basierten AR-Anleitung sowie einer Papieranleitung verglichen. Die Papieranleitung stellt den Status Quo in der Montage dar. Die Umsetzung mit einer Durchsichtdatenbrille ist zwar technisch aufwändiger, hat aber gegenüber dem bereits untersuchten Smartphone den Vorteil, dass die Hände für die Montage frei sind. Die untersuchten Hypothesen sind, dass die Anleitungsart einen Einfluss auf die Montagezeit, die Anzahl der Fehler, die Gebrauchstauglichkeit und die Beanspruchung der montierenden Personen hat. Jede Anleitungsart wurde mit acht Probanden und fünf Montagedurchgängen in einem Between-Subject-Design untersucht. Zwischen Durchsichtdatenbrille und Smartphone zeigte sich bei der Gesamtmontagedauer kein signifikanter Unterschied. Eine vergleichbare Fehleranzahl bei geringerer Beanspruchung macht die Durchsichtdatenbrille dennoch interessant für die Wissensvermittlung.

Keywords: Erweiterte Realität, Head-Mounted Display, Anlernen, Fertigung, Digitalisierung

1 Einleitung

Trotz fortschreitender Digitalisierung in Produktionsbetrieben und vermehrtem Einsatz von Robotern, insbesondere kollaborativen Robotern in der automatisierten Fertigung [Ja23], gibt es nach wie vor Montagearbeitsplätze, die eine rein manuelle Montage vorsehen. Die Gründe dafür sind oft eine zu hohe Variantenvielfalt in Verbindung mit geringen Stückzahlen. Dies hat steigende Anforderungen an die Mitarbeiter zur Folge [Fr98]. Montagearbeitsplätze haben üblicherweise eine hohe Fluktuation und es müssen oft neue Mitarbeiter eingelernt werden [HSS19].

Einen Ansatz zur Verbesserung der Verständlichkeit von Arbeitsanweisungen und Montageanleitungen kann die Verwendung von innovativen Methoden wie Augmented Reality (AR) und Virtual Reality (VR) darstellen [Sc05]. Diese sind in der Industrie noch nicht weit verbreitet, es wird jedoch erwartet, dass die Verbreitung in den nächsten Jahren zunehmen wird [KS22].

Der Begriff Augmented Reality beschreibt die Ergänzung oder Erweiterung der Realität um virtuelle Inhalte [Dö13], die in Echtzeit, interaktiv und räumlich korrekt registriert werden

1 Universität Kassel, Fachgebiet Mensch-Maschine-Systemtechnik, Mönchebergstraße 7, 34125 Kassel, d.mack@uni-kassel.de, https://orcid.org/0009-0005-6557-6224

doi:10.18420/delfi2024_04

[Az97]. Neben der Darstellung von virtuellen Inhalten auf einem transparenten Bildschirm einer Durchsichtdatenbrille (Optical See-Through) [Tö10] können Bildschirme verwendet werden, um ein Kameravideo mit Inhalten zu erweitern (Video See-Through) [HWZ21]. Durch die Kombination mit dem optischen Trackingverfahren des Markertrackings lassen sich einfache AR-Anwendungen erstellen, die beispielsweise mit dem Smartphone oder einem Tablet verwendet werden können. Neben schwarz-weißen Mustern können auch andere Bilder als Marker verwendet werden. Diese werden von der Kamera des Geräts erkannt und genutzt, um virtuelle Inhalte in Bezug zum jeweiligen Marker lagerichtig einzublenden [Tö10]. Markerbasiertes AR ist eine kostengünstige Option für den Einsatz von AR, da lediglich eine Kamera zur Erfassung der Marker und meist kostenfreie Software zum Programmieren benötigt werden [Tö10]. Hinzu kommt die einfache Handhabung von Markern, die auch für Laien leicht zu erlernen ist. Die Marker können schnell erstellt, ausgedruckt und frei platziert werden [Dö13].

AR zur Montageunterstützung ist schon seit längerer Zeit Gegenstand der Forschung [CM92]. Oftmals haben die Veröffentlichungen jedoch starke Limitationen, wenn Montageaufgaben mit Spielzeug-Bausteinen nachgestellt oder ein monitorbasiertes AR mit fixierter Perspektive eingesetzt wird, die nicht der Kopfbewegung folgt [Ya19].

Bezüglich der Unterstützung von AR-Anleitungen mit Durchsichtdatenbrillen gibt es in der Literatur widersprüchliche Ergebnisse. Eine Veröffentlichung, welche die Unterstützung bei der Bedienung einer Maschine untersuchte, kam zu dem Ergebnis, dass die Probanden bei visueller Unterstützung mittels der *HoloLens 1* langsamer arbeiteten und mehr Fehler machten als die Vergleichsgruppen, welche durch eine Papieranleitung, eine visuelle Anleitung auf einem Bildschirm oder eine mündliche Anleitung unterstützt wurden [He19]. Ebenfalls mit einer *HoloLens 1* wurde eine Studie mit einer Anwendung zur Montageunterstützung durchgeführt, die keinen signifikanten Unterschied in der Montagegeschwindigkeit im Vergleich zu einer Papieranleitung ergab [DK18]. Beide Studien bemängeln die Gebrauchstauglichkeit des verwendeten Gerätes, insbesondere das geringe Sichtfeld der *HoloLens 1* wird genannt [DK18]. Andere Studien mit der *HoloLens 1* zeigen hingegen eine Verbesserung der Fehlerrate und der Montagegeschwindigkeit [La19]. In einer Studie, in der die *HoloLens 2* zur Unterstützung der visuellen Qualitätskontrolle eingesetzt wurde, konnte gezeigt werden, dass das verbesserte Sichtfeld in Kombination mit dem deutlich genaueren Tracking im Vergleich zur *HoloLens 1* die Beschleunigung der Qualitätskontrolle durch die AR-Anwendung erklären kann [SCN23]. Ansonsten ist die Zahl der Veröffentlichungen mit der *HoloLens 2*, die explizit die Montage mit realen Bauteilen untersuchen, sehr begrenzt bzw. diese haben andere Schwerpunkte. Eine Veröffentlichung, die die Montage von Spielzeug-Bausteinen untersucht, kommt zu dem Ergebnis, dass das Endergebnis mit der *HoloLens 2* besser als mit einer Papieranleitung ist, aber das Erlernen der einzelnen Schritte reduziert wird [Ge22].

Andere Darstellungsformen von AR zeigen, dass insbesondere bei den ersten Montagedurchgängen eine Zeitersparnis mit einer monitorbasierten AR-Anleitung erreicht werden konnte [Ho15]. Ein Literaturüberblick kommt zu dem Schluss, dass eine verbesserte Darstellung

im dreidimensionalen Raum eine bessere Interaktion und Unterscheidung zwischen real und virtuell ermöglichen könnte [WON16]. Eine weitere Literaturübersicht zeigt, dass die kognitive Beanspruchung bei der Verwendung von Durchsichtdatenbrillen im Vergleich zu anderen Formen von AR, wie z. B. einer Smartphone-basierten Lösung, gleich bleibt oder sogar reduziert wird, während die Leistung steigt [BBK22].

Diese teilweise widersprüchlichen Ergebnisse in Verbindung mit den Hinweisen auf die technischen Grenzen der verwendeten Durchsichtdatenbrillen führen dazu, dass ein realitätsnahes Training in einem komplexen Montageszenario mit einer technisch aktuellen Durchsichtdatenbrille untersucht werden sollte. Zu dem in der Studie verwendeten, kostengünstigen Modell *NREAL Light* mit einem Gewicht von 106 g und einem Blickfeld von 52° sind bislang keine vergleichbaren Studien bekannt, wie eine Recherche mit den Suchbegriffen „NREAl Light“, „Assembly Training“ und „Montagetraining“ bei *Google Scholar* und *Semantic Scholar* zeigte.

Da es widersprüchliche Aussagen in den bisherigen Studien gibt, wurde nachfolgend mit ungerichteten Hypothesen gearbeitet: Die Art der Instruktion beeinflusst die Montagezeit (H1), die Anzahl der Montagefehler (H2), die Gebrauchstauglichkeit (H3) und die Beanspruchung der Probanden (H4). Der Schwerpunkt liegt dabei auf dem Einfluss der Darstellungsart der AR-Anleitung.

In einer vorangegangenen Studie wurde ein Montagearbeitsplatz mithilfe einer Smartphone-gestützten AR-Anleitung untersucht [FS20]. Die AR-Gruppe montierte zu Beginn langsamer als die Papier-Gruppe, war jedoch im fünften Durchgang signifikant schneller. Demgegenüber wurde hier die Darstellungsform der AR-Anleitung von einem Smartphone zu einer Durchsichtdatenbrille geändert. Es wurde bewusst auf den Einsatz einer High-End-Durchsichtdatenbrille verzichtet, die mit Tiefensensoren ein genaueres Tracking ermöglichen würde, um einen kostengünstigen Einstieg mit einer AR- Durchsichtdatenbrille zu ermöglichen. Dementsprechend wurde ein leicht umsetzbares kamerabasiertes Tracking verwendet. Um die Vergleichbarkeit mit den Daten der vorherigen Studie zu gewährleisten, wurde das Szenario unverändert gelassen. Es wird der gleiche Notgriff zur Türentriegelung von Schienenfahrzeugen verwendet, das gleiche Werkzeug sowie die gleichen Montageblöcke.

Im Vergleich zu anderen Studien zeichnet sich diese durch den realen Anwendungsfall aus, da sie an einem originalgetreu nachgebauten Montageplatz mit Originalteilen durchgeführt wurde. Die kostengünstige AR-Durchsichtdatenbrillenumsetzung bietet sich in der Praxis an, um Novizen in der Montage anzuleiten und zu trainieren. Angesichts der neuesten Initiativen der Bundesregierung, ausländische Fachkräfte trotz Sprachbarrieren in Arbeit zu bringen, ist der Bedarf an Anwendungen, die sprachliche Barrieren reduzieren oder sogar ganz vermeiden, aktueller denn je [Cr24].

2 Methoden

2.1 Montageaufgabe

In der vorliegenden Studie wird die Montage eines Notgriffs zur manuellen Entriegelung von Bahntüren betrachtet. Die Baugruppe ist in bis zu 30 verschiedenen Varianten erhältlich und wird derzeit manuell montiert. Sie besteht aus 15 verschiedenen Bauteilen und insgesamt 16 Einzelteilen, die während der Montage zusammengesetzt werden müssen. Für die Montage sind folgende Schritte erforderlich: 'Positionieren', 'Verschrauben', 'Einlegen', 'Einhängen' und 'Verpressen'. In Abb. 1 links ist das vollständige Modell der Baugruppe zu sehen.

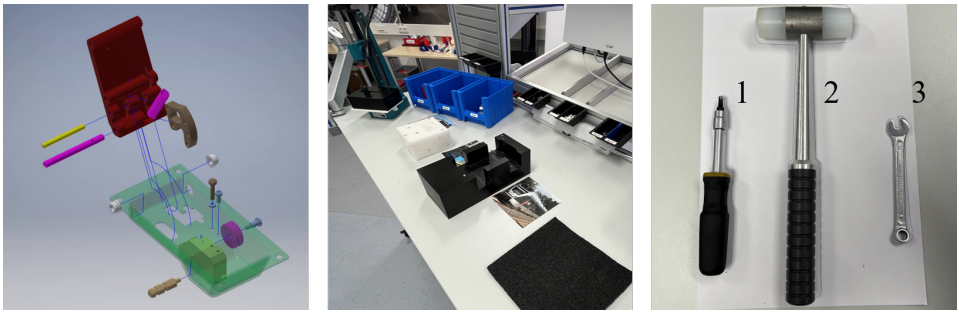


Abb. 1: 3-D-Modell des zu montierenden Notgriffs (links); Montagearbeitsplatz mit Teilelagern, Montagevorrichtungen und einer Kniehebelpresse (mittig); für die Montage benötigtes Werkzeug (rechts): Innensechskant-Schraubendreher SW 4 (1), Kunststoffschonhammer (2), Ring-Maulschlüssel SW 8 (3)

Die Baugruppe wird stückweise montiert, indem sie schrittweise zusammengebaut wird. Erst wenn alle Montagevorgänge einer Baugruppe abgeschlossen sind, wird die nächste Baugruppe montiert [LW12]. Im Praxisunternehmen erfolgt die Montage nach einer Anlernphase mithilfe einer Papieranleitung selbständig. Für die vorliegende Studie wird ein Montagetisch im Labor mit den Maßen 1800 mm x 750 mm verwendet, der dem originalen Vorbild entspricht. Die Montage erfolgt mit Hilfe von Montagevorrichtungen, einer Kniehebelpresse und handgeführtem mechanischem Werkzeug, wie in Abb. 1 dargestellt.

2.2 AR-Montageanleitung per Smartphone

In der vorherigen Studie wurde eine markerbasierte AR-Montageanleitung für den Notgriff als Video-See-Through-Smartphone-Anwendung genutzt. Hierzu diente eine bestehende Papieranleitung als Grundlage, welche auf die AR-Anwendung übertragen wurde. Der Montageprozess wurde in sieben Teilschritte unterteilt und mit je einem Marker versehen. Diese Marker werden von der Anwendung erkannt und verwendet, um passende Informationen lagerichtig anzuzeigen. In jedem der sieben Teilschritte werden dem Mitarbeiter 3-D-Modelle

der benötigten Bauteile sowie ein Textfeld mit den dazugehörigen Positionsnummern der Teilelager angezeigt. Gegebenenfalls werden weitere Informationen über den auszuführenden Arbeitsschritt bereitgestellt. Die Teilschritte bestehen teilweise aus mehreren aufeinander aufbauenden Montageschritten. In die Anwendung wurden zwei Schaltflächen integriert, um zwischen den angezeigten Informationen vor- und zurückzuschalten. In Abb. 2 ist die Darstellung auf dem Smartphone zu sehen.

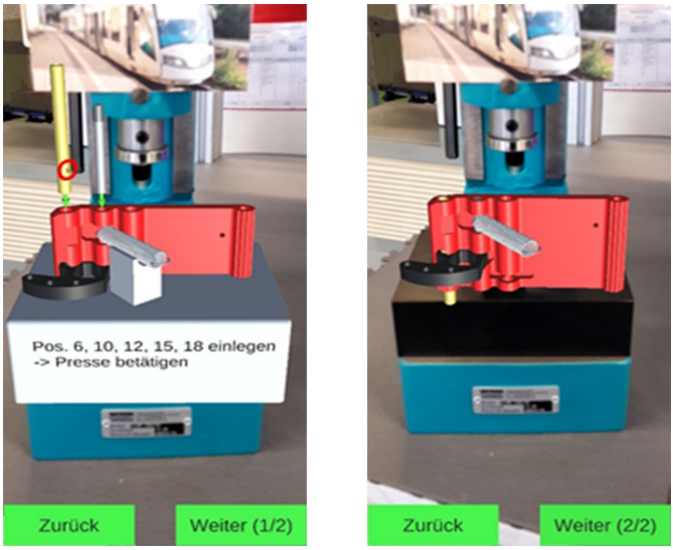


Abb. 2: Smartphone-AR-Anleitung für einen Montageschritt mit Angabe der Teilpositionsnummern sowie der Arbeitsaufforderung „Presse betätigen“, Schaltflächen „Zurück“ und „Weiter“ zum selbständigen Durchblättern von Informationen (links) sowie Anzeige der korrekt montierten Teile (rechts)

Die AR-Anwendung wurde mithilfe der 3-D-Entwicklungsumgebung *Unity* (Version 2019.1.12f1) und dem *Vuforia Kit* für markerbasiertes AR erstellt. In *Unity* wurden die 3 D-Modelle der Baugruppe mit den Markern der entsprechenden Teilschritte verknüpft und teilweise mit Text für die Positionsnummer und weiteren Hinweisen ergänzt.

2.3 AR-Montageanleitung per Durchsichtdatenbrille

Die Montageanleitung mittels Durchsichtdatenbrille wurde mit der Smartphone-Anwendung inhaltlich vergleichbar ausgestaltet. Der Montageprozess wurde in 21 Teilschritte unterteilt, wie es auch in der ursprünglichen Papieranleitung der Fall war. Allerdings ist die Durchsichtdatenbrille im Gegensatz zum Smartphone nicht auf einzelne Marker angewiesen; sie orientiert sich ebenfalls an Markern, diese dienen jedoch der Positionierung des gesamten Overlays.

Die Umsetzung wurde mithilfe der 3-D-Entwicklungsumgebung *Unity* (Version 2021.3.0f1) und dem *NRSDK Unity Package* (Version 1.10.2) erstellt. Im Gegensatz zum Smartphone ermöglicht die Durchsichtdatenbrille das „Wegweisen“ mittels leuchtender Kugeln, ähnlich einem Pick-by-light-System, an den einzelnen Vorratskisten. Wenn eine Vorratskiste nicht im Blickfeld ist, wird ein Richtungspfeil eingeblendet, der eine schnellere Orientierung der Probanden ermöglicht. Die Datenbrille wird mittels eingebautem Handtracking bedient. Die Darstellung auf der Durchsichtdatenbrille ist in Abb. 3 zu sehen.

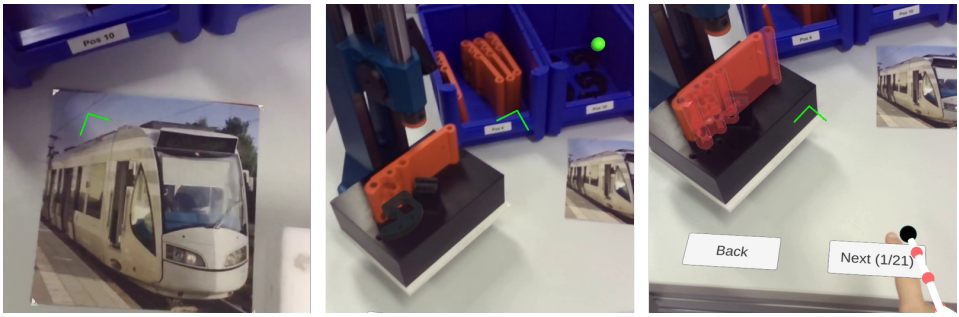


Abb. 3: Blick durch die Datenbrille mit Markierung des Markers zur Rückmeldung, dass die Anwendung kalibriert ist (weiße Ecken am Straßenbahnfoto) (links), Pfeil zur Position des nächsten Bauteils (grüner Richtungspfeil zum grünen Punkt) (mittig), Handtracking und leicht verschobenes Bauteil (rechts)

2.4 Messung

Zu Beginn der Studie wurde die Technikaffinität der Probanden mittels des TA-EG-Fragebogens gemessen. Der Fragebogen besteht aus 19 Items auf einer fünfgliedrigen Likert-Skala von 1 (trifft gar nicht zu) bis 5 (trifft voll zu) [Ka09]. Die negativen Items werden negiert und die Mittelwerte der vier Skalen „Begeisterung“, „Kompetenz“, „negative Einstellung“ und „positive Einstellung“ gebildet. Anschließend wird aus diesen Skalen ein Gesamtmittelwert ermittelt. Die Werte des TA-EG-Fragebogens liegen zwischen 1 (sehr niedrige Technikaffinität) und 5 (sehr hohe Technikaffinität).

Die Montagezeiten und die Anzahl der Montagefehler (aufgeteilt in „falsches Bauteil“ und „Bauteil falsch montiert“) wurden mittels Videoaufzeichnung der Probanden und der internen Kameras der Durchsichtdatenbrille ermittelt.

Zur Ermittlung der Gebrauchstauglichkeit der Anwendung wurde der System-Usability-Scale-Fragebogen (SUS) verwendet [Br95]. Der Fragebogen besteht aus zehn Items, die auf einer fünfstufigen Likert-Skala von 1 („Ich stimme gar nicht zu.“) bis 5 („Ich stimme voll zu.“) bewertet werden. Die Hälfte der Items ist positiv und die andere Hälfte ist negativ formuliert. Zur Auswertung werden die negativ formulierten Items negiert und im Anschluss alle Werte in eine Skala von null bis vier umgerechnet. Die Summe der Items wird mit

2,5 multipliziert [Br95]. Das Ergebnis liegt auf einer Skala von 0 bis 100, wobei 0 für eine sehr schlechte und 100 für eine sehr gute Gebrauchstauglichkeit steht.

Die Beanspruchung der Probanden wurde mithilfe des NASA-TLX-Fragebogens ermittelt [HM88]. Der Fragebogen besteht aus der Bewertung einer 20-stufigen Skala von „gering“ bis „hoch“ bei sechs verschiedenen Subskalen: „geistige Anforderungen“, „körperliche Anforderungen“, „zeitliche Anforderungen“, „Leistung“, „Anstrengung“ und „Frustration“. Anschließend wird in mehreren Paarvergleichen angegeben, welcher Punkt für die Arbeitsaufgabe wichtiger war. Abschließend wird der Gesamtwert zwischen 0 (keine Beanspruchung) und 100 (extreme Beanspruchung) gewichtet.

2.5 Versuchsaufbau

Die Durchsichtdatenbrille *NREAL Light* erlaubt eine Visualisierung mit sechs Freiheitsgraden. Die Brille hat pro Auge eine Displayauflösung von 1080p, eine Helligkeit von 1000 Nits und eine Sichtfeldweite von 52°. Zudem verfügt sie über drei Kameras: zwei Graustufenkameras mit einer Auflösung von 640p x 480p und einem Field of View (FoV) von 120° sowie eine RGB-Kamera mit einer Auflösung von 2592p x 1944p und einem FoV von 74°. Die Brille hat einen kompakten Formfaktor und erinnert an eine etwas größere Sonnenbrille (siehe Abb. 4).

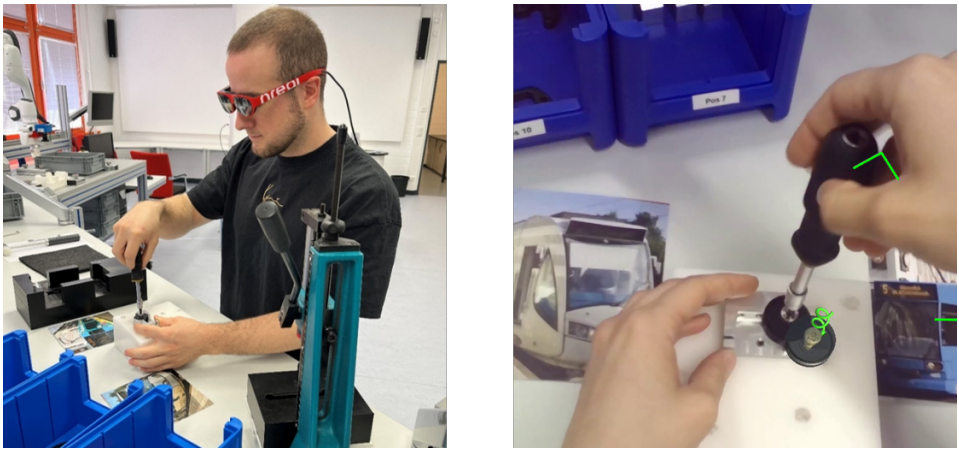


Abb. 4: Proband während der Montage des Notgriffes mit der Durchsichtdatenbrille

Die kompakte Form wird durch das Auslagern des Akkus und des Prozessors in ein separates Gehäuse ermöglicht. Das Gehäuse ist über ein Kabel mit der Kamera verbunden und verfügt über einen Touch-Controller, der aufgrund des verwendeten Hand-Trackings jedoch nicht benötigt wird.

2.6 Versuchsablauf

Zu Beginn des Versuchs, der einem Between-Subject-Design folgte, wurden die Probanden über den Inhalt und den Ablauf der Studie aufgeklärt und sie unterzeichneten eine Einverständniserklärung inklusive Datenschutzerklärung. Die Studie begann mit dem Ausfüllen eines demografischen Fragebogens, der Angaben zum Alter, Geschlecht, zur Sehhilfe und Vorerfahrung mit AR-Anwendungen enthielt. Anschließend erhielten die Teilnehmer eine Einführung in die Nutzung der Durchsichtdatenbrille mittels Handgesten und das Scannen eines Markers zur Positionierung der Anwendung. Nach Klärung eventueller offener Fragen begann die Montageaufgabe, bei der fünf Notgriffe am gleichen Versuchsaufbau wie in der vorherigen Studie montiert werden mussten.

Die fünf Montagedurchläufe erfolgten direkt hintereinander. Der Versuchsleiter beobachtete die Montage und gab mittels eines Schildes mit der Aufschrift „falsches Bauteil verwendet“ oder „Bauteil falsch montiert“ entsprechende Rückmeldung.

Nach den fünf Montagedurchgängen füllten die Probanden die Fragebögen TA-EG, SUS und NASA-TLX aus und hatten die Möglichkeit, Feedback zur verwendeten Anleitung zu geben.

2.7 Stichprobe

Die Durchsichtdatenbrille wurde von acht Probanden genutzt, die keine Vorerfahrung in der manuellen Montage einer komplexen Baugruppe hatten. Im Gegensatz zur vorherigen Stichprobe befinden sie sich nicht in einer technischen Ausbildung. Es nahmen zwei Frauen und sechs Männer teil, deren Durchschnittsalter 28,1 Jahre (SD = 3,2 Jahre) betrug. Die Smartphone-Nutzer aus der vorherigen Studie (8 Männer) waren im Durchschnitt 19,3 Jahre (SD = 1,7 Jahre) alt. Ein zweiseitiger ungepaarter t-Test zwischen den beiden Gruppen, geprüft auf Normalverteilung mit dem Shapiro-Wilk-Test und auf Varianzhomogenität mit dem Levene-Test, zeigte einen signifikanten Unterschied im Alter der Probanden ($t(14{,}00) = 5{,}741$; $p < 0{,}001$). Es wurden keine signifikanten Gruppenunterschiede in den durchschnittlichen TA-EG-Werten zwischen den Datenbrillen- ($M = 3{,}79$, $SD = 0{,}22$) und Smartphone-Nutzern ($M = 3{,}86$, $SD = 0{,}45$) festgestellt ($U = 22{,}00$, $Z = -1{,}05$, $p = 0{,}328$). Da diese Daten nicht normalverteilt sind, wurde der Mann-Whitney-U-Test durchgeführt.

3 Ergebnisse

Vor der Auswertung wurden alle Daten auf Normalverteilung mittels des Shapiro-Wilk-Tests und auf Varianzhomogenität mittels des Levene-Tests geprüft. Entweder wurde ein t-Test bzw. eine Varianzanalyse oder ein Welch-Test für unabhängige Stichproben durchgeführt. Die Montagezeiten der fünf Durchgänge sind in Abb. 5 dargestellt. Bei Hypothese H1 werden die Unterschiede in den Montagezeiten zwischen der Durchsichtdatenbrillen- und der Smartphone-Gruppe untersucht. Um die Übersichtlichkeit zu erhöhen, wurde hier auf eine Darstellung der Werte für die Papieranleitung verzichtet. Bezogen auf den ersten Durchgang ($t(14{,}00) = 3{,}755$; $p = 0{,}002$) und die Gesamtdauer ($t(14{,}00) = 2{,}770$; $p = 0{,}015$) war diese mit durchschnittlich 8,53 min (SD = 1,31 min) bzw. 24,98 min (SD = 3,43 min) gegenüber der Durchsichtdatenbrille signifikant schneller.

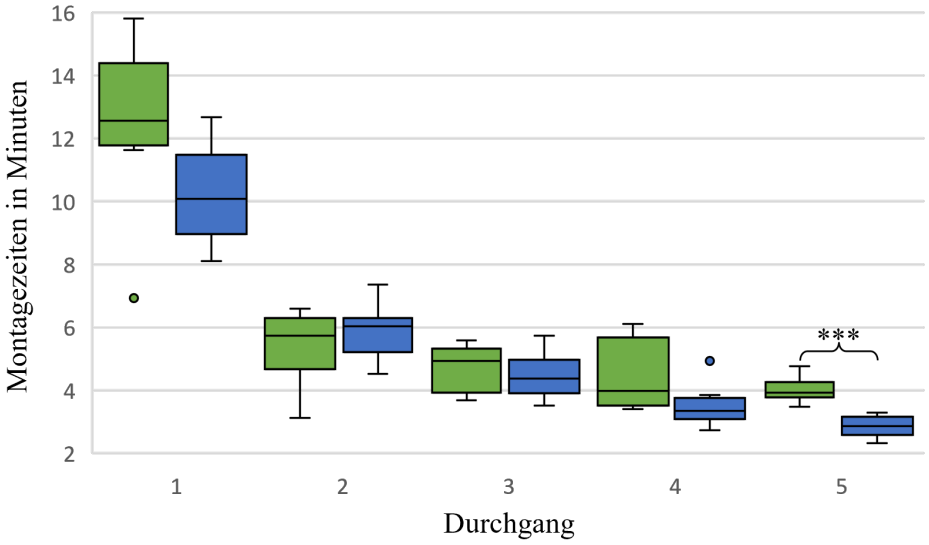


Abb. 5: Boxplots der Montagezeiten der Durchsichtdatenbrillen- (grün) und Smartphone-Gruppe (blau) [FS20] gruppiert nach den Montagedurchgängen

In Tab. 1 sind die Mittelwerte und Standardabweichungen dargestellt sowie die Ergebnisse des t-Tests bzw. des Welch-Test, da in Montage 4 keine Varianzgleichheit gegeben ist.

	Durchsicht-datenbrille	Smartphone [FS20]	t-test bzw. Welch-Test
	M [min] (SD [min])		
Durchgang 1	12,44 (2,63)	10,16 (1,52)	t(14,00) = 2,12; p = 0,053
Durchgang 2	5,37 (1,14)	5,89 (0,85)	t(12,99) = -1,05; p = 0,312
Durchgang 3	4,71 (0,74)	4,46 (0,70)	t(14,00) = 0,68; p = 0,507
Durchgang 4	4,45 (1,10)	3,50 (0,67)	t(11,52) = 2,09; p = 0,060
Durchgang 5	4,02 (0,39)	2,85 (0,34)	t(14,00) = 6,42; p < 0,001
Gesamtdauer	30,98 (5,09)	26,88 (3,06)	t(14,00) = 1,957; p = 0,071

Tab. 1: Mittelwerte und Standardabweichung der Montagezeiten der Durchsichtdatenbrillen- und der Smartphone-Gruppe über fünf Durchgänge sowie die Ergebnisse der t-Tests

Bezüglich Hypothese H2 wird überprüft, ob sich die Anzahl der Fehler je nach verwendeter Anleitungsart unterscheidet. Zur Zählung der Fehler pro Proband und Montagedurchgang wurden die Videoaufzeichnungen ausgewertet. Die Durchsichtdatenbrillen-Gruppe machte im Durchschnitt 3,38 Fehler (SD = 1,06 Fehler) und die Smartphone-Gruppe im Durchschnitt 3,13 Fehler (SD = 1,13 Fehler) [FS20]. Der t-Test zeigte keine signifikanten Unterschiede (t(14,00) = 0,457; p = 0,655). In Abb. 6 sind die durchschnittlichen Montagefehler der fünf Durchgänge zu sehen.

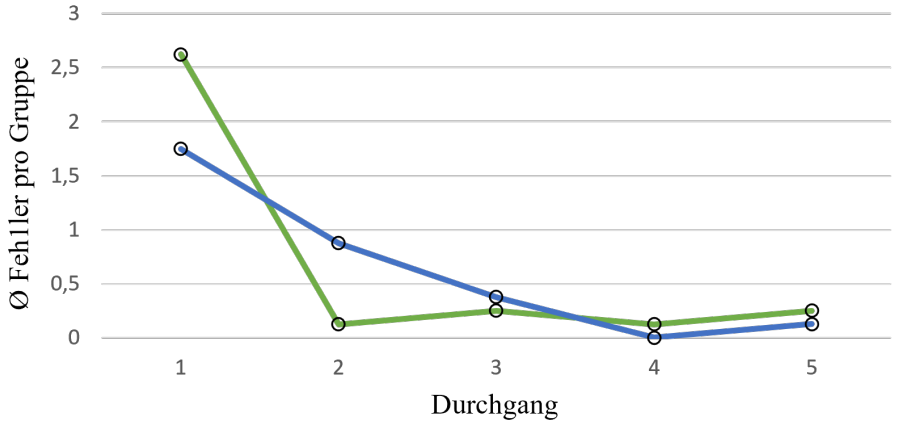


Abb. 6: Durchschnittliche Montagefehler der Durchsichtdatenbrillen- (grün) und der Smartphone-Gruppe (blau) [FS20] über die Montagedurchgänge

Bei den Hypothesen H3 (Gebrauchstauglichkeit) und H4 (Beanspruchung) wird überprüft, inwieweit sich die Anleitungsart Durchsichtdatenbrille, Smartphone und Papier auswirkt. Die Ergebnisse sind in Abb. 7 zu sehen.

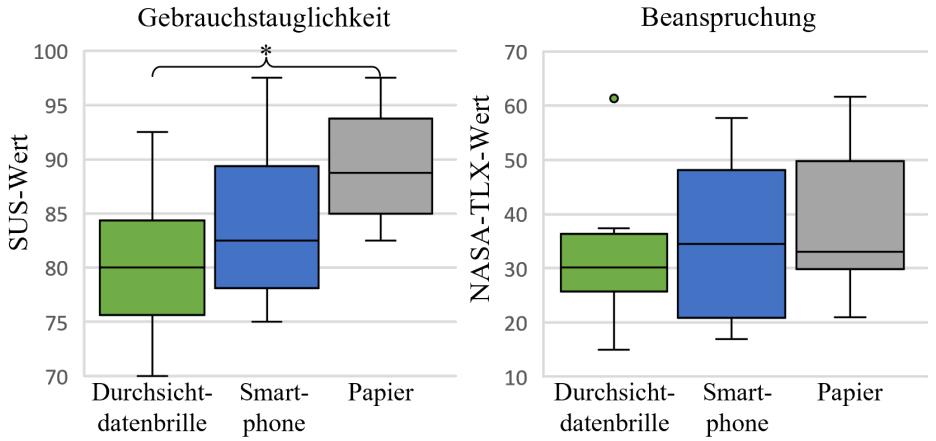


Abb. 7: Gebrauchstauglichkeit (SUS) (links) und Beanspruchung (NASA-TLX) (rechts)

Es wurde mittels Varianzanalyse untersucht, ob es Unterschiede in der Gebrauchstauglichkeit und Beanspruchung zwischen den drei Gruppen gibt. Die Daten waren für jede Gruppe normalverteilt (Shapiro-Wilk-Test, $p > 0{,}05$) und die Varianzhomogenität wurde durch den Levene-Test bestätigt ($p > 0{,}05$). Die durchschnittliche Gebrauchstauglichkeit, die mit dem SUS erhoben wurde, zeigt bei der Durchsichtdatenbrillen-Gruppe ($M = 80{,}31$, $SD = 6{,}87$), der Smartphone-Gruppe ($M = 84{,}06$, $SD = 7{,}31$) [FS20] und der Papier-Gruppe ($M = 89{,}06$, $SD = 5{,}16$) [FS20] statistisch signifikante Unterschiede ($F(2{,}21) = 3{,}63$, $p = 0{,}044$). Der Tukey-Post-Hoc-Test ergab einen signifikanten Unterschied in den SUS-Werten zwischen Durchsichtdatenbrille und Papier. Die durchschnittliche Beanspruchung, die mit dem NASA-TLX erhoben wurde, betrug für die Gruppe, die die Durchsichtdatenbrille verwendete, $M = 32{,}62$, $SD = 13{,}35$, für die Gruppe, die das Smartphone verwendete, $M = 34{,}70$, $SD = 14{,}53$, [FS20] und für die Gruppe, die Papier verwendete, $M = 37{,}37$, $SD = 13{,}61$ [FS20]. Diese Unterschiede in der Beanspruchung bei den verschiedenen Anleitungsarten sind statistisch nicht signifikant ($F(2{,}21) = 0{,}237$, $p = 0{,}791$).

4 Diskussion und Fazit

Beim Vergleich der Werte der AR-Anleitungen zeigt sich die Tendenz, dass die Gruppe mit Durchsichtdatenbrille langsamer ist als die Smartphone-Gruppe. Eine mögliche Erklärung könnte sein, dass es – anders als beispielsweise bei Hořejší – keine Unterstützung in Form von Zusatzinformation gab [Ho15]. Es ist anzunehmen, dass der Zeitvorteil der Smartphone-Gruppe einerseits auf der Vorerfahrung im Umgang mit einem alltäglichen Gerät beruht und andererseits darauf zurückzuführen ist, dass die Verwendung einer Durchsichtdatenbrille noch eher ungewohnt ist und der Reifegrad der Technologieumsetzung noch nicht so hoch ist, was sich auch in der Gebrauchstauglichkeit zeigt. Die Positionsgenauigkeit könnte

hinsichtlich der Geschwindigkeit eine wichtige Rolle gespielt haben. Zudem ist als Limitation zu berücksichtigen, dass die Stichprobe der Durchsichtdatenbrillen-Anleitung nicht über die gleichen Vorerfahrungen in der manuellen Montage von Baugruppen verfügt wie Industriemechaniker in der Ausbildung.

Die Ergebnisse erlauben keine eindeutigen Aussagen bezüglich der Fehler. Die Gruppe mit Durchsichtdatenbrille machte im ersten Durchgang mehr Fehler, jedoch in den anderen Durchgängen gleich viele oder sogar weniger Fehler. Die durchschnittliche Gesamtzahl an Fehlern war vergleichbar. Man könnte vermuten, dass die Durchsichtdatenbrille nach einer längeren Eingewöhnungsphase eine bessere Kontrollfunktion bietet.

Die Gebrauchstauglichkeit beider AR-Anwendungen wurde mit „gut“, die Papierunterstützung mit „exzellent“ bewertet [BKM09]. Die signifikant schlechtere Gebrauchstauglichkeit der Durchsichtdatenbrille im Vergleich zur Papieranleitung deutet darauf hin, dass die Durchsichtdatenbrillen-AR-Anwendung noch nicht das Reifegradniveau einer Papieranleitung erreicht hat. Die durchschnittliche subjektive Beanspruchung mit Durchsichtdatenbrille liegt jedoch mit 32,62 unter den Bewertungen der Smartphone-Gruppe mit 34,70 und der Papieranleitung mit 37,37, auch wenn keine signifikanten Unterschiede zwischen den Gruppen festgestellt wurden. Zur Einordnung der Beanspruchungswerte wurden Ergebnisse anderer Studien herangezogen: In einer Meta-Analyse wurden die NASA-TLX-Werte von 237 wissenschaftlichen Studien mit insgesamt 1173 Datensätzen ausgewertet. Der Median aller Werte beträgt 49,93. 25% der Werte liegen unter 36,77 [Gr15]. Somit lassen sich die vorliegenden Werte einer eher geringen Beanspruchung zuordnen.

Zusammenfassend lässt sich sagen, dass es bei der Verwendung von AR-Montageanleitungen eine Tendenz zu schlechterer Gebrauchstauglichkeit gibt (möglicherweise bedingt durch den technischen Reifegrad der Hardware und Software) sowie zu einer niedrigeren Beanspruchung; die niedrigere Beanspruchung deckt sich mit [BBK22]. Dies zeigt das Potential von AR-Durchsichtdatenbrillen-Anleitungen auf, jedoch reicht der Status-Quo noch nicht aus, um herkömmliche Montageanleitungen vollständig zu ersetzen. Gerade im Hinblick auf den Bedarf an sprachbarrierefreien Trainingsmethoden für anspruchsvolle Montagetätigkeiten, die mit einer Papieranleitung schwieriger umzusetzen sind, sollte mit zunehmender Marktdurchdringung von Datenbrillen und deren steigender Verbreitung an Frameworks zur einfachen Umsetzung von Montagetrainings weiterhin gearbeitet werden.

Die immer günstiger werdende Technik von Durchsichtdatenbrillen, gekoppelt mit immer leistungsfähigeren Tracking-Methoden, könnte bei fortschreitendem Reifegrad eine sinnvolle Alternative in der Montage darstellen. Die nächsten Schritte sind – mithilfe leistungsfähigerer Hardware – das Tracking und die Positionierung zu präzisieren und den Reifegrad von AR-Durchsichtdatenbrillen-Anleitungen zu verbessern. Eine Möglichkeit wäre, das Tracking direkt an den Montageblöcken und Bauteilen mittels Bilderkennung zu orientieren. Dies würde ein präziseres Tracking ermöglichen und wäre nicht auf vorab definierte Montagevorrichtungen und -schritte angewiesen. Gerade im Hinblick auf die Implementierung im Realbetrieb wäre eine solche Umsetzung wünschenswert.

Literaturverzeichnis

[Az97] Azuma, R. T.: A Survey of Augmented Reality. Presence: Teleoperators and Virtual Environments 6 (4), S. 355–385, 1997.

[BBK22] Buchner, J.; Buntins, K.; Kerres, M.: The impact of augmented reality on cognitive load and performance: A systematic review. Journal of Computer Assisted Learning 38 (1), S. 285–303, 2022.

[BKM09] Bangor, A.; Kortum, P.; Miller, J.: Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale: Adding an Adjective Rating Scale. Journal of Usability Studies 4 (3), S. 114–123, 2009.

[Br95] Brooke, J.: SUS: A quick and dirty usability scale. Usability Eval. Ind. 189, 1995.

[CM92] Caudell, T. P.; Mizell, D. W.: Augmented reality: an application of heads-up display technology to manual manufacturing processes. In: Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences. IEEE, 659–669 vol.2, 1992.

[Cr24] Creutzburg, D.: Der „Job-Turbo“ läuft noch nicht auf Hochtouren, hrsg. von Braunberger, G.; Kaube, J.; Knop, C.; Kohler, B., 2024, URL: https://www.faz.net/aktuell/wirtschaft/fluechtlinge-und-buergergeld-job-turbo-laeuft-nicht-auf-hochtouren-19620117.html, Stand: 30. 03. 2024.

[DK18] Deshpande, A.; Kim, I.: The effects of augmented reality on improving spatial problem solving for object assembly. Advanced Engineering Informatics 38, S. 760–775, 2018.

[Dö13] Dörner, R.; Broll, W.; Grimm, P.; Jung, B., Hrsg.: Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität. Springer Vieweg, Berlin und Heidelberg, 2013.

[Fr98] Franke, H.-J.: Variantenvielfalt und resultierende Komplexität, Ursachen und Methoden zu ihrer Bewältigung. 1998.

[FS20] Funk, J.; Schmidt, L.: Evaluation einer Augmented-Reality-Anleitung für eine komplexe Montageaufgabe: Vergleich einer Smartphone-basierten Augmented-Reality-Anleitung mit einer klassischen Papieranleitung für die Anlernphase in der manuellen Montage. In (Zender, R.; Ifenthaler, D.; Leonhardt, T.; Schumacher, C., Hrsg.): DELFI 2020. GI-Edition - Lecture Notes in Informatics Proceedings, Gesellschaft für Informatik, Bonn, S. 25–35, 2020.

[Ge22] Generosi, A.; Agostinelli, T.; Mengoni, M.; Ceccacci, S.: Augmented Reality for assembly operation training: does immersion affect the recall performance? In: 2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE). IEEE, S. 58–63, 2022.

[Gr15] Grier, R. A.: How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 59 (1), S. 1727–1731, 2015.

[He19] He, S.; Jia, Y.; Sun, Z.; Yu, C.; Yi, X.; Shi, Y.; Xu, Y.: AR Assistive System in Domestic Environment Using HMDs: Comparing Visual and Aural Instructions. In (Chen, J. Y.; Fragomeni, G., Hrsg.): Virtual, Augmented and Mixed Reality. Multimodal Interaction. Bd. 11574, Lecture Notes in Computer Science, Springer International Publishing, Cham, S. 71–83, 2019.

[HM88] Hancock, P. A.; Meshkati, N., Hrsg.: Advances in Psychology : Human Mental Workload. North-Holland, 1988.

[Ho15] Hořejší, P.: Augmented Reality System for Virtual Training of Parts Assembly. Procedia Engineering 100, S. 699–706, 2015.

[HSS19] Hammermann, A.; Schmidt, J.; Stettes, O.: Fluktuation auf dem deutschen Arbeitsmarkt, 2019.

[HWZ21] Haoming, L.; Wanting, H.; Zhanping, F.: Overview of the development of video see through technology in mixed reality technology and simulation application. In: 2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT). IEEE, S. 609–613, 2021.

[Ja23] Janson, M.: Mensch und Roboter arbeiten zunehmend Hand in Hand, 2023, URL: https://de.statista.com/infografik/24129/absatz-von-konventionellen-und-kollaborativen-industrierobotern-weltweit, Stand: 01. 04. 2024.

[Ka09] Karrer, K.; Glaser, C.; Clemens, C.; Lichtenstein, A.; Stößel, C.: Technikaffinität erfassen – der Fragebogen TA-EGDer Mensch im Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme. VDI Verlag GmbH, Düsseldorf, 2009.

[KS22] Klöß, S.; Streim, A.: Unternehmen nutzen VR oder AR vor allem für die Weiterbildung, 2022, URL: https://www.bitkom.org/Presse/Presseinformation/Unternehmen-nutzen-VR-AR-Weiterbildung, Stand: 20. 06. 2024.

[La19] Lampen, E.; Teuber, J.; Gaisbauer, F.; Bär, T.; Pfeiffer, T.; Wachsmuth, S.: Combining Simulation and Augmented Reality Methods for Enhanced Worker Assistance in Manual Assembly. Procedia CIRP 81, S. 588–593, 2019.

[LW12] Lotter, B.; Wiendahl, H.-P.: Montage in der industriellen Produktion. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[Sc05] Schmidt, L.; Wiedenmaier, S.; Oehme, O.; Luczak, H.: Benutzerzentrierte Gestaltung von Augmented Reality in der Produktion. In (Stary, C., Hrsg.): Mensch & Computer 2005. Oldenbourg, München, 2005.

[SCN23] Seeliger, A.; Cheng, L.; Netland, T.: Augmented reality for industrial quality inspection: An experiment assessing task performance and human factors. Computers in Industry 151, S. 103985, 2023.

[Tö10] Tönnis, M.: Augmented Reality: Einblicke in die Erweiterte Realität. Springer, Berlin und Heidelberg, 2010.

[WON16] Wang, X.; Ong, S. K.; Nee, A. Y. C.: A comprehensive survey of augmented reality assembly research. Advances in Manufacturing 4 (1), S. 1–22, 2016.

[Ya19] Yang, Z.; Shi, J.; Jiang, W.; Sui, Y.; Wu, Y.; Ma, S.; Kang, C.; Li, H.: Influences of Augmented Reality Assistance on Performance and Cognitive Loads in Different Stages of Assembly Task. Frontiers in psychology 10, S. 1703, 2019.

Analyzing Chat Protocols of Novice Programmers Solving Introductory Programming Tasks with ChatGPT

Andreas Scholl [1], Daniel Schiffner [2], and Natalie Kiesler [3]



Abstract: Large Language Models (LLMs) have taken the world by storm, and students are assumed to use related tools at a great scale. In this research paper we aim to gain an understanding of how introductory programming students chat with LLMs and related tools, e.g., ChatGPT-3.5. To address this goal, computing students at a large German university were motivated to solve programming exercises with the assistance of ChatGPT as part of their weekly introductory course exercises. Then students (n=213) submitted their chat protocols (with 2335 prompts in sum) as data basis for this analysis. The data was analyzed w.r.t. the prompts, frequencies, the chats' progress, contents, and other use pattern, which revealed a great variety of interactions, both potentially supportive and concerning. Learning about students' interactions with ChatGPT will help inform and align teaching practices and instructions for future introductory programming courses in higher education.

Keywords: ChatGPT-3.5, large language models, students, interaction pattern, application, chat protocols, introductory programming, higher education

1 Introduction

The increasing need for competent computing graduates proficient in programming, software development, and related technical competencies [Ca17] is one of the factors exacerbating pressure on higher education institutions to offer high quality, competency-based education [Ra21]. However, the latter requires extensive resources, mentoring, and, for example, formative feedback for learners, especially in introductory programming classes [Je22; Lo24]. This is due to the fact that novices experience a number of challenges in the process, which have been subject to extensive research in the past decades [Du86; Lu18; SS86]. Among them are cognitively demanding competencies [Ki20; Ki24], such as problem understanding, designing and writing algorithms, debugging, and understanding error messages [Du86; ER16; Ki20; Lu18; SS86]). Educators' expectations towards novice learners and what they can achieve in their first semester(s) seem to be too high and unrealistic [Lu16; Lu18; WCL07]. Moreover, the student-educator ratio in introductory programming classes keeps increasing in German higher education institutions, thereby limiting resources to provide feedback and hints, and adequately address heterogeneous prior knowledge and diverse educational biographies [Pe16; SB22].

1 Nuremberg Tech, Computer Science, Keßlerplatz 12, 90489 Nuremberg, Germany, andreas.scholl@th-nuernberg.de, https://orcid.org/0009-0009-1275-1374
2 DIPF | Leibniz Institute for Research and Information in Education, Information Centre Education, Rostocker Straße 6, 60323 Frankfurt am Main, Germany, d.schiffner@dipf.de, https://orcid.org/0000-0002-0794-0359
3 Nuremberg Tech, Computer Science, Keßlerplatz 12, 90489 Nuremberg, Germany, natalie.kiesler@th-nuernberg.de, https://orcid.org/0000-0002-6843-2729

doi:10.18420/delfi2024_05

Fortunately, educational technologies based on Large Language Models (LLMs) have the potential to support novice learners of programming seeking help, if carefully used. LLMs and related tools have been shown to pass introductory programming tasks and courses [Ge23; KS23a; Sa23], enhance programming error messages [Le23a; Le23b; Ma23; Sa22], and generate formative feedback [ADS23; AKS24; BK23; KLK24; RKJ23] for learners. It is therefore not surprising that the emergence of LLMs triggered an extensive discussion within the computing education community regarding learning objectives, curricula, assessments and ethical questions [Pr23b]. Due to the novelty of LLMs and their recent broad availability, most studies investigated LLMs from the educator's perspective, hypothesizing about (students') application scenarios (e.g., [Al24; Jo24]). The student perspective (e.g., their trust or attitude towards LLMs [Am24; RHG24]), let alone their actual use of LLMs in a classroom setting, has only been addressed in a few very recent studies (e.g., [Li24b]). Understanding students' use of LLMs and respective interactions, however, is crucial to support them and adequately instruct them regarding the critical and reflective application of LLMs in their studies.

Therefore, the present work investigates how students in an introductory programming class actually work with LLMs, such as ChatGPT, when doing their coursework. Specifically, it is guided by the research question (RQ): *How do students chat with ChatGPT in the context of introductory programming course assignments?* To answer this question, an exercise sheet was developed as part of an introductory programming course at Goethe University in the winter term 2023/24. Students were instructed to solve programming tasks with the assistance of ChatGPT-3.5 (i.e., the freely available version at the time, December 2023), and to submit their chat protocols, which is the data basis of this work. The quantitative and qualitative analysis of the chat protocols provides the first insights of its kind into students' authentic interactions with the LLM in a curricular course setting. The contribution of this paper is thus a collection of students' interaction with an LLM, such as ChatGPT. These insights have implications for educators considering the use of LLMs in their introductory programming courses, as they can help inform the development of new forms of instructions with the goal of fostering a conscious and critical use.

2 Related Work

In 2019, a literature review of 146 articles on artificial intelligence applications in higher education [Za19] pointed out the rapidly increasing relevance of Artificial Intelligence (AI) in computing education and related fields. The review identified four main areas of interest: "adaptive systems and personalization, assessment and evaluation, profiling and prediction and intelligent tutoring systems" [Za19, p. 11]. Notably, Zawacki-Richter et al. recognized "the dramatic lack of critical reflection" [Za19, p. 21] of the challenges and risks of AI tools, including pedagogical and ethical considerations.

Five years later, in 2024, Generative AI (GenAI) and related tools have already taken the world by storm [Pr23b; Pr24], and implications for higher education have become the focus

of the discourse in many disciplines, including computing education research [Ma24]. This development is due to the broad availability of GenAI tools, which gained popularity with the launch of OpenAI's ChatGPT in late November 2022. Its impressive performance in, for example, introductory programming tasks and exams [Ge23; KS23a; Sa23] initiated discussions on LLMs' potential and how it would affect computing curricula, learning objectives, but also teaching, learning, and assessing in general [Be23; Ki23; Pr23a; Pr23b]. Prather et al. further emphasize the fast pace of GenAI tools and their developments, indicating the need for ongoing research to keep up with recent technological advances. This is especially true if research data (e.g., for benchmarking these tools) are not available [KS23b; Pr23b].

General concerns and limitations of LLMs and GenAI tools comprise the accuracy and reliability in educational settings, as ChatGPT is susceptible to (re-)producing biased or inaccurate information [Gi24]. The tool's knowledge base, which may lack recent information, can further lead to inaccuracies. Hallucinations add to this problem. Finally, it can bypass plagiarism detection tools. All of these aspects challenge the integrity of academic work and the reliability of AI as an educational tool, indicating the need for action by educational institutions [Gi24; Pr23b; Zh22a].

In the context of introductory programming education, studies revealed the potential of GenAI tools for several application scenarios. Among them are the effective generation of code explanations [Ma23; Sa22], enhanced programming error messages [Le23b], but also the analysis of student code with the goal of fixing students' errors [Ph23; Zh22b], or providing (elaborate) feedback [AKS24; KLK24]. All of these studies focus on students' (potential) use of LLMs.

For example, MacNeil et al. investigated students' perceptions towards automatically generated line-by-line code explanations by OpenAI's Codex and GPT-3 as part of an e-book. A majority of students evaluated the generated explanations as helpful [Ma23]. Similarly, Leinonen et al. found that students rate code explanations generated by GPT-3 better on average than explanations from their peers. They also note that students showed no aversion towards LLM-generated feedback, and that student prefer line-by-line code explanations [Le23a]. It is therefore not surprising that LLMs are being integrated into educational tools and environments to generate novice-friendly explanations tailored to each error [Ta24], without revealing code solutions [Ka24], or to provide guardrails [Li24a].

Related to that is the use of LLMs to analyze students' solutions and fix errors, or provide various types of elaborated feedback [Na06; Sh08] for novice programmers [KJH18]. Phung et al. investigated the use of LLMs to fix syntax errors in Python programs and developed a technique to receive high precision feedback. Other studies qualitatively explored the feedback generated by ChatGPT [ADS23; AKS24; KLK24]. Kiesler et al. characterized the feedback generated by ChatGPT-3.5 in response to authentic student solutions for introductory programming tasks. They found stylistic elements, textual explanations of the cause of errors and their fix, illustrating examples, meta-cognitive and motivational

elements, but also misleading information, uncertainty in the model's response, and requests for more information by the LLM. Azaiz et al. [ADS23] noted difficulties of GPT-3.5 with the formatting of its output, recognizing correct solutions, and it hallucinated errors. Roest et al. [RKJ23] conclude that the feedback generated by the GPT-3.5-turbo model seems to lack sufficient detail towards the end of an assignment. Nonetheless, a recent qualitative evaluation of GPT4 Turbo's feedback shows notable improvements, as the outputs are more structured, consistent, and always personalized [AKS24].

While some of the aforementioned studies utilize authentic student data (e.g., students' solutions as an input to LLMs), others directly involve students to survey their perspective, or usability of LLMs. One example is the study by Prather et al. [Pr23c] who investigated students' use of GitHub Copilot in an introductory programming assignment. Their observations and interviews explore students' perceptions regarding the (usability) challenges and potential benefits of this technology, resulting in design implications. Similarly, Vaithilingam et al. explored the usability of LLMs and their code generation abilities from the student perspective, concluding that Copilot's design should be improved [VZG22]. Jayagopal et al. explored the learnability of program synthesizers (including Copilot) by observing and interviewing students, which resulted in a set of lessons learned regarding system design. They also found the possibility to write a prompt to be perceived as more exciting and less confusing by students, compared to tools requiring format specifications [JLC22]. In another usability study, Barke et al. presents a grounded theory analysis of programmers' interactions with Copilot, based on observations of 20 experienced programmers. They distinguish users in acceleration mode (i.e., Copilot is used to consciously achieve a goal faster) and exploration mode (i.e., Copilot is used to explore options). Moreover, they show that over-reliance on Copilot can reduce the task completion rate, and cognitive load can result from having to choose from multiple suggestions [BJP23].

In another study, Denny et al. [De23] investigated first year students' interactions (n=54) with *Promptly* in the context of a Python class. The tool utilizes the concept of "Prompt Problems", aiming at teaching students how to prompt LLMs based on predefined problems. It automatically evaluates (i.e., executes and tests) code generated by an LLM (text-davinci-003 and gpt-3.5-turbo). In their analysis, however, Denny et al. focus on the quantification of students' interactions with Promptly (e.g., number of words per prompt, prompts per problem) but not the follow-up interactions. Moreover, students' responses to reflective questions were used as a data basis.

Even though many of these studies have investigated potential applications of LLMs by computing students, and specifically, novice programmers [AKS24; KLK24; Le23a; Ma23], they are conducted from an educator's perspective, meaning all use cases had been predefined. So, they are not necessarily authentic. The presented usability studies [BJP23; JLC22; Pr23c; VZG22] were based on observations and interviews, and did not focus on novice programmers and their application of LLMs and related tools. The only study on students' interactions [De23] emphasizes quantitative aspects, and does not analyze follow-up prompts and interaction patterns.

In addition, few very recent studies so far have been dedicated to computing students' trust in LLMs [Am24], and their attitude towards related tools [RHG24]. Even fewer studies investigated students' use of LLMs based on actual interaction data. Liu et al. investigated students' use of ChatGPT-3.5 as virtual teaching assistants in a classroom setting. However, they evaluated the tools' effectiveness (which was not based on students' interaction patterns) [Li24b]. Grande et al. evaluated the student perspectives on using ChatGPT for an assignment on professional ethics, but they used the students' submissions to the task as a data basis [GKR24]. As a result, there is a lack of authentic interaction data of students using an LLM, i.e., their complete chat protocols, and their (qualitative) evaluation in the context of an introductory programming class.

3 Methodology

To answer the RQ *How do students chat with ChatGPT in the context of introductory programming course assignments?*, the study leverages empirical data collected from students (n=213) enrolled in an introductory programming course. Students were asked to complete a series of programming exercises with the assistance of ChatGPT-3.5, and to submit their chat protocols as part of the assignment. In this section, we introduce the course context, instructions, the selected tasks, and the methodology applied for data analysis.

3.1 Course Context Used for Data Gathering

The context of this study was an introductory programming course for first-year computing students (n=790) at Goethe University Frankfurt (Germany) during the winter term 2023/24. The majority were computer science students, with some individuals from other majors who pursue computer science as a minor. The course is designed for novice learners or programmers, with no prerequisites. The class was accompanied by a Moodle course with learning materials. It comprised a 2-hour lecture per week for all 790 students, and a 2-hour tutorial session in groups of 20-30 students. A crucial components of the tutorial are the weekly or bi-weekly exercise sheets with programming tasks. Students are awarded points for (mostly individually) submitting their solutions, which contribute towards the exam held at the end of the semester.

For the present study, a specific exercise sheet was developed for the tutorials in the week of December 6, 2023. Students were expected to submit their individual solutions two weeks later, group work was not permitted. Students were further instructed to "complete the tasks using ChatGPT via the free version" (3.5) on the web interface, and to submit "all prompts and responses" as paired entries in a spreadsheet via the Moodle course. However, no specific instructions were provided on how to interact with ChatGPT-3.5, except for a reference to OpenAI's guide on prompt engineering [Op23]. This approach was deliberately chosen to avoid influencing students' interactions with ChatGPT-3.5. The exercise sheet

was available in both German and English to accommodate the diverse student body of the course. Using ChatGPT for this exercise sheet was voluntary, but two extra points were offered as an incentive. Moreover, students had been introduced to the present study's objective and the process during the lecture preceding the tutorial.

3.2 Selected Tasks

The exercise sheet addressed the concept of *recursion*, *functions*, *lists*, *conditionals*, *string manipulation*, and *documentation*. The exercise comprised two main tasks, while each one included sub-tasks. The first task, consisting of four sub-tasks, presented code snippets with recursive elements for several operations: (1a) summation of the digits of a number, (1b) reversing a list, (1c) performing multiplication, and (1d) computing the Ackermann function. Students were asked to read and interpret the given code snippets with the goal of determining the output of the code, the number of function calls, and identifying the implemented type of recursion.

The second task presented a more complex challenge, namely to implement a function that determines the number of "happy strings" within all sub-strings of a given string. A "happy string" was defined as a string that can either be rearranged into (or already is) a repetition of some string. To solve this task, students were required to (2a) test a given string for the "happy" property, and (2b) test all possible sub-strings of the given string. Students were encouraged to adopt a recursive approach by awarding additional points.

Prior to their integration into the exercise sheet, the authors evaluated ChatGPT-3.5's capability to solve the selected tasks. For task 1, ChatGPT demonstrated a high success rate in identifying the correct functions but frequently miscalculated the number of calls, often overlooking the base case. Task 2 proved to be more challenging for ChatGPT-3.5. Although the responses included the necessary steps for problem solving, it never included a correct solution (we tried 10 regenerations). This may have been due to a degree of ambiguity, and, for example, the uncommon term "happy string" in the task description.

3.3 Data Analysis

To evaluate students' interactions with ChatGPT-3.5, and how they seek assistance from the GenAI tool, the collected chat protocols were quantitatively and qualitatively analyzed. For the **quantification** of students' use patterns, the following aspects were investigated:

- *Number of prompts per student*: The total count of prompts given to ChatGPT-3.5, offering insights into the extent of interaction from students with the GenAI tool for the given tasks.
- *Number of words per prompt*: Identifying the verbosity or conciseness of students' prompts, to help characterize students' descriptions of the problem or task (and its complexity).

- *Follow-up interactions per student*: This characteristic aims to capture the iterative nature of problem-solving interactions with ChatGPT-3.5. It is supposed to help identify students' use patterns.

In addition, the chat protocols were **qualitatively** analyzed [Ma00] by two of the authors to identify the type of follow-up interactions, the contents or issues addressed by the students' prompts, and the overall (interaction) strategy students seem to pursue or apply.

- *Solution requests (SR)*: Determines whether the student's prompt explicitly seeks a direct solution, shedding light on the intention behind the query and the expected outcome of the interaction (e.g., if only the task description is prompted). Frequencies emphasize the extend to which students apply this kind of request.
- *Type of follow-up interaction*: Categorizes whether a prompt is a standalone query (STA), a follow-up to a previous prompt (PRE), a response to a ChatGPT answer (RES), or a correction of a previous response (COR). We further add the frequencies of these types. Each prompt was categorized into only one of these four types.
- *Issue or problem solving step addressed by the prompt*: This classification identifies the issue addressed by the prompt. These categories include common problem solving steps and well-known issues for novice learners based on the literature [Du86; ER16; Ki20; Lu18; SS86]: problem understanding (PU), conceptual understanding (CU), code generation (CG), debugging (DE), runtime analysis (RA), syntax/style (SY), documentation (DO), test cases (TC), and other categories (OT). It should be noted that a single prompt can reflect multiple issues within the problem solving process.
- *Categories describing students' interactions*: Describes how students navigate and modify their prompts based on the generated output. Categories reflecting these interaction pattern were inductively built based on the material.

4 Results

A total of 360 students successfully completed the exercise sheet. Of these, 305 students engaged with ChatGPT-3.5 as part of the exercise. Due to the absence of the required template usage or submission in non-processable file formats, the data from 92 students could not be included in the final analysis. Therefore, the final dataset comprises 213 students, providing a collection of 2335 prompts for detailed examination. In total, we received 1668 German and 426 English prompts (The data is available online [SK24]).

Students engaged in approximately 10.96 prompts on average. The median during all interactions is 7 prompts per student. Based on these numbers, we categorize students into three groups, as shown in Table 1: Group A, comprising individuals who use 0-5 prompts; Group B, including those who use 6-11 prompts; and Group C, consisting of students who use 12 or more prompts. The distinction yields almost equal sized groups.

Regarding the evaluation of word counts per prompt, it should be noted that the number of words for task 1 alone is 146, and 175 for task 2. This partially explains the decrease in

Range of prompts per student	No. of Students	No. of Prompts	Avg word count	Solution Requests (SR)	Follow-Up Interactions			
					STA	PRE	RES	COR
A: 0 to 5	66	262	73.74	213	172	57	25	8
B: 6 to 11	79	612	51.49	416	240	205	111	56
C: 12 and more	68	1461	40.46	603	358	462	322	319

Tab. 1: Quantification of students' interactions with ChatGPT-3.5, including some of the qualitative aspects of their follow-up interactions.

the average word count from Group A to Group B by 30%, followed by a further decline from Group B to Group C by 21% to 40.46 words on average per prompt. Furthermore, the number of solution requests (SR) shows a declining trend across the groups. In Group A, approximately 81% of prompts are SR, which decreases to 68% in Group B and 41% in Group C (see Table 1).

The nature of follow-up interactions (FIs) also varies among the groups (see Table 1). In Group A, the majority of interactions had no follow-up (STA, in 172 cases). In many cases, the prompts resembled the task descriptions, which had been used as input. At the same time, 90 instances of FIs were observed, comprising 57 follow-ups to a previous prompt (PRE), 25 responses to a generated answer by ChatGPT (RES), and 8 corrections of responses (COR), indicating a correction rate of approximately 3%.

Group B reveals an increase of FIs, with 372 instances of FIs, and somewhat fewer instances (240) without an FI (STA). Notably, there is a greater number of students referring back to their own prompts (PRE, 205 cases), while 56 corrections of a previous response (COR) were identified (9%). In contrast, Group C demonstrates a significant increase of FIs, with the vast majority of follow-ups (1103 is sum). In Group 3, the correction rate (22%) was particularly high, with 319 correction instances (COR).

The analysis of student prompts relating to programming problem-solving categories reveals several trends, summarized in Table 2. For example, the proportion of initial prompts aimed at understanding the problem (PU), including solution requests, is smaller in longer conversations. In group A, these prompts constitute 76% of the interactions, dropping to 50% in group B, and further declining to 34% in Group C. Conversely, there is an increase of prompts requesting help in debugging (DE) in longer conversations. This category sees a rise from 6% in group A to 17% in group B, and a further increase to 24% in group C. The number of prompts referring to runtime analyses (RA) is higher for group A, accounting for 15% of interactions. This may be due to fewer number of prompts in group A and runtime analysis being an explicit part of Task 1. In terms of concept understanding (CU), code generation (CG), and documentation (DO), there is a slight increase from group A to group B and C, while the percentage difference between group B and C is minimal. Additionally, the analysis shows a relatively stable percentage of prompts related to syntax/style (SY) from group A to C, and a relatively consistent ratio for test case prompts (TC) to overall prompts observed across group A, B, and C (about 10%).

Range of prompts per student	No. of Prompts	Problem-solving category								
		PU	CU	CG	DE	RA	SY	DO	TC	OT
A: 0 to 5	262	198	25	34	16	40	3	24	26	3
B: 6 to 11	612	303	87	121	101	55	10	83	58	10
C: 12 and more	1461	497	185	273	354	115	36	186	143	29

Tab. 2: Distribution of students' prompts across the various problem-solving categories (multiple categories per student and prompt).

After qualitatively analyzing the prompts, we were able to identify two common patterns of interactions with the GenAI tool: (1) the *Task Description Prompts* pattern, and (2) the *Prompts in Own Words* pattern. For (1), students seem to use the given task description at least as part of their initial prompt. In the other pattern (2), students rephrase the task description or immediately focus on specific parts of the problem statement. We provide a summary of both pattern and sub-pattern in Table 3 and Table 4, whereas the example ID refers to the student ID in the chat protocols. The order of the prompting patterns in the tables is supposed to reflect on the increasing specificity. Table 3 summarizes respective use pattern for *Task Description Prompts*, showing different approaches to generate a solution using ChatGPT after having used the task description as an initial prompt.

Example ID	**Follow-ups after initial task description prompt:**
1012	No or few additional instructions.
1044	Giving direct orders / requests including style, documentation or corrections.
1199	Short, but extensive prompting for explanations to gain understanding.
1135	Specific prompts requesting explanations and corrections.
1070	Testing of initially generated code by the LLM. Providing incorrect (console) output followed by "Correct the code" instructions. No additional input or adaptation.
1195	Testing of initially generated code by the LLM. Providing incorrect (console) output followed by "Correct the code" instructions, resulting in disappointment.
1006	Testing of initially generated code by the LLM. Providing incorrect (console) output followed by "Correct the code" instructions. Followed by instruction to restart, and adaptation of instruction.
1156	Requesting (conceptual) explanations. Follow-up to create text and not bulletpoints.

Tab. 3: Examples for the *Task Description Prompts*-pattern encountered in the dataset.

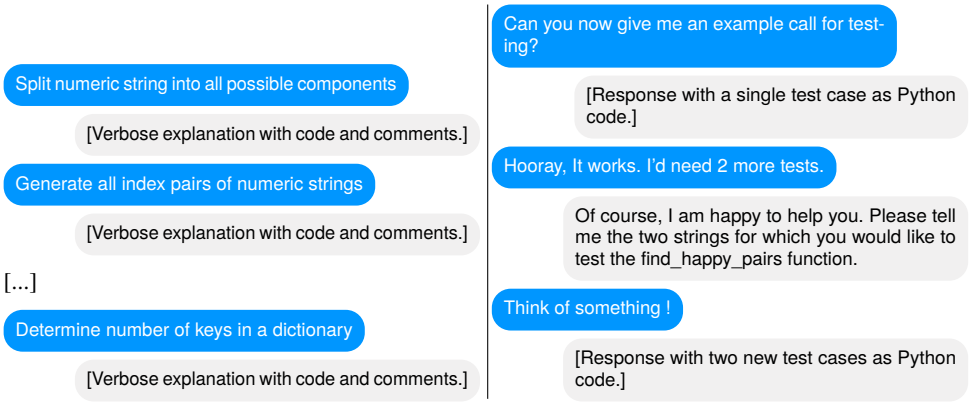
For the *Prompt in Own Words* pattern (summarized in Table 4), students exhibit different approaches to achieve the correct solution or responses to their questions. These pattern seem more directed towards a specific aspect of the problem, or occurred only after the student had created a solution.

In addition to these two patterns, we observed some other interesting aspects of students' interactions. This includes switching the language from German to English as soon as the GenAI tool produced English outputs. Some prompts started in English, but students

Example ID	Prompts and follow-ups in students' own words
1014	Demanding explanations for specific concepts, e.g. indices, dictionaries.
1058	Adding own solution and asking specific questions.
1106	Providing additional task constraints. Task description is provided as a follow up. Requests specific aspects, e.g., test cases, documentation, while providing examples.

Tab. 4: Examples of *Prompts in Own Words*-pattern found in the dataset.

asked to create a documentation in German. Also, as part of the OT categorization, we found interjections, e.g., "Hooray", "Thank You!", which resemble human-to-human communication (ID 1037). In Figure 1, we present two exemplary excerpts from students and their ChatGPT interactions. Figure 1a illustrates that the student did not write complete sentences, but assigned explicit tasks to be executed. Figure 1b shows a human-to-human interaction with the tool.



(a) Excerpt from chat protocol 1014 with brief and direct commands.

(b) Excerpt from chat protocol 1037 resembling human-to-human interaction.

Fig. 1: Excerpts of two example chat protocols illustrating different interaction patterns (translated into English by the authors).

5 Discussion

The analysis of students' chat protocols indicates diverse interactions with GenAI tools, such as ChatGPT, when being applied as part of an introduction programming course in higher education. For example, students engaged differently with ChatGPT in terms of the number of prompts used, their length, and regarding (the content of) follow-up interactions.

One group of students (n=66) only used 0-5 prompts while working on the selected tasks. This group may not have been thoroughly engaged in the interaction with ChatGPT. There is also a large number of solution requests (SR) present in this group. This may be an indicator for students looking for a quick solution to the given problems.

The second group of students (n=79) used ChatGPT differently. For example, we see an increasing number of prompts in this group. Still, many students requested the correct solution (SR). What is interesting though is that there are many follow-up interactions. This might indicate that either the initial response of ChatGPT was not considered good enough, or that the generated response actually fostered an interest in (an aspect of) the topic or problem. The increasing number of follow-ups on previous prompts (PRE) and responses to a ChatGPT answer (RES) disambiguate this group from students who only submitted 0-5 prompts.

The third group of students (n=68) used the GenAI tool most frequently, but for different purposes. For example, a smaller percentage of solution requests was identified. Especially due to the high number of corrections (COR), students seemed to be more reflective of the generated answers when trying to generate valid solutions. Especially interesting is the increase in responses to a ChatGPT answer (RES), which show that students actually used the tool as a "learning partner" instead of a simple query tool. This is also supported by the high numbers of prompts requesting help for understanding (PU) and debugging (DE).

The qualitative data analysis further reveals a difference in the categorization of problem-solving steps, the identified follow-up interactions, and prompting patters (*Task Description Prompts* and *Prompts in Own Words*). While the former are well known especially in the context of feedback generation [KJH18; KLK24], the latter represent entirely new categories aiming to describe students' interactions and prompting pattern with ChatGPT. Educators can use these pattern as a basis for designing instructions on how to (not) use LLMs and related tools, and to reflect on their limitations and potential benefits.

To answer our research question *How do students chat with ChatGPT in the context of introductory programming course assignments*, we conclude that there is a certain variety and range of students' applications of the tool. There are at least three very different behaviors present. One seems to focus on problem solving by generating a single prompt to generate the best answer possible. Other students revealed a more "peer-like" interaction with the GenAI tool. They used more follow-up prompts, focused on problem solving step-by-step, tried to use the tool to identify and correct errors, and communicated respectively (we thus observed signs of anthropomorphism towards ChatGPT in this study). The third recognized use pattern is more straight forward and focuses on code generation, but utilizes the option of having an interactive tool. We assume that there are even more facets to these behaviors.

6 Limitations

A limitation of this study is the context of the data collection, which was a single class taking place at one university. Even though the high number of students participating in the study strengthens its representative character, the results may not be transferable to other educational contexts. Another limitation is that due to their knowledge of the study, students may have interacted differently with the tool than they would normally have in the context of an assignment, which is known as observer's paradox [RD39].

7 Conclusions and Future Work

This work presents the first-of-its kind exploratory study on students' use of ChatGPT as a tool to support problem-solving in the early stages of programming education at university level. To investigate students interactions, an exercise sheet had been designed and students were instructed to use ChatGPT, but not how to do so. Students' chat protocols were quantitatively and qualitatively analyzed to characterize students' use and interaction patterns. The high participation rate (n=213) allows us to draw several conclusions. For example, the results reveal three groups of students interacting differently with the tool. Among these groups are students seeking an immediate solution, but also students seeking help with problem understanding, conceptual knowledge, debugging, syntax error or style and documentation. Some students also used the chat more extensively than others, thereby resembling a "studdy buddy" or "virtual tutor". The chat protocols further revealed new patterns of interactions, e.g., students starting with *Task Description Prompts* and specifying their requests later, and those using *Prompts in Own Words* with more specific requests. However, there is a huge diversity in students' use of ChatGPT.

Implications for teaching and learning are manifold. For example, the diversity of use patterns reflects the heterogeneity of students' needs for support, and feedback. GenAI tools have the potential to offer that degree of personalization [AKS24]. Regardless, educators need to provide the necessary tools and advice on how to use them successfully, e.g., prompt patterns [Wh23]. Moreover, it is important to communicate the tools' limitations and potential pitfalls to students so they do not end up frustrated. The computing education community also needs to continue their discussion on the competencies required for successful graduates, and the role of GenAI tools in that process.

As our approach has been executed with a general LLM (ChatGPT-3.5), it would be interesting to see how a specialized AI tool for coding would increase its support for learners as part of future work. Moreover, a replication study at another institution or country would further strengthen the results. With the world facing new challenges and new tools every day, the potential in using AI for learning computer science is not yet fully understood. And to quote ChatGPT-3.5: "In this perpetual cycle of learning and teaching, computers enlighten students who, in turn, empower computers, perpetuating the beautiful recursion of knowledge."

References

[ADS23] Azaiz, I.; Deckarm, O.; Strickroth, S.: AI-enhanced Auto-Correction of Programming Exercises: How Effective is GPT-3.5? International Journal of Engineering Pedagogy (iJEP) 13 (8), pp. 67–83, 2023, DOI: 10.3991/ijep.v13i8.45621.

[AKS24] Azaiz, I.; Kiesler, N.; Strickroth, S.: Feedback-Generation for Programming Exercises With GPT-4, 2024, DOI: 10.48550/arXiv.2403.04449, arXiv: 2403.04449 [cs.AI].

[Al24] Al-Hossami, E.; Bunescu, R.; Smith, J.; Teehan, R.: Can Language Models Employ the Socratic Method? Experiments with Code Debugging. In: Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1. SIGCSE 2024, ACM, Portland, OR, USA, pp. 53–59, 2024, DOI: 10.1145/3626252.3630799.

[Am24] Amoozadeh, M.; Daniels, D.; Nam, D.; Kumar, A.; Chen, S.; Hilton, M.; Srinivasa Ragavan, S.; Alipour, M. A.: Trust in Generative AI among Students: An exploratory study. In: Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1. ACM, New York, pp. 67–73, 2024, DOI: 10.1145/3626252.3630842.

[Be23] Becker, B. A.; Craig, M.; Denny, P.; Keuning, H.; Kiesler, N.; Leinonen, J.; Luxton-Reilly, A.; Prather, J.; Quille, K.: Generative AI in Introductory Programming. 2023, URL: https://csed.acm.org/wp-content/uploads/2023/12/Generative-AI-Nov-2023-Version.pdf.

[BJP23] Barke, S.; James, M. B.; Polikarpova, N.: Grounded Copilot: How Programmers Interact with Code-Generating Models. Proc. ACM Program. Lang. 7 (OOPSLA1), 2023, DOI: 10.1145/3586030.

[BK23] Bengtsson, D.; Kaliff, A.: Assessment Accuracy of a Large Language Model on Programming Assignments, 2023, URL: https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-331000.

[Ca17] Camp, T.; Adrion, W. R.; Bizot, B.; Davidson, S.; Hall, M.; Hambrusch, S.; Walker, E.; Zweben, S.: Generation CS: the growth of computer science, New York, 2017, DOI: 10.1145/3084362.

[De23] Denny, P.; Leinonen, J.; Prather, J.; Luxton-Reilly, A.; Amarouche, T.; Becker, B. A.; Reeves, B. N.: Promptly: Using Prompt Problems to Teach Learners How to Effectively Utilize AI Code Generators, 2023, arXiv: 2307.16364 [cs.HC].

[Du86] Du Boulay, B.: Some difficulties of learning to program. Journal of Educational Computing Research 2 (1), pp. 57–73, 1986, DOI: 10.2190/3LFX-9RRF-67T8-UVK9.

[ER16] Ebert, M.; Ring, M.: A presentation framework for programming in programing lectures. In: Proc. EDUCON. IEEE, pp. 369–374, 2016.

[Ge23] Geng, C.; Zhang, Y.; Pientka, B.; Si, X.: Can ChatGPT Pass An Introductory Level Functional Language Programming Course?, 2023, arXiv: 2305.02230 [cs.CY].

[Gi24] Gill, S. S.; Xu, M.; Patros, P.; Wu, H.; Kaur, R.; Kaur, K.; Fuller, S.; Singh, M.; Arora, P.; Parlikad, A. K.; Stankovski, V.; Abraham, A.; Ghosh, S. K.; Lutfiyya, H.; Kanhere, S. S.; Bahsoon, R.; Rana, O.; Dustdar, S.; Sakellariou, R.; Uhlig, S.; Buyya, R.: Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots. Internet of Things and Cyber-Physical Systems 4, pp. 19–23, 2024, DOI: 10.1016/j.iotcps.2023.06.002.

[GKR24] Grande, V.; Kiesler, N.; Rodriguez, M. A. F.: Student Perspectives on Using a Large Language Model (LLM) for an Assignment on Professional Ethics. In: Proceedings of the 2024 Conference on Innovation and Technology in Computer Science Education V. 2. ITiCSE 2024, ACM, Milan, Italy, 2024, DOI: 10.1145/3649217.3653624.

[Je22] Jeuring, J.; Keuning, H.; Marwan, S.; Bouvier, D.; Izu, C.; Kiesler, N.; Lehtinen, T.; Lohr, D.; Peterson, A.; Sarsa, S.: Towards Giving Timely Formative Feedback and Hints to Novice Programmers. In: Proceedings of the 2022 Working Group Reports on Innovation and Technology in Computer Science Education. ITiCSE-WGR '22, ACM, Dublin, Ireland, pp. 95–115, 2022, DOI: 10.1145/3571785.3574124.

[JLC22] Jayagopal, D.; Lubin, J.; Chasins, S. E.: Exploring the learnability of program synthesizers by novice programmers. In: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology. Pp. 1–15, 2022.

[Jo24] Joshi, I.; Budhiraja, R.; Dev, H.; Kadia, J.; Ataullah, M. O.; Mitra, S.; Akolekar, H. D.; Kumar, D.: ChatGPT in the Classroom: An Analysis of Its Strengths and Weaknesses for Solving Undergraduate Computer Science Questions. In: Proc. of the 55th ACM Technical Symposium on Computer Science Education V. 1. ACM, New York, pp. 625–631, 2024, DOI: 10.1145/3626252.3630803.

[Ka24] Kazemitabaar, M.; Ye, R.; Wang, X.; Henley, A. Z.; Denny, P.; Craig, M.; Grossman, T.: CodeAid: Evaluating a Classroom Deployment of an LLM-based Programming Assistant that Balances Student and Educator Needs. arXiv preprint arXiv:2401.11314, 2024.

[Ki20] Kiesler, N.: Towards a Competence Model for the Novice Programmer Using Bloom's Revised Taxonomy – An Empirical Approach. In: Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education. ITiCSE '20, ACM, Trondheim, Norway, pp. 459–465, 2020, DOI: 10.1145/3341525.3387419.

[Ki23] Kiesler, N.: Beyond the Textbook: Rethinking Students' Competencies in the LLM Era, Generative AI: Implications for Teaching and Learning, Uppsala, Sweden, 2023, DOI: 10.13140/RG.2.2.28355.37922/1.

[Ki24] Kiesler, N.: Modeling Programming Competency: A Qualitative Analysis. Springer International Publishing, Cham, 2024, ISBN: 978-3-031-47148-3.

[KJH18] Keuning, H.; Jeuring, J.; Heeren, B.: A Systematic Literature Review of Automated Feedback Generation for Programming Exercises. ACM Trans. Comput. Educ. 19 (1), 2018, DOI: 10.1145/3231711.

[KLK24] Kiesler, N.; Lohr, D.; Keuning, H.: Exploring the Potential of Large Language Models to Generate Formative Programming Feedback. In: 2023 IEEE Frontiers in Education Conference (FIE). Pp. 1–5, 2024, DOI: 10.1109/FIE58773.2023.10343457.

[KS23a] Kiesler, N.; Schiffner, D.: Large Language Models in Introductory Programming Education: ChatGPT's Performance and Implications for Assessments, 2023, DOI: 10.48550/arXiv.2308.08572, arXiv: 2308.08572 [cs.SE].

[KS23b] Kiesler, N.; Schiffner, D.: Why We Need Open Data in Computer Science Education Research. In: Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1. ITiCSE 2023, ACM, Turku, Finland, pp. 348–353, 2023, DOI: 10.1145/3587102.3588860.

[Le23a] Leinonen, J.; Denny, P.; MacNeil, S.; Sarsa, S.; Bernstein, S.; Kim, J.; Tran, A.; Hellas, A.: Comparing Code Explanations Created by Students and Large Language Models. In: Proc. ITiCSE. Pp. 124–130, 2023, DOI: 10.1145/3587102.3588785.

[Le23b] Leinonen, J.; Hellas, A.; Sarsa, S.; Reeves, B.; Denny, P.; Prather, J.; Becker, B. A.: Using Large Language Models to Enhance Programming Error Messages. In: Proc. SIGCSE. ACM, 2023, DOI: 10.1145/3545945.3569770.

[Li24a] Liffiton, M.; Sheese, B. E.; Savelka, J.; Denny, P.: CodeHelp: Using Large Language Models with Guardrails for Scalable Support in Programming Classes. In: Proceedings of the 23rd Koli Calling International Conference on Computing Education Research. Koli Calling '23, ACM, Koli, Finland, 2024, DOI: 10.1145/3631802.3631830.

[Li24b] Liu, R.; Zenke, C.; Liu, C.; Holmes, A.; Thornton, P.; Malan, D. J.: Teaching CS50 with AI: Leveraging Generative Artificial Intelligence in Computer Science Education. In: Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1. SIGCSE 2024, ACM, New York, pp. 750–756, 2024, DOI: 10.1145/3626252.3630938.

[Lo24] Lohr, D.; Kiesler, N.; Keuning, H.; Jeuring, J.: "Let Them Try to Figure It Out First" – Reasons Why Experts (Do Not) Provide Feedback to Novice Programmers. In: Proceedings of the 2024 Conference on Innovation and Technology in Computer Science Education V. 1. ITiCSE 2024, ACM, Milan, Italy, 2024, DOI: 10.1145/3649217.3653530.

[Lu16] Luxton-Reilly, A.: Learning to Program is Easy. In: Proc. ITiCSE. Pp. 284–289, 2016, DOI: 10.1145/2899415.2899432.

[Lu18] Luxton-Reilly, A.; Simon; Albluwi, I.; Becker, B. A.; Giannakos, M.; Kumar, A. N.; Ott, L.; Paterson, J.; Scott, M. J.; Sheard, J.; Szabo, C.: Introductory Programming: A Systematic Literature Review. In: Proc. ITiCSE. ACM, New York, pp. 55–106, 2018, DOI: 10.1145/3293881.3295779.

[Ma00] Mayring, P.: Qualitative content analysis forum qualitative sozialforschung. Forum: qualitative social research 1 (2), 2000.

[Ma23] MacNeil, S.; Tran, A.; Hellas, A.; Kim, J.; Sarsa, S.; Denny, P.; Bernstein, S.; Leinonen, J.: Experiences from Using Code Explanations Generated by Large Language Models in a Web Software Development E-Book. In: Proc. SIGCSE TS. Pp. 931–937, 2023, DOI: 10.1145/3545945.3569785.

[Ma24] MacNeil, S.; Leinonen, J.; Denny, P.; Kiesler, N.; Hellas, A.; Prather, J.; Becker, B. A.; Wermelinger, M.; Reid, K.: Discussing the Changing Landscape of Generative AI in Computing Education. In: Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2. ACM, New York, p. 1916, 2024, DOI: 10.1145/3626253.3635369.

[Na06] Narciss, S.: Informatives Tutorielles Feedback: Entwicklungs- und Evaluationsprinzipien auf der Basis instruktionspsychologischer Erkenntnisse. Waxmann Verlag, Münster, 2006.

[Op23] OpenAI: Prompt Engineering, https://platform.openai.com/docs/guides/prompt-engineering, Accessed: 2024-04-08, 2023.

[Pe16] Petersen, A.; Craig, M.; Campbell, J.; Tafliovich, A.: Revisiting why students drop CS1. In: Proc. Koli Calling. Pp. 71–80, 2016, DOI: 10.1145/2999541.2999552.

[Ph23] Phung, T.; Cambronero, J.; Gulwani, S.; Kohn, T.; Majumdar, R.; Singla, A.; Soares, G.: Generating High-Precision Feedback for Programming Syntax Errors using Large Language Models, 2023, arXiv: 2302.04662 [cs.PL].

[Pr23a] Prather, J.; Denny, P.; Leinonen, J.; Becker, B. A.; Albluwi, I.; Caspersen, M. E.; Craig, M.; Keuning, H.; Kiesler, N.; Kohn, T.; Luxton-Reilly, A.; MacNeil, S.; Petersen, A.; Pettit, R.; Reeves, B. N.; Savelka, J.: Transformed by Transformers: Navigating the AI Coding Revolution for Computing Education: An ITiCSE Working Group Conducted by Humans. In: Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 2. ACM, New York, pp. 561–562, 2023, DOI: 10.1145/3587103.3594206.

[Pr23b] Prather, J.; Denny, P.; Leinonen, J.; Becker, B. A.; Albluwi, I.; Craig, M.; Keuning, H.; Kiesler, N.; Kohn, T.; Luxton-Reilly, A.; MacNeil, S.; Petersen, A.; Pettit, R.; Reeves, B. N.; Savelka, J.: The Robots Are Here: Navigating the Generative AI Revolution in Computing Education. In: Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education. ACM, New York, pp. 108–159, 2023, DOI: 10.1145/3623762.3633499.

[Pr23c] Prather, J.; Reeves, B. N.; Denny, P.; Becker, B. A.; Leinonen, J.; Luxton-Reilly, A.; Powell, G.; Finnie-Ansley, J.; Santos, E. A.: "It's Weird That It Knows What I Want": Usability and Interactions with Copilot for Novice Programmers. ACM Trans. Comput.-Hum. Interact. 2023, DOI: 10.1145/3617367.

[Pr24] Prather, J.; Leinonen, J.; Kiesler, N.; Gorson Benario, J.; Lau, S.; MacNeil, S.; Norouzi, N.; Opel, S.; Pettit, V.; Porter, L.; Reeves, B.; Savelka, J.; Smith IV, D. H.; Strickroth, S.; Zingaro, D.: How Instructors Incorporate Generative AI into Teaching Computing. In: Proceedings of the 2024 Conference on Innovation and Technology in Computer Science Education V. 2. ITiCSE 2024, ACM, Milan, Italy, 2024, DOI: 10.1145/3649405.3659534.

[Ra21] Raj, R.; Sabin, M.; Impagliazzo, J.; Bowers, D.; Daniels, M.; Hermans, F.; Kiesler, N.; Kumar, A. N.; MacKellar, B.; McCauley, R.; Nabi, S. W.; Oudshoorn, M.: Professional Competencies in Computing Education: Pedagogies and Assessment. In: Proceedings of the 2021 Working Group Reports on Innovation and Technology in Computer Science Education. ACM, New York, pp. 133–161, 2021, DOI: 10.1145/3502870.3506570.

[RD39] Roethlisberger, F. J.; Dickson, W. J.: Management and the Worker. Harvard University Press, Cambridge, 1939.

[RHG24] Rogers, M. P.; Hillberg, H. M.; Groves, C. L.: Attitudes Towards the Use (and Misuse) of ChatGPT: A Preliminary Study. In: Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1. SIGCSE 2024, ACM, Portland, OR, USA, pp. 1147–1153, 2024, DOI: 10.1145/3626252.3630784.

[RKJ23] Roest, L.; Keuning, H.; Jeuring, J.: Next-Step Hint Generation for Introductory Programming Using Large Language Models, 2023, arXiv: 2312.10055 [cs.CY].

[Sa22] Sarsa, S.; Denny, P.; Hellas, A.; Leinonen, J.: Automatic Generation of Programming Exercises and Code Explanations Using Large Language Models. In: Proc. ICER. ACM, 2022, DOI: 10.1145/3501385.3543957.

[Sa23] Savelka, J.; Agarwal, A.; Bogart, C.; Sakr, M.: Large Language Models (GPT) Struggle to Answer Multiple-Choice Questions about Code, 2023, arXiv: 2303.08033 [cs.CL].

[SB22] Strickroth, S.; Bry, F.: The Future of Higher Education is Social and Personalized! Experience Report and Perspectives. In: Proc. CSEDU. Vol. 1, pp. 389–396, 2022, DOI: 10.5220/0011087700003182.

[Sh08] Shute, V. J.: Focus on formative feedback. Review of Educational Research 78 (1), 2008.

[SK24] Scholl, A.; Kiesler, N.: Data: Analyzing Chat Protocols of Novice Programmers Solving Introductory Programming Tasks with ChatGPT, 2024, DOI: 10.17605/OSF.IO/WBKQV, URL: osf.io/wbkqv.

[SS86] Spohrer, J. C.; Soloway, E.: Novice mistakes: Are the folk wisdoms correct? Communications of the ACM 29 (7), pp. 624–632, 1986, DOI: 10.1145/6138.6145.

[Ta24] Taylor, A.; Vassar, A.; Renzella, J.; Pearce, H.: dcc –help: Transforming the Role of the Compiler by Generating Context-Aware Error Explanations with Large Language Models. In: Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1. ACM, New York, pp. 1314–1320, 2024, DOI: 10.1145/3626252.3630822.

[VZG22] Vaithilingam, P.; Zhang, T.; Glassman, E. L.: Expectation vs. Experience: Evaluating the Usability of Code Generation Tools Powered by Large Language Models. In. ACM, New York, pp. 1–7, 2022.

[WCL07] Whalley, J.; Clear, T.; Lister, R.: The many ways of the Bracelet project. BACIT, 2007.

[Wh23] White, J.; Fu, Q.; Hays, S.; Sandborn, M.; Olea, C.; Gilbert, H.; Elnashar, A.; Spencer-Smith, J.; Schmidt, D. C.: A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT, 2023, arXiv: 2302.11382 [cs.SE].

[Za19] Zawacki-Richter, O.; Marin, V. I.; Bond, M.; Gouverneur, F.: Systematic review of research on artificial intelligence applications in higher education–where are the educators? International Journal of Educational Technology in Higher Education 16 (1), pp. 1–27, 2019.

[Zh22a] Zhai, X.: ChatGPT User Experience: Implications for Education. 2022, DOI: http://dx.doi.org/10.2139/ssrn.4312418.

[Zh22b] Zhang, J.; Cambronero, J.; Gulwani, S.; Le, V.; Piskac, R.; Soares, G.; Verbruggen, G.: Repairing Bugs in Python Assignments Using Large Language Models. arXiv preprint arXiv:2209.14876, 2022.

Informiertes Zustimmungsmanagement für Learning Analytics in Lernmanagementsystemen

Sven Judel[1], Christopher Zeise[2] und Ulrik Schroeder[1]



Abstract: Dieser Beitrag präsentiert ein Konzept um Lernende angemessen über Learning Analytics zu informieren und ihnen das Management von Zustimmungen zu Datensammlungen und -verarbeitungen einfach aber vollständig zu ermöglichen. Drei Komponenten übernehmen die Aufklärung, die Zustimmungsabfrage beim ersten Aufruf des Lernmanagementsystems sowie jedes Lernraums und die Anpassung der Zustimmungen zu jeder Zeit. Die Komponenten zum Management wurden beispielhaft für Moodle umgesetzt und mit Studierenden evaluiert. Die Ergebnisse deuten auf eine gute und sichere Nutzbarkeit hin, welche den Lernenden mehr Kontrolle über ihre Daten gibt.

Keywords: Trusted Learning Analytics, Zustimmung, Lernmanagementsystem, Moodle

1 Einleitung

Ziele von Trusted Learning Analytics beinhalten die Kontrolle der Datenspender, meist Lernende, über ihre eigenen Daten sicherzustellen und größtmögliche Transparenz darüber zu schaffen, welche Daten gesammelt und wie diese verarbeitet werden [SKRA16]. Um diesem Ziel näher zu kommen, wurde in die bestehende Learning Analytics Infrastruktur Excalibur LA [JS22] eine Komponente eingeführt, die das Management von Zustimmungen zu Datensammlung und -verarbeitung in Quellsystemen, hier konkret Lernmanagementsysteme (LMS), ermöglicht. Dazu wurden anhand von Guidelines zu Trusted Learning Analytics und gutem UI und UX Design von Cookie Zustimmungs-Interfaces 13 Anforderungen an ein Zustimmungs-Interface für Learning Analytics abgeleitet und ein Schema zur Abbildung der Zustimmungen exploriert. Drei Komponenten wurden konzeptioniert, welche Lernende über Learning Analytics informieren und aufklären, initiale Zustimmungen einholen und eine Anpassung der Zustimmungen, bis hin zum kompletten Opt-Out, zu jeder Zeit ermöglichen. Im Rahmen dieser Arbeit wurden zunächst die beiden Komponenten zum Zustimmungsmanagement betrachtet. Mit Paper Prototypes wurden erste Entwürfe mit Studierenden evaluiert und mit dem gesammelten Feedback eine Implementierung für das LMS Moodle umgesetzt. Dieser Software-Prototyp wurde ebenfalls evaluiert und die Implementierung basierend auf den Ergebnissen weiterentwickelt.

2 Related Work und Anforderungsanalyse

Konkrete Vorarbeit zum Verwalten von Zustimmungen zu Learning Analytics (LA) lassen sich zum aktuellen Zeitpunkt nicht in der Literatur finden. So liefert die Suche nach den

1 RWTH Aachen, Lerntechnologien, Ahornstraße 55, 52074 Aachen, {judel,schroeder}@cs.rwth-aachen.de
2 RWHT Aachen, Templergraben 55, 52062 Aachen, christopher.zeise@rwth-aachen.de

doi:10.18420/delfi2024_06

beiden Schlüsselbegriffen *consent management* und *learning analytics* in Kombination in Titel, Abstract oder Keywords von Publikationen in keiner der großen Datenbanken (ACM Digital Library, Scopus, IEEE Xplore, Web of Science) ein Ergebnis. Mit weniger strengen Suchanfragen lassen sich Publikationen, wie z. B. [Jon19] oder zu verschiedenen Ethical LA Prinzipien, wie in [CPM21] zusammengefasst, finden, welche die Problemstellung detailliert aufzeigen und Lösungen skizzieren. Praktische Erfahrungen fehlen jedoch. Entsprechend werden Grundlagen für die Arbeit aus verwandten Forschungsfeldern wie Trusted Learning Analytics und dem Design von Cookie Zustimmungs-Interfaces gezogen.

2.1 Trusted Learning Analytics

Vertrauen in Learning Analytics, einschließlich dem Vertrauen, dass Daten korrekt, sicher und ethisch verwendet werden, aufzubauen, ist Teil des Forschungsfeld der Trusted Learning Analytics (TLA). Dieses Vertrauen kann für Lernende eine Grundlage für die Bereitschaft, (mehr) Daten von und über sich sammeln und verarbeiten zu lassen, darstellen. Ein Teil dieses Vertrauens in LA stellt die Kontrolle der Lernenden über die eigenen Daten dar. Dies beinhaltet das Erteilen von Zustimmungen zu Datensammlung und -verarbeitung, die Einsicht in Verarbeitungsprozesse, sowie die Optionen, Zustimmungen jederzeit widerrufen und gesammelte Daten über die eigene Person einsehen und löschen zu können. [SKRA16] Dieses Vorgehen wird nicht nur von Experten im Feld der TLA empfohlen, sondern ist z. B. durch die DSGVO innerhalb der Europäischen Union rechtlich vorgegeben [Eur16].

Acht Punkte, die bei der Einführung von (T)LA berücksichtigt werden sollten, werden durch die DELICATE Checklist vorgegeben. Diese beinhalten kommunikative Aspekte wie *D*etermination - Warum möchte man LA betreiben, *E*xplain - Transparenz bei Intentionen und Zielen, *L*egitimate - Warum darf man die Daten sammeln, als auch organisatorische und technische Maßnahmen wie *A*nonymise - Verhinderung der Wiederfindbarkeit von Individuen, *T*echnical - Prozeduren zur Sicherstellung der Privatsphäre und *E*xternal - Wenn mit externen Stellen zusammengearbeitet wird. Während die kommunikativen Aspekte auf die Aufklärung und Informierung von Stakeholdern wie Lernende abzielt, fordern die Punkte *I*nvolve - Einbezug von Stakeholdern und Datenspendern und *C*onsent - Vertragsschluss mit den Datenspendern eine stärkere Auseinandersetzung mit diesen. [DG16]

Tsai und Gasevic [TG17] reviewten neben der Arbeit von Steiner et al. [SKRA16] und Drachsler und Geller [DG16] sechs weitere Learning Analytics Policies und bereiteten auf, welche Dimensionen von Learning Analytics diese behandeln. Aus diesen werden keine neuen Anforderungen an das Zustimmungsmanagement abgeleitet, sie heben jedoch die Relevanz von Datenschutz, informierter Zustimmung vor Beginn der Datensammlung, Opt-Out Optionen, Transparenz und Datenhoheit für die Einführung und Nutzung von LA weiter hervor [TG17].

2.2 Vergleich zu Cookie Zustimmungs-Interfaces

Während aus TLA konzeptionell Funktionen abgeleitet werden können, welche ein Zustimmungsmanagement bereitstellen und erfüllen soll, lassen sich Anforderungen an das Design des Interfaces aus Good und Bad Practices des Designs von Cookie Zustimmungs-Interfaces ziehen. Analysen solcher Interfaces, wie z. B. von Parrilli und Hernández-Ramírez [PHR22] oder Habib et al. [Hab+22], identifizieren die Verwendung von Dark Patterns und stellen literaturbasiert Guidelines auf, um diesen entgegen zu wirken. Darunter fallen u. a., dass keine Zustimmungsoptionen vorausgewählt und alle Buttons gleich gestaltet sein sollen.

Andere Guidelines, wie z. B. dass das Interface nicht den Blick auf den Inhalt einer Webseite verhindern soll [PHR22], werden hier vernachlässigt. Es wird davon ausgegangen, dass der Zugriff auf ein LMS bzw. einen Lernraum nicht zum Zweck des *Reinschnupperns* stattfindet und die Gestaltung des Lernraums nicht darauf ausgerichtet wird, Lernende langfristig auf der Webseite halten zu wollen. Stattdessen legt die Studie von Habib et al. nahe, dass Zustimmungs-Interfaces markant bzw. blockierend gestaltet werden sollen, um das Bewusstsein der vorhandenen Datensammlung und -verarbeitung bei den Nutzenden zu fördern [Hab+22]. Auch Singh et al. [Sin+22] untersuchten die Präferenzen der Interface-Gestaltung. Der häufigste angegebene Grund, aus dem Nutzenden ein Design missfiel, war die Darstellung zu vieler Informationen. Dem folgten der zeitliche Aufwand der Nutzung und fehlende Informationen bzw. Transparenz. [Sin+22]

2.3 Anforderungen an das Zustimmungsmanagement

Aus den in [TG17] zusammengetragenen Guidelines und Policies zu TLA werden sechs funktionale Anforderungen (F1-F6) an das Interface des Zustimmungsmanagements abgeleitet. Dieses soll ...

F1	... den Zweck der Datensammlung und -verarbeitung vermitteln
F2	... die Hoheit der Lernenden über die eigenen Daten verstärken
F3	... Transparenz erhöhen
F4	... informierte Zustimmung verbessern
F5	... Lernenden jederzeit einen Opt-Out ermöglichen
F6	... Lernende über Änderungen der Datensammlung und -verarbeitung informieren und bei Bedarf neue Zustimmungen erfragen

Tab. 1: Funktionale Anforderungen an das Zustimmungsmanagement

Zusätzlich lassen sich Design-Anforderungen (D1-D7) aus den UI und UX Guidelines zu gutem Design von Cookie-Interfaces ableiten, welche die Nutzung und Nutzbarkeit des Zustimmungsmanagement fördern sollen (siehe Tabelle 2).

D1	Das Interface soll die Nutzung der Seite blockieren um die Auseinandersetzung mit dem Zustimmungsmanagement zu erzwingen [Hab+22]
D2	Es sollen präzise und klare Texte zur Erklärung genutzt werden, was Cookies (hier xAPI Statements (siehe Abschnitt 3) und Learning Analytics) sind und für welchen Zweck sie verwendet werden [Hab+22; PHR22; TG17]
D3	Das Interface soll visuell ansprechend gestaltet sein: Es soll Nutzende anziehen und ihnen helfen den Nutzen zu verstehen [Hab+22; PHR22; Sin+22]
D4	Dark Patters sollen vermieden werden [Hab+22; PHR22]
D5	Anwendung etablierter UI und UX Guidelines, wie z. B. Morvilles UX Honeycomb oder der Gestalt Laws [Hab+22; PHR22]
D6	Zustimmungen sollen in Gruppen zusammengefasst werden [Hab+22; Sin+22]
D7	Aktionen sollen rückgängig gemacht werden können [Hab+22]

Tab. 2: Design-Anforderungen an das Zustimmungsmanagement

3 Technische Infrastruktur

Zusätzlich zu den 13 Anforderungen an die Funktionalität und das Design des Interfaces muss die technische Infrastruktur, in die sich das Interface integrieren soll, berücksichtigt werden. Die technische Grundlage für die betrachteten Learning Analytics, stellen Moodle und Excalibur LA dar. Moodle ist das zentrale Lernmanagementsystem der RWTH Aachen, in dem pro Semester über 47.000 Studierende in 3500 bis 3700 Kursen teilnehmen. In ausgewählten Kursen werden seit Oktober 2021 ein Plugin zum Logging von Studierenden-Interaktionen für Learning Analytics [JSS22] und ein Dashboard für die Anzeige der Analyseergebnisse eingebunden. Verarbeitet werden die Logs außerhalb von Moodle, in der Learning Analytics Infrastruktur Excalibur LA [JS22]. Diese nimmt die Logs von Moodle als xAPI Statements entgegen und speichert sie in einem Learning Record Store (LRS). Alle 24 Stunden werden die Statements der letzten 24 Stunden von verschiedenen Analytics Engines verarbeitet, die ihre Ergebnisse in einen Result Store schreiben, auf welchen Plattformen wie Moodle über eine REST API zugreifen können. Diese authentifiziert anfragende Plattformen und gibt die freigegebenen Ergebnisdaten aus.

3.1 Rights Engine und Zustimmungsschema

Das Konzept von Excalibur sieht neben LRS, Analytics Engines und Result Store samt API eine Rights Engine vor, welche u. a. die Zustimmung von Studierenden zu Datensammlung und -verarbeitung verwaltet [JS22]. Diese war in der ersten Version von Excalibur noch nicht umgesetzt und im Rahmen des Projekts POLARIS wurde eine erste Implementierung in Zusammenarbeit von Ruhr-Universität Bochum und der RWTH Aachen durch die Digital Learning GmbH umgesetzt [Leh23]. Anhand eines Schemas kann darin die Zustimmung von Studierenden zur Datensammlungen modelliert werden. Dieses orientiert sich an der Kombination von xAPI Verben und Objekten, welche in Gruppen zusammengefasst werden und damit eine Grundlage für Anforderung D6 bereitstellen.

List. 1 zeigt einen Ausschnitt eines Schemas für Moodle, für die neuste Version der Implementierung der Rights Engine. Dieses besteht aus einer ID sowie einem Namen und einer Beschreibung. Letztere sind Objekte mit Sprach-Tags nach den IETF Language Tags [PD09] als Schlüsseln und lokalisierten Texten als Werte. Dies orientiert sich an der Definition von xAPI Ressourcen in Registries[3] und liefert so Frontends nötige Optionen zur Lokalisierung mit, ohne dass die einzelnen Anwendungen diese selbst bereitstellen und warten müssen. Das vierte Attribut im Beispiel beschreibt die drei Gruppen des Schemas. Es sind Schemas für die Interaktionen mit und Beiträgen zu Kursmaterialien und -aktivitäten (*course_module_interactions*), Quiz Analytics und Media Analytics gegeben. Jede Gruppe hat eigene Label, Beschreibungen und eine Angabe des Zwecks der Datensammlung. Abschließend werden die Verb-Objekt-Kombinationen gelistet. Auf oberster Ebene stehen die Verben, unter denen Objekte entweder anhand von Typen oder ihrer ID aufgeführt werden. So ist im Beispiel das Verb *finished* mit den zwei Objekttypen *audio* und *video* verknüpft und Lernende können diesen Kombinationen einzeln zustimmen oder das Verfolgen ihrer fertig abgespielten Medien ablehnen.

```
"id": "moodle-schema",
"name": { "de-DE": "Moodle Schema", "en-US": "Moodle Schema" },
"description": {
 "de-DE": "Schema für Learning Analytics in Moodle.",
 "en-US": "Schema for Learning Analytics in Moodle."
},
"groups": [
 { "id": "course_module_interactions", ... },
 { "id": "quiz_analytics", ... },
 {
  "id": "media_analytics",
  "label": { "de-DE": "Media Analytics", "en-US": "Media Analytics" },
  "description": {
   "de-DE": "Die Analyse von Video und Audio Interaktionen ...",
   "en-US": "..."
  },
  "purposeOfCollection": {
   "de-DE": "Die Identifikation oft abgespielter Medieninhalte ...",
   "en-US": "..."
  },
  "verbs": [{
   "id": "https://xapi.elearn.rwth-aachen.de/definitions/media/verbs/
       finished",
   "label": { "de-DE": "beendete", "en-US": "finished" },
   "description": { ... },
   "allowAnonymizedCollection": true,
```

3 Beispiel: `https://xapi.elearn.rwth-aachen.de`

```
    "objects": [{
     "id": "https://xapi.elearn.rwth-aachen.de/definitions/media/
         activities/audio",
     "label": { "de-DE": "Audio", "en-US": "Audio" },
     "matching": "definitionType",
     "definition": { ... }
     }, {
      "id": "https://xapi.elearn.rwth-aachen.de/definitions/media/
          activities/video",
      ...
     }]
    }, {
     "id": "https://xapi.elearn.rwth-aachen.de/definitions/media/verbs/
         interacted",
    }, ...
   ]
  }
]
...
```

List. 1: Auszug eines Schemas für die Datensammlung in Moodle

Im Fall von Studien, in denen die Datensammlung abgestimmt und erlaubt ist, können Verb-Objekt-Kombinationen durch das *essential* Attribut im Objekt als nicht abwählbar markiert werden. Diese werden dann stets gesammelt, da das gesetzte Attribut impliziert, dass keine Zustimmung dafür abgefragt werden muss. Das *allowAnonymizedCollection* Attribut gibt an, dass auch wenn das Loggen einer Kombination nicht erlaubt ist, diese trotzdem mit einem anonymen Akteur geloggt wird. Dies setzt in einigen Institutionen die Erfüllung weiterer Bedingungen, wie z. B. eine definierte Gruppengröße, voraus, die durch zusätzliche Attribute im Schema abgebildet werden können.

3.2 Zentrale Zustimmungsverwaltung und Limitation

Mit der ersten Version der Rights Engine wurde auch ein Frontend entwickelt, in dem Lernende ihre Zustimmungen für alle gegebenen Schemas verwalten können. Gleichzeitig dient dieses Interface auch für die Bereitstellung von Datenauskünften und das Anstoßen von Datenlöschaufträgen durch Lernende. Diese Möglichkeiten sind wünschenswert, wenn Lernende sich explizit mit dem Thema Datenschutz, Learning Analytics und ihren Zustimmungen auseinandersetzen möchten. Ein zukünftiges Feature in diesem Frontend könnte auch das Erteilen der Erlaubnis zum Verknüpfen von Daten verschiedener Plattformen sein, damit plattformübergreifende Analysen ermöglicht werden.

Doch um schnell Zustimmungen zu verwalten und einen Systemwechsel zu vermeiden, soll eine Verwaltungsschnittstelle in die jeweiligen Plattformen integriert werden, in denen Daten für LA anfallen.

4 Drei Komponenten für informierte Zustimmungen

Die Anforderungen F1, F3 und D3 verlangen nach klaren und vollständigen Informationen, die Lernenden vermittelt werden müssen. Diese sollten jedoch nicht das Erteilen bzw. Anpassen von Zustimmungen behindern. Daraus entstand der Ansatz, das Zustimmungsmanagement in mehrere Komponenten aufzuteilen, von denen zwei, *Initial Consent Component* (ICC) und *Adjust Consent Component* (ACC), für das Management der Zustimmungen verantwortlich sind und eine dritte, *Educational Component* (EC), für die ausführliche Aufklärung der Lernenden. Die **ICC** soll Lernenden beim ersten Aufruf der Plattform und jedes neuen Lernraums angezeigt werden und die Zustimmungen für die Datensammlung abfragen. Sie soll einfach und schnell zu bedienen sein und die Zustimmungen anhand der definierten Gruppen abfragen. Um die Lernenden auf das Vorhandensein der Learning Analytics auf der Plattform hinzuweisen, soll die ICC beim ersten Aufruf des LMS blockierend dargestellt werden (D1) und Standard bzw. Plattform-globale Einstellungen einfordern. Diese globalen Einstellungen können später per Klick angewandt werden, wenn zum ersten Mal auf einen Kurs zugegriffen wird. Feinere Einstellungen sollen über einen Link auf die **ACC** ermöglicht werden. Diese soll an einer zentralen Stelle im LMS bereitgestellt und jederzeit von Lernenden erreicht werden können. Es wird erwartet, dass diese Anpassungen mit der Bereitschaft, sich intensiver mit den Paketen auseinander zu setzen, durchgeführt wird. Entsprechend sieht die ACC mehr Informationen, wie z. B. eine ausklappbare Liste der Verb-Objekt-Kombinationen pro Paket, vor, in welcher zudem einzelne Kombinationen abgewählt werden können. Dies soll jedoch nicht den vollen Umfang der EC ersetzen. In der **EC** sollen u. a. Verarbeitungsprozesse vollständig wiedergegeben werden, sodass klar kommuniziert werden kann, wie einzelne Verb-Objekt-Kombinationen zu einem Analyseergebnis führen oder beitragen. Eine Möglichkeit der Darstellung könnte durch eine Integration der EC in ein Dashboard erfolgen [Jon19]. Für jede Visualisierung könnte eine Info-Box bereitgestellt werden, welche wiedergibt, welche Daten von Lernenden für diese Darstellung verwendet wurden oder, falls nicht erteilt, welche weiteren Zustimmungen benötigt werden, um die Darstellung zu ermöglichen.

4.1 Paper Prototypes und Evaluation

Im ersten Schritt dieser Arbeit wurden zunächst die ICC und ACC umgesetzt. Um initiales Feedback von potentiellen Nutzenden einzuholen, wurden Paper Prototypes (PaPr) erstellt und 13 Studierenden zur Evaluation vorgelegt. Alle Teilnehmenden gaben an, Moodle nahezu täglich zu nutzen, berichteten aber unterschiedliche Interaktionen mit Cookie Interfaces. Dabei wurde die ganze Spanne vom Setzen der privatsphärefreundlichsten Option bis

hin zur schnellsten, selbst wenn dies die privatsphäreunfreundlichste Option bedeutet, abgedeckt. Da das ICC beim erstmaligen Aufruf des LMS anders dargestellt werden soll, um Standard-Einstellungen einzufordern, als bei erstmaligen Zugriffen auf einen Lernraum, wurden für beide Szenarien je zwei Prototypen erstellt. Die blauen Dreiecke stellen im Folgenden Platzhalter für Tooltips dar. Abbildungen 1 und 2 zeigen die Varianten für den erstmaligen LMS Zugriff. Um die initialen Annahmen zum Aufstellen von Anforderung D1 zu evaluieren, ist der Dialog in Abbildung 1 blockierend, während in Abbildung 2 dieser nur in einer Ecke dargestellt wird und auf größeren Bildschirmen noch weniger Inhalte der Webseite überdecken würde. Dieses zweite Konzept wurde für die Darstellung im Lernraum übernommen und noch zusätzlich eine stärker vereinfachte Version, wie in Abbildung 3 dargestellt, ausgehändigt. Die Prämisse war, dass Lernende beim erstmaligen LMS Zugriff über die LA informiert werden, ihre Zustimmung nicht mehrfach wiederholt eingeben wollen und eine einfache Möglichkeit, die Standard-Einstellungen zu übernehmen, gegeben werden soll.

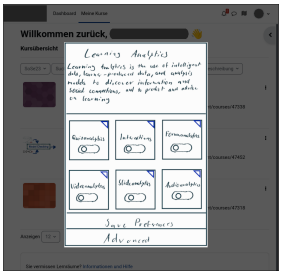


Abb. 1: ICC PaPr für erstmaligem LMS Zugriff nach D1

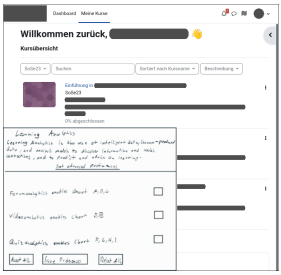


Abb. 2: ICC Alternativer PaPr für erstmaligem LMS Zugriff

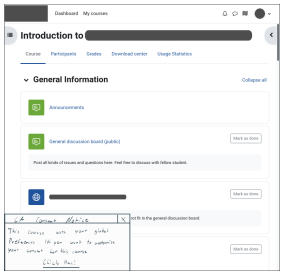


Abb. 3: ICC Alternativer PaPr für erstmaligem Lernraum Zugriff

Für die ACC waren zunächst auch zwei Ansichten geplant: Eine Basis-Ansicht, in welcher wie in der ICC Zustimmungen nur auf Gruppen-Ebene verwaltet werden können, und eine erweiterte Ansicht, in der bis zu den einzelnen Verb-Objekt-Kombinationen Zustimmungen angepasst werden können. Abbildungen 4 und 5 zeigen die beiden Prototypen für die erweiterte Ansicht. Verschiedene Elemente werden hier mit der Intention, Wünsche der Studierenden über die Inklusion der verschiedenen Elemente zu sammeln, gegenübergestellt. So können z. B. in Abbildung 4 Einstellungen für einen Kurs auch direkt global, also für die Standard-Einstellungen, übernommen werden, während in Abbildung 5 die Gruppierung angepasst werden kann, sodass Lernende Objekte als übergeordnete Kategorie einstellen und die zugehörigen Verben betrachten können. Gemeinsam haben die Ansichten, dass Pakete unten eine Fortschrittsleiste darstellen, welche angeben soll, wie viele Verb-Objekt-Kombinationen gesammelt und verarbeitet werden dürfen. Dies kann als Indikator genutzt werden um Lernenden zu kommunizieren, ob das Paket bzw. die darin enthaltenen Analysen in vollem Umfang genutzt werden können oder sie nochmal die gegebenen Zustimmungen betrachten sollten.

Die Evaluationssessions dauerten zwischen 30 und 45 Minuten, in denen die Studierenden vier Aufgaben, beginnend beim Setzen von Standard-Einstellungen, bis zum nachträglichen

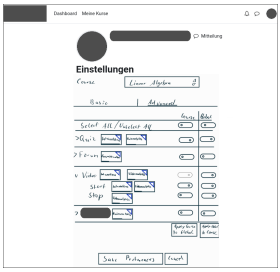


Abb. 4: ACC PaPr



Abb. 5: ACC Alternativer PaPr

Zustimmen zu fehlenden Verb-Objekt-Kombinationen für ein vorgegebenes Paket, gestellt bekamen. Unter Anwendung der Think Aloud Methode wurde qualitatives Feedback durch Notizen und Audioaufnahmen [JZ24] gesammelt. Während den Studierenden bei der Aufgabenbearbeitung nur eine, aber immer die gleiche, Version der beiden Prototypen pro Ansicht ausgegeben wurde, wurden im Anschluss beide Versionen ausgegeben und die Studierenden gebeten, ihre Gedanken zu äußern und eine für sie ideale Darstellung, ggf. als Kombination der beiden Versionen, zu beschreiben.

Insgesamt wurden die meisten Prototypen als intuitiv verständlich bewertet. Dies bestätigte sich durch die gute Task Performance der Studierenden. Fast alle Aufgaben wurden korrekt gelöst und die meisten Studierenden waren sich sicher, wann sie die Aufgabe erfüllt hatten. Probleme gab es nur bei der erweiterten Ansicht der ACC. Da in den Prototypen nur grobe Beschreibungen von Learning Analytics verwendet wurden, war xAPI nicht allen Studierenden ein Begriff, was das Verständnis der Darstellung erschwerte. Der Wunsch nach mehr Informationen wurde auch allgemein geäußert. Ein Vorschlag zur Umsetzung, welche die gegebenen Darstellungen nicht überlädt, war die Verlinkung einer separaten Webseite mit ausführlichen Informationen zu Learning Analytics und xAPI Statements bzw. wie die Verb-Objekt-Kombinationen zu verstehen sind. Weiterhin bevorzugten die Studierenden die Darstellung des ICC wie in Abbildung 1 (was die Relevanz von Anforderung D1 bestätigt), wobei für die Darstellung in einem Lernraum die Version aus Abbildung 3 bevorzugt wurde, sofern der Text des Links angepasst würde. *Click Here!* wurde als "scammy" und "ähnlich wie in einer Spam Mail" beschrieben und stattdessen *Customize* vorgeschlagen. Weiterhin wurden die Wünsche geäußert, Standard-Einstellungen direkt auf den Kurs anwenden zu können, dass das ICC in einem Lernraum nach einer Zeit verschwindet, wenn man nicht damit interagiert und Standard-Einstellungen gegeben sind und mit diesen auch angeben zu können, dass sie in jedem Lernraum genutzt werden, sodass beim ersten Zugriff gar kein ICC mehr erscheint. Zugriff auf das ACC wurde sich an verschiedenen Stellen gewünscht. Am häufigsten wurde ein Link auf dem Moodle Dashboard, welches nach einem Login angezeigt wird, oder innerhalb der Einstellungen eines Accounts genannt. Die Einbindung der Standard-Einstellungen im Dropdown-Element in der ACC wurde der Nebeneinanderstellung (Abbildung 4) vorgezogen. Im Gegensatz zur ICC wurde in der ACC die Option der *Select All* dem direkten Bestätigen mit *Accept all* vorgezogen.

4.2 Umsetzung als Moodle Plugin

Mit dem Feedback der Evaluation wurde ein erster Software-Prototyp erstellt, welcher den geäußerten Bedürfnissen an Funktionalität und Gestaltung entspricht (D3, D5). Abbildung 6 zeigt die ICC beim ersten Zugriff auf das LMS. Diese ist als Modal umgesetzt, für welches Moodle bereits ein Modul bereitstellt, welches die weitere Nutzung der Seite blockiert, bis Einstellungen vorgenommen wurden (D1). Weitere graphische Elemente, wie Buttons und die Toggle-Elemente werden ebenfalls aus Moodle Stylings entnommen, um das Verständnis der Nutzung zu erhöhen. Oben ist weiterhin ein kompakter Informationstext (D2) angedacht und die Möglichkeit gegeben, über einen Link auf eine Webseite oder die Educational Component mit weiteren Informationen zu gelangen (F1 bis F4). Die Darstellung der Pakete ist aus dem Paper Prototypen übernommen (D6) und die Buttons um die Optionen allen direkt zuzustimmen oder diese abzulehnen erweitert. Alle Buttons sind gleich gestaltet und keine Pakete vorausgewählt (D4). Beim ersten Zugriff auf einen Lernraum wird mittels eines Bootstrap Alerts das in Abbildung 7 dargestellte Popup eingeblendet. Die direkte Übernahme erfolgt durch das Klicken des *Okay* Buttons oder automatisch nach 15 Sekunden. Mit einem Klick auf den *Customize* Button wird ein Modal ähnlich zu dem in Abbildung 6 geöffnet, hier lediglich um einen Button *Use default preferences* erweitert.

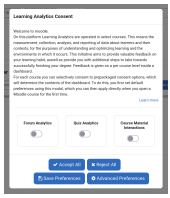


Abb. 6: ICC Modal für Standard-Einstellungen

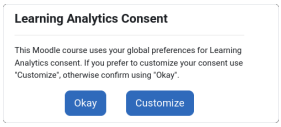


Abb. 7: ICC Popup im Lernraum

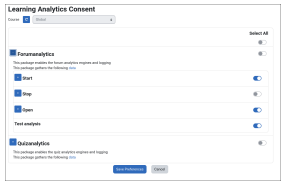


Abb. 8: Neugestaltete ACC

Die beiden Ansichten der ACC wurden basierend auf Vorschlägen aus der Evaluation in einer einzigen Ansicht zusammengeführt. In Abbildung 8 ist oben in der Ansicht ein Dropdown zur Auswahl eines Kurses oder der Standard-Einstellungen gegeben, dem die Auflistung der Gruppen des genutzten Schemas folgen. Gruppen können aufgeklappt werden, was die genutzten Verben auflistet. Diese können ebenfalls aufgeklappt werden um die Objekte anzuzeigen. Das Deaktivieren eines Verbs deaktiviert alle zugehörigen Objekte. Mit einem Toggle können alle Gruppen an- oder abgewählt werden, was jedoch noch mit dem *Save Preferences* Button bestätigt werden muss. Die ACC ist über die Account Einstellungen zu erreichen, was beim verwendeten Moodle-Theme zwei Klicks entspricht, egal, auf welcher Unterseite man sich befindet. Somit können Lernende ihre Zustimmungen jederzeit anpassen und zurückziehen (F5, D7). Zudem wurde eine mögliche Darstellung von Datenverarbeitungen (in Abbildung 8 die *Test analysis*) exploriert.

5 Evaluation des Plugins und Nachbereitung

Der erstellte Software-Prototyp wurde mit zwölf Studierenden evaluiert, von denen neun bereits an der Evaluation der Paper Prototypes teilgenommen hatten. Erneut wurden Aufgaben gestellt, welche die Nutzung aller Funktionen des Plugins fordern sollten. Dabei wurden Task Performance, Anzahl der Fehler und Time to Recover gemessen. Durchschnittlich wurden fünf der sechs Aufgaben pro Session erfolgreich abgeschlossen. Problematisch stellte sich die Suche nach der ACC dar. Es wurde im Kurs selbst, dem Moodle Dashboard und der Liste der eigenen Kurse danach gesucht. Da diese Fehlnavigation als Fehler gezählt wurden, war die Fehlerzahl in zwei Aufgaben höher als in den restlichen. Die Time to Recover war jedoch stets gering. Abschließend wurde noch der System Usability Scale (SUS) Fragebogen ausgegeben. Auch wenn die darin enthaltene Frage nach einer möglichen Steigerung der eigenen Produktivität durch die Nutzung des zu evaluierenden Tools nicht gänzlich in den Kontext des Plugins passt, stellen vergleichbare Messwerkzeuge eine ähnliche Frage, sodass dennoch dieser Fragebogen benutzt wurde. Erwartungsgemäß wurde die Usability durch die Person, die bei der Bearbeitung der Aufgaben Schwierigkeiten hatte, schlechter bewertet (>70). Alle quantitativen Messwerte dieser Evaluation können in [JZ24] eingesehen werden. Dennoch ist ein Durchschnittswert von 85,625 gegeben, welcher als Indikator einer guten Nutzbarkeit des Plugins gewertet wird [BKM08; Bro95].

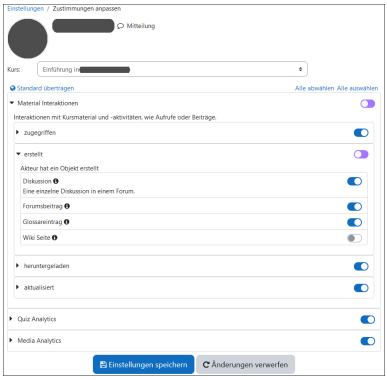


Abb. 9: Nachbereitete ACC

Lag das Plugin während der Entwicklung noch einsprachig in Englisch vor, wurden Deutsche Texte nach der Evaluation eingebunden. Aufgrund der Integration von Learning Analytics in mehreren Plattformen an der RWTH (z. B. [BJS22]) wurde das Wording der *globalen Einstellungen* zu *Standard-Einstellungen* geändert. So soll eine falsche Interpretation vermieden werden, mit der Lernende annehmen, sie würden durch das gestaltete Interface in Moodle Einstellungen für alle Plattformen gleichzeitig vornehmen. Das Design der ICC Ansichten wurde beibehalten, während die ACC eine Feinabstimmung erfuhr (siehe Abbildung 9). Statt der Buttons werden Icons zum Aufklappen verwendet und die Zustimmung zu Datenverarbeitungen wurde zunächst wieder entfernt, bis eine entsprechende Modellierung abgestimmt ist. Eine visuelle Anzeige teilweise zugestimmter Gruppen wurde durch die Einfärbung der

Toggles in Lila eingeführt, sowie das Toggle zum An- und Abwählen aller Gruppen auf zwei klickbare Textelemente aufgeteilt, um explizit Befehle darzustellen. Der Text *Standard übertragen* links ermöglicht es, die Standard-Einstellungen auf einen Kurs anzuwenden, was jedoch auch durch einen Klick des *Speichern* Buttons bestätigt werden muss. Abschließend wurde der Text des Buttons zum Zurücksetzen angepasst um die Funktionalität deutlicher zu kommunizieren. Der Code ist unter [JZ24] abrufbar.

6 Zusammenfassung, Limitationen und Weiterführende Arbeiten

Dieser Beitrag präsentierte ein Konzept, bestehend aus drei Komponenten, um Lernende über Learning Analytics zu informieren und von ihnen Zustimmungen zu Datensammlung und -verarbeitung einzuholen und verwalten zu lassen. 13 Anforderungen an Funktionalität und Design wurden von Trusted Learning Analytics und dem Design von Cookie Zustimmungs-Interfaces abgeleitet und bei der Konzeption berücksichtigt. Die beiden Komponenten zum Zustimmungsmanagement wurden als Paper und Software-Prototyp mit Studierenden evaluiert. Während in der Evaluation der Paper Prototypes qualitatives Feedback zur Gestaltung der Komponenten gesammelt wurden, fokussierte die Evaluation des Software-Prototypen die Usability. Es wurde eine gute Usability festgestellt und die prototypische Implementierung für das Lernmanagementsystem Moodle mit dem gesammelten Feedback nachbereitet, jedoch sollte das Plugin noch weiter evaluiert werden. Hauptsächlich Studierende der Informatik nahmen an der Evaluation teil, sodass Studierende anderer Disziplinen, ggf. mit anderem Verständnis solcher Interfaces andere Rückmeldungen geben könnten, die jedoch auch bei der Entwicklung berücksichtigt werden müssen, um allen Studierenden die vollständige Kontrolle der eigenen Daten gleichermaßen zu gewährleisten. Dabei wird auch Barrierefreiheit eine wichtige Rolle spielen und sollte in der nächsten Iteration berücksichtigt werden. Die genutzten Texte für die Kurzbeschreibungen der Pakete im Zustimmungsschema und auch im Plugin selbst sollten nochmal reflektiert und evaluiert werden. Dies sollte auch in die Entwicklung der Educational Component einfließen. Zudem muss eine Strategie für die Aktualisierung von Schemas und die Informierung der Lernenden darüber (F6) entwickelt werden. Sobald das Schema auch die Zustimmung zu Datenverarbeitungen abbilden kann, sollte die ACC auch dahingehen angepasst werden, z. B. wie in Abbildung 8 exploriert. Das vorgestellte Konzept, sowie die davon umgesetzten Komponenten stellen eine Grundlage für die Sicherstellung von informierten Zustimmungen für Plattformen, in denen Learning Analytics genutzt wird, dar und sollten mit den beschrieben weiteren Arbeiten fortgeführt werden.

Literaturverzeichnis

[BJS22] Brocker, A.; Judel, S.; Schroeder, U.: Integration of Gamification and Learning Analytics in Jupyter. eleed 14 (2), 2022, URL: http://nbn-resolving.de/urn:nbn:de:0009-5-55339.

[BKM08] Bangor, A.; Kortum, P. T.; Miller, J. T.: An Empirical Evaluation of the System Usability Scale. International Journal of Human–Computer Interaction 24 (6), S. 574–594, 2008, DOI: 10.1080/10447310802205776.

[Bro95] Brooke, J.: SUS: A quick and dirty usability scale. Usability Eval. Ind. 189, 1995.

[CPM21] Cerratto Pargman, T.; McGrath, C.: Mapping the Ethics of Learning Analytics in Higher Education: A Systematic Literature Review of Empirical Research. Journal of Learning Analytics 8 (2), S. 123–139, 2021, ISSN: 19297750, DOI: 10.18608/jla.2021.1.

[DG16] Drachsler, H.; Greller, W.: Privacy and Analytics: it's a DELICATE Issue: A Checklist for Trusted Learning Analytics. In: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16. ACM Press, Edinburgh, United Kingdom, S. 89–98, 2016, DOI: 10.1145/2883851.2883893.

[Eur16] Europäisches Parlament, Rat der Europäischen Union: EU Verordnung 2016/679, https://eur-lex.europa.eu/legal-content/DE/ALL/?uri=CELEX:32016R0679, Letzter Zugriff: 21.12.2023, 2016.

[Hab+22] Habib, H.; Li, M.; Young, E.; Cranor, L.: "Okay, whatever": An Evaluation of Cookie Consent Interfaces. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22, Association for Computing Machinery, New York, NY, USA, S. 1–27, 2022, DOI: 10.1145/3491102.3501985.

[Jon19] Jones, K. M. L.: Learning analytics and higher education: a proposed model for establishing informed consent mechanisms to promote student privacy and autonomy. en, International Journal of Educational Technology in Higher Education 16 (1), S. 24, 2019, ISSN: 2365-9440, DOI: 10.1186/s41239-019-0155-0.

[JS22] Judel, S.; Schroeder, U.: EXCALIBUR LA - An Extendable and Scalable Infrastructure Build for Learning Analytics. In: 2022 International Conference on Advanced Learning Technologies (ICALT). IEEE Computer Society, Bucharest, Romania, S. 155–157, 2022, DOI: 10.1109/ICALT55010.2022.00053.

[JSS22] Judel, S.; Schnell, E.; Schroeder, U.: Performantes xAPI Logging in Moodle. In: DELFI 2022: die 20. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V. S. 159–164, 2022, DOI: 10.18420/DELFI2022-028.

[JZ24] Judel, S.; Zeise, C.: Digitaler Anhang der Ausarbeitung bei OSF, https://doi.org/10.17605/OSF.IO/B4EQA, Letzter Zugriff: 21.06.2024, 2024.

[Leh23] Lehr- und Forschungsgebiet Informatik 9: Besuch der LearningAID 2023, https://learntech.rwth-aachen.de/go/id/bdjdjj?#aaaaaaaaabdjdma, Letzter Zugriff: 19.12.2023, 2023.

[PD09] Phillips, A.; Davis, M.: Tags for Identifying Languages, RFC 5646, 2009, DOI: 10.17487/RFC5646, URL: https://www.rfc-editor.org/info/rfc5646.

[PHR22] Parrilli, D. M.; Hernández-Ramírez, R.: Building a Privacy Oriented UI and UX Design: An Introduction to Its Foundations and Potential Developments. In (Martins, N.; Brandão, D., Hrsg.): Advances in Design and Digital Communication II. Springer Series in Design and Innovation, Springer International Publishing, Cham, S. 16–30, 2022, DOI: 10.1007/978-3-030-89735-2_2.

[Sin+22] Singh, A. K.; Upadhyaya, N.; Seth, A.; Hu, X.; Sastry, N.; Mondal, M.: What Cookie Consent Notices Do Users Prefer: A Study In The Wild. In: Proceedings of the 2022 European Symposium on Usable Security. Association for Computing Machinery, New York, NY, USA, S. 28–39, 2022.

[SKRA16] Steiner, C. M.; Kickmeier-Rust, M. D.; Albert, D.: LEA in Private: A Privacy and Data Protection Framework for a Learning Analytics Toolbox. Journal of Learning Analytics 3 (1), 2016, DOI: 10.18608/jla.2016.31.5.

[TG17] Tsai, Y.-S.; Gasevic, D.: Learning Analytics in Higher Education — Challenges and Policies: A Review of Eight Learning Analytics Policies. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference. ACM, Vancouver British Columbia Canada, S. 233–242, 2017, DOI: 10.1145/3027385.3027400.

Best Paper Kandidaten Kurzbeiträge

Evaluating Task-Level Struggle Detection Methods in Intelligent Tutoring Systems for Programming

Jesper Dannath [1], Alina Deriyeva [1] und Benjamin Paaßen [1]

Abstract: Intelligent Tutoring Systems require student modeling in order to make pedagogical decisions, such as individualized feedback or task selection. Typically, student modeling is based on the eventual correctness of tasks. However, for multi-step or iterative learning tasks, like in programming, the intermediate states towards a correct solution also carry crucial information about learner skill. We investigate how to detect learners who struggle on their path towards a correct solution of a task. Prior work addressed struggle detection in programming environments on different granularity levels, but has mostly focused on preventing course dropout. We conducted a pilot study of our programming learning environment and evaluated different approaches for struggle detection at the task level. For the evaluation of measures, we use downstream Item Response Theory competency models. We find that detecting struggle based on large language model text embeddings outperforms chosen baselines with regard to correlation with a programming competency proxy.

Keywords: Intelligent Tutoring Systems, Item Response Theory, Struggle, Large Language Models

1 Introduction

Programming education is in high demand due to the ongoing digitization in many fields. Additionally, programming contributes to acquiring computational thinking skills [LK14], which are valuable for many educational and societal aspects [Bo16; SSA17; SSS21]. However, the demand for programming education is not met completely due to limited resources, particularly due to the need of extensive hands-on experience. Intelligent Tutoring Systems (ITSs) can be an efficient way to increase the supply [LHL23; Mi21] and bring added value to programming education [CLW18; KF16; SSS20].

To guide task- and hint selection in an ITS, Knowledge Tracing models like Performance Factor Analysis are used to track students' competence development over time [AWN22; PCK09]. However, in their default form, such models only use the (eventual) correctness of an attempt as indicator of student competence. In programming, such indicators are very sparse as programming tasks tend to be multi-step and long (multiple minutes) compared to simpler tasks like multiple choice questions, which often take only a few seconds to complete. Moreover, there are many external resources like online forums or chat bots enabling learners to reach the correct solution with a broader spectrum of knowledge levels. Therefore, we aim to take more fine-grained information on task performance into

1 University of Bielefeld, Faculty of Technology, Inspiration 1, 33619 Bielefeld, Germany,
jdannath@techfak.uni-bielefeld.de, https://orcid.org/0009-0008-5174-8365;
aderiyeva@techfak.uni-bielefeld.de, https://orcid.org/0009-0003-6757-5722;
bpaassen@techfak.uni-bielefeld.de, https://orcid.org/0000-0002-3899-2450

doi:10.18420/delfi2024_07

account, namely indicators of struggle. In addition to evaluating existing struggle indicators, we propose a novel struggle indicator utilizing LLM-based embeddings of code-snippets. We observe that the competence estimator resulting from this indicator shows higher correlation with self-reported programming skill compared to other indicators. We evaluate all indicators on data from a classroom study with a digital learning environment for text-based programming.

2 Background and Related Work

Intelligent Tutoring Systems already have shown that providing step-level feedback can be effective [Va11]. However, regarding indicators of struggle, previous research often looked at a granularity of sessions or whole courses [Ah15; EKC17; MZ18; Ta22]. In order to use struggle indicators to give proactive feedback or to estimate competency, it would be necessary to detect struggle in real-time at the level of tasks [Do21]. Still, insights from the evaluation of struggle indicators on a higher level should not be neglected. For instance, Tabarsi et al. [Ta22] validated different struggle indicators for a two week long project in a block-based programming language. They operationalized several detectable behaviours that might indicate struggle and compared their frequency with self-reported experience of struggle by study participants. They found that unusually long idle times of more than five minutes correlate with negative self-assessments and not knowing immediately what to work on. A low number of average blocks per minute was correlated with negative feelings towards programming and a slow progress rate.

Some researchers have used a heuristic approach to determine struggle at task level based on learner submission count, but have not evaluated this approach empirically [Al18; EL16; Zh23]. In more advanced approaches to measure task struggle, researchers have determined the task-progress of a learner over time [Do21; EL16; Mc21]. Dong et al's approach in particular is based on parsing the abstract syntax tree of learner code and building a graph representation of all student attempt-progressions on tasks. The closest solution is then derived using the graph mapping costs with the current program state as introduced by the SourceCheck algorithm [PZB17]. Absolute task progress is defined as the minimal inverse distance to the assigned solution state up to a certain time t. The authors define struggling as not achieving the 25th percentile of absolute progress within two minutes of solving a task. They achieved a 77% overlap between the measures prediction of struggling and experts opinions on when to intervene. However, this approach cannot be used for text-based programming as it requires syntactically correct programs which are guaranteed in the block-based setting, but not in the text-based setting. Moreover, states of syntactical incorrectness might be especially interesting for giving feedback. Therefore, a generalization is needed and advances in the LLM-based calculation of textual embeddings might provide a suitable framework to achieve this.

Large Language Models like CodeLlama, have shown great results in many coding-related benchmarks [Ro23]. One advantage of open source models compared to closed source

coding LLMs like Codex [Li22] and AlphaCode [Ch21] is that latent embeddings of word-sequences can be generated more flexibly. We used PromptEOL, a method for generating sentence embeddings proposed by Jiang et al. [Ji23]. It showed a proficiency on Semantic Textual Similarity tasks superior to BERT, which was the previous state of the art for text embeddings with foundational models [De19]. Fine-tuned models showed better results than PromptEOL on semantic similarity tasks. However, due to the scope of our pilot study, we rely solely on PromptEOL.

3 Method

In line with prior work, we define struggle as a continued failure to progress on a task [Do21]. We hypothesize that struggle is caused by a lack of competency in task-relevant skills. We aim to provide detectors for struggle in digital learning environments for text-based programming. We adapted core concepts from prior work for the domain of beginner Python programming. We use a contrapositive approach to validate the struggle indicators. For each task-level struggle indicator, we create a scale using a two parameter logistic Item Response Theory (IRT) model [Ba01]. The IRT-model parameters are estimated using marginal maximum likelihood estimation and the girth package[2]. Under the assumption that a good indicator of struggle should allow for fitting a plausible competency model, we can deduce that an implausible competency model likely means that the underlying struggle indicator is sub-optimal as well. Implausibility of the model is evaluated based on the IRT ability parameter. As baseline struggle indicators, we adopt the eventual correctness of tasks and the lower-than-median submission count [Al18]. For all indicators, we derive a quantity indicative of task performance and turn it into an indicator of struggle via a global cutoff vale. The cutoff is always the median over all learners attempts on all tasks. We found that a per-task cutoff would be problematic, as task difficulty is then neglected.

3.1 Different struggle indicators

As a heuristic struggle indicator we used the elapsed **time until a learner first submits a correct solution**. If the time is longer than for the median of all learners on all tasks (02:01 minutes) we assume the learner is struggling. Derived from Tabarsi at al. [Ta22], we used the number of **overly long idle times** of learners as indicators of struggle. We adapted the indicator to the task level and therefore choose to lower the threshold for long idle times to two minutes. We also adopted the approach of measuring the average output of learners. Since Tabarsi et al. used a block-based programming language, we use **lines per minute** instead of blocks.

The recent advances in program synthesis and understanding achieved by open source LLMs like CodeLlama [Ro23], encouraged us to extend Dong et al.'s approach. We measure

2 https://eribean.github.io/girth/

struggle based on **LLM progress embeddings**. We calculate embeddings of the student code-states on the text-level, using PromptEOL on a quantized CodeLlama-7B model [Ji23]. The dimensionality of the derived embeddings was reduced from 4096 to 50 using a separate PCA-model for each task. The prompt for deriving the embeddings according to PromptEOL was: *[task description]\n Given the above programming task, consider this partial solution to the task:\n [code snippet]\n The current progression-state of this solution can be described in one word as: [output]*. As a second step, we used a greedy approach to match a correct solution state to each learner attempt. If a learner had reached a correct solution, we used it as solution state. Otherwise, we used the general solution with the smallest Euclidean embedding-space-distance to one of the attempt states. General solutions were derived by agglomerative clustering with single linkage of all valid solutions based on the euclidean distance between embeddings. A representative candidate is derived via the minimal distance from the cluster center. Figure 1 displays an exemplary attempt of a learner. The progress that is directly inferred from the LLM embeddings exhibits fluctuation, nevertheless, it increases alongside with the absolute progress on longer time periods. Whether a learner struggled with a particular task is then determined by whether the normalized absolute progress within the first two minutes was lower-then-median.

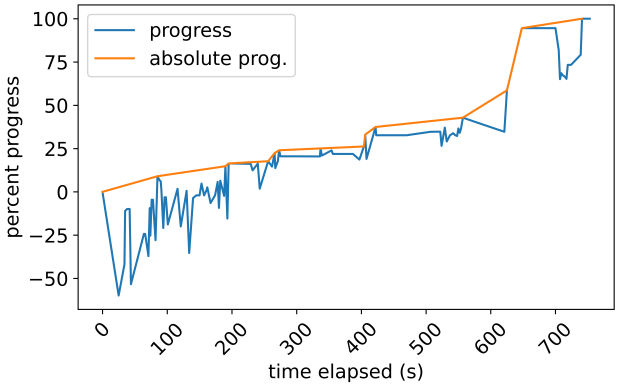


Fig. 1: Normalized progression-plot of a learners attempt at a task using LLM embeddings

Dong et al.'s distance measure is an edit distance on block-based programs, which is not directly applicable to (syntactically invalid) source code. Hence, we use a simpler edit distance baseline, namely the **absolute progress in terms of the Levenshtein distance** on the source code. In this indicator, we still use the general solutions derived from the LLM-based clustering, but the individual progress-curves are calculated solely via the Levensthein distance, utilising the edist package [PMH15].

4 Results

We evaluate the aforementioned struggle indicators on data from a classroom study conducted with a digital learning environment for programming. After obtaining a confirmation from

the local ethics committee, we recruited 97 participants from the university courses „Introduction to Machine Learning“ and „Introduction to Data Mining“, both of which are designated for bachelor level. 84 of the participants gave their consent for the scientific usage of their data. Because both of the courses require Python programming, but do not include it as a necessary prerequisite, we designed an introductory Python course consisting of 22 tasks. In addition, 55 students volunteered to fill in a pre-questionnaire, which included questions about their overall learning preferences and preliminary knowledge.

Measure	Model fit			Correlation with proxies	
Measure	log likelihood	Acc.	F1	General knowledge in programming	Used Python before?
correct eventually	-189.02	0.95	0.97	0.12	0.11
submission count	-280.65	0.86	0.89	0.26	0.16
time	-290.79	0.89	0.90	0.36**	**0.35****
overly idle	-284.73	0.88	0.91	0.22	0.32*
embedding progress	-281.72	0.89	0.91	**0.5*****	0.24
edit progress	-271.70	0.88	0.90	0.39**	0.21
lines per minute	-331.85	0.85	0.87	0.34*	0.32*

Tab. 1: Model fit of different competency measures and their correlations with competency proxies. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The left side of Table 1 shows the model fit of the IRT models on the respective struggle indicator. All of the models exhibit high accuracy, which supports the claim of their internal validity. Note, that these results are not suited for the comparison between models since all models only predict the struggle indicator they receive as input. For example, while the accuracy is the highest for „correct eventually“, this is likely due to the fact that this indicator is particularly simple to predict: most attempts (83%) were correct eventually. The right two columns of Table 1 report the Pearson correlation between the generated IRT-scales resulting from the different struggle indicators with self-reported programming skill from the survey questions. We asked the students to rate their prior programming knowledge on a four-level scale („None“, „Basic“, „Intermediate“, „Experienced“). Also, we asked them if they had used Python before on a three-level scale („No“, „Yes, a bit“, „Yes, a lot“). Both questions should to some extent relate to prior competency in the target domain (beginner Python programming). However, we believe that the first question is more indicative of competency, because it directly asks for a self-assessment of knowledge, while the Python-question is only about general usage, regardless of context, point in time, or depth. The „correct eventually“ and „submission number“ measures do not have any significant correlation with the competency proxies. Also, the „overly idle“ measure only yields a low-significance correlation with the Python usage proxy. The progress-based measures of struggle are highly related to prior programming knowledge. The LLM-embedding progress measure displays the highest correlation of all the competency measures, also exceeding the Levenshtein progress baseline by 0.11. We note that the original version of PromptEOL that features a very simple baseline-prompt yielded a 0.42 correlation with self-reported programming knowledge, showing the relevance of the chosen prompt.

5 Conclusion

We evaluated different ways of measuring struggle within digital programming learning environments at the tasks-level. Classical indicators for task assessment like eventual task correctness and submission count performed sub-optimally within our system and study population. We believe that these results align with general problems that these indicators display for the domain of programming. The time-based measure displays a good alignment with different competency proxies (Table 1). It has the advantage that it is real-time capable, enabling proactive feedback. It also mitigates the cold-start problem because one would just need to set a single initial cutoff value for the time-median. We therefore believe that using the time-based indicator would be a pragmatic choice for programming ITSs with limited resources. Furthermore, we found that our novel method which defines struggle as the lack of progress towards a goal in embedding space can potentially be a superior measure of struggle. However, measuring progress toward the correct solution requires a prediction which solution the current learner is aiming for, which may be a limitation in practice and a direction for future work.

Because the pilot study was conducted in a real-world classroom setting, there arise further limitations regarding the generalisation and scope of our results. The recruitment of our study population within two courses of a single university is not representative for programming learners in general. Therefore, our methods should only be applied with careful consideration. Next, there was no objective evaluation of competency of the learners. Our results regarding the comparison of struggle indicators rely on self-reported programming knowledge and could therefore be biased. However, we believe that a good competency measure should significantly correlate with the self-reported programming-knowledge scale. Additionally, we could not use the full potential of optimizations for the LLM-based progress measurement and struggle detection due to the limited scope of the study. In future work, we aim to apply LLM fine-tuning to improve the alignment of the text embeddings with the domain of progress measurement [GYC22]. Also, further research in struggle detection could focus on a per task self-evaluation of a sense of struggle by the learners, or incorporate sensory information. Moreover, we would like to use the struggle indicators to support automatic pedagogical decisions in ITSs.

6 Acknowledgements

This work is part of the “SAIL: SustAInable Lifecycle of Intelligent Socio-Technical Systems” project. SAIL is funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia (grant no. NW21-059A). We gratefully acknowledge this support.

Literaturverzeichnis

[Ah15] Ahadi, A.; Lister, R.; Haapala, H.; Vihavainen, A.: Exploring Machine Learning Methods to Automatically Identify Students in Need of Assistance. In: Proceedings of the eleventh annual International Conference on International Computing Education Research. ACM, Omaha Nebraska USA, S. 121–130, 2015, ISBN: 978-1-4503-3630-7, DOI: 10.1145/2787622.2787717.

[Al18] Alzahrani, N.; Vahid, F.; Edgcomb, A.; Nguyen, K.; Lysecky, R.: Python Versus C++: An Analysis of Student Struggle on Small Coding Exercises in Introductory Programming Courses. In: Proceedings of the 49th ACM Technical Symposium on Computer Science Education. ACM, Baltimore Maryland USA, S. 86–91, 2018, ISBN: 978-1-4503-5103-4, DOI: 10.1145/3159450.3160586.

[AWN22] Abdelrahman, G. M.; Wang, Q.; Nunes, B. P.: Knowledge Tracing: A Survey. en, ACM Computing Surveys 55, S. 1–37, 2022.

[Ba01] Baker, F. B.: The basics of item response theory. ERIC Clearinghouse on Assessment und Evaluation, College Park, Md., 2001, ISBN: 978-1-886047-03-7.

[Bo16] Bocconi, S. et al.: EXPLORING THE FIELD OF COMPUTATIONAL THINKING AS A 21ST CENTURY SKILL. In: EDULEARN16 Proceedings. 8th International Conference on Education and New Learning Technologies, IATED, Barcelona, Spain, S. 4725–4733, 2016, ISBN: 978-84-608-8860-4, DOI: 10.21125/edulearn.2016.2136.

[Ch21] Chen, M. et al.: Evaluating Large Language Models Trained on Code, en, arXiv:2107.03374 [cs], 2021, URL: http://arxiv.org/abs/2107.03374.

[CLW18] Crow, T.; Luxton-Reilly, A.; Wuensche, B.: Intelligent tutoring systems for programming education: a systematic review. In: Proceedings of the 20th Australasian Computing Education Conference. ACE '18, Association for Computing Machinery, Brisbane, Queensland, Australia, S. 53–62, 2018, ISBN: 9781450363402, DOI: 10.1145/3160489.3160492.

[De19] Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: NAACL-HLT 2019. Bd. 1, Minneapolis, MN, USA, S. 4171–4186, 2019.

[Do21] Dong, Y.; Samiha, M.; Preya, S.; Price, T.; Barnes, T.: Using Student Trace Logs To Determine Meaningful Progress and Struggle During Programming Problem Solving. In: Proceedings of the 14th International Conference on Educational Data Mining. Paris, France, S. 439–445, 2021.

[EKC17] Estey, A.; Keuning, H.; Coady, Y.: Automatically Classifying Students in Need of Support by Detecting Changes in Programming Behaviour. In: Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education. ACM, Seattle Washington USA, S. 189–194, 2017, ISBN: 978-1-4503-4698-6, DOI: 10.1145/3017680.3017790.

[EL16] Edwards, S.; Li, Z.: Towards progress indicators for measuring student programming effort during solution development. In: Proceedings of the 16th Koli Calling International Conference on Computing Education Research. ACM, Koli Finland, S. 31–40, 2016, ISBN: 978-1-4503-4770-9, DOI: 10.1145/2999541.2999561.

[GYC22] Gao, T.; Yao, X.; Chen, D.: SimCSE: Simple Contrastive Learning of Sentence Embeddings. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Punta Cana, Dominican Republic, S. 6894–6910, 2022.

[Ji23] Jiang, T.; Huang, S.; Luan, Z.; Wang, D.; Zhuang, F.: Scaling Sentence Embeddings with Large Language Models, en, arXiv:2307.16645 [cs], 2023, URL: http://arxiv.org/abs/2307.16645.

[KF16] Kulik, J. A.; Fletcher, J. D.: Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. Review of Educational Research 86 (1), S. 42–78, 2016, DOI: 10.3102/0034654315581420.

[LHL23] Lin, C.-C.; Huang, A. Y. Q.; Lu, O. H. T.: Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. Smart Learning Environments 10, 41, 2023, DOI: 10.1186/s40561-023-00260-y.

[Li22] Li, Y. et al.: Competition-Level Code Generation with AlphaCode. en, Science 378 (6624), arXiv:2203.07814 [cs], S. 1092–1097, 2022, ISSN: 0036-8075, 1095-9203, DOI: 10.1126/science.abq1158.

[LK14] Lye, S. Y.; Koh, J. H. L.: Review on teaching and learning of computational thinking through programming: What is next for K-12? Computers in Human Behavior 41, S. 51–61, 2014, ISSN: 0747-5632, DOI: 10.1016/j.chb.2014.09.012.

[Mc21] McBroom, J.; Paassen, B.; Jeffries, B.; Koprinska, I.; Yacef, K.: Progress Networks as a Tool for Analysing Student Programming Difficulties. In: Proceedings of the 23rd Australasian Computing Education Conference. ACM, Virtual SA Australia, S. 158–167, 2021, ISBN: 978-1-4503-8976-1, DOI: 10.1145/3441636.3442366.

[Mi21] Miranda, J. et al.: The core components of education 4.0 in higher education: Three case studies in engineering education. Computers & Electrical Engineering 93, S. 107278, 2021, ISSN: 0045-7906, DOI: https://doi.org/10.1016/j.compeleceng.2021.107278.

[MZ18] Munson, J. P.; Zitovsky, J. P.: Models for Early Identification of Struggling Novice Programmers. In: Proceedings of the 49th ACM Technical Symposium on Computer Science Education. ACM, Baltimore Maryland USA, S. 699–704, 2018, ISBN: 978-1-4503-5103-4, DOI: 10.1145/3159450.3159476.

[PCK09] Pavlik, P.; Cen, H.; Koedinger, K. R.: Performance Factors Analysis – A New Alternative to Knowledge Tracing. In: 14th International Conference on Artificial Intelligence in Education. Brighton, England, 2009.

[PMH15] Paassen, B.; Mokbel, B.; Hammer, B.: A Toolbox for Adaptive Sequence Dissimilarity Measures for Intelligent Tutoring Systems, en, https://pypi.org/project/edist/, 2015.

[PZB17] Price, T.; Zhi, R.; Barnes, T.: Evaluation of a Data-driven Feedback Algorithm for Open-ended Programming. In: Proceedings of the 10th International Conference on Educational Data Mining. Wuhan, China, 2017.

[Ro23] Rozière, B. et al.: Code Llama: Open Foundation Models for Code, en, arXiv:2308.12950 [cs], 2023.

[SSA17] Shute, V. J.; Sun, C.; Asbell-Clarke, J.: Demystifying computational thinking. en, Educational Research Review 22, S. 142–158, 2017, ISSN: 1747938X, DOI: 10.1016/j.edurev.2017.09.003, URL: https://linkinghub.elsevier.com/retrieve/pii/S1747938X17300350.

[SSS20] Scherer, R.; Siddiq, F.; Sánchez Viveros, B.: A meta-analysis of teaching and learning computer programming: Effective instructional approaches and conditions. Computers in Human Behavior 109, S. 106349, 2020, ISSN: 0747-5632, DOI: 10.1016/j.chb.2020.106349.

[SSS21] Scherer, R.; Siddiq, F.; Sánchez-Scherer, B.: Some Evidence on the Cognitive Benefits of Learning to Code. Frontiers in Psychology 12, 2021, DOI: 10.3389/fpsyg.2021.559424.

[Ta22] Tabarsi, B. T. et al.: How to Catch Novice Programmers' Struggles: Detecting Moments of Struggle in Open-Ended Block-Based Programming Projects using Trace Log Data. In: Proceedings of the 6th Educational Data Mining in Computer Science Education (CSEDM) Workshop. Durham, UK, S. 57–65, 2022, DOI: 10.5281/ZENODO.6983260.

[Va11] VanLehn Kurt, K.: The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. en, Educational Psychologist 46 (4), S. 197–221, 2011, ISSN: 0046-1520, 1532-6985, DOI: 10.1080/00461520.2011.611369, URL: http://www.tandfonline.com/doi/abs/10.1080/00461520.2011.611369.

[Zh23] Zhang, Y.; Pinto, J. D.; Fan, A. X.; Paquette, L.: Using Problem Similarity- and Order-based Weighting to Model Learner Performance in Introductory Computer Science Problems. en, Journal of Educational Data Mining 15 (1), S. 63–99, 2023.

Problem-Specific Visual Feedback in Discrete Modelling

Maurice Herwig [1], Norbert Hundeshagen [1], John Hundhausen[1], Stefan Kablowski[1], and Martin Lange [1]

Abstract: Discrete modelling as the basis of problem solving is an essential skill for computer scientists, but the correct use of formal languages like propositional logic for such purposes remains a big challenge for undergraduate students. The DiMo tool provides support for the acquisition of formal modelling competencies using propositional logic. We extend the tool by generic capabilities to generate problem-specific feedback to students. This allows them to visualise the result of their modelling attempts in terms of the modelled problem at hand, thus helping students to initiate corresponding learning cycles.

Keywords: propositional logic, teaching modelling, feedback systems, visual feedback, error-driven learning

1 Motivation: Formal Modelling in Propositional Logic

An integral part of being a computer scientist is the ability to do problem solving: to devise computer-aided solutions for real-world problems. This implies some central competencies to be taught in respective courses at university level, which should bring out the ability to transfer properties of a given (real-world) problem into machine-readable form – aka formal modelling.

Here we are concerned with tool-supported learning of formal modelling in propositional logic (PL), a simple language that is widely used in computer science, from Boolean circuits to planning in A.I. [KS92], computer-aided verification [Cl01], cryptanalysis [MM00] and many more, mainly due to nowadays' access to highly efficient SAT solvers.[2]

There are many high-quality learning tools that aim at basic PL competencies such as writing, evaluating and normalising single formulas. An older overview can be found in [Hu11], a recent tool is [Ge18]. Application-oriented formal modelling via PL mainly enjoys rudimentary support only, for instances through logical quizzes, with very few notable exceptions, cf. [Fe20, GSS11].

A PL modelling problem is typically defined as an instance of the following scheme:

1 Theoretical Computer Science / Formal Methods, University of Kassel, Germany, maurice.herwig@uni-kassel.de, https://orcid.org/0009-0005-5519-3043; hundeshagen@uni-kassel.de, https://orcid.org/0009-0003-5900-7821; johnhundhausen@gmail.com; stefan.kablowski@gmail.com; martin.lange@uni-kassel.de, https://orcid.org/0000-0002-1621-0972

2 See, for instance, `http://www.satlive.org/` .

doi:10.18420/delfi2024_08

given: *a (real-world) problem A whose instances $A(\overline{p})$ depend on parameters $\overline{p}$ and are either solvable or unsolvable*
task: *construct a propositional formula (scheme) $\Phi_A(\overline{p})$ that is satisfiable* ***if and only if*** *$A(\overline{p})$ is solvable*

An exemplary exercise adhering to this definition is the so-called *n-Queens Problem*:

> ***Exercise:*** *Given some $n \geq 1$, is it possible to place n queens on the chessboard of size $n \times n$ s.t. no queen can capture another according to the rules of chess?*
> *Construct a propositional formula $\Phi(n)$ that is satisfiable if and only if the n-Queens Problem has a solution.*

Let Q be this problem. Its only parameter is n. The instances $Q(2), Q(3)$ are unsolvable, whereas $Q(1)$ and $Q(n)$ for $n \geq 4$ are known to be solvable. Logical problems other than satisfiability (validity, equivalence, model counting, etc.) can of course also serve as target problems.

We use DiMo [HLS21] for the purpose of teaching and supporting students in solving exercises like the n-Queens Problem. DiMo essentially is acting as a SAT solver interface for parametrised formulas (so-called *formula schemes*) and feedback is given in form of evaluations of propositional variables, and thus, is by all means problem-*unspecific*. Fig. 1 shows the major components of the user interface highlighted by coloured boxes. The blue box contains a text editor field for writing PL formula schemes in machine-readable format with IDE-like syntax-highlighting. In the green box, DiMo sets the formula scheme(s) using the traditional mathematical format that students would typically know from lectures. Here, it provides learning support on a syntactical level, for the competency of writing syntactically correct formula schemes. The part marked with the yellow box provides feedback on the correctness of the formula scheme, listing satisfying variable interpretations for successive formulas of the scheme – the semantical level of learning formal modelling.

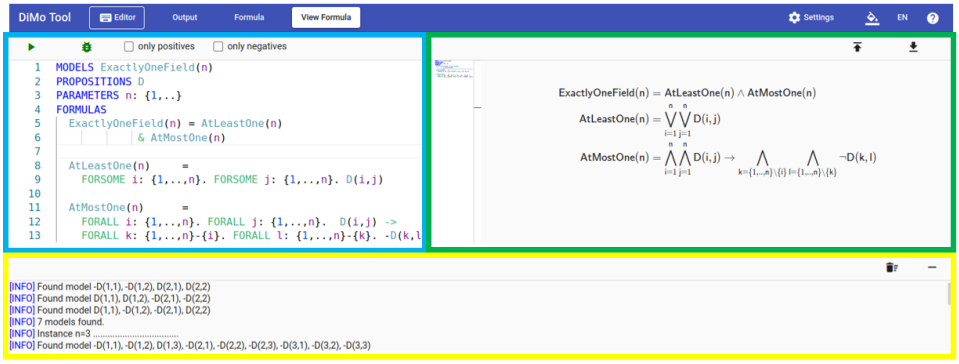


Fig. 1: The frontend of DiMo's first version.

From a didactical perspective, the provision of (tool-)assistance for the aforementioned semantical level incorporates the computation of information that concerns the ability to validate the quality of a model (here: the formula scheme). The validation of such a PL model typically means to verify that a formula (scheme) like some $\Phi_Q(n)$, modelling the n-Queens Problem, is correct in that $\Phi_Q(n)$ is satisfiable iff $Q(n)$ has a solution, validated typically by the fact that propositional evaluations of the formula instance encode particular placements of the queens.

The approach of validation in formal modelling is well researched in the field of science, technology, engineering and mathematics (cf. [BBF09, GM02]). The former describes the modelling process as a cycle that involves a repeated comparison of the model with the real-world problem; more precisely, the validation of a model-based solution in the real world is one of the major difficulties for students and, as pointed out above, one of the most important steps in modelling in PL. Admittedly, the data for the latter results are based on studies in school mathematics. We believe, based on own experience and the generality of the competencies needed for formal modelling, that these results carry over to the undergraduate setting in computer science. Thus, to support learning in this area the requirements for tool-generated valuable feedback are "[. . .] maximal students independence [. . .]" (cf. [BBF09, p.54]) and implied "strategic interventions" (cf. [BBF09, p.52]) aimed at the validation step. Typical teachers' interventions for the latter are questions that encourage students to visualise their model w.r.t. the real-world problem. Therefore, replacing this textual representation of PL models in DiMo by problem-*specific* output in the form of visual feedback specific to the modelling task is the main goal of this paper. Generally, such visual feedback plays a key role in understanding formal principles (cf. [Ar03]) and is meant to initiate the modelling cycle mentioned above.

2 Extending DiMo

DiMo's main Specs. We only provide a brief description of DiMo's essential features. A complete tool description is given in [HLS21]. From a software architecture point of view, DiMo is a classical web-application with a backend implemented mainly in OCaml[3] and a TypeScript/Angular[4] frontend. Formula schemes are written in the so-called DiMo-language.

DiMo's Novel Extension: Problem-Specific Visualisation of PL Models. The main idea leading to the ability to generate problem-specific visualisations – instead of plain propositional models – is based on an extension of DiMo's language by small imperative programs that can iterate over propositional models and print information in some interpretable output format. See the green box in Fig. 2 for an example program to produce output for the n-Queens Problem from above.

3 Objective Caml: `https://ocaml.org/`

4 `https://www.typescriptlang.org`, `https://angular.io`

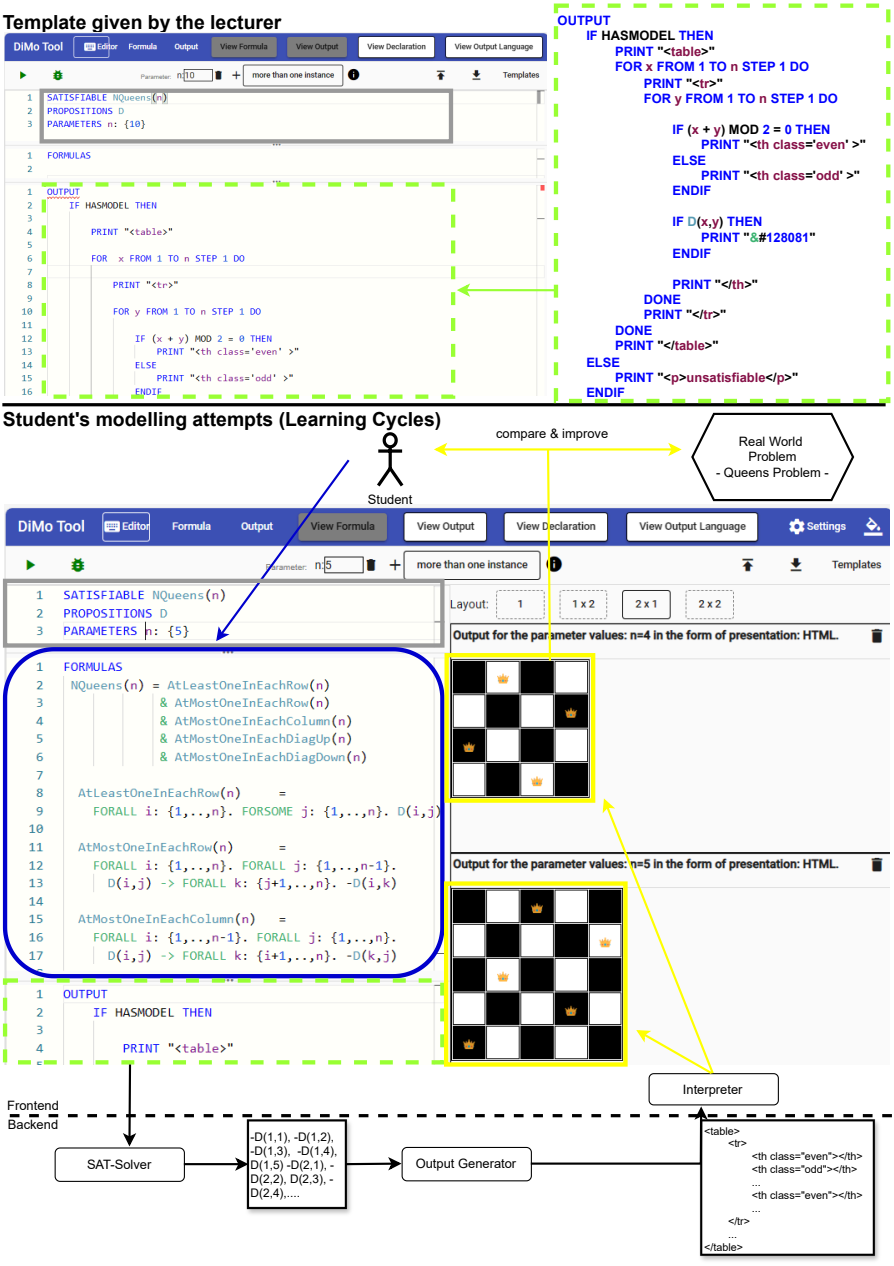


Fig. 2: How to use DiMo: (top) lecturer's creation of templates as parameter declarations etc. (grey box) and output programme (green dotted box). (Middle) Students' attempts at modelling by writing formulas in the formula scheme editor (blue box with rounded corners) and receiving graphical feedback for different instances (yellow boxes), here $n = 4$ and $n = 5$. Both are easily seen to be valid solutions. (Bottom) DiMo's technical process for creating graphical feedback.

The output is generated using for-loops over the domain of parameters (cf. lines 4 and 6 of the green box), conditional statements over Boolean expressions, in particular, arithmetic relations (line 7) and truth values of atomic PL variables (line 12). The evaluation of such programs is done in three-valued logic, as the output program may refer to PL variables that do not have truth values, for instance because they did not occur in the formula scheme created by a student. The core of the output language is a simple print-statement that allows arbitrary strings to be created (line 5 or 13).

This technically simple extension enables the use of different follow-on interpreters for the strings created by DiMo's output program for various instances of the formula scheme. The lower part of Fig. 2 shows how problem-specific visualisation can be given for the n-Queens Problem: (I) the formula scheme $\Phi_Q(n)$ (blue box with rounded corners) is instantiated in the background for some parameter value like $n = 5$, (II) a SAT solver is used to compute a satisfying variable assignment ϑ for $Q(5)$ if there is some, (III) the output program (green dotted box) is run with the same parameter evaluation $n = 5$ and ϑ to create, e.g., HTML code which (IV) is used to display the model in the foreground as feedback about the correctness of the formula scheme (yellow box).

DiMo currently supports two more output types besides HTML: plain text and VisSat, an internal development aimed at easing the visualisation of matrix-like structures. Fig. 3 shows examples of feedback given in different output types for different problems: HTML for the well-known River Crossing Puzzle,[5] plain text for the n-Queens Problem, and VisSat-output for the problem of finding a knight's tour on a chess board.[6]

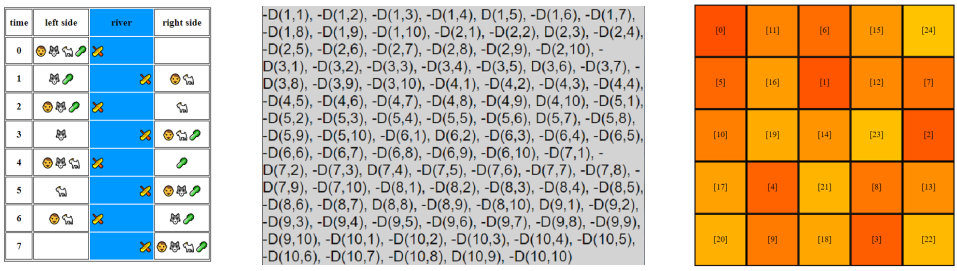


Fig. 3: Examples for displaying feedback using three different output types: HTML, plain text and VisSat.

HTML output is particularly useful for modelled problems that come with a straight-forward graphical view, like the n-Queens Problem. DiMo also allows HTML output to be equipped with additional CSS code in order to enhance visibility etc. VisSat provides a simpler output format whenever this view is based on a representation of matrices or graphs, in order to avoid hard-coding tables using *tr*- and *td*-tags. Plain text can be used to tweak the difficulty of the learning process, for instance by *not* providing a direct visualisation but requiring students to form this visualisation themselves. From a technical point of view, adding new

5 https://en.wikipedia.org/wiki/River_crossing_puzzle
6 https://en.wikipedia.org/wiki/Knight%27s_tour

interpreters for output languages to DiMo is fairly simple, only requiring their integration as an Angular component.

Usage. DiMo can be used as a self-learning tool or integrated into a course teaching formal modelling with PL. It offers a concise language to provide the audience with reasonable complex examples of PL formula schemes and live demonstrations of the effect of different modelling attempts on the problem at hand.

Even though implementing new visualisations in DiMo's output language is reasonably easy, we recommend the use of DiMo's template system, which allows instructors to provide a set of given output programs for modelling problems (cf. top of Fig. 2).

In a typical usage scenario students would be given a modelling exercise and would be required to use DiMo to create and check their solutions by using a given output program. This only works best, though, when the PL variables used to model the problem are given in the exercise, as the output program makes use of them. This has pros and cons. Picking out suitable variables is essential for formal PL modelling, and this part of the task is taken out of the students' hands. On the other hand, when the propositional variables are known, DiMo can also be used to semi-automatically verify the quality of a solution by checking it for equivalence against a master solution.

Evaluation The extension of DiMo presented in this paper is yet to be used in a wider course setting. However, to evaluate the graphical-feedback system, we conducted a small qualitative trial with 9 undergraduate and graduate students.

A two-level task was given: first, find the mistake in given propositional models; second, suggest how to fix it in the formula schemes. The models were given in textual or graphical representations. Finally, two questions were asked about the perceived difficulty of the given exercises and the helpfulness of the presentation form.

The main result taken from the latter question can be subsumed by the following quote from the answers: "*The error [in the propositional model] was obvious in the graphical representation. The more information is given in textual form, the more difficult it gets to find the error.*" Seven out of nine students answered the question accordingly, while one student found the textual representation more helpful. Moreover, some sparse data on the time spent to actually "see" the error in the representation suggests a huge impact on the time needed to validate a given model. The latter cannot be explained by a focus effect,[7] as the study group was divided into two halves that received the feedback in different orders.

Conducting a quantitative study with a larger number of participants is planned for the future in the context of an undergraduate course on formal logic.

7 Mistakes found in the first task lead to the mistakes in the second task.

Acknowledgments: The work is partially funded by *Stiftung Innovation in der Hochschullehre* within the project *UKS_digi*.

Bibliography

[Ar03] Arcavi, A.: The role of visual representations in the learning of mathematics. Educational Studies in Mathematics, 52(3), 215-241. Educational Studies in Mathematics, 52:215–241, 04 2003.

[BBF09] Blum, W.; Borromeo Ferri, R.: Mathematical Modelling: Can It Be Taught And Learnt? Journal of Mathematical Modelling and Application, 1, 12 2009.

[Cl01] Clarke, E. M.; Biere, A.; Raimi, R.; Zhu, Y.: Bounded Model Checking Using Satisfiability Solving. Formal Methods in System Design, 19(1):7–34, 2001.

[Fe20] Fernandez, J.; Gasquet, O.; Herzig, A.; Longin, D.; Lorini, Emiliano; Maris, F.; Régnier, P.: TouIST: a Friendly Language for Propositional Logic and More. In: Proc. of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020. ijcai.org, pp. 5240–5242, 2020.

[Ge18] Geck, G.; Ljulin, A.; Peter, S.; Schmidt, J.; Vehlken, F.; Zeume, T.: Introduction to Iltis: an interactive, web-based system for teaching logic. In: Proc. 23rd Annual ACM Conf. on Innovation and Technology in Computer Science Education, ITiCSE'18. ACM, pp. 141–146, 2018.

[GM02] Greca, I. M.; Moreira, M. A.: Mental, physical, and mathematical models in the teaching and learning of physics. Science Education, 86(1):106–121, 2002.

[GSS11] Gasquet, O.; Schwarzentruber, F.; Strecker, M.: Satoulouse: The Computational Power of Propositional Logic Shown to Beginners. In: Proc. of the Third International Congress on Tools for Teaching Logic, TICTTL 2011. volume 6680 of Lecture Notes in Computer Science. Springer, pp. 77–84, 2011.

[HLS21] Hundeshagen, N.; Lange, M.; Siebert, G.: DiMo - Discrete Modelling Using Propositional Logic. In: Proc. 24th Int. Conf. on Theory and Applications of Satisfiability Testing, SAT'21. volume 12831 of Lecture Notes in Computer Science. Springer, pp. 242–250, 2021.

[Hu11] Huertas, M Antonia: A classification of tools for learning logic, 2011. `http://hdl.handle.net/10609/6501`.

[KS92] Kautz, H. A.; Selman, B.: Planning as Satisfiability. In: Proc. 10th European Conf. on Artificial Intelligence, ECAI'92. pp. 359–363, 1992.

[MM00] Massacci, F.; Marraro, L.: Logical Cryptanalysis as a SAT Problem. J. of Automated Reasoning, 24(1/2):165–203, 2000.

Beyond Realism: Rethinking Presence in Virtual Environments for Abstract Concept Learning



David Baberowski [1], Thiemo Leonhardt [2], Lanea Lilienthal [1] und Nadine Bergner [2]

Abstract: Presence, as the feeling of actually being in a virtual place, is an important metric for VR learning applications due to its positive impact on learning. In the literature, the level of immersion a system can provide is assumed to be the main factor influencing presence. This paper investigates this assumption for virtual environments in abstract learning contexts that are not based on reality. This is done by building a physical replica of a VR learning game and comparing both in terms of learning outcomes and presence. The study shows that subjects do not have a consistent view of what can be described as real in abstract contexts, suggesting that the distinction between reality-based virtual environments and those in abstract contexts is important in terms of presence.

Keywords: Virtual Reality, Learning Games, Immersion, Presence, Computer Science

1 Introduction

Virtual reality (VR) makes it easier for learners to become immersed in what they're learning. [Bu23]. Immersion can be defined as an objective measure of a system's sensory fidelity and tracking ability [Sl03] and is recognized as a contributing factor to presence. There is currently no standard definition of presence. In this paper, we follow Dengel's argumentation, who describes presence similarly to Slater [SS16] as the subjective feeling of being present in a virtual environment in non-augmented environments as the essence of the current research discussion [De20]. Research suggests a positive correlation between presence, motivation and learning [Ma21], therefore making presence an important metric in the design and evaluation of VR learning applications [Ma23]. When VR is used for abstract content learning the virtual environment differs from applications that are designed to practice practical skills or simulate real environments. In abstract content learning, the virtual environment does not recreate real things and situations, but rather provides an experiential analogy that may not be based on reality at all [BLB24]. When thinking about presence, the question arises whether the learner feels present in the artificial analogy or in the abstract learning content itself. To answer this question, a study design was developed in which an abstract VR learning game from computer science was replicated in the real world as identically as possible. The difference in immersion between the real tangible

1 TU Dresden, Didaktik der Informatik, Nöthnitzer Str. 42, 01187 Dresden,
david.baberowski@tu-dresden.de, https://orcid.org/0000-0001-6308-4334;
lanea.lilienthal@tu-dresden.de, https://orcid.org/0000-0002-8508-9483
2 RWTH Aachen, Lehr- und Forschungsgebiet Didaktik der Informatik, Mies-van-der-Rohe-Str. 15, 52074 Aachen, leonhardt@informatik.rwth-aachen.de, https://orcid.org/0000-0003-4725-9776;
bergner@informatik.rwth-aachen.de, https://orcid.org/0000-0003-3527-3204

doi:10.18420/delfi2024_09

environment and the virtual one offers the opportunity to gain a better understanding of presence in abstract learning scenarios.

2 Measuring Presence

Presence measurement instruments have been the subject of research in the context of the unclear definition. In a comparison of 15 different presence measurement instruments, Schwind et al. recommended the Igroup Presence Questionnaire (IPQ) as offering the highest reliability within a reasonable time frame [Sc19]. They also noted that further research is needed to adapt presence questionnaires for VR applications. The IPQ was developed by Schubert et al. [Sc03] as a means to measure the sense of presence. To capture the multidimensionality of presence, the instrument was divided into three subscales: Spatial Presence (SP) measures the user's sense of being physically located and acting within the virtual environment. Involvement (INV) measures the user's attentional focus on the virtual environment and the involvement experienced. It assesses the extent to which the user focuses on the virtual environment rather than the real world. Experienced Realism (EXP) measures how realistic the user perceives the virtual environment. It reflects how much the virtual experience is subjectively similar to the real world. While questions related to SP and INV can be applied to both realistic and abstract concepts, questions related to EXP may be more tailored to realistic settings. In particular, items REAL2 („How much did your experience in the virtual environment seem consistent with your real world experience?“) and REAL3 („How real did the virtual world seem to you?“, with answers ranging from „about as real as an imagined world“ to „indistinguishable from the real world“) may be problematic because they don't take into account that abstract concepts are not directly experienced in the real world. The response to items on the EXP subscale depends, among other things, on whether the learner refers to the virtual environment or the represented abstract concept. Therefore the IPQ is well suited to answer the question posed about presence in abstract contexts.

3 Educational Design

For the study, two versions of an educational game were designed and implemented: a virtual one using head-mounted displays and controllers (VR) and a tangible physical replica (TG). Both are based on the existing VR game „Inside the Router“ [LBB22; Le23]. Here the learner is placed inside a router and has to perform basic routing tasks with data packages displayed as capsules in a tube post system. For the study the VR and TG versions should be as similar as possible, in order to reduce the influence of design differences on the measurements. To this end, the aspects of *objects*, *interactions* and *feedback* were analyzed and adapted as the main factors influencing the learning game mechanics.

As part of the *in-game object* analysis, physical equivalents for the assets (3D models of the capsules and tubes) were created using 3D printers based on the original version to reproduce

shape and color as faithfully as possible (see 1). Deviations exist in the scaling of the objects, as the distortion of size perception for objects in VR is a well-known phenomenon [ONH17]. At the same time, scaling has an impact on the complexity of manufacturing the TG version, so different scaling options were tested during the design phase in order to minimize the manufacturing effort while maintaining the subjective comparability of the virtual and physical interaction objects. In both versions, the tubes were arranged in a semicircle, naturally defining the playing area. While the rest of the room was kept simple and dark in the VR version, black curtains were used to shield the participants from the rest of the room in the tangible version. This meant that the instructor, who was controlling the physical version from behind the scenes, was not visible to the participants.



Abb. 1: Comparison of the assets of the VR version (left) and the tangible version (right)

To make the *interactions* as similar as possible, the task of routing is executed by simply placing the capsules in the tubes. In the VR version, an animation was used to make the capsules appear. The capsules could then be taken out of the tube with a grab interaction using hand controllers and placed in another tube. In the tangible version, the animation was implemented using a miniature conveyor belt behind the scenes that transported the capsule out of the tube and into the play area. The conveyor belt was controlled by a microcontroller equipped with a distance sensor that could detect the removal and insertion of capsules into each tube. The sequence and addresses of the data packets were identical in the VR and TG versions. During the game, it was the task of the instructor behind the scenes to load the tubes with the next capsules. However, the conveyor belt meant that the participants never came into direct contact with the instructor, so the feeling of interacting with an automated machine (a router) as in the VR version was maintained.

A central component of the learning application is the *direct feedback* on the player's actions. Participants receive immediate feedback whenever a capsule is placed in a tube. Since there is no other external intervention during the learning task, the feedback is the only control mechanism and therefore an important aspect of the pedagogical design. Participants receive feedback via red and green lights after placing a capsule, combined with a positive

or negative sound, depending on whether the placement was correct. In addition, if a capsule is placed in the wrong tube, the tube rejects the package and sends it back to the learner to try again. These features were implemented in both the VR and TG versions. However, due to limitations in the design of the tubes, the feedback from the lights was much less pronounced than in the VR version. However, since the ejection of the packages via the conveyor belt and the output of the sounds via the loudspeakers provided clear feedback, the two versions are probably comparable in this respect.

4 Study Results

The study was conducted in a cross-over design with 23 students (convenience sampling). The study procedure is shown in figure 2. First, a video was shown to the test subjects to activate their prior knowledge of networks. This was followed by the pre-test, after which the first intervention in VR or TG was randomly selected and carried out. The post-test followed directly after the intervention. Then the second intervention with the other version and another post-test concluded the study. Since the learning effects were not independent, only the first intervention was evaluated for competencies making this part into a classical pre-post design.

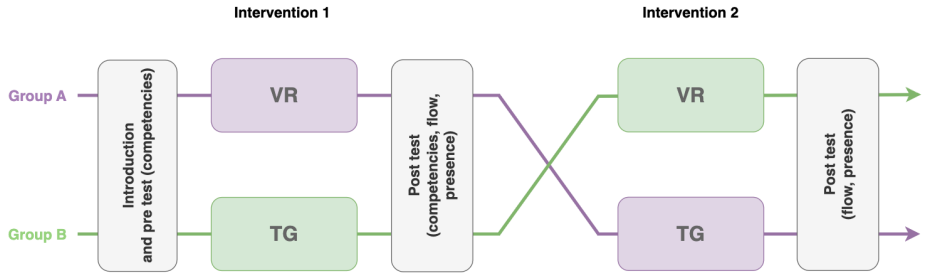


Abb. 2: Study design with three questionnaires and two interventions

Learning Effect: The items, used to measure learning effects were developed and tested by experts for their alignment with the learning objectives of Inside the Router. The questionaire including these tasks can be found together with the study data[3]. The test subjects answered around 79% of the tasks correctly in the pre-test, but this proportion increased significantly in the post-test to 90% for the VR application (N=11, one-sided Wilcoxon signed rank test, p=0.003) and 84% for the TG version (N=12, not significant, p=0.12). This indicates a saturation and that the tasks were not challenging enough for the test subjects (see fig. 3). While some tasks focus on partial and transfer competencies, task 6 was most similar to the actual activity during the game: assigning given packages to the correct clients, using their IP addresses. This task showed the biggest effect, with increases in correct answers for the

3 https://osf.io/2kvnm/?view_only=b57c19b40fbd4cbbaafd63b72dbc5544

group of VR players from 50% to 74% (r=0.89, p=0.004) and for the TG players from 51% to 68% (r=0.70, p=0.037).

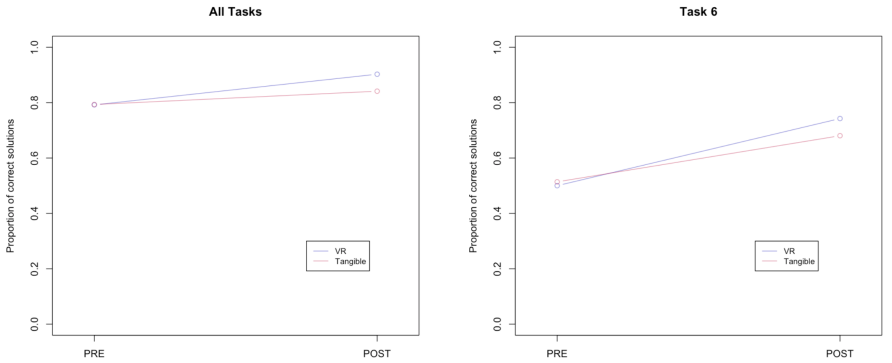


Abb. 3: Differences in correct solved tasks for VR and tangible versions.

Flow: The Flow-Kurz-Skala (FKS) according to Rheinberg [RVE03] was used to measure flow. This scale refers to the experience of flow on a 7-point scale (0 to 6). Since both versions were build to be as similar as possible in terms of learning experience, they were expected to be similar. The flow values of VR (4.8) and TG (4.6) do not differ significantly (N=23, Wilcoxon signed-rank test, $\alpha = 0.25$, $1 - \beta = 0.86$), even though there could be an effect of the game version on the flow, which we were not able to measure due to the small sample size.

Presence: Presence was measured, using the IPQ [Sc03] and was asessed on a 7-point scale (0 to 6). The versions differ significantly and show strong effects in the SP (VR: 4.67 ± 0.98, TG: 3.86 ± 1.22, Wilcoxon signed-rank test, p=0.01, r=0.55) and INV (VR: 3.66 ± 1.21, TG: 2.23 ± 1.61, Wilcoxon signed-rank test, p=0.002, r=0.66) subscales, however not in terms of EXP (VR: 2.24 ± 1.17, TG: 2.73 ± 1.38). When looking at the distribution of shifts between VR and TG the EXP subscale does not show a normal distribution which could suggest an inconsistent interpretation of the items by the probands (see fig. 4).

5 Discussion

The presented study argues that the implemented learning environments VR and TG, that both use the same educational design, are comparable in terms of learning effect and flow (learning experience). While they convey the same content, the environments differ in terms of realism.

In the study we measured strong effects of the different versions on presence for the SP and INV subscales. The TG version was in many regards more realistically but showed significantly less presence on these scales. This means that established assumptions about

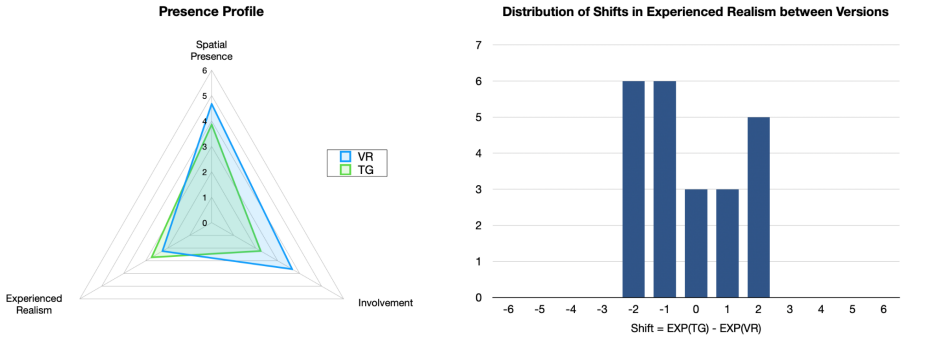


Abb. 4: Presence profiles for both applications (left) and the distribution of the shift in experienced realism between the two versions.

the relationship between realism, immersion and presence for abstract learning content in environments without direct reference to reality, should be questioned.

The results for the EXP subscale can be explained in two ways: On one side lerners may percieve the TG environment as more real than VR, because they don't use a device to percieve it. On the other hand, learners may focus on the abstract concept (the functionality of a router) and therefore percieve the TG environment as a stage, possibly percieving it as less real than the VR. The distribution of shifts between VR and TG (see fig. 4) suggests that both perspectives are present in learners. Future research should test these assumptions with qualitative studies and develop an adapted scale for presence in environments without a direct reference to reality.

Literatur

[BLB24] Baberowski, D.; Leonhardt, T.; Bergner, N.: Inside the System - Designing VR Serious Games for Computer Science Education. In (Dondio, P.; Rocha, M.; Brennan, A.; Schönbohm, A.; De Rosa, F.; Koskinen, A.; Bellotti, F., Hrsg.): Games and Learning Alliance. Bd. 14475, Springer Nature Switzerland, Cham, S. 311–320, 2024.

[Bu23] Buchner, J.: Wie Augmented Und Virtual Reality Lernen Bewirken Können. In: Didaktik in Einer Kultur Der Digitalität. Wirkmächtige Mediendidaktik, Zukunftsorientierte Pädagogik. Klinkhardt, S. 71–90, 2023.

[De20] Dengel, A.: Effects of Immersion and Presence on Learning Outcomes in Immersive Educational Virtual Environments for Computer Science Education, Diss., Universität Passau, 2020.

[LBB22] Leonhardt, T.; Baberowski, D.; Bergner, N.: Inside the Router: An interactive VR learning application to practice routing and network address translation. Gesellschaft für Informatik e.V., 2022.

[Le23] Leonhardt, T.; Lilienthal, L.; Baberowski, D.; Bergner, N.: The tension between abstract and realistic visualization in VR learning applications for the classroom./ , 2023.

[Ma21] Makransky, G.: The Immersion Principle in Multimedia Learning. In (Mayer, R. E.; Fiorella, L., Hrsg.): The Cambridge Handbook of Multimedia Learning. 3. Aufl., Cambridge University Press, S. 296–303, 2021.

[Ma23] Matovu, H.; Ungu, D. A. K.; Won, M.; Tsai, C.-C.; Treagust, D. F.; Mocerino, M.; Tasker, R.: Immersive virtual reality for science learning: Design, implementation, and evaluation. Studies in Science Education 59/2, S. 205–244, 2023.

[ONH17] Ogawa, N.; Narumi, T.; Hirose, M.: Distortion in perceived size and body-based scaling in virtual environments. In: Proceedings of the 8th Augmented Human International Conference. AH '17: 8th Augmented Human International Conference. ACM, Silicon Valley California USA, S. 1–5, 2017.

[RVE03] Rheinberg, F.; Vollmeyer, R.; Engeser, S.: Die Erfassung des Flow-Erlebens. Diagnostik von Motivation und Selbstkonzept, Tests und Trends N.F. 2/, hrsg. von J. Stiensmeier-Pelster & F. Rheinberg, S. 261–279, 2003.

[Sc03] Schubert, T. W.: The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. Zeitschrift für Medienpsychologie 15/2, S. 69–71, 2003.

[Sc19] Schwind, V.; Knierim, P.; Haas, N.; Henze, N.: Using Presence Questionnaires in Virtual Reality. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, Glasgow Scotland Uk, S. 1–12, 2019.

[Sl03] Slater, M.: A Note on Presence Terminology. Presence Connect 3/, 2003.

[SS16] Slater, M.; Sanchez-Vives, M. V.: Enhancing Our Lives with Immersive Virtual Reality. Frontiers in Robotics and AI 3/, 2016.

KI Einsatz in der Hochschule I

Entwicklung und Evaluation einer KI-Assistenz zur didaktisch-pädagogischen Unterstützung des Lernprozesses mit Programmieraufgaben

Peter Kießling [1], Florian Funke [1] und Sven Hofmann [1]



Abstract: Der Einsatz von Chat-Systemen mit KI-generierten Antworten nimmt in allen Lebensbereichen kontinuierlich zu. Auch im Schulkontext wird vermehrt auf KI als Unterstützungssystem bspw. für die Erstellung von Aufgaben zurückgegriffen [Ma23]. Chatbots wie ChatGPT liefern hierbei oft vollständige Lösungen für gestellte Aufgaben. Dadurch lösen Lernende grundlegende Aufgaben mithilfe eines Werkzeugs, welches den Anspruch an die eigene kognitive Leistung der Lernenden verringert.
Gleichzeitig eröffnet dies neue Lernansätze. Im Schulkontext ergeben sich jedoch zwei Probleme: Erstens steht die Generierung von vollständigen Lösungen einem produktiven Lernprozess entgegen. Zweitens verfügen die Lernenden unter Umständen über keinen eigenen Zugang zu solchen Systemen. Um diesen Problemen entgegenzuwirken, wird in diesem Beitrag die Entwicklung und Evaluierung eines eigenen Unterrichtswerkzeugs vorgestellt. Der sogenannte Programmier-Copilot beantwortet Fragen der Schülerinnen und Schüler und versucht dabei gleichzeitig ein hohes Maß an kognitiver Aktivierung zu erhalten. Im folgenden Beitrag werden die Entwicklung sowie die Erprobung der Wirksamkeit eines solchen Tools im Unterricht dargestellt.

Keywords: KI, Informatikunterricht, Lernbegleitung, Programmierunterstützung, personalisiertes Lernen

1 Einleitung

Eine effektive und zeitgemäße Vermittlung von Lerninhalten ist seit jeher eine wesentliche Motivation der Fachdidaktik. Hierzu gehört für den Informatikunterricht nicht nur die Betrachtung von Technologien in ihrer funktionalen Perspektive, sondern auch deren gesellschaftliche Auswirkungen sowie eine reflektierte Nutzung als Werkzeug [Br16]. Unterricht zum Thema Programmierung stellt hierbei eine Herausforderung dar, die sowohl Schülerinnen und Schüler als auch Lehrkräfte vor vielfältige Schwierigkeiten stellt [Bl04]. Die Komplexität und der Abstraktionsgrad des Lerninhaltes sowie das Eingehen auf individuelle Lerngeschwindigkeiten sind zu bewältigende Herausforderungen, für einen lernförderlichen Informatikunterricht. Lernsysteme, welche eine Rückmeldung zu einer Frage im Lernprozess geben, bieten die Möglichkeit zur Senkung der intrinsischen Belastung und damit zur Vermeidung einer Überlastung des Arbeitsgedächnisses. Dies hat nach der *Cognitive Load Theory* einen potenziell positiven Einfluss auf den Lernprozess [SAK11].

1 Universität Leipzig, Didaktik der Informatik, Augustusplatz 10, 04109 Leipzig, Deutschland,
peter.kiessling@informatik.uni-leipzig.de, https://orcid.org/0000-0002-1820-6328;
florian.funke@uni-leipzig.de, https://orcid.org/0000-0003-4043-894X;
sven.hofmann@uni-leipzig.de, https://orcid.org/0000-0003-3845-1125

doi:10.18420/delfi2024_10

Zudem werden traditionelle Formen des Lernens und Prüfens, wie z.B. eine klassische Hausarbeit, durch KI-basierte Sprachmodelle zunehmend in Frage gestellt [GK23]. Besagte Sprachmodelle können unterstützend auch als persönlicher interaktiver Lerncoach fungieren [Al23a]. Zudem können ergänzende Informationen angeboten werden, wenn die Lernenden Hilfe benötigen [He23].
In diesem Kontext gewinnen KI-Technologien zunehmend an Bedeutung, um den Lernprozess zu optimieren und individuelle Unterstützungen zu geben. Daraufhin entstand die Idee, einen KI-basierten Programmierassistenten zu entwickeln. Dieser soll die Schülerinnen und Schülern beim Erlernen der textbasierten Programmierung unterstützen und individaulisierte Rückmeldungen bieten. Gleichzeitig können Lernende dadurch einen sinnvollen Einsatz von KI als Werkzeug erproben. Durch die Integration dieser Technologie können personalisierte Lernerfahrungen geschaffen werden, die die Problemlösekompetenz der Schülerinnen und Schüler fördern.
Trotz des großen Potentials von KI in der Bildung mangelt es bisher an effektiven Werkzeugen zur Unterstützung von Schülerinnen und Schülern im Informatikunterricht. Der Einsatz von chatbasierten KI-Systemen wie ChatGPT von OpenAI bedarf i. d. R. einen personalisierten Account. Dies macht die individuelle Nutzung im Unterricht schwierig bis unmöglich. Sprachmodelle wie ChatGPT generieren häufig vollständige Lösungen zu Programmieraufgaben, so wie sie im Schulkontext gefordert sind. Dabei könnten die Lernenden die Hilfesysteme derart intensiv in Anspruch nehmen, dass der individuelle Lernprozess behindert wird und persönliche Kompetenzen zu wenig oder nicht gefördert werden [Al23b; De23]. Ein unreflektierter Einsatz beeinflusst dadurch den Lernprozess eher negativ, da dieser zu sehr verkürzt wird und Phasen einer lernförderlichen kognitiven Aktivierung vollständig übersprungen werden können.
Daraus ergibt sich der Bedarf nach einem KI-basierten Unterstützungssystem für die Schule, welches die genannten Vorteile bietet und ohne personalisierte Schüleraccounts genutzt werden kann.
Im vorliegenden Beitrag wird zunächst die Entwicklung des *Programmier-Copilots* für den Informatikunterricht dargestellt. Darüber hinaus werden Vorgehen und Ergebnisse der Wirksamkeitsuntersuchung beim Einsatz des Copilots im Informatikunterricht präsentiert.

2 Entwicklung des KI-basierten Programmierassistenten

Der *Programmier-Copilot* wurde in Python verfasst und sollte zunächst den Zugang zu Sprachmodellen von OpenAI ohne personenbezogene Schüleraccounts ermöglichen. Hierzu präsentiert das Tool ein eigenes Chatfenster, in dem Anfragen gestellt werden können. Diese werden anschließend im Hintergrund an das Modell *gpt-3.5-turbo* weitergeleitet[2]. Eine solche zunächst trivial erscheinende Konzeption hat zwei zentrale Vorteile.
Zum Einen haben Nutzerinnen und Nutzer damit Zugang zu besagtem Sprachmodell. Es ist lediglich das Eintragen einer gültigen API erforderlich, so dass technisch gesehen nur

2 Programmier-Copilot in der Version 0.1, April 2024

ein gemeinsamer Nutzeraccount benötigt wird. Dieser kann von einer Lehrkraft bzw. der Schule angelegt werden. Die Anfragen laufen entsprechend über die API des verwalteten Nutzeraccounts. Zum Anderen besteht damit die Möglichkeit die Anfrage an das Sprachmodell im Hintergrund zu bearbeiten. Hierfür wurden verschiedene *Beschränkungsstrings* erprobt und schließlich in eine einheitliche Formulierung überführt. Die Beschränkung wird vom Programm im Hintergrund mit dem Anfragestring konkateniert. Dadurch kann die Unterstützung des Sprachmodells in verschiedenen Parametern eingeschränkt werden, welche im Folgenden dargestellt sind.
Der Beschränkungsstring orientiert sich zunächst an Empfehlungen zum lernförderlichen Einsatz eines CAS bzw. GTR[3] im Mathematikunterricht. Hierfür ist es unter anderem wesentlich Beziehungen zwischen mentalen Abbildern und realen Darstellungen weiterzuentwickeln und damit die kognitive Aktivierung aufrecht zu erhalten [We12].
Eine Hilfestellung im Sinne einer lernförderlichen Rückmeldung besteht dementsprechend aus der Beschreibung eines Lösungsansatzes, ohne diesen zu detailliert vorzugeben. Gleichzeitig soll eine Antwort das Vorwissen der Lernenden adressieren – hierfür sollten beispielsweise grundlegende Fachbegriffe nicht unaufgefordert erklärt werden. Um dem gerecht zu werden, wurde der folgende Beschränkungsstring implementiert: *„Das folgende Gespräch soll meine Fähigkeiten in Python verbessern. Gib mir bitte keinen fertigen Programmcode vor, arbeite mit grundlegenden Fachbegriffen ohne diese unaufgefordert zu erklären und beschränke deine Antworten auf wenige kurze Anregungen für mögliche Lösungen, auf die ich anschließend selber kommen muss. Meine Frage ist: “*. Die Spezifikation auf die Sprache Python folgte den innerhalb der Untersuchung gegebenen Rahmenbedingungen. Eine Änderung der Sprache innerhalb der Einstellungen des Programms ist Teil der geplanten Weiterentwicklung.
Der Einsatz des *Programmier-Copiloten* als Hilfsmittel soll von den Schülerinnen und Schülern als begrenzte Ressource erkannt werden. Dazu wurde das Programm mit einem Counter versehen, mit dem in der aktuellen Version eine Beschränkung auf fünf Anfragen gesetzt ist. Das vollständige Programm ist Open Source und über den angegebenen Link[4] erreichbar.

3 Wirksamkeitsuntersuchung

Im Anschluss an die Entwicklung des *Programmier-Copilots* erfolgte die Erprobung im schulischen Kontext sowie die Untersuchung der Einflüsse auf den Lernerfolg von Schülerinnen und Schülern. Die Forschungsfrage lautete demnach: *Inwieweit hat die Verwendung des Tools einen messbaren Einfluss auf die Lernleistung der Schülerinnen und Schüler?* Hierzu wurde eine geeignete staatliche Schule in Sachsen gewählt, in der möglichst homogene Rahmenbedingungen vorlagen. Die Untersuchung fokussiert sich somit zunächst auf die Quantifizierung von Einflüssen und nicht auf formale statistische Signifikanz.

3 Abkürzung für Computer Algebra System bzw. Grafiktaschenrechner

4 Programmdatein, Quellcode, Unterrichtsmaterial und Musterlösungen erreichbar über `t1p.de/ddil-software-programmier-copilot`

3.1 Untersuchungsdesign

Die Untersuchung wurde angelehnt an ein Gruppen-Prä-Post-Design konzeptioniert [Ro04]. Dabei wurden drei Gruppen für die Untersuchung ausgewählt, die die gleiche Intervention in unterschiedlicher Ausprägung durchliefen (vgl. Tab. 1). Eine genaue Aufschlüsselung der Items sowie die entsprechend erhobenen Forschungsdaten der Untersuchung sind ebenfalls dem genannten Link zu entnehmen.
Die Probanden bestehen aus drei Kursen der Klassenstufe 11, die von der gleichen Lehrkraft unterrichtet wurden. In allen drei Grundkursen erfolgte die Erprobung des Copiloten in der ersten Unterrichtseinheit zur textbasierten, objektorientierten Programmierung innerhalb des Lernbereiches 2 „Algorithmierung und Programmierung“ [SM22].

Gruppe	Umsetzung
Gr. A	ohne Programmier-Copilot
Gr. B	mit beschränktem Programmier-Copilot
Gr. C	mit unbeschränktem Programmier-Copilot

Tab. 1: Testbedingungen in den Gruppen

Als Messpunkte für Prä- und Posttest wurden Beginn und Ende der Unterrichtseinheit ausgewählt, insbesondere um mögliche Effekte durch Nacharbeitung und Austausch über die Kurse hinweg zu minimieren. Prä- und Posttest enthielten dabei jeweils identische Aufgabenstellungen, die sich auf unterschiedliche Beispiele bezogen. Geprüft wurden die Grundlagen des Themenbereichs Klassen und Objekte sowie deren Implementierung, fachlichen Fragen zur Programmierung sowie Anwendungen im Python-Quellcode. Anschließend wurden Prä- und Posttests bewertet und die Entwicklung der Ergebnisse zwischen den einzelnen Stichproben verglichen.

3.2 Unterrichtseinheit

In der Unterrichtseinheit fanden neben Prä- und Posttest mit jeweils 15 Minuten Bearbeitungszeit die Vermittlung erster Inhalte und die Bearbeitung von Aufgaben zur objektorientierten Programmierung statt. Die Materialien und Abläufe sowie äußere Rahmenbedingungen bspw. eine Durchführung im gleichen Raum wurden hierbei standardisiert vorgegeben, um Vergleichbarkeit zwischen den Gruppen zu gewährleisten. Anschließend an den Prätest folgte eine klassenübergreifende Arbeitsphase von 30 Minuten. Als ein erstes Beispiel wurde eine Programmvorlage bereitgestellt. Diese beinhaltet eine Struktur bestehend aus den Klassen *Ferienjobvermittlung* als Institution und *Person*. Das Material wurde für eine angeleitete Übung im Plenum verwendet. Die Schülerinnen und Schüler sollten fehlende Attribute innerhalb des Programms korrigieren, Methoden vervollständigen und durch das Erzeugen eines Objektes und dem Aufrufen von Methoden das gesamte Programm auf Funktionalität überprüfen.

Die sich anschließende Arbeitsphase beinhaltete vergleichbare Aufgaben in einem anderen Setting. Das Material bezog sich hier auf die Klassen *Bibliothek* und *Buch*. Gegebene Methoden sollten auch hier vervollständigt und auf ihre Funktionalität geprüft werden.
Die Schülerinnen und Schüler waren damit gefordert, den demonstrierten Lerninhalt selbstständig auf einen anderen Kontext anzuwenden. Die vorbereiteten Materialien sowie damit verbundene Musterlösungen sind ebenfalls dem genannten Link zu entnehmen. Abschließend erfolgte der Posttest in Verbindung mit dem Einreichen der entstandenen Lösungen aus der individuellen Arbeitsphase sowie eine kurze Gruppenbefragung zur Nutzung des Tools, sodass auch Eindrücke bei der Arbeit mit dem Tool erfasst wurden.

4 Ergebnisse

Zur Analyse der Ergebnisse wurden die Bewertungen einzelner Schülerinnen und Schüler zunächst hinsichtlich des Anteils richtiger Antworten untersucht und anschließend ein Gruppen-Mittelwert bestimmt. Anschließend wurde das Verhältnis der Gruppen-Mittelwerte aus Post- und Prätest ermittelt und daraus die interventionsspezifische Verbesserungsrate berechnet. Außerdem wurde die Differenz der Gruppenmittelwerte bestimmt, um die nominale Verbesserung abschätzen zu können.
In Tab. 2 ist ersichtlich, dass die Verbesserungsraten bei der Verwendung des Werkzeugs unabhängig von der Beschränkung der Anfragen-Zahl mehr als doppelt so hoch wie bei Durchführung der Einheit ohne den *Programmier-Copiloten* sind. Dies schlägt sich nicht nur im Verhältnis von Post- und Prätest nieder, sondern auch in der nominalen Verbesserung zwischen den Testungen. Auffällig ist auch, dass zunächst die Verbesserungsrate im unbeschränkten Einsatz des Copiloten etwas höher ist, als im beschränkten Einsatz. Die nominelle Verbesserung unterscheidet sich hingegen kaum.

Gr.	P.-Copilot	N	Verhältnis Post/Prä	Verbesserung	Differenz Post-Prä	SD[5]
A	ohne	14	1.15	14.6%	.07	.16
B+C	mit	22	1.39	38.8%	.16	.19
B	beschränkt	13	1.35	34.9%	.16	.22
C	unbeschränkt	9	1.46	45.7%	.17	.15

Tab. 2: Entwicklung der Ergebnisse in den Gruppen

Bei der Auswertung der Daten wiesen einige wenige Probanden eine negative Verbesserungsrate auf. Dabei ließ sich erkennen, dass zwar der Prätest umfänglich beantwortet wurde, im Posttest aber häufiger keine Antwort auf einzelne Aufgabenstellungen gaben. Aus diesem Grund wurden in der nachfolgenden Analyse in Tab. 3 Datensätze eliminiert, bei denen kein vollständig bearbeiteter Posttest vorlag. Dadurch sind in den Berechnungen nur Werte enthalten, bei denen von einer gewissenhaften Beantwortung in Prä- und Posttest ausgegangen werden kann.

5 Standard Deviation (SD) als Abkürzung für Standardabweichung

Gr.	P.-Copilot	N	Verhältnis Post/Prä	Verbesserung	Differenz Post-Prä	SD
A	ohne	9	1.31	30.9%	.15	.11
B	beschränkt	10	1.59	58.8%	.24	.17
C	unbeschränkt	8	1.51	50.6%	.19	.15

Tab. 3: Entwicklung der Ergebnisse in den bereinigten Gruppen

Hier wird sichtbar, dass die Verbesserungsraten bei Verwendung des *Programmier-Copiloten* ebenfalls deutlich höher sind, als in Gr. A. Die Werte zwischen beschränktem und unbeschränktem Copiloten unterscheiden sich jedoch im Vgl. zu Tab. 2. Die Verbesserungsrate bei Lernenden mit dem beschränkten Tool ist höher als in der unbeschränkten Version. Insbesondere aber auch die nominelle Änderung der Ergebnisse in der beschränkten Version ist nun größer. Dies entspricht den oben genannten Erwartungen an das Anwendungsgebiet des Tools.
In einer abschließenden offenen Gruppenbefragung gaben die Schülerinnen und Schüler an, dass das Tool zu wenig Hilfestellung gab. Grob formulierte Fragen mündeten häufig in einer Paraphrasierung der Aufgabenstellung. Für konkrete Nachfragen konnte der *Programmier-Copilot* aber wertvolle Hinweise geben.
Um umfassende und allgemeingültige Erkenntnisse zu gewinnen, bedarf es weiterer Untersuchungen mit größerer Stichprobe und den Einsatz des entwickelten Werkzeugs in unterschiedlichen Unterrichtssituationen. Als Antwort auf die Forschungsfrage konnten in der Untersuchung sowohl erste positive Tendenzen hinsichtlich des Lernzuwachses sowie Erkenntnisse für die Weiterentwicklung des *Programmier-Copilots* mit Hilfe der dargestellten Untersuchungsmethodik gewonnen werden.

5 Fazit und Ausblick

Mit dem *Programmier-Copiloten* wurde ein Werkzeug entwickelt, dass in der Stichprobe eine höhere Verbesserungsrate bei Nutzung des Tools im Vergleich zum Bearbeiten der Aufgaben ohne dieses Werkzeug nachweisen konnte. Auch im Vergleich zum unbeschränkten Einsatz des Werkzeugs konnten erste positive Tendenzen entsprechend der zu Beginn deutlich gemachten didaktisch-pädagogischen Implikationen gezeigt werden.
Zukünftig soll eine Weiterentwicklung des *Programmier-Copilots* im Bezug auf die Rückmeldungen erfolgen und bspw. durch eine Überarbeitung des User-Interfaces, anpassbare Beschränkungsstrings und variable Anfrage-Anzahlen vielfältigere Einsatzmöglichkeiten und eine intuitivere Nutzung ermöglichen. Ebenso ist eine umfangreichere Evaluation des Werkzeugs angedacht.
Durch die Bearbeitung des Beschränkungsstrings wären verschiedene Stufen von Unterstützung und damit verbundener kognitiver Entlastung möglich. Die trotz ihrer Auswahlmöglichkeiten einen kognitiv aktivierenden Lernprozess erhalten.

Literaturverzeichnis

[Al23a] Albrecht, S.: ChatGPT und andere Computermodelle zur Sprachverarbeitung. Grundlagen, Anwendungspotenziale und mögliche Auswirkungen. Büro für Technikfolgeabschätzung im Bundestag, Berlin, 2023.

[Al23b] Alouani, N.: Why You're the Biggest Loser in the AI Wars, Zugriff: 2024-06-06, 2023, URL: https://nabilalouani.substack.com/p/why-youre-the-biggest-loser-in-the.

[Bl04] Blackwell, A. F.: End-user developers at home. Communications of the ACM 47 (9), S. 65–66, 2004, DOI: 10.1145/1015864.1015892.

[Br16] Brinda, T.; Diethelm, I.; Gemulla, R.; Romeike, R.; Schöning, J.; Schulte, C.: Dagstuhl-Erklärung – Bildung in der digital vernetzten Welt, 2016, URL: https://dagstuhl-dreieck.de, Stand: 01. 04. 2024.

[De23] Deutscher Ethikrat: Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz. Stellungnahme, Berlin, 2023.

[GK23] Gutbrod, J.; Klopsch, B.: Chatbots as a pedagogical challenge for schools and teaching. TATuP - Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis 32 (3), S. 72–78, 2023, DOI: 10.14512/tatup.32.3.72, URL: https://www.tatup.de/index.php/tatup/article/view/7086.

[He23] Heidt, A.: Arms race with automation: professors fret about AI-generated coursework. Nature Technology Feature, Zugriff: 2024-06-06, 2023, URL: https://www.nature.com/articles/d41586-023-00204-z.

[Ma23] Mainzer, K.: Epilog: Grenzen der KI–theoretisch, praktisch, ethisch. In: Philosophisches Handbuch Künstliche Intelligenz. Springer Fachmedien Wiesbaden, Wiesbaden, S. 1–32, 2023.

[Ro04] Rost, J.: Lehrbuch Testtheorie - Testkonstruktion. OCLC: 248943003, Huber, Bern, 2004, ISBN: 978-3-456-83964-6.

[SAK11] Sweller, J.; Ayres, P.; Kalyuga, S.: Cognitive Load Theory. Springer, New York, 2011.

[SM22] SMK: Lehrplan Gymnasium, Informatik, Zugriff auf: https://www.schulportal.sachsen.de/lplandb/, Dresden, 2022.

[We12] Weigand, H.: Fünf Thesen zum Einsatz digitaler Technologien im zukünftigen Mathematikunterricht. In (Blum, W.; Borromeo F., R.; Maaß, K., Hrsg.): Mathematikunterricht im Kontext von Realität, Kultur und Lehrerprofessionalität: Festschrift für Gabriele Kaiser. Vieweg+Teubner Verlag, Wiesbaden, S. 315–324, 2012, DOI: 10.1007/978-3-8348-2389-2_32.

A Multidisciplinary Approach to AI-based self-motivated Learning and Teaching with Large Language Models

Thomas Ranzenberger[1], Carolin Freier[1], Luca Reinold[1], Korbinian Riedhammer[1], Fabian Schneider[1], Christopher Simic[1], Claudia Simon[1], Steffen Freisinger[2], Munir Georges[2], and Tobias Bocklet[1]

Abstract: We present a learning experience platform that uses machine learning methods to support students and lecturers in self-motivated online learning and teaching processes. The platform is being developed as an agile open-source collaborative project supported by multiple universities and partners. The development is guided didactically, reviewed, and scientifically evaluated in several cycles. Transparency, data protection and the copyright compliant use of the system is a central part of the project. The system further employs large language models (LLMs). Due to privacy concerns, we utilize locally hosted LLM instances and explicitly do not rely on available cloud products. Students and lecturers can interact with an LLM-based chatbot in the current prototype. The AI-generated outputs contain cross-references to the current educational video's context, indicating if sections are based on the lectures context or world knowledge. We present the prototype and results of our qualitative evaluation from the perspective of lecturers and students.

Keywords: Artificial Intelligence in Education, Learning Experience Platform, Open Source Software, Large Language Models

1 Introduction

Nine universities and three cross-institutional initiatives have teamed up in the collaborative project Hochschul-Assistenz-System (HAnS). The project is funded by the German Federal Ministry of Education and Research (BMBF) to design and implement an open-source learning experience platform with an artificial intelligence (AI) tutor for higher education. The involved universities allow a multidisciplinary project execution. They participate by evaluating the features and user interactions from a didactic perspective, participate in the general data protection regulation (GDPR) compliant project execution, consider ethic aspects whether certain groups are disadvantaged, or conduct qualitative and quantitative evaluations and acceptance surveys together with students and lecturers. Influenced by the Corona pandemic, our platform focuses on podcasts and video based teaching material with supplemental slides. It enables keyword searches via indexing the meta data and the transcript provided by an automatic speech recognition (ASR) system. As the capabilities

1 Technische Hochschule Nürnberg Georg Simon Ohm, Zentrum für Künstliche Intelligenz, Keßlerplatz 12, 90489 Nürnberg, Germany, thomas.ranzenberger@th-nuernberg.de, https://orcid.org/0009-0007-2837-3553; carolin.freier@th-nuernberg.de, https://orcid.org/0000-0002-1700-2403; luca.reinold@th-nuernberg.de; korbinian.riedhammer@th-nuernberg.de, https://orcid.org/0000-0003-3582-2154; fabian.schneider@th-nuernberg.de; christopher.simic@th-nuernberg.de; claudia.simon@th-nuernberg.de; tobias.bocklet@th-nuernberg.de, https://orcid.org/0009-0008-7780-8821

2 Technische Hochschule Ingolstadt, AImotion Bavaria, Esplanade 10, 85049 Ingolstadt, Germany, steffen.freisinger@thi.de; munir.georges@thi.de, https://orcid.org/0000-0002-5542-149X

doi:10.18420/delfi2024_11

of large language models (LLMs) are constantly improving and evolving, we use them to realize our basic AI tutor framework to generate exercises. In this work-in-progress, we describe the current state of our prototype using an LLM-based chatbot for question answering and generation of multiple-choice questions in our learning experience platform. We evaluate how the use of LLMs in our prototype can support students and lecturers and what risks and expectations they have.

2 Prototype

The current prototype of our platform processes the uploaded teaching material, mainly videos, using a machine learning pipeline. The pipeline extracts the raw audio and uses an open-source Whisper [Ra23a] based automatic speech recognition system (ASR) to transcribe the spoken words into text. The used ASR system is capable of transcribing 1 hour of audio in less than 1 minute [Ra24]. In addition to the original transcript, the subtitles for German and English for the prototype video player are created. The resulting transcript is processed using an unsupervised topic segmentation approach [Fr23b] to detect the main topics of the lecture material. For each topic a heading and a short summary is generated by prompting an LLM using the text from the transcript of each topic interval. The topics are visualized as cards in a topics tab on the right tab bar as shown on the left of Fig. 2. A full summary in a separate tab is available which was generated by prompting the LLM with the complete transcript text.

As the platform is developed open-source and GDPR compliant we use open-source LLMs and host them on the universities infrastructure. The currently used model is a finetuned Llama2-based Vicuna-13b v1.5 model with 16000 tokens input length including the provided prompt context [Zh23]. The video transcript is displayed below the streaming video player and could be searched via the search bar supporting auto completion at the top, see left hand side of Fig. 2. Students or lecturers could directly jump to the position of the spoken word in the video by clicking on the corresponding search result. Search results are displayed for each search in a separate tab. The slides and video size and the transcript font size could be adjusted, which allows students or lecturers to keep focus on their desired modality. Further details are described by [Ra23b]. The current prototype focus is on the AI tutor which generates exercises and supports the students in their learning journey.

2.1 Chatbot with multiple prompt modes

The first version of the AI tutor in our current prototype is based on LLMs. To provide meaningful and intended responses in relation to the lecture material we use a retrieval augmented generation (RAG) approach [Le20]. The Llama2-based LLM prompt context is provided by a Generative T5 Retriever model [Ni22]. The retrieval model is hosted on premise together with an open-source translation models from Facebook FAIR's WMT19 submission [Ng19]. The user has the option to translate the German input prompt for the LLM to English [Ra24]. If the translation mode is used, the response of the LLM is

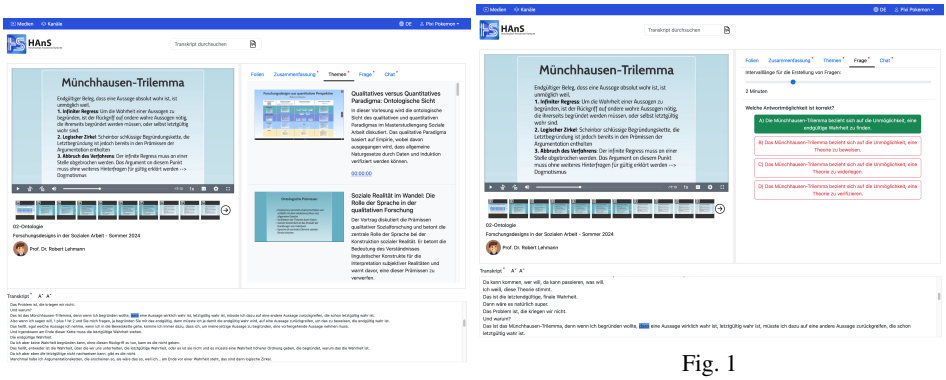


Fig. 1

Fig. 2: *Left:* Segmented topics with headings and short summaries generated by the LLM. *Right:* Multiple-choice question generation with adjustable context size.

translated back to German. This allows to evaluate and explore the language dependent response capabilities of the used LLM. The translations can be displayed using translation buttons of each user or bot message or by changing the locale at the top of the web page, see Fig. 3.

The chatbot offers different prompt modes to the user [Ra24]. The user could explore to prompt without using the context of the lecture material. This allows to evaluate how the LLM behaves without any additional provided context by only relying on the training data of the model itself (known as model or world knowledge). In contrast it can be evaluated how the answer changes if the prompt mode is switched to use context mode or to use context and cite mode. Fig. 3 uses context and cite mode. The mode uses a system prompt which instructs the LLM to cite the context which was provided by the retrieval model. The retrieval model selects 4 sections of the transcript which are in the closest relation to the users input prompt. These sections are used as the LLMs context and can be displayed by clicking on a stack button of each bot message. The LLM is instructed to cite these sections and the user can display the transcript sections by clicking on the citations as shown on the left in Fig. 3. This allows to verify and detect hallucinations in the generated response directly with the transcript which is based on the spoken words of the lecturer.

Our current prototype offers an additional mode which allows the user to select the LLMs context without using the retrieval model. This allows the students and lecturers to select the context which should be used to generate the response for their instructions. On the right of Fig. 3, the chatbot is asked for a more concrete example of an abstract explanation of the lecturer, helping the student to better understand certain aspects of a lecture topic. Lecturers can use the prompt mode to gather inspiration for exam tasks or creating learning objective tests.

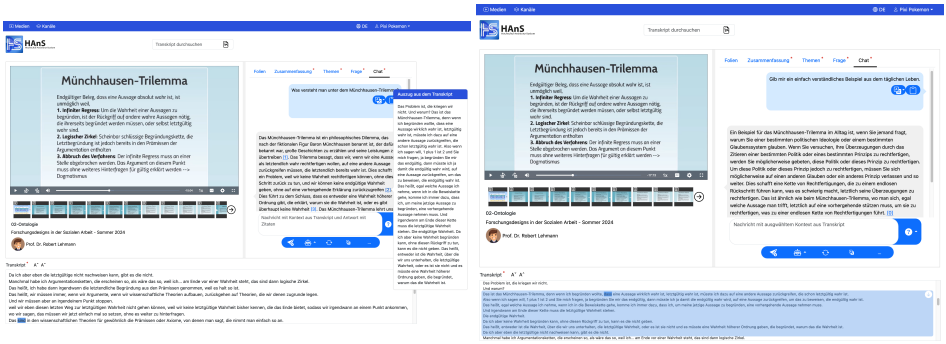


Fig. 3: *Left:* Prompt mode with citations allows the users to check the correctness of the chatbot response. *Right:* The new prompt mode allows the selection of the prompt context by the user and cites it in the chatbot response.

2.2 LLM-based multiple-choice questions

In order to further support the students on their learning journey, the prototype creates multiple-choice questions during video playback, see right hand side of Fig. 2. The questions are generated using the text of the transcript section corresponding to the selected context size in minutes. Our previous prototype had a fixed context size of 3 minutes [Ra24]. The control of the context size helps to handle different speaking styles and flow of speech of the lecturers. The parsed question and answer choices of the LLM are displayed on a separate workbench tab in the video player view as shown on the right of Fig. 2. The LLM is requested to generate a question with one single correct answer and three additional answers. The response also contains the correct answers index. The user can select an answer and the user interface indicates if the selected answer was correct, allowing students to interactively check and recap their current state of knowledge. In the next section, we present findings of the qualitative evaluation of the learning experience platform.

3 Evaluation

To achieve the goal of dialogical and transparent self-motivated online learning using LLM-based features, the platform is evaluated using a multidisciplinary approach [Fr23a]. This section presents findings from the qualitative evaluation of the current state of the prototype. The qualitative evaluation examines the expectations of students and lecturers as well as the learning practices of students in the prototype development phases. The aim of this evaluation is to provide continuous feedback for the further improvement of the system.

Online group discussions with students and lecturers are used to reconstruct existing experiences and attitudes towards digital learning materials in general and learning with the LLM-based features and tools presented in Sect. 2 in particular [KK18]. Students'

learning practices with videos integrated into the learning experience platform are examined ethnographically by means of participant observation with subsequent guided interviews [Br15]. Both forms of data are analyzed using the documentary method [BPS09].

Since July 2022, a total of 65 students and 10 lecturers have taken part in the qualitative evaluation [No17]. This work-in-progress presents results from the evaluation period (03/2023 - 02/2024) with a total of twenty surveys involving 32 people. In this system status, lecturers can use the LLM-based features in their teaching. However, the students discuss the features without having their own application beyond the evaluation.

3.1 Potentials and challenges: The student's perspective

The students surveyed see the potential of the learning experience platform particularly in asynchronous scenarios, improved organization, and support from artificial intelligence. They see the possibility of asynchronous self-motivated online learning with the prototype as positive, as it enables them to catch up on missed teaching content in a targeted manner and to watch teaching videos several times if they have problems of understanding: "in the course it is then also easier for me to learn because I can just watch it several times if something was not understood." [3]

Another finding of the surveyed students is the more effective work with the learning materials and better organization of self-learning that was achieved by bundling instructional videos, transcripts, and slide sets in the platform. The AI-based subdivision of the instructional videos into thematic chapters with subheadings was also found to be helpful in improving the accessibility of the content: "the most practical (...) [is] this structure. I have it in the form that I can see that minute XY deals with this topic, so that I don't have to watch a 30-minute video and wait until (...) the point that interests me comes up."

The students have a predominantly positive attitude towards support from artificial intelligence in system-supported self-motivated online learning with the learning experience platform. They prefer a factually sober presentation of search results and would like to be able to clearly distinguish which sources the search results refer to. They hope that the integration of automated control questions will enable them to better monitor their own learning progress. Whether the results of the control questions should also be visible to lecturers is, however, the subject of critical debate. Students agree that they would like the AI to provide explanations and information on the correct solution in the event of incorrect answers.

3.2 Potentials and challenges: The lecturer's perspective

Lecturers see two main potentials in the use of the platform in their courses. Firstly, it enables temporal and geographical flexibility in teaching and learning. Students can prepare

3 The quotes have been linguistically smoothed for the translation.

for courses and examinations regardless of time and place. This enables them to optimize their learning and study in a self-determined way. Intrinsically motivated students, who are well organized, benefit from this form of self-motivated online learning.

Secondly, lecturers see two different effects of AI functions. On the one hand, lecturers assume that content is conveyed more effectively. Although the accuracy of the speech recognition in the transcripts and the topic structure using AI were sometimes perceived as not yet fully developed, the AI functions nevertheless contribute to a better understanding of the content. Lecturers positively emphasize the automatically generated transcripts, as they support students in creating their own learning materials. The searchability of the learning materials also facilitates the learning process. An improvement in exam results is partly attributed to the fact that students can listen to and memorize content repeatedly: "the results of exams have improved and that's not because my teaching is different, but because they have the chance to hear it 100 times and do it."

On the other hand, lecturers also see the risk that with the help of AI functions, students do not learn content in its entirety, but only engage episodically with the content they consider relevant to their exams: "so the [students] goes and gets the slides or just the text that has been generated via AI down there. Why should you watch the whole video and then I just read through the text and prepare for the exam that way and that is also a reduction that I find very dangerous."

4 Conclusion

The learning experience platform HAnS offers an LLM-based chatbot with different prompt modes including citation to cross-check the correctness of responses with the used context sections of the video transcript. The generation of multiple-choice questions has the potential to support students in self-motivated online learning. Students can use the LLM-based chatbot for their studies from April 2024 onward. Further evaluation will show how students actually work with it. However, the initial results already show that both students and lecturers are prepared to make targeted use of transparent LLM-based features. At the same time, the results reveal that the learning environment and chatbot usage in the course must be designed in a way that both well-organized and less independent learners have equal learning opportunities.

References

[BPS09] Bohnsack, R.; Przyborski, A.; Schäffer, B.: Das Gruppendiskussionsverfahren in der Forschungspraxis. Verlag Barbara Budrich, Opladen, 2009, ISBN: 9783866491779.

[Br15] Breidenstein, G.; Hirschauer, S.; Kalthoff, H.; Nieswand, B.: Ethnografie: Die Praxis der Feldforschung. UTB GmbH, 2015, ISBN: 978-3-8252-4497-2.

[Fr23a] Freier, C.; Bocklet, T.; Helten, A.-K.; Hoffmann, F.; Hunger, M.; Kovács, L.; Richter, F.; Riedhammer, K.; Schmohl, T.; Simon, C.: Wie kann videogestütztes Lernen die Erwartungen Studierender und Dozierender erfüllen? Soziale Passagen 15 (2), pp. 631–635, 2023, ISSN: 1867-0199, DOI: 10.1007/s12592-023-00478-0, URL: https://doi.org/10.1007/s12592-023-00478-0.

[Fr23b] Freisinger, S.; Schneider, F.; Herygers, A.; Georges, M.; Bocklet, T.; Riedhammer, K.: Unsupervised Multilingual Topic Segmentation of Video Lectures: What can Hierarchical Labels tell us about the Performance? In: Proc. 9th Workshop on Speech and Language Technology in Education (SLaTE). Pp. 141–145, 2023, DOI: 10.21437/SLaTE.2023-27.

[KK18] Kühn, T.; Koschel, K.-V.: Gruppendiskussionen. Springer Fachmedien, Wiesbaden, 2018, ISBN: 978-3-658-18936-5 978-3-658-18937-2.

[Le20] Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.-t.; Rocktäschel, T.; Riedel, S.; Kiela, D.: Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In (Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; Lin, H., eds.): Advances in Neural Information Processing Systems. Vol. 33, Curran Associates, Inc., pp. 9459–9474, 2020, URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf.

[Ng19] Ng, N.; Yee, K.; Baevski, A.; Ott, M.; Auli, M.; Edunov, S.: Facebook FAIR's WMT19 News Translation Task Submission. In (Bojar, O.; Chatterjee, R.; Federmann, C.; Fishel, M.; Graham, Y.; Haddow, B.; Huck, M.; Yepes, A. J.; Koehn, P.; Martins, A.; Monz, C.; Negri, M.; Névéol, A.; Neves, M.; Post, M.; Turchi, M.; Verspoor, K., eds.): Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1). Association for Computational Linguistics, Florence, Italy, pp. 314–319, 2019, DOI: 10.18653/v1/W19-5333, URL: https://aclanthology.org/W19-5333.

[Ni22] Ni, J.; Qu, C.; Lu, J.; Dai, Z.; Hernandez Abrego, G.; Ma, J.; Zhao, V.; Luan, Y.; Hall, K.; Chang, M.-W.; Yang, Y.: Large Dual Encoders Are Generalizable Retrievers. In (Goldberg, Y.; Kozareva, Z.; Zhang, Y., eds.): Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, pp. 9844–9855, 2022, DOI: 10.18653/v1/2022.emnlp-main.669, URL: https://aclanthology.org/2022.emnlp-main.669.

[No17] Nohl, A.-M.: Interview und Dokumentarische Methode: Anleitungen für die Forschungspraxis. Springer Fachmedien, Wiesbaden, 2017, ISBN: 978-3-658-16079-1 978-3-658-16080-7.

[Ra23a] Radford, A.; Kim, J. W.; Xu, T.; Brockman, G.; Mcleavey, C.; Sutskever, I.: Robust Speech Recognition via Large-Scale Weak Supervision. In (Krause, A.; Brunskill, E.; Cho, K.; Engelhardt, B.; Sabato, S.; Scarlett, J., eds.): Proceedings of the 40th International Conference on Machine Learning. Vol. 202. Proceedings of Machine Learning Research, PMLR, pp. 28492–28518, 2023, URL: https://proceedings.mlr.press/v202/radford23a.html.

[Ra23b] Ranzenberger, T.; Bocklet, T.; Freisinger, S.; Frischholz, L.; Georges, M.; Glocker, K.; Herygers, A.; Peinl, R.; Riedhammer, K.; Schneider, F.; Simic, C.; Zakaria, K.: The Hochschul-Assistenz-System HAnS: An ML-Based Learning Experience Platform. In (Draxler, C., ed.): Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2023. TUDpress, Dresden, pp. 168–169, 2023, ISBN: 978-3-95908-303-4, URL: https://www.essv.de/pdf/pdf/2023_168_169.pdf.

[Ra24] Ranzenberger, T.; Bocklet, T.; Freisinger, S.; Georges, M.; Glocker, K.; Herygers, A.; Riedhammer, K.; Schneider, F.; Simic, C.; Zakaria, K.: Extending HAnS: Large Language Models for Question Answering, Summarization, and Topic Segmentation in an ML-based Learning Experience Platform. In (Baumann, T., ed.): Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2024. TUDpress, Dresden, pp. 219–224, 2024, ISBN: 978-3-95908-325-6, DOI: 10.35096/othr/pub-7103, URL: https://www.essv.de/pdf/2024_219_224.pdf.

[Zh23] Zheng, L.; Chiang, W.-L.; Sheng, Y.; Zhuang, S.; Wu, Z.; Zhuang, Y.; Lin, Z.; Li, Z.; Li, D.; Xing, E.; Zhang, H.; Gonzalez, J. E.; Stoica, I.: Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. In: Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track. 2023, URL: https://openreview.net/forum?id=uccHPGDlao.

Evaluation von LLM- und Intent-basierten Ansätzen zur Umsetzung eines Chatbots für die Unterstützung bei der Studienorganisation

Andre Cordes[1]



Abstract: Conversational User Interfaces wie Chatbots bieten großes Potential, Studierende ergänzend zu bestehenden Beratungsangeboten bei der Studienorganisation zu unterstützen. Insbesondere durch die Fortschritte im Bereich der Large Language Models (LLMs) eröffnen sich neue Herangehensweisen an die Konstruktion solcher Chatbots. Diese sind jedoch mit Chancen und Risiken verbunden, so dass die Wahl eines geeigneten Ansatzes sorgsam abgewogen werden muss. In diesem Beitrag werden drei Ansätze zur Erstellung solcher Chatbots untersucht und miteinander verglichen: ChatGPT mit Retrieval Augmented Generation (RAG), das Open-Source LLM Mistral mit RAG und ein Intent-basierter Chatbot. Die Ansätze werden hinsichtlich Qualität der Antworten und Risiken (z.B. Halluzinationen) verglichen. Insgesamt zeigt sich, dass alle Ansätze potenziell Anwendung für die Unterstützung bei der Studienorganisation finden können. Aus den gewonnenen Erkenntnissen lässt sich jedoch keine klare Empfehlung für einen Ansatz ableiten, weshalb in weiteren Arbeiten ein hybrider Chatbot untersucht werden sollte.

Keywords: Chatbot, Künstliche Intelligenz, Evaluation, Large Language Models, Retrieval-Augmented Generation

1 Einleitung

Vielen Studierenden fällt es am Anfang ihres Studiums schwer, die große Menge an notwendigen organisatorischen Informationen für einen strukturierten Studienverlauf zu überblicken. Diese Informationen umfassen u.a. Aufbau, Angebote und Abläufe der Hochschule sowie Besonderheiten des jeweiligen Studienganges [SW19]. Viele Informationen werden erst bei konkretem Bedarf für Studierende relevant. Ein Chatbot kann Studierenden hier einen zentralen, niederschwelligen und rund um die Uhr verfügbaren Kanal bieten [Ne20].

Im Forschungsprojekt *IPPOLIS*[2] der Fachhochschule Dortmund wird ein Chatbot für die Unterstützung von Studierenden bei der Studienorganisation entwickelt. Der Chatbot mit Namen *Info-Bot* wird am Fachbereich Informatik eingesetzt und in mehreren Iterationen evaluiert und verbessert. Der Info-Bot basiert auf dem Mentoring-Programm am Fachbereich Informatik. Dieses verfolgt das Ziel, Studierende über den Verlauf ihres Studiums durch Mentor*innen zu begleiten und als erste Ansprechpartner zu organisatorischen Fragen zu fungieren. Etwa 40% aller regelmäßigen Anfragen an das Mentoring können i.d.R. durch

1 Fachhochschule Dortmund, Fachbereich Informatik, Emil-Figge-Straße 42, 44227 Dortmund, Deutschland, andre.cordes@fh-dortmund.de

2 (Intelligente Unterstützung projekt- und problemorientierter Lehre und Integration in Studienabläufe) https://fh.do/ippolis

doi:10.18420/delfi2024_12

eine kurze Erklärung einfach beantwortet werden, weshalb sich hierfür ein Chatbot anbietet, um diese automatisiert zu beantworten.

Anfänglich wurde der Info-Bot mit dem Chatbot-Framework Botpress[3] als Intent-basierter Chatbot [Qi18] entwickelt und im April 2023 für Studierende veröffentlicht. Intent-Klassifizierung zielt darauf ab, die Absicht einer Nutzendenanfrage zu ermitteln und einem Frage-Antwort Paar zuzuordnen. Die jüngsten technischen Fortschritte von Large Language Models (LLM) in populären Chatbots wie ChatGPT [Op22] haben die Frage aufgeworfen, welche Vorteile ein LLM-basierter Chatbot im Vergleich zum bisherigen Ansatz für diesen Anwendungsfall bringen könnte. Insbesondere sind die Erwartungen von Nutzenden zu erfüllen, welche vor allem durch ChatGPT stark geprägt wurden und natürlichsprachliche Kommunikation mit Chatbots mit hoher Antwortqualität erwarten [CG21; Ro23]. Weiterhin können LLMs mittels Retrieval-Augmented Generation (RAG) [Le20] besonders gut mit großen Mengen strukturierter Informationen umgehen [Ch24a].

Das Ziel dieses Beitrags ist es, drei Ansätze zur Umsetzung von Chatbots zur Studienorganisation zu untersuchen und hinsichtlich der Qualität der Antworten und Risiken (z.B. Halluzinationen) zu vergleichen. Die Evaluation erfolgt, indem allen Ansätzen die gleiche Wissensbasis zur Verfügung gestellt wird und die Antworten auf eine Testmenge von Fragen zur Wissensbasis empirisch verglichen werden. Untersucht werden die folgenden Ansätze:

1. Das proprietäre LLM ChatGPT von OpenAI, welches viel untersucht ist, bei vielen Aufgabentypen und in vielen Domänen gute Ergebnisse erzielt, jedoch datenschutztechnisch kritisch zu betrachten ist [Ya23].
2. Weiterhin wird das Open-Source LLM Mistral 7B [Ji23] untersucht, welches u.a. in Inferenz-Benchmarks bessere Ergebnisse erzielte als die besten Open-Source 13B und 34B Modelle [Mi24].
3. Als letzter Ansatz wird ein Intent-basierter Chatbot untersucht, welcher besonders geeignet ist für strukturierte und vorhersehbare Interaktionen, wie z.B. ein FAQ [Ch24b]. Darüber hinaus ist dieser Ansatz transparenter und interpretierbarer als LLMs in der Funktionsweise, aber aufwändiger in der redaktionellen Wartung [Ka23].

Im Folgenden beschreibt dieser Beitrag die gewählten Chatbot-Ansätze und den Testaufbau (Abschnitt 2 und Abschnitt 3). Anschließend werden die Ergebnisse der Evaluation dargestellt und die Zusammenhänge und Unterschiede mit anderen Arbeiten erläutert (Abschnitt 4 und Abschnitt 5). Abschließend werden die Ergebnisse kritisch betrachtet Abschnitt 6).

2 Chatbot-Ansätze

Für die Umsetzung der drei Ansätze stellt der Info-Bot den Intent-basierten Ansatz dar und dient auch als Grundlage für die zwei LLM-Ansätze. Der Info-Bot dient als FAQ für eine

3 https://github.com/botpress/v12, genutzt wurde Version 12.30.1

Vielzahl von Fragen zur Studienorganisation und verfügt über eine umfassende Wissensbasis, welche 37 Themen mit 49 Intents abdeckt. Die Informationen der Wissensbasis liegen in Textform vor und können einfach für die LLM-Ansätze mit RAG genutzt werden (siehe Abb. 1). Die Dokumente des Info-Bots werden über einen Chunker in semantisch zugehörige Absätze unterteilt und mittels des Embedding-Models all-MiniLM-L6-v2[4] in Vektoren konvertiert, ebenso wie die Anfragen an das System. Zu jeder Anfrage werden mittels des K-Nearest-Neighbors Algorithmus [Pe09] die zehn semantisch ähnlichsten Absätze ermittelt. Die Anfrage und die ermittelten Absätze werden anschließend an die LLMs ChatGPT (GPT-3.5-Turbo-0125), bzw. Mistral (mistral-7b-openorca Q8_0) in Form eines Prompts übergeben. Abschließend wird die generierte Antwort ausgegeben.

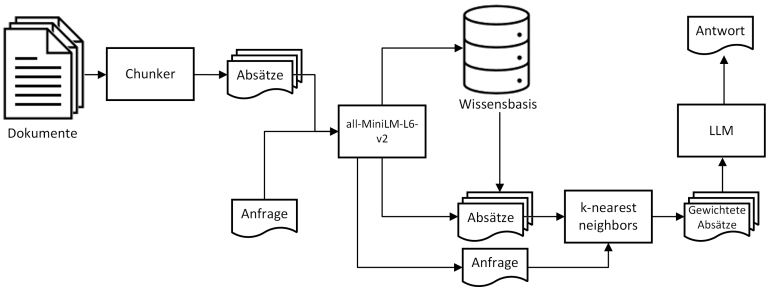


Abb. 1: RAG-Pipeline der LLMs. Angelehnt an [Ga23].

3 Testdurchführung

Die Tests werden durchgeführt mit einem Intel Core i7-11800H, 64 GB RAM, Windows 10 Pro Version 22H2, Python 3.9.0, LangChain 0.1.6[5] und Ollama 0.1.30[6] . Alle Chatbots erhalten die gleiche Wissensbasis mit 37 Themen. Die Themen wurden im Rahmen der Anforderungsanalyse des Info-Bots im Jahr 2022 erhoben und stetig in späteren Iterationen des Info-Bots verbessert. Je Thema kann die Wissensbasis stark in ihrer Komplexität variieren, bspw. ist für das Thema *Fundsachen* nur die Information enthalten, wo das Fundbüro sich befindet, während für das Thema *Bachelorarbeit* eine komplexe Beschreibung des Anmeldevorgangs, der Rahmenbedingungen und weitere Informationen enthalten sind.

Allen Chatbots wird die gleiche Testmenge von 125 Fragen gestellt (siehe Anhang), im Schnitt drei Fragen pro Thema. Diese wurden händisch erstellt und basieren teilweise auf real gestellten Fragen von Studierenden an den Info-Bot. Es wurde darauf geachtet, dass keine Frage bereits in den Trainingsdaten des Intent-Ansatzes enthalten ist. Die Fragen unterteilen sich in drei Typen unterschiedlicher Anzahl:

- Bekannte (50): Die Antworten zu diesen Fragen sind jeweils in einzelnen Absätzen der Wissensbasis direkt enthalten.

4 https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2
5 https://www.langchain.com/
6 https://ollama.com

- Inferenz (25): Die Antworten sind nicht direkt in der Wissensbasis enthalten und müssen entweder aus mehreren Absätzen der Wissensbasis kombiniert werden oder mithilfe von Allgemeinwissen bestimmt werden.
- Unbekannte (50): Die Antworten sind nicht in der Wissensbasis enthalten und können nur bedingt mithilfe von Allgemeinwissen inferiert werden.

Die Antworten der Ansätze auf die Fragen werden durch einen Experten (human feedback) einem der folgenden drei Typen zugeordnet:

- Korrekt: Die Frage wird umfassend beantwortet. Dabei ist zu beachten, dass eine nicht beantwortete Frage auch als korrekt beantwortet gilt, wenn die Frage zum Typ der unbekannten Fragen gehört.
- Teilweise korrekt: Die Frage teilweise beantwortet oder enthält zumindest Verweise auf Quellen, in welchen die korrekte Antwort gefunden werden kann (z.B. URLs oder Informationen zu Kontaktpersonen). Eine korrekte Antwort gilt als nur teilweise korrekt, wenn neben den notwendigen Informationen auch irrelevante oder falsche Informationen wie z.B. Halluzinationen ausgegeben werden.
- Falsch: Wenn die Antwort unter keinen der ersten Typen fällt wird sie als falsch kategorisiert.

4 Evaluation

Abb. 2 zeigt die korrekt und teilweise korrekt beantworteten Fragen je Fragentyp jedes Ansatzes und jeden Antworttyp. GPT-3.5 schnitt bei allen Fragentypen am besten ab. Der Intent-Ansatz schnitt am schlechtesten bei den bekannten Fragen ab (34) und Mistral konnte von den unbekannten Fragen nur 33 korrekt, bzw. teilweise korrekt beantworten. Relativ zu allen 125 Fragen konnte der Intent-Ansatz 71,2% aller Fragen korrekt oder teilweise korrekt beantworten, Mistral 72,8% und GPT-3.5 85,6%. Bei den Antworttypen schnitt GPT-3.5 ebenfalls am besten ab. Die Ergebnisse von Mistral und dem Intent-Ansatz sind sehr ähnlich. Auffällig sind die Fragen, bei welchen keine Antwort gegeben wurde. Von den 50 nicht beantworteten Fragen vom Intent-Ansatz waren 40 unbekannte Fragen, welche somit korrekt beantwortet wurden und 10 wurden fälschlicherweise nicht beantwortet, während Mistral und GPT-3.5 2 bzw. 3 Fragen nicht beantworteten.

Besonders zu beachten sind die Fälle der False Positives und False Negatives (siehe Abb. 4). Ein False Positive liegt vor, wenn das Modell fälschlicherweise die Antwort zu kennen glaubt, während bei einem False Negative eine falsche Antwort oder keine Antwort ausgegeben wird, obwohl die Antwort bekannt sein sollte oder hergeleitet werden kann. Für Mistral und GPT-3.5 liegen beide Fälle gleich hoch, wobei Mistral generell mehr aufweist (17 bzw. 9 für GPT-3.5). Der Intent-Ansatz liegt bei den False Positives knapp über GPT-3.5 (10) und erzielte die meisten False Negatives (26).

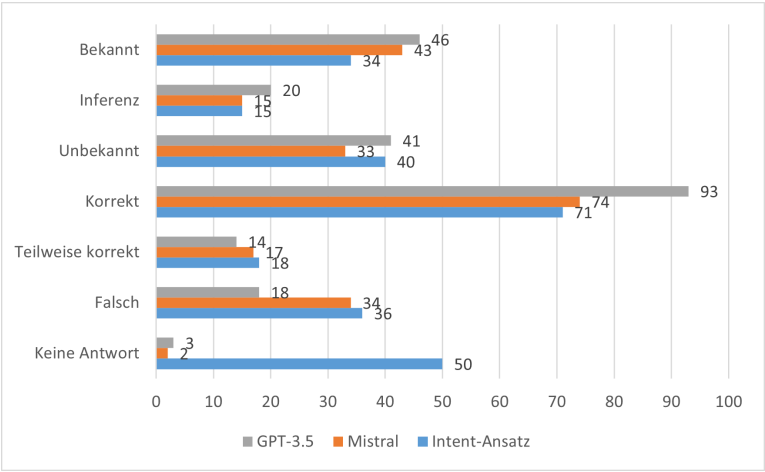


Abb. 2: Summe der korrekt und teilweise korrekt beantworteten Fragen je Ansatz je Fragentyp und Anzahl der Antworttypen je Ansatz.

Mistral und GPT-3.5 haben in 22 bzw. 9 Fällen Halluzinationen ausgegeben. Bei diesen handelte es sich u.a. um nichtexistierende URLs, falsche Annahmen und Folgefehler oder falsche Tatsachen, z.B. fiktive Raumnummern, Öffnungszeiten oder falsche Interpretationen von Abkürzungen. In 4 Fällen wies die Antwort von Mistral Rechtschreib-, Grammatikfehler oder eine schlechte Wortwahl auf. Auf GPT-3.5 traf dies nicht zu.

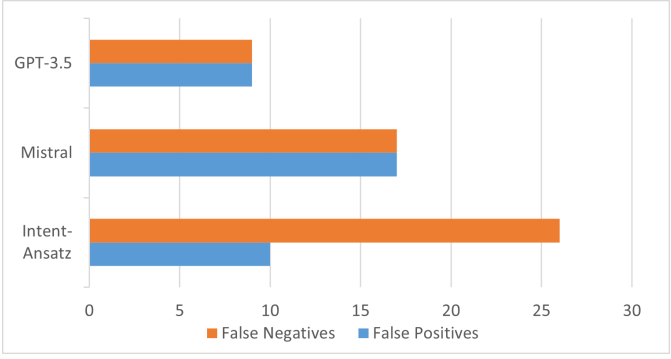


Abb. 3: Anzahl der False Positives und False Negatives je Ansatz.

5 Verwandte Arbeiten

Der von Cherumanal et al. entwickelte FAQ-Chatbot zur Beantwortung von Fragen zu Studienangeboten [Ch24b] wurde Intent-basiert und mittels des LLMs Falcon-7B und RAG entwickelt. Ein FAQ dient als Wissensbasis für beide Ansätze. Diesen wurden

106 Fragen gestellt, welche bekannte, unbekannte und Inferenz-Fragen beinhalten (siehe Abschnitt 3). Bei den bekannten Fragen schnitt der Intent-basierte Ansatz besser ab, während für Inferenz-Fragen das LLM signifikant besser abschnitt. Für die unbekannte Fragen gab der Intent-basierte Ansatz in 80% aller Fälle korrekterweise keine Antwort aus, während das LLM zu etwa 80% eine falsche Antwort ausgab. Die Ergebnisse konnten in diesem Beitrag nur bedingt bestätigt werden (siehe Abschnitt 4). Des Weiteren bearbeiten die Chatbots in diesem Beitrag ein deutlich breiteres Spektrum an Themen als in [Ch24b].

Die in [FWN24] evaluierten zwei Chatbots wurden ebenfalls mit Intents, bzw. LLMs und RAG entwickelt und unterstützen Fabrikarbeiter mit Arbeitsanweisungen und FAQs mit Expertenwissen. Beide Chatbots wurden mit standardisierten Fragebögen hinsichtlich Effizienz der Interaktion, User Experience, Arbeitsbelastung und Benutzerfreundlichkeit untersucht. Das LLM schnitt in vielen Dimensionen der untersuchten Eigenschaften besser ab, während der Intent-basierte Ansatz in keiner Dimension bessere Ergebnisse erzielte. Die Erkenntnisse aus [FWN24] sind nur bedingt vergleichbar mit den Ergebnissen dieses Beitrags, da in letzterem die Ansätze im Vordergrund stehen und nicht die Nutzenden.

6 Diskussion, Fazit und Ausblick

In diesem Beitrag konnte die Effektivität von ChatGPT mit RAG, Mistral mit RAG und einem Intent-basierten Ansatz zur Umsetzung von Chatbots zur Studienorganisation gezeigt werden. Alle Ansätze konnten zwischen etwa 70 und 85% aller Fragen korrekt oder teilweise korrekt beantworten, wobei GPT-3.5 mit 85,6% dominierte. Trotzdem schöpfen alle Ansätze noch nicht ihr volles Potential bei der Verwendung der Wissensbasis aus. Besonders die Halluzinationen der LLMs sind kritisch zu betrachten, weil diese Antworten häufig plausibel und realistisch klingen, weshalb Studierende falsche Antworten möglicherweise nicht als solche erkennen können. Für den Intent-Absatz stellen die knapp 34,66% an False Negatives verschenktes Potential dar, während die 80% korrekterweise nicht beantworteten Fragen für die Verlässlichkeit des Ansatzes sprechen.

Die Ergebnisse legen nahe, im nächsten Schritt einen hybriden Ansatz zu untersuchen, welcher die Stärken der jeweiligen Ansätze kombiniert. Bspw. könnte der Intent-Ansatz primär genutzt werden und falls keine zufriedenstellende Antwort gefunden werden kann, wird eine Antwort durch ein LLM generiert. In einer weiteren Arbeit werden weiterhin Aspekte wie Einfachheit der redaktionellen Wartung und Skalierbarkeit untersucht. Zusätzlich wird ein Augenmerk auf solche Themen gelegt, die nicht durch den Chatbot beantwortet sollen. Diese umfassen z.B. Fragen zum Studienabbruch, psychologische Beratung und Fragen zur Prüfungsordnung, die zum Schutz von Studierenden besser von kompetenten Menschen beantwortet werden sollten.

IPPOLIS wird durch das Bundesministerium für Bildung und Forschung (BMBF) sowie durch das Ministerium für Kultur und Wissenschaft NRW unter dem Förderkennzeichen 16DHBKI050 gefördert. https://fh.do/ippolis

Literaturverzeichnis

[CG21] Chaves, A. P.; Gerosa, M. A.: How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. International Journal of Human–Computer Interaction 37 (8), S. 729–758, 2021.

[Ch24a] Chen, X.; Gao, P.; Song, J.; Tan, X.: HiQA: A Hierarchical Contextual Augmentation RAG for Massive Documents QA. arXiv preprint arXiv:2402.01767, 2024.

[Ch24b] Cherumanal, S. P.; Tian, L.; Abushaqra, F. M.; de Paula, A. F. M.; Ji, K.; Hettiachchi, D.; Trippas, J. R.; Ali, H.; Scholer, F.; Spina, D.: Walert: Putting Conversational Search Knowledge into Action by Building and Evaluating a Large Language Model-Powered Chatbot. arXiv preprint arXiv:2401.07216, 2024.

[FWN24] Freire, S. K.; Wang, C.; Niforatos, E.: Chatbots in Knowledge-Intensive Contexts: Comparing Intent and LLM-Based Systems. arXiv preprint arXiv:2402.04955, 2024.

[Ga23] Gao, Y.; Xiong, Y.; Gao, X.; Jia, K.; Pan, J.; Bi, Y.; Dai, Y.; Sun, J.; Wang, H.: Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997, 2023.

[Ji23] Jiang, A. Q.; Sablayrolles, A.; Mensch, A.; Bamford, C.; Chaplot, D. S.; Casas, D. d. l.; Bressand, F.; Lengyel, G.; Lample, G.; Saulnier, L. et al.: Mistral 7B. arXiv preprint arXiv:2310.06825, 2023.

[Ka23] Karanikolas, N.; Manga, E.; Samaridi, N.; Tousidou, E.; Vassilakopoulos, M.: Large Language Models versus Natural Language Understanding and Generation. In: Proceedings of the 27th Pan-Hellenic Conference on Progress in Computing and Informatics. S. 278–290, 2023.

[Le20] Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.-t.; Rocktäschel, T. et al.: Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in Neural Information Processing Systems 33, S. 9459–9474, 2020.

[Mi24] Minaee, S.; Mikolov, T.; Nikzad, N.; Chenaghlu, M.; Socher, R.; Amatriain, X.; Gao, J.: Large language models: A survey. arXiv preprint arXiv:2402.06196, 2024.

[Ne20] Neumann, A. T.; de Lange, P.; Klamma, R.; Pengel, N.; Arndt, T.: Intelligent Mentoring Bots in Learning Management Systems: Concepts, Realizations and Evaluations. In: International Symposium on Emerging Technologies for Education. Springer, S. 3–14, 2020.

[Op22] OpenAI: Introducing chatgpt, 2022, URL: https://openai.com/blog/chatgpt, Stand: 08. 04. 2024.

[Pe09] Peterson, L. E.: K-nearest neighbor. Scholarpedia 4 (2), S. 1883, 2009.

[Qi18] Qiu, L.; Chen, Y.; Jia, H.; Zhang, Z.: Query intent recognition based on multi-class features. IEEE Access 6, S. 52195–52204, 2018.

[Ro23] Ross, S. I.; Martinez, F.; Houde, S.; Muller, M.; Weisz, J. D.: The programmer's assistant: Conversational interaction with a large language model for software development. In: Proceedings of the 28th International Conference on Intelligent User Interfaces. S. 491–514, 2023.

[SW19] Schulze-Reichelt, F.; Wippermann, M.: Die Bedeutung von Fachschaftsräten in der Studieneingangsphase am Beispiel der Universität Potsdam. Alles auf Anfang! Befunde und Perspektiven zum Studieneingang, S. 149–164, 2019.

[Ya23] Yang, J.; Jin, H.; Tang, R.; Han, X.; Feng, Q.; Jiang, H.; Zhong, S.; Yin, B.; Hu, X.: Harnessing the power of llms in practice: A survey on chatgpt and beyond. ACM Transactions on Knowledge Discovery from Data, 2023.

Anhang A

LLM-Prompt: Du bist Info-Bot, ein hilfreicher Assistent für Studierende an der Fachhochschule Dortmund. Benutze ausschließlich die folgenden Informationen, um Fragen zu beantworten. Wenn eine Frage nicht mit Hilfe der vorliegenden Informationen beantwortet werden kann, dann sag es:

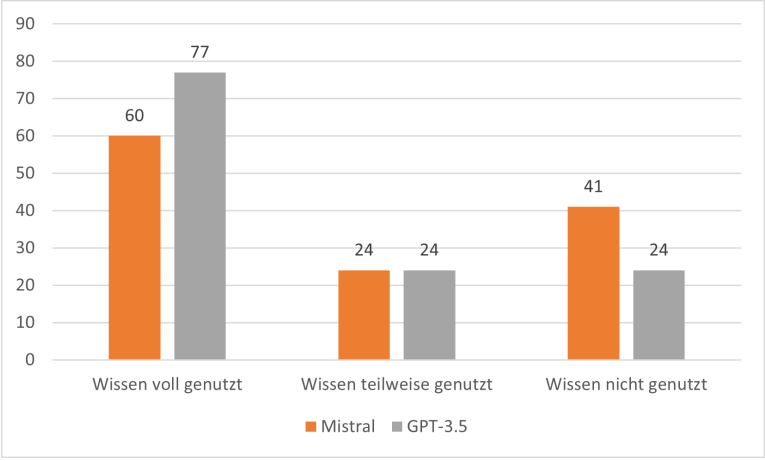


Abb. 4: Anzahl der Antworten unter Verwendung, teilweiser Verwendung und nicht Nutzung der Wissensbasis für Mistral und GPT-3.5. Wenn die Wissensbasis voll genutzt wurde, ist die Antwort entweder wortwörtlich übernommen oder stark an das Wissen aus der Wissensbasis angelehnt worden, analog dazu die teilweise und Nichtnutzung der Wissensbasis. Auffällig ist, dass GPT-3.5 häufiger als Mistral die Wissensbasis voll nutzt (70 bzw. 60 Antworten). Insgesamt wurden von Mistral und GPT-3.5 46 bzw. 60 der bekannten und Inferenzfragen korrekt und unter Verwendung der Wissensbasis beantwortet. Von den unbekannten Fragen konnten 62% von Mistral, bzw. 76% von GPT-3.5 korrekt oder teilweise korrekt beantwortet worden, obwohl das nötige Wissen nicht in der Wissensbasis vorhanden war. Für bekannte und Inferenzfragen wurde in 6 bzw. 4 Fällen das korrekt vorliegende Wissen aus der Wissensbasis von Mistral, bzw. GPT-3.5 ignoriert und stattdessen eine eigene falsche Antwort generiert.

Anhang B

Tab. 1: Fragenkatalog der 125 gestellten Fragen an die Ansätze.

Anfang Tabelle		
Frage	Thema	Typ
Was ist ein Auslandssemester?	Auslandssemester	bekannt
Wo finde ich Informationen zu Auslandssemestern?	Auslandssemester	bekannt
Wie fängt man seine Bachelorarbeit an?	Bachelorarbeit	Inferenz
Wo finde ich das Merkblatt für die Abschlussarbeit?	Bachelorarbeit	Inferenz
Wie melde ich meine Bachelorarbeit an?	Bachelorarbeit	unbekannt
Was benötige ich alles für die Anmeldung meiner Bachelorthesis?	Bachelorarbeit	Inferenz
Wie lange habe ich Zeit für die Bachelorthesis?	Bachelorarbeit	bekannt
Wie finde ich den passenden Professor für meine Bachelorarbeit?	Bachelorarbeit	bekannt
Wie finde ich ein Thema für meine Bachelorthesis?	Bachelorarbeit	bekannt
Was muss ich für die Abgabe meiner Bachelorarbeit machen?	Bachelorarbeit	bekannt
Was macht der Career Service?	CareerService	bekannt
Was für Kurse werden vom Career Service angeboten?	CareerService	bekannt
Wo kann ich mich für die Veranstaltungen des Career Service anmelden?	CareerService	bekannt
Can you speak English?	EnglischSprechen	bekannt
Which languages can you speak?	EnglischSprechen	bekannt
An wen kann ich mich wenden, wenn ich den Kurs Rechnernetze von meiner alten Hochschule anrechnen lassen möchte?	FaecherAnrechnen	bekannt
Ich möchte meine bestandenen Prüfungen aus meinem alten Studium anrechnen lassen	FaecherAnrechnen	bekannt
Wie kann ich mir Präsentationstechniken für LuA anrechnen lassen?	FaecherAnrechnen	Inferenz
Wie kann ich mir Präsentationstechniken für Lern und Arbeitstechniken anrechnen?	FaecherAnrechnen	Inferenz
Wo finde ich das Formular für ein Urlaubssemester?	Formulare	Inferenz
Wo finde ich die Formulare des Studienbüros?	Formulare	bekannt
Wo werden Fundsachen abgegeben?	Fundsachen	bekannt
Ich habe meine Maus im Praktikum vergessen	Fundsachen	Inferenz
Ich hab meinen Regenschirm verloren, wo kann ich den abholen?	Fundsachen	Inferenz
Wie lange hat die kostbar heute auf?	Gastronomie	unbekannt
In welcher Gruppe bin ich?	Gruppe	bekannt
Wie erfahre ich meine Gruppennummer?	Gruppe	bekannt
Wie lautet mein Gruppenbuchstabe?	Gruppe	bekannt
Welche Gruppe bin ich im Stundenplan?	Gruppe	bekannt
Wie melde ich mich beim Hochschulsport an?	Hochschulsport	unbekannt
Wie teuer ist eine Sportkarte?	Hochschulsport	bekannt
Wann beginnt das Sportangebot des Hochschulsports?	Hochschulsport	unbekannt

Fortsetzung Tabelle 1		
Frage	Thema	Typ
Wann ist die Schnupperwoche des Hochschulsports?	Hochschulsport	bekannt
Wann bekomme ich eine Bestätigung das ich auf einen NC beschränkten Studiengang angenommen wurde?	Immatrikulation	unbekannt
Wo finde ich Informationen für Erstsemester?	Erstsemester	bekannt
Wo finde ich den Fachschaftsrat?	Erstsemester	bekannt
Wie richte ich das VPN richtig ein?	ItDienste	Inferenz
Wo finde ich Hilfe zu den IT-Diensten?	ItDienste	bekannt
Wie lauten die Anmeldebedingungen für eine Nachschreibeklausur?	Klausuren	unbekannt
Wie komme ich an die Notenschnitte meiner geschriebenen Klausuren?	Klausuren	unbekannt
Was mache ich, wenn meine Klausur falsch bewertet wurde?	Klausuren	unbekannt
Wie bestehe ich die Mathe 2 Prüfung?	Klausuren	unbekannt
Theaterkarten für Studierende	Kulturangebote	unbekannt
Welcher Professor kennt sich mit Medientechnik aus?	Kontaktpersonen	Inferenz
Welcher Prof macht LinA2?	Kontaktpersonen	Inferenz
Wie funktioniert Lern- und Arbeitstechniken?	LuA	bekannt
Wann muss ich Lern- und Arbeitstechniken machen?	LuA	bekannt
Wann ist die Einführungsveranstaltung von Lern- und Arbeitstechniken?	LuA	bekannt
Kann ich Präsentationstechniken vom Studium Generale für LuA anrechnen lassen?	LuA	Inferenz
Gibt es eine Anwesendheitspflicht für LuA?	LuA	unbekannt
Wer macht Lern- und Arbeitstechniken?	LuA	unbekannt
Wie fängt man mit seiner Masterthesis an?	Masterarbeit	unbekannt
Wo melde ich meine Masterthesis an?	Masterarbeit	unbekannt
Wie lange habe ich Zeit für die Masterthesis?	Masterarbeit	bekannt
Wie finde ich den passenden Professor für meine Masterarbeit?	Masterarbeit	bekannt
Wie finde ich ein Thema für meine Masterthesis?	Masterarbeit	bekannt
Was muss ich für die Abgabe meiner Masterarbeit machen?	Masterarbeit	bekannt
Was macht das Mentoring?	Mentoring	bekannt
Wo finde ich Informationen zum Mentoring?	Mentoring	bekannt
Muss ich am Mentoring teilnehmen?	Mentoring	Inferenz
Ich habe bis jetzt keine Einladung zum Mentoring Gespräch erhalten	Mentoring	unbekannt
Ich will kein Mentoring-Gespräch durchführen. Was soll ich tun?	Mentoring	unbekannt
Ich habe das Mentoring Angebot nicht wahrgenommen. Was soll ich tun?	Mentoring	unbekannt
Worüber wird im Mentoring-Gespräch gesprochen?	Mentoring	unbekannt
Wie funktioniert die Einladung zum ersten Mentoring-Gespräch?	Mentoring	unbekannt
Wie kann ich ein Praxissemester machen?	Praxissemester	unbekannt

Fortsetzung Tabelle 1

Frage	Thema	Typ
Bei welchen Unternehmen kann ich ein Praxissemester machen?	Praxissemester	unbekannt
Praxissemester Ergebnisse	Praxissemester	unbekannt
Wie fang ich meine Projektarbeit an?	Projektarbeit	unbekannt
Wie lange habe ich Zeit für meine Projektarbeit?	Projektarbeit	unbekannt
Wann ist die Abgabe für meine Projektarbeit?	Projektarbeit	bekannt
Bis wann muss ich meine Projektarbeit fertig haben?	Projektarbeit	Inferenz
In welchem Semester muss ich die Projektarbeit schreiben?	Projektarbeit	bekannt
Wie kann ich meine Projektarbeit abgeben?	Projektarbeit	bekannt
Wie läuft die Projektarbeit in meinem Studiengang ab?	Projektarbeit	unbekannt
Kann ich die Projektarbeit und die Bachelorarbeit direkt hintereinander machen?	Projektarbeit	unbekannt
Muss der Betreuer für die Projektarbeit ein Professor sein oder geht auch ein wissenschaftlicher Mitarbeiter?	Projektarbeit	unbekannt
Wie wird die Projektarbeit bewertet?	Projektarbeit	unbekannt
Wann läuft die Prüfungsordnung aus für Studierende der Wirtschaftsinformatik im Praxissemester die ihr Studium vor dem Wintersemester 2018/19 aufnahmen?	Pruefungsordnung	unbekannt
Wie ist der Link zum Prüfungsplan?	Pruefungsplan	bekannt
Wann kann ich eine Klausur nachschreiben?	Pruefungsplan	Inferenz
Wann finden die Prüfungen statt?	Pruefungsplan	Inferenz
Wann finden mündliche Prüfungen statt?	Pruefungsplan	Inferenz
Ist der neue Prüfungsplan schon raus?	Pruefungsplan	unbekannt
Kann ich vom Semesterbeitrag befreit werden?	Semesterbeitrag	unbekannt
Ich habe ein Auto, kann ich vom Semesterticket befreit werden?	Semesterticket	unbekannt
Wie funktionieren Seminare?	Seminare	bekannt
Kriegt man eine Note im Seminar?	Seminare	unbekannt
Wird Seminar Methodik benotet?	Seminare	unbekannt
Ich suche den Link zum Seminarplan	Seminare	bekannt
Wie kann ich mein Studium finanzieren?	Studienfinanzierung	bekannt
Wie bekomme ich BAföG?	Studienfinanzierung	unbekannt
Wie bekommt man ein Stipendium?	Studienfinanzierung	unbekannt
Wo sehe ich meinen Stundenplan?	Stundenplan	bekannt
Wo finde ich die Vorlesungszeiten?	Stundenplan	Inferenz
Wann ist die Übung von Web-Tech?	Stundenplan	unbekannt
Wo ist das Praktikum am Donnerstag von Rechnernetze?	Stundenplan	unbekannt
Wann kommt der neue Stundenplan?	Stundenplan	unbekannt
Wo findet das Seminar Robotik im Fachbereich Informatik statt	Stundenplan	unbekannt
Wie funktioniert Technisches Englisch?	TechnischesEnglisch	bekannt
Wann findet technisches Englisch statt?	TechnischesEnglisch	Inferenz
Kann ich den Einstufungstest zu Technisches Englisch nachschreiben?	TechnischesEnglisch	Inferenz

Fortsetzung Tabelle 1		
Frage	Thema	Typ
Ich habe den Einstufungstest zu Technisches Englisch verpasst	TechnischesEnglisch	bekannt
Wann ist der Einstufungstest zu Technisches Englisch?	TechnischesEnglisch	Inferenz
Wo finde ich Informationen zum Verbundstudiengang?	Verbundstudiengang	unbekannt
Welche Wahlpflichtmodule kann ich wählen?	Wahlpflichtmodule	Inferenz
Wie funktionieren Wahlpflichtfächer?	Wahlpflichtmodule	unbekannt
Wo finde ich eine Studentenwohnung?	Wohnen	bekannt
Wo finde ich die Zulassung für PK2?	Zulassung	bekannt
Ich bin im 8. Semester und mir fehlt noch eine Klausur aus dem 3. Semester, darf ich in diesem Semester Klausuren schreiben?	Zulassung	unbekannt
Wie läuft das SOPRA ab?	Verschiedenes	unbekannt
Wie läuft das Software-Praktikum ab?	Verschiedenes	unbekannt
Kannst du mir sagen, welche Veranstaltungen ich konkret habe?	Verschiedenes	unbekannt
Was muss ich im sechsten Semester belegen?	Verschiedenes	Inferenz
Wie kann ich Zweithörer werden?	Verschiedenes	unbekannt
Was bedeutet QDL?	Abkürzungen	unbekannt
Wofür steht ECTS?	Abkürzungen	unbekannt
Wie ist dein Name?	Smalltalk	bekannt
Wer bist du?	Smalltalk	bekannt
Wer hat dich entwickelt?	Smalltalk	bekannt
Wer hat dich programmiert?	Smalltalk	Inferenz
Arbeitest du auf Basis von chatGPT?	Smalltalk	Inferenz
Speicherst du meinen Text?	Smalltalk	unbekannt
Wie geht es dir?	Smalltalk	bekannt
Wie fühlst du dich?	Smalltalk	bekannt
Ende Tabelle 1		

ChatGPT erzähl mir eine Geschichte: Die Verwandlung von Lernwelten durch KI-gestützte Erzählungen

Rebecca Finster [1], Linda Grogorick [1] und Susanne Robra-Bissantz [1]

Abstract: Geschichten, gut erzählt, schaffen Neugierde, Aufmerksamkeit und Interesse. Im Bildungsbereich ist Storytelling bereits seit vielen Jahren ein beliebtes Mittel, um Wissen weiterzugeben. Dabei können Geschichten manuell oder automatisiert entwickelt werden. Das automatisierte Storytelling hat sich insbesondere durch die Weiterentwicklung von Large Language Models sowie dem damit einhergehenden Aufkommen von ChatGPT gewandelt. In diesem Beitrag wird nach dem Action Design Research Ansatz eine Instanziierung von automatisiertem Storytelling in der begleitenden Übung einer Grundlagenvorlesung durchgeführt, um die transformativen Potenziale beim Lernen zu untersuchen. Ziel ist es zu verstehen, wie praxisnahe Erzählungen das Verständnis und die Motivation der Studierenden fördern können. Ergebnisse der Studie zeigen, dass sorgfältig auf die Lerninhalte abgestimmte, automatisiert generierte Erzählungen nicht nur das Verständnis erleichtern, sondern auch als effizientes Werkzeug zur Entwicklung von Lernmaterialien dienen können, sofern eine sorgfältige Qualitätskontrolle der Inhalte sichergestellt ist.

Keywords: Storytelling, ChatGPT, Hochschullehre

1 Einleitung

Das Erzählen von Geschichten ist für die Bildungsforschung von großer Bedeutung, da sie einen narrativen Kontext für die Verarbeitung von Lerninhalten bieten [Ga11]. Obwohl sie Teil der realen Welt sind, schaffen sie eine immersive fiktionale Welt in den Köpfen der Menschen [Lu17]. Eigene Vorstellungen der erzählten Ereignisse werden lernförderlich durch die Verknüpfung mit Erfahrungen und Kenntnissen kreiert [Sp09].

Um die Entwicklung von Geschichten zu vereinfachen, können diese automatisch generiert werden. Wir möchten in einem Action Design Research (ADR) Ansatz in einer Lehrveranstaltung die transformativen Chancen von automatisiertem Storytelling erforschen und verstehen, inwiefern praxisnahe Erzählungen das Verständnis und die Motivation der Lernenden und Lehrenden fördern. ADR ist ein praxisnaher gestaltungsorientierter Forschungsansatz, bei dem mit Stakeholdern in Iterationen nach einer Problemlösung gesucht wird, der sich auch im lernzentrierten Kontext anbietet [Se11][RS20]. Die Forschungsfrage – *Inwiefern verändert automatisiertes Storytelling das Lernen und Lehren?* – reflektiert die Annahme, dass narrative Elemente das Verständnis abstrakter Konzepte erleichtern und die Anwendung theoretischen Wissens in praktischen Szenarien fördern.

1 Technische Universität Braunschweig, Inst. f. Wirtschaftsinformatik - Service-Informationssysteme, Mühlenpfordtstraße 23 , 38106 Braunschweig, Deutschland,
r.finster@tu-braunschweig.de, https://orcid.org/0000-0002-1686-9566;
l.grogorick@tu-braunschweig.de, https://orcid.org/0000-0003-0270-2666;
s.robra-bissantz@tu-braunschweig.de, https://orcid.org/0000-0001-9972-3935

doi:10.18420/delfi2024_13

Zwar wurde schon 2017 der Einsatz von automatisiertem Storytelling in der Lehre diskutiert [Za17], doch konzentriert sich die Forschung hauptsächlich auf theoretisch-konzeptionelle Ansätze, ohne praktische Anwendung. Mit der Entwicklung von Chatbots, besonders ChatGPT [Br20], haben sich die Möglichkeiten des automatisierten Storytelling weiterentwickelt. So wurden beispielsweise schon die Personalisierung der Lehre durch ChatGPT [JG23] [BA23] und dessen Einfluss auf die Storytelling-Fähigkeiten von Lernenden [Sh23] [Ch23] [Li24] untersucht. Besonders hervorzuheben sind eine philippinische Studie, die die Vermittlung von Werten durch von ChatGPT generierte Geschichten mittels Prompt Engineering untersucht [Gr23a], und eine Untersuchung zu Serious Games über Klimawandel, die sich auf Kohärenz, Lesbarkeit und Komplexität der Narrative konzentriert [Gu23]. Im Gegensatz zu bisherigen Arbeiten konzentrieren wir uns auf die praktische Anwendung von ChatGPT als reiner Geschichtenerzähler im Hochschulbereich, wobei wir die Bedürfnisse der Lernenden und Lehrenden gleichermaßen betrachten.

Im Folgenden gehen wir zunächst auf automatisiertes Storytelling ein, ehe wir unseren Praxisfall und unser Vorgehen vorstellen. Im Anschluss präsentieren wir unsere Evaluation und Reflexion ehe wir ein Fazit ziehen.

2 Automatisiertes Storytelling

Sobald das Erzählen von Geschichten mit einer Mischung aus digitalen Medien, einschließlich Text, Bilder, Videos oder Audioaufnahmen, kombiniert wird, ist die Rede von digitalem Storytelling [Ro06]. Heutzutage wird es vielfältig verwendet, häufig jedoch zum Erlernen wesentlicher Kompetenzen (z.B. Informationskompetenz) des 21. Jahrhunderts [Ro08] [Gr23b]. Digitales Storytelling kann insbesondere dazu verwendet werden abstrakte und konzeptionelle Inhalte verständlicher zu machen. Durch die Einbindung von digitalen Geschichten sparen Lehrende nicht nur Zeit, sondern erleichtern auch den Diskussionseinstieg in behandelte Themen [Ro16][Ro08]. Die Entwicklung von digitalen Geschichten kann manuell oder durch automatisierte Prozesse erfolgen, wobei Letztere die Herausforderung haben, Handlung konsistent zu gestalten und Abhängigkeiten zu wahren [WSR17]. Früher erforderten automatisch erstellte Geschichten erheblichen manuellen Aufwand oder spezialisierte Algorithmen [Ge09]. Heute ermöglichen Large Language Models, wie ChatGPT [Br20], das Generieren von individualisierten Geschichten und ihre Einbettung in neue Kontexte.

Beim Blick in ein Bücherregal mit Werken wie „Herr der Ringe“ und „Der Hobbit“ von J.R.R. Tolkien sehen wir mehrere Bücher innerhalb der kohärenten Welt Mittelerdes, die große Erzählstränge (z.B. „Ring muss in den Vulkan“) und Einzelgeschichten (z.B. „Zwerg und Elb freunden sich an“) enthalten. So eine konsistente Erzählwelt ist auch in automatisiertem Storytelling wichtig und kann in fünf automatisierbaren Phasen, die manuell ergänz- oder ersetzbar sind, kreiert werden [WFB14]. Tab. 1 beschreibt diese fünf Phasen und wie ChatGPT dort unterstützen kann.

Phase	Beschreibung	ChatGPT (GPT-4)
Erkennung und Extraktion	Erkennung und Extraktion von Geschichten aus umfangreichen textuellen Datenquellen, z.B. Social Media	Prompts denkbar, wo GPT-4 aus vorgegebenen Datenquellen narrative Strukturen extrahiert, aber keine Stärke des GPT.
Repräsentation	Darstellung der extrahierten Geschichten oft in graphen- oder tupelbasierten Datenstrukturen, die kausale oder temporale Beziehungen zwischen Ereignissen abbilden [RB13]	GPT-4 kann Zusammenfassungen generieren, Kernaspekte identifizieren und die Beziehungen zwischen Ereignissen oder Charakteren in natürlicher Sprache statt in Tupeln darstellen.
Merging / Matching	Matching oder Merging von verschiedenen Geschichten innerhalb eines vorgegebenen narrativen Rahmens oder einer Welt	GPT-4 kann Inhalte verschiedener Quellen „verstehen“, in Bezug setzen und auf Erzählwelten übertragen.
Generierung	Erzeugung einer spezifischen Geschichte aus verschiedenen Möglichkeiten, abhängig von definierten Zielen, Nutzerinteraktionen oder anderen Kriterien.	Stärke von GPT-4 [Al23]: Es kann auf Anfrage neue Geschichten basierend auf Stichworten erstellen und interaktiv dynamische Anpassungen vornehmen.
Storytelling	Präsentation der generierten Geschichte in einer ansprechenden Weise, oft durch Erzählstrukturen oder -techniken, die die Geschichte lebendig wirken lassen.	Die Sprachfähigkeiten von GPT-4 ermöglichen ihm verschiedene Erzählstile und -strukturen anzuwenden, um die Erzählung nach Vorgabe zu optimieren.

Tab. 1: GPT-4 als Geschichtenerzähler [WFB14][Br20]

Wir sehen das Potenzial von ChatGPT als Geschichtenerzähler bei dem konkrete Kontexte, Regeln und Lerninhalte vorgegeben und überprüft werden, um das kreative Potenzial des GPT auszuschöpfen ohne dabei die Korrektheit der Lerninhalte zu gefährden und damit Lernenden eine fesselndere anpassbare Lernwelt zu schaffen. Die natürlichsprachliche Interaktion öffnet auch technisch weniger versierten Lehrkräften neue Wege der Unterrichtsgestaltung.

3 Instanziierung: EWI

Im Rahmen der Vorlesung „Einführung in die Wirtschaftsinformatik“(EWI), für ca. 300 Studierende, wurde die Implementierung von automatisiertem Storytelling mit ChatGPT bei der Erstellung von Lehr-/Lernmaterialien untersucht. Aktuell wird die Veranstaltung nach mehreren pandemiebedingten Online-Semestern in einem Hybrid-Konzept umgesetzt, bei dem der Vorlesungs- und Übungsbetrieb wieder in Präsenz stattfindet und durch digitale Courseware-Materialen ergänzt wird. Seit 2020 führen wir Online-Prüfungen im Open-Book-Format durch, bei denen die Studierenden anhand einer Fallstudie die Lerninhalte

erklären und anwenden sollen. Bei den quantitativen Lehrevaluationen ist mehrfach ein fehlender roter Faden und fehlender Praxisbezug angemerkt worden, so dass wir hier jetzt nachgebessert haben: In der begleitenden Übung setzten wir ChatGPT (GPT-4) ein, um eine fortlaufende Geschichte zu erstellen, die die verschiedenen Lerninhalte mit einem praxisnahen roten Faden verbindet.

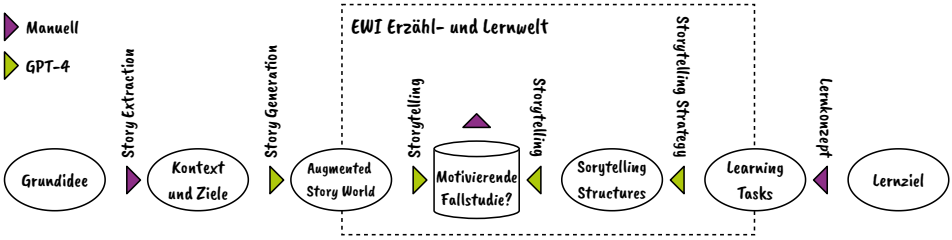


Abb. 1: Storytelling in der EWI

In Abb. 1 haben wir die Struktur unseres Storytellingansatzes visualisiert und erläutern nun unser Vorgehen anhand der verschiedenen In- und Outputs (Ovale). Unser Ansatz kreiert zunächst die Grundgeschichte (links in Abb. 1) und damit eine Lern- und Erzählwelt (gestrichelter Kasten), um dann Lernziele in konkrete Aufgaben und Storys zu übersetzen, die in diese Welt eingebettet sind (rechts in Abb. 1).

Grundidee | Lernziel Die Lernziele sind durch unser Curriculum festgelegt, während die Grundidee unserer Geschichte aus einem detaillierten Datenflussplan einer Brauerei aus einer älteren Übung entwickelt wurde.

Kontext und Ziele | Learning Tasks Manuelle Story Extraction und Definition des Lernkonzepts, in dem Kontext zur Brauerei, wie Craftbeer, deren Geschichte, und der Lehrkontext definiert sowie konkrete Methoden, z.B. OLAP, zum Üben ausgewählt wurden.

Augmented Story World | Storytelling Structures Mit den Kontexten aus der letzten Phase wurde ChatGPT gebeten eine Geschichte zu erstellen, die als roter Faden für alle Übungen dienen sollte. Auf der anderen Seite wurden passende Storyfragmente zu den angedachten Übungsaufgaben generiert. Nachfolgend ein Beispiel für einen Promptverlauf, in dem der kohärente Kontext gezeigt wird:

```
You: erinnerst du dich an unsere fallstudie zur brauerei?
ChatGPT: Ja, ich erinnere mich an die Fallstudie zur Brauerei,
die wir zuvor diskutiert haben. [...]
You: ich brauche zur weiterführung der fallstudie einen text aus dem
studierende informationen zu personas, user stories und erm-
diagrammen extrahieren können.
```

Storytelling In diesem Schritt wurde die Basisstory ausgegeben, die dann für alle Übungen wiedererkennbar gezeigt wurde (s. Abb. 2, oben rechts) und die jeweiligen Einzelgeschichten

für die Übungsthemen (s. Abb. 2, unten rechts). Bei Bedarf wurden die Stories von Hand angepasst. Um sich besser an die Geschichte zu erinnern wurde immer dasselbe Bild (s. Abb. 2, links) in den Folien gezeigt, wenn wir Aufgaben der Brauerei bearbeitet haben. Auf den Folgefolien waren die weiteren Abschnitte der Geschichte mit den notwendigen Details für die Übungsaufgaben, z.B. für das Thema OLAP: *Die Brauerei bietet eine beeindruckende Vielfalt an Biersorten, darunter Pale Ales, IPAs, [...] und Pilsner, die in verschiedenen Größen [...] erhältlich sind, angefangen bei klassischen 12-Unzen-Flaschen bis hin zu 2-Liter-Flaschen und Six Packs. Die Leidenschaft für erstklassige Zutaten und das handwerkliche Können der Braumeister spiegeln sich in jedem Schluck Bier wider.*

Diese Methode eröffnete uns neue Wege, theoretische Inhalte praktisch motivierend zu vermitteln und bereitete Studierende auf die fallstudienbasierte Prüfung vor, bei der analog zur Bierbrauerei eine Gärtnerei Thema war. Die Integration eines durchgängigen roten Fadens und die abwechslungsreichen, aber kapazitätssparenden, Möglichkeiten boten uns die Chance, die Veranstaltung eingängiger und unterhaltsamer zu gestalten.



Die "BieerMeister Brauerei" ist eine renommierte Brauerei, die eine breite Palette hochwertiger Biere produziert und vertreibt. Mit einer über hundertjährigen Geschichte und einem starken Engagement für Qualität und Tradition hat sich BierMeister einen festen Platz in der Welt der Craft-Biere erobert.

Die "BieerMeister Brauerei" ist bestrebt, ihre finanzielle Datenanalyse weiter zu verbessern und eine umfassende Übersicht über ihre Geschäftstätigkeit zu erhalten. Ein OLAP-Würfel wird derzeit entwickelt, um diese Analyse zu erleichtern und den Entscheidungsträgern in der Finanzabteilung wertvolle Einblicke zu liefern. Durch die Nutzung dieser Daten wird die Brauerei ihre Position als führender Anbieter von handwerklich gebrautem Bier weiter festigen und ihr Wachstum vorantreiben.

Abb. 2: Beispielfolie mit unserer Fallstudie

4 Evaluation und Reflexion

Die Evaluation unseres Ansatzes zum automatisierten Storytelling in der Lehre basiert auf einem qualitativen Teaching Analysis Poll (TAP), den wir mit etwa 20 Freiwilligen im letzten Übungstermin durchgeführt haben. Der, unter externer Aufsicht durchgeführte, Online-Fragebogen zielte darauf ab erste Einblicke in die Wirksamkeit des Storytelling-Ansatzes aus Sicht der Lernenden, gerade in Bezug auf Lernunterstützung, Rolle der KI sowie Vor- und Nachteile der KI-Unterstützung, zu gewinnen.

Studierende empfanden die Geschichte als hilfreich für das Verständnis komplexer Modelle und die Verbindung von Theorie mit Praxis, wobei die ansprechende Darstellung und der Praxisbezug besonders hervorgehoben wurden, wie diese Zitate belegen: *„Die Themen der Vorlesung wurden damit in der Praxis verständlich aufgegriffen. Das hat mir geholfen zu verstehen, was in der Vorlesung gemeint war“*, *„Hat mir komplexe Modelle/Sachverhalte*

anschaulich dargestellt“ und *„...war schon recht gut da man trockene Theorie anwenden konnte. Es hat geholfen das es ein Thema war was viele Studierende kennen, Bier. Das hat es auch etwas lustig gestaltet“*. Kritisch angemerkt wurde die Notwendigkeit einer Überprüfung der KI-generierten Inhalte auf Richtigkeit und Relevanz, wie das Zitat *„Es ist sicherlich gut, wenn da immer noch jemand drüber schaut“* zeigt. Vorschläge zur Verbesserung umfassen eine Erhöhung der Interaktivität und eine umfassendere Integration in den Lehrplan. Die Effizienzsteigerung bei der Erstellung von Lehrmaterialien und die Möglichkeit, durch KI neue Perspektiven einzubringen, wurden als Vorteile gesehen. Es wird das Potenzial gesehen, dass automatisiertes Storytelling die Lehre durch praxisnahe und interaktive Inhalte bereichert, sofern eine sorgfältige Überprüfung und Anpassung der Inhalte gewährleistet ist.

Der Einsatz von automatisiertem Storytelling im Lehrkonzept hat sowohl bei der Erstellung von Lernmaterialien entlastet, als auch eine tiefere Auseinandersetzung mit den Lerninhalten ermöglicht, indem es eine Fokussierung auf die fachliche Korrektheit erleichterte. Die Prompterstellung erforderte anfangs ein bisschen Übung und neben der Korrektheit der Lerninhalte musste auch die Kohärenz und ein unerwünschter Bias der Story immer wieder überprüft und angepasst werden. Die Flexibilität dieser Methode erlaubte es, auf aktuelle Ereignisse reagieren und die Lehrinhalte dynamisch anpassen zu können, was zu einer vielseitigen Lehr-/Lernerfahrung führte. Obwohl wir nur eine einzige Veranstaltung mit begrenzter Teilnehmerzahl untersuchen konnten, zeigte sich eine deutliche Verbesserung in der Kohärenz des Lehrmaterials und eine intensivere Beschäftigung der Studierenden mit den Inhalten. Wie in der Evaluation gewünscht wird zukünftig die Entwicklung diverser Fallstudien für unterschiedliche Lerngruppen angedacht, um die Anwendbarkeit und den Transfer der Lerninhalte zu fördern. Die Einbeziehung der Studierenden in die Erstellung von Geschichten könnte die Interaktivität weiter verstärken und den Lernprozess bereichern.

5 Fazit

In diesem Beitrag haben wir den Einsatz von ChatGPT als Geschichtenerzähler in einer Grundlagenveranstaltung der Wirtschaftsinformatik vorgestellt, haben unser Vorgehen evaluiert und reflektiert und sind auf den Mehrwert für Lehrende und Lernende sowie zukünftige Einsatzszenarien eingegangen. ChatGPT eignet sich hervorragend als Assistent bei der Erstellung von Lernmaterialien mit erzählerischer Komponente in nahezu allen Fachdisziplinen, solange die fachlichen Inhalte weiterhin von Lehrenden geliefert und zusammen mit der Kohärenz der Story überprüft werden. Dadurch entsteht nicht nur eine Entlastung der zeitlichen und kognitiven Kapazitäten der Lehrenden, sondern Lehrende und Lernende können mit neuem Interesse und immer neuen Blickwinkeln an die Lerninhalte herantreten und in einem relevanten Kontext kennenlernen. Dabei sehen wir nicht nur Potenziale in der Hochschulbildung, sondern möchten diese in Zukunft auch in der (personalisierten) Weiterbildung untersuchen. Mit ChatGPT als Geschichtenerzähler öffnen wir ein neues Kapitel in unserer Lehre, in dem KI-gestützte Erzählungen Lernwelten nicht nur verändern, sondern immersiv verwandeln.

Literatur

[Al23] Alvarez, A.: ChatGPT as a Narrative Structure Interpreter. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 14384 LNCS/, S. 113–121, 2023, ISSN: 0302-9743.

[BA23] Baidoo-anu, D.; Ansah, L. O.: Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. Journal of AI 7/1, S. 52–62, 2023, ISSN: 3023-4018.

[Br20] Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D. M.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; Amodei, D.: Language Models Are Few-Shot Learners, 2020, arXiv: 2005.14165 [cs].

[Ch23] Chen, X.; Xie, H.; Zou, D.; Wang, F. L.: ChatGPT for Generating Stories and Mind-Maps in Storytelling. In: 2023 10th International Conference on Behavioural and Social Computing (BESC). S. 1–8, 2023.

[Ga11] Gallagher, K. M.: In Search of a Theoretical Basis for Storytelling in Education Research: Story as Method. International Journal of Research & Method in Education 34/1, S. 49–61, 2011, ISSN: 1743-727X, 1743-7288.

[Ge09] Gervas, P.: Computational Approaches to Storytelling and Creativity. AI Magazine 30/3, S. 49–49, 2009, ISSN: 2371-9621.

[Gr23a] Gregorio, A.; Manuel, S.; Palmares, A.; Tang, S.; Ong, E.: Generating Interactive Stories with ChatGPT to Teach Filipino Values. In: 31st International Conference on Computers in Education, ICCE 2023 - Proceedings. Bd. 2, S. 52–61, 2023, ISBN: 9786269689026.

[Gr23b] Grogorick, L.; Rohde, M.; Khosrawi-Rad, B.; Robra-Bissantz, S.: Fiction or Reality – Which Game Story Promotes Learning Outcome More? 36th Bled eConference. In: Digital Economy and Society: The Balancing Act for Digital Innovation in Times of Instability, Conference Proceedings. S. 441–456, 2023.

[Gu23] Gursesli, M. C.; Taveekitworachai, P.; Abdullah, F.; Dewantoro, M. F.; Lanata, A.; Guazzini, A.; Lê, V. K.; Villars, A.; Thawonmas, R.: The Chronicles of ChatGPT: Generating and Evaluating Visual Novel Narratives on Climate Change Through ChatGPT. In (Holloway-Attaway, L.; Murray, J. T., Hrsg.): Interactive Storytelling. Springer Nature Switzerland, Cham, S. 181–194, 2023, ISBN: 978-3-031-47658-7.

[JG23] Jauhiainen, J. S.; Guerra, A. G.: Generative AI and ChatGPT in School Children's Education: Evidence from a School Lesson. Sustainability 15/18, S. 14025, 2023, ISSN: 2071-1050.

[Li24] Li, R.: A "Dance of Storytelling": Dissonances between Substance and Style in Collaborative Storytelling with AI. Computers and Composition 71/, 2024, ISSN: 8755-4615.

[Lu17] Lugmayr, A.; Sutinen, E.; Suhonen, J.; Sedano, C. I.; Hlavacs, H.; Montero, C. S.: Serious Storytelling – a First Definition and Review. Multimedia Tools and Applications 76/14, S. 15707–15733, 2017, ISSN: 1573-7721.

[RB13] Riedl, M.; Bulitko, V.: Interactive Narrative: An Intelligent Systems Approach. In: AI Magazine. Bd. 34, AI Access Foundation, S. 67–77, 2013.

[Ro06] Robin, B.: The Educational Uses of Digital Storytelling. In (Crawford, C.; Carlsen, R.; McFerrin, K.; Price, J.; Weber, R.; Willis, D., Hrsg.): Proceedings of SITE 2006–Society for Information Technology & Teacher Education International Conference. S. 709–716, 2006.

[Ro08] Robin, B. R.: Digital Storytelling: A Powerful Technology Tool for the 21st Century Classroom. Theory Into Practice 47/3, S. 220–228, 2008, ISSN: 0040-5841, 1543-0421.

[Ro16] Robin, B. R.: The Power of Digital Storytelling to Support Teaching and Learning. Digital Education Review/30, S. 17–29, 2016.

[RS20] Robra-Bissantz, S.; Strahringer, S.: Wirtschaftsinformatik-Forschung für die Praxis. HMD Praxis der Wirtschaftsinformatik 57/2, S. 162–188, 2020, ISSN: 1436-3011, 2198-2775.

[Se11] Sein, M. K.; Henfridsson, O.; Purao, S.; Rossi, M.; Lindgren, R.: Action Design Research. MIS Quarterly 35/1, S. 37–56, 2011, ISSN: 0276-7783.

[Sh23] Shidiq, M.: The Use of Artificial Intelligence-based Chat-GPT and Its Challenges for the World Od Education; from the Viewpoint of the Development of Creative Writing Skills. Proceedings of International Conference on Education, Society and Humanity 1/1, S. 353–357, 2023, ISSN: 2986-5832.

[Sp09] Speer, N. K.; Reynolds, J. R.; Swallow, K. M.; Zacks, J. M.: Reading Stories Activates Neural Representations of Visual and Motor Experiences. Psychological Science 20/8, S. 989–999, 2009, ISSN: 0956-7976, 1467-9280.

[WFB14] Wille, P.; Finster, R.; Balke, W. T.: A Model for Interactive Csr Campaigns Using Storytelling. CEUR Workshop Proceedings 1198/, 2014, ISSN: 16130073.

[WSR17] Wiseman, S.; Shieber, S. M.; Rush, A. M.: Challenges in Data-to-Document Generation, 2017, arXiv: 1707.08052 [cs].

[Za17] Zampolli, S.: Verb+s Is Looking for Love: Towards a Meaningful Narrativization of Abstract Content. In (Nunes, N.; Oakley, I.; Nisi, V., Hrsg.): Interactive Storytelling. Bd. 10690, Springer International Publishing, Cham, S. 126–134, 2017, ISBN: 978-3-319-71026-6 978-3-319-71027-3.

Bildungstechnologien in Lehramt und Schule

A project-based AI and Data Education Concept for STEM Teacher Education in Teaching-Learning-Labs

Moritz Kreinsen [1], Maimon Thiems [1], Sandra Sprenger [2], and Sandra Schulz [1]



Abstract: This practical article reports on the experiences of the Teacher AID Lab, which was set up at the University of Hamburg in partnership with a student research center. It is a teaching-learning lab aimed at pre-service teachers who are trained to subsequently teach school students on selected topics. The Lab focuses on a project-based introduction of AI and data management concepts for pre-service teachers of various subjects. The course was conducted with 13 pre-service teachers and 18 pupils (8th grade). Due to the positive feedback from all students, it is now to be set up permanently and the material is to be made freely accessible.

Keywords: artificial intelligence, teacher education, computer science education, teaching-learning lab, stem education

1 Introduction

Artificial Intelligence (AI) effects our lives every day and for novices, it is not always transparent in which technologies AI is implemented and in which it is not, how the underlying concepts of machine learning or artificial neural networks work, how data plays a role in AI systems and how AI systems can be used to research and solve domain-specific problems [MRS23; TGS23]. As these matters and the relevance of AI education are growing, a number of different federal states in Germany have already implemented corresponding content in the curriculum [VP23]. In the United States, even comprehensive national standards for K-12 AI education are being developed [TGS23]. Further countries such as China, India or South Korea have also already integrated AI education into their school curricula [Ki21]. Since it is therefore seen as a task of K-12 education to convey the basics of artificial intelligence in an understandable way, teachers in particular must be trained for this content and its educational reconstruction.

As we see, there is already a wealth of materials for school education, especially in the field of computer science [DOK22; Hi23; LSR19; MB23]. We are particularly interested in the question: How can we train pre-service teachers of various subjects for this at university? What content and methods (from computer science education) are appropriate?

1 Universität Hamburg, Computer Science Education, Von-Melle-Park 8, 20146 Hamburg, Germany,
moritz.kreinsen@uni-hamburg.de, https://orcid.org/0000-0002-2156-5946;
maimon.thiems@uni-hamburg.de, https://orcid.org/0009-0007-1556-0560;
sandra.schulz@uni-hamburg.de, https://orcid.org/0000-0002-2254-6579
2 Universität Hamburg, Geography Education, Von-Melle-Park 8, 20146 Hamburg, Germany,
sandra.sprenger@uni-hamburg.de, https://orcid.org/0000-0002-9599-8103

doi:10.18420/delfi2024_14

2 Related Work

One approach to promote the development of teaching skills in combination with practical teaching experience as early as possible in teacher training, i.e. during university education, are so-called teaching-learning labs (TLLs). Here, pre-service teachers can try out new methods and teaching approaches in an open laboratory format, reflect on them in a theory-based way, work together with school students and thus build up professional skills for their later teaching career [DSS17; St20]. Standl et al. [St20] describe the development of a TLL for students of computer science teaching, use *plugged* and *unplugged* teaching approaches and develop teaching-learning activities for working in this TLL. The main aim is to successfully integrate practical teaching components into the computer science teaching degree program, but also to promote students' computer science education. For example, the studies by Demarle-Meusel, Sabitzer and Sylle [DSS17] focus on the topics of computational thinking and digital literacy in an interdisciplinary TLL, which is being researched in parallel and also offers opportunities for student (educational) research.

For teaching AI in schools different materials and approaches are recently created. For example, Lindner, Seegerer and Romeike [LSR19] developed teaching material for *unplugged* AI activities. The authors aim to bring important concepts of AI into schools with tangible learning materials (without computers). According to teaching large language models, Hielscher [Hi23] published material to dive deeper into their functioning using a specially developed, low-threshold web application for educational purposes called *SoekiaGPT*.

Addressing teacher education, Marx and Bergner [MB23] created a seminar to teach machine learning to computer science pre-service teachers. Here, an asynchronous MOOC is used to provide domain knowledge for the students. In attendance phases, pre-service teachers continue to gain domain-specific, educational and practical teaching qualifications, including the construction of short video clips. The seminar evaluation shows that the structure of the seminar and the examination form were positively emphasized by them.

Druga, Otero and Ko [DOK22] have investigated, summarized and classified such and similar teaching resources in the field of K-12 AI education in a systematic review. They propose to "move from singular activities and demos to more holistic designs that include support, guidance, and flexibility for how AI technology, concepts, and pedagogy play out in the classroom" [DOK22, p. 96].

When focusing on teaching AI to school students, it is important to take research results regarding students' conceptions on AI into account. Various student conceptions are uncovered, e.g. the conception of pre-programmed behaviour, anthropomorphism and understanding AI as a black-box [KS21; MLB22]. Taking these conception into account it gets visible that they need to be adjusted by scientifically correct concepts of AI [Ra23].

However, material and courses for teacher education are rare so far. The *Interdisciplinary AI and Data Lab in Teacher Education* (Teacher AID Lab) aims to develop, evaluate and publish teaching material for pre-service teachers studying various subjects.

3 Design of the Teaching-Learning-Lab

Course Organization
A teaching-learning lab is a methodological approach in which pre-service teachers were trained by their lecturers regarding teaching and material design. This course was led by lecturers from the chair of Computer Science Education and the chair of Geography Education. In accordance with the study regulations, this was a compulsory elective course for the pre-service teachers, which extends over one semester of 90 minutes each. This could be chosen by all students studying in a Master of Education program. The course is organized in three phases, with the second and third phases overlapping. At the beginning of the semester in the first phase, they receive an introduction to terminology and concepts, learn project and teaching methods in order to teach these to school students and study the basic theory of project-, problem- and research-oriented learning. In the second phase, the pre-service teachers are then expected to pass on their newly acquired knowledge to school students. To do this, they go to an extracurricular learning location, in this case the student research center (SFZ) Hamburg, where they can teach whole classes or subgroups. In total, the students work on AI and data projects with the school students over five sessions of 105 minutes each (due to the overlap in the timetables of the two target groups). The aim of the phase is to be able to present a product or result. Halfway through and at the end of the practical phase in the student research center, the pre-service teachers come together for a reflection session. This third phase consists primarily of reflecting on the experiences in the project lessons with the school student's groups, exchanging ideas about challenges and working out solutions and the next steps. At the end of the semester, the pre-service teachers present their completed projects, experiences, teaching concepts and approaches to the colloquium and once again engage in an exchange about interdisciplinary and project-oriented concepts for teaching AI and data concepts.

Pre-service teachers receive a total of 3 ECTS points for completing the seminar, the practical work in the students research center with the school students and the presentation in the colloquium. They receive a further 3 ECTS points in the final exam in the format of a written report, in which they (further) develop a teaching sequence, workshop, project week or course based on the experiences and content of the project carried out in terms of subject education and theory, phase it, name structural and material resources and classify and discuss it in the light of their own subject.

Course Contents
For the development and selection of the course content, it was important for us to take into account the findings of Long and Magerko [LM20] with regard to the skills to be acquired, Michaeli, Romeike and Seegerer [MRS23] as well as Druga, Otero and Ko [DOK22] according to the design of the learning resources. The main part of the course is a project, where the students can choose from four different contexts to implicitly learn about data and AI: 1.) data-driven determination of the quality of bicycle paths, 2.) plant detection, 3.) chemicals detection, and 4.) how *ChatGPT* works. For all the contexts learning and teaching material [Kr24] was developed in advance by the lecturers, used and modified in

the seminar by the pre-service teachers. In this practical article we will only present the course material regarding the *plant detection*.

Within the *plant detection* project the school students start with the investigation of relevant steps and resources for image classification with artificial neural networks. The next step is the collection of training data on plants. For this the school students can collect the data in the nature or in flower stores. Afterwards the students get an introduction into *Machine Learning for Kids*[3], they implement their training data and test the machine learning model. For applying the model in the field the students need to design a smartphone app using *MIT App Inventor*[4] and implementing simple functions for capturing/retrieving images and applying the pre-trained model using an extension. In the last step the students should test the app in field excursions.

Course Objectives
The main goals of the course are:

1. **Design and test course material**: to design differentiated and self-explanatory course material to teach basic concepts of AI and data management to pre-service teachers and school students. A particular challenge here is that the students come from different disciplines and have very different prior knowledge of computer science. Consequently, the material needs to offer a low-threshold introduction into AI.
2. **Enhance student motivation**: The construction of a course for school students that motivates to engage with AI is targeted. Based on the work that deals with students' conceptions of AI, it is important to expand or replace their existing conceptions with the scientifically correct conceptions.

4 Implementation of the Teaching-Learning-Lab

In winter semester 2023/2024 the Teacher AID Lab run with 13 pre-service teachers (who studied the subjects: biology, chemistry, physical education, geography, mathematics, physics, philosophy, english, social sciences, spanish, music, german) and 18 school students from a grammar school in 8th grade. A group of seven people (three pre-service teachers and four school students) decided to implement the *plant detection* app. The other prepared topics were chosen as well by the rest of the group.

Some results of the group are visualized in the following Fig. 1. The left side shows the program the students constructed during the course. The program allows the user to take photos. These photos are recognized with the "Recognize" button. The "Gallery" button is used to perform the afterpicking method. This means that all images are selectively selected in "Gallery" and recognized. After the first part, the program calls the extension *ML4KComponent1*. This extension uses the *Machine Learning for Kids* API to get the

3 https://machinelearningforkids.co.uk/
4 https://code.appinventor.mit.edu/

pre-trained model, which is trained using the existing photos (data), and the photos are classified in the same way. This process of categorization makes the prediction clearer, more familiar in the sense of probability.



Fig. 1: Left side: program created with *MIT App Inventor*; right side: interface of the App

On the right side of Fig.1 the designed app interface is presented. *MIT App Inventor* offers developers a viewer where they can create a prototype for the *plant detection* app project. *Plant variety*, *confidence* in percent and *status of the model* are displayed at the top using 3 horizontal alignments and 4 labels, each labeled with text. In the middle the pictures take or selected are presented. At the bottom the buttons *train*, *gallery*, *camera* and *recognize* are implemented, necessary for the functionality of the *Machine Learning for Kids* Component.

5 Report of Practical Experiences

For evaluating the Teacher AID Lab we gave the pre-service teachers an online questionnaire and a standardized course evaluation form (from the University of Hamburg). In the course evaluation form they were asked what they particularly liked about the course. Here, the pre-service teachers mentioned the AI unplugged learning activities and the topicality of the subject. Specifically liked was to practical work with the school students. This is very valuable for pre-service teachers, because the distance to the practical profession is a

frequent criticism of teacher training studies. Furthermore one participant mentioned: “In addition, working with students in a course was new to me and I enjoyed it, even though I was more or less a foreign to the subject lesson. Nevertheless, I felt well prepared professionally for the most part”. To better understand what activities were helpful for the pre-service teachers to learn about AI they were asked to rate the activities in the online questionaire. It was possible to rate between 1 (does not apply at all) to 5 (fully applies). The agreement for ten activities is $M = 3.7$ ($Md = 4$). This means that students tend to perceive the activities and methods used as helpful for their conceptual understanding of AI. We see this as a great approval of our course concept, but also a need for revision and evaluation in some areas.

Afterwards the form asked what they perceived as negative about the course. Some pre-service teachers wished to have more time, e.g. to learn more about AI. We interpret these answers as positive indication for reaching our goal to motivate for engaging with AI. Another pre-service teacher described the desire for a joint opening and closing event with the students. Furthermore, a participant reported a high workload for this course. Preparing material for school students, holding lessons with the students and a written report is a lot of effort for the pre-service teachers. Because of the practical phase with the school students it is difficult to reduce the workload for the course. For future courses we will try to change the examination form to a portfolio instead of a written report to address this critique. It would be also possible to change the ECTS points of the course to have a better balance between the workload and the ECTS points. Besides the questionnaires the educators observed that the participating school students were highly motivated by working with pre-service teachers as equals during the project work. However, we have also observed that both encounter major challenges, especially when programming or specific hardware is required. This is obviously mainly due to a lack of previous experience or subject knowledge, but could even be observed in block-based programming, where the participants only had to correctly integrate prefabricated code blocks into suitable control structures.

6 Conclusion

Based on the pre-service teachers’ answers in the questionnaires and the observations of the educators (first and second authors) this course was successful in respect to its objectives. We conclude that the low-threshold teaching methods (e.g. “unplugged” activities) for introducing concepts of AI and data management are very suitable for teacher education at universities. Because of the high rating of the AI activities it is likely the students were very engaged with the material and had a learning gain regarding AI knowledge. We reached to goal to create a course with material that is also suitable for pre-service teachers far from computer science. Even though they seem challenged when using specific hardware or programming. For future courses it is necessary to reduce the pre-service teachers’ workload by another examination type and prepared code snippets to reduce the cognitive load during the programming phase.

References

[DOK22] Druga, S.; Otero, N.; Ko, A. J.: The Landscape of Teaching Resources for AI Education. In: Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 1. ITiCSE '22, Association for Computing Machinery, New York, NY, USA, pp. 96–102, 2022, DOI: 10.1145/3502718.3524782.

[DSS17] Demarle-Meusel, H.; Sabitzer, B.; Sylle, J.: The Teaching-Learning-Lab - Digital Literacy and Computational Thinking for Everyone. In: Proceedings of the 9th International Conference on Computer Supported Education - Volume 2: CSEDU, INSTICC, SciTePress, pp. 166–170, 2017, DOI: 10.5220/0006367001660170.

[Hi23] Hielscher, M.: SoekiaGPT - ein didaktisches Sprachmodell. Informatische Bildung in Schulen (IBiS) 1 (1), 2023, DOI: 10.18420/ibis-01-01-04.

[Ki21] Kim, S.; Jang, Y.; Choi, S.; Kim, W.; Jung, H.; Kim, S.; Kim, H.: Analyzing Teacher Competency with TPACK for K-12 AI Education. KI - Künstliche Intelligenz 35 (2), pp. 139–151, 2021, DOI: 10.1007/s13218-021-00731-9.

[Kr24] Kreinsen, M.; Thiems, M.; Sprenger, S.; Schulz, S.: Interdisciplinary AI and Data Lab in Teacher Education: Teaching-Learning Material, 2024, DOI: 10.5281/zenodo.12063385.

[KS21] Kreinsen, M.; Schulz, S.: Students' conceptions of artificial intelligence. In: Proceedings of the 16th Workshop in Primary and Secondary Computing Education. Association for Computing Machinery, New York, NY, USA, 2021, DOI: 10.1145/3481312.3481328.

[LM20] Long, D.; Magerko, B.: What is AI Literacy? Competencies and Design Considerations. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20, Association for Computing Machinery, New York, NY, USA, pp. 1–16, 2020, DOI: 10.1145/3313831.3376727.

[LSR19] Lindner, A.; Seegerer, S.; Romeike, R.: Unplugged Activities in the Context of AI. In: Informatics in Schools. New Ideas in School Informatics. Springer International Publishing, Cham, pp. 123–135, 2019, DOI: 10.1007/978-3-030-33759-9_10.

[MB23] Marx, E.; Bergner, N.: Seminarkonzept zur fachlichen und fachdidaktischen Qualifizierung von Informatiklehramtsstudierenden zum Maschinellen Lernen. HDI – Hochschuldidaktik Informatik, p. 65, 2023.

[MLB22] Marx, E.; Leonhardt, T.; Bergner, N.: Brief Summary of Existing Research on Students' Conceptions of AI. In: Proceedings of the 17th Workshop in Primary and Secondary Computing Education. 2022, DOI: 10.1145/3556787.3556872.

[MRS23] Michaeli, T.; Romeike, R.; Seegerer, S.: What Students Can Learn About Artificial Intelligence – Recommendations for K-12 Computing Education. In: Towards a Collaborative Society Through Creative Learning. Springer Nature Switzerland, pp. 196–208, 2023, DOI: 10.1007/978-3-031-43393-1_19.

[Ra23] Rabe, A.; Kreinsen, M.; Grospietsch, F.; Schulz, S.: How to Deal with Students' Conceptions: Conceptual Change Texts on Artificial Intelligence. In: Proceedings of the 18th WiPSCE Conference on Primary and Secondary Computing Education Research. Association for Computing Machinery, New York, NY, USA, 2023, DOI: 10.1145/3605468.3609781.

[St20] Standl, B.; Bentz, A.; Ulbrich, M.; Vielsack, A.; Wagner, I.: Design- and Evaluation-Concept for Teaching and Learning Laboratories in Informatics Teacher Education. In: Informatics in Schools. Engaging Learners in Computational Thinking. Springer International Publishing, pp. 133–145, 2020, DOI: 10.1007/978-3-030-63212-0_11.

[TGS23] Touretzky, D.; Gardner-McCune, C.; Seehorn, D.: Machine Learning and the Five Big Ideas in AI. International Journal of Artificial Intelligence in Education 33 (2), pp. 233–266, 2023, DOI: 10.1007/s40593-022-00314-1.

[VP23] Vo, G. M.; Pancratz, N.: AI Education in German K-10 Computer Science Curricula. In: Proceedings of the 18th WiPSCE Conference on Primary and Secondary Computing Education Research. Association for Computing Machinery, New York, NY, USA, 2023, DOI: 10.1145/3605468.3605471.

Generative KI zur Lernenbegleitung in den Bildungswissenschaften: Implementierung eines LLM-basierten Chatbots im Lehramtsstudium

Hassan Soliman [1], Milos Kravcik [1], Alexander Tobias Neumann [2], Yue Yin [2], Norbert Pengel [3], Maike Haag [3] und Heinz-Werner Wollersheim [3]

Abstract:

Lehramtsstudierende nehmen an einer Vielzahl von Lernaktivitäten in verschiedenen Fachbereichen teil, u.a. in den Bildungswissenschaften. Sie benötigen angemessene Unterstützung, u.a. in Form zeitnahen Feedbacks. Da ihre Anzahl steigt, ist es ein wichtiges Ziel, ihnen eine skalierbare Unterstützung zu bieten, was auch eine Herausforderung für Forschende und Entwickelnde darstellt. Als eine mögliche Lösung stellen wir einen Chatbot vor, der auf einem Large Language Model (LLM) basiert und durch bestehende Lern- und Informationsmaterialien sowie hochschuldidaktische Prozessbeschreibungen ergänzt wird. In diesem Praxisbeitrag werden Erfahrungen aus der didaktischen Konzeption und technischen Umsetzung sowie erste Ergebnisse zur Qualität von LLM-basierten Chatbot-Antworten auf inhaltliche und organisatorische Fragen eines technologiegestützten bildungswissenschaftlichen Moduls im Lehramtsstudium vorgestellt. Der Bot ist in der Lage, Kursmaterial abzurufen, zu analysieren und umfassende Antworten auf spezifische Fragen zu geben. Vorläufige Ergebnisse zeigen, dass es möglich ist, zwischen verschiedenen Kontexten, in denen die Studierenden arbeiten, zu unterscheiden und ihnen eine schnelle Antwort zu geben, die das relevante Material berücksichtigt.

Keywords: Chatbot, Large Language Model, Skalierbares Mentoring

1 Einleitung

Skalierbares Mentoring ist eine Herausforderung im Bildungswesen. Die Entwicklung entsprechender Technologien, insbesondere der Künstlichen Intelligenz (KI), eröffnet neue Perspektiven in der digitalen Hochschulbildung. Unser Ziel ist es, einen Chatbot zu entwickeln, der Studierende in bildungswissenschaftlichen Modulen unterstützt. Vorhandene Lern- und Informationsmaterialien sowie hochschuldidaktische Prozessbeschreibungen können das Potenzial in Bezug auf Textverständnis und -generierung erhöhen. Hier wurde generative KI eingesetzt, um einen bereits implementierten Chatbot zu verbessern [Ne21].

1 DFKI, Berlin, Germany, hassan.soliman@dfki.de, https://orcid.org/0009-0003-4574-9074;
milos.kravcik@dfki.de, https://orcid.org/0000-0003-1224-1250
2 RWTH Aachen University, Aachen, Germany,
neumann@dbis.rwth-aachen.de, https://orcid.org/0000-0002-9210-5226;
yue.yin@rwth-aachen.de, https://orcid.org/0009-0006-8369-8396
3 Leipzig University, Leipzig, Germany,
norbert.pengel@uni-leipzig.de, https://orcid.org/0000-0002-3263-6877;
maike.haag@uni-leipzig.de, https://orcid.org/0009-0001-6161-2022;
wollersheim@uni-leipzig.de, https://orcid.org/0000-0002-4690-5839

doi:10.18420/delfi2024_15

Zunächst wird auf verwandte Arbeiten verwiesen, danach werden die didaktischen Inhalte erläutert. Im Hauptteil wird der technologische Hintergrund vorgestellt, einschließlich des Chatbot-Designs und der Implementierung des LLM-Prototyps. Abschließend werden erste Ergebnisse präsentiert und die nächsten Schritte skizziert.

2 Verwandte Arbeiten

KI kann eine zentrale Rolle im Bildungswesen spielen, da sie in der Lage ist, große Mengen an Daten im Zusammenhang mit dem Lernprozess zu analysieren [Ki23]. Der Aufbau sozialer generativer KI für Bildung erfordert die Entwicklung leistungsfähiger KI-Systeme, die mit Menschen kommunizieren können. Wir müssen überlegen, wie man eine sozial-generative KI für den Bildungsbereich entwickelt und einschränkt [Sh23].

Forschende haben mit generativen KI-Modellen experimentiert, um herauszufinden, welche Aufgaben sie unterstützen können. Beispiele hierfür sind automatisch generierte Online-Kurse zum Sprachenlernen [RP23], die Bewertung der Korrektheit von Schülerantworten und die Erzeugung von Feedback in einem digitalen Lernspiel [Ng23] und die Bewertung von Aufsätzen zur Stärkung der Schreibfähigkeiten von Studierenden [Se23]. Solche Ergebnisse werden zeigen, welche Aufgaben an KI delegiert werden können und welche auch in Zukunft echte menschliche Fähigkeiten erfordern werden.

3 Lernunterstützung in den Bildungswissenschaften

Die Szenarien in unserem Testbed der Universität Leipzig fokussieren sich didaktisch auf die Unterstützung selbstgesteuerten Lernens sowie auf Mentoring und Beratung im Rahmen eines bildungswissenschaftlichen Moduls des Lehramtsstudiums [Pe23]. Zur Unterstützung von Lernprozessen gilt Mentoring als effektives Mittel [ED15]. Ein zentrales Merkmal ist die dyadische Beziehung zwischen Mentor/in und Mentee [Zi09], welche in Lehrveranstaltungen mit vielen Studierenden aufgrund der erforderlichen Ressourcen nicht realisierbar ist. Daher stellt sich die Frage, wie mentorielle Prozesse skaliert werden können, um die Lernunterstützung nach dem Vorbild des Mentorings umzusetzen. Zur Beantwortung dieser Fragestellung wurden lernbegleitende Szenarien in der Mentoring Workbench (MWB), einer digitalen Arbeitsumgebung, die verschiedene Tools bündelt [ZBJ23] für die Unterstützung des Selbststudiums entwickelt. Eine zentrale Komponente der MWB ist der Chatbot, der als virtueller Ansprechpartner rund um die Uhr verfügbar ist und zwei Funktionen erfüllen soll: Experte und Lernbegleiter [DRS18]. Als Lernbegleiter unterstützt er den individuellen Lernprozess der Studierenden. Er gibt Feedback zu den eingereichten Schreibaufgaben und ermutigt die Studierenden, ihren eigenen Lernprozess zu planen, zu überwachen und zu reflektieren. Als Experte beantwortet der Chatbot Fragen zum Inhalt und zur Organisation Moduls im Rahmen des Lehramtsstudiums an. In diesem Beitrag wird letzteres thematisiert.

4 Design und Implementierung

Bis Ende 2022 konzentrierte sich unsere Chatbot-Entwicklung auf Open-Source-Lösungen wie RASA[4], ein Framework für maschinelles Lernen, das explizit für die Entwicklung konversationeller KI konzipiert wurde. Teile des Frameworks wurden genutzt, um das Social Bot Framework (SBF) zu integrieren. Dies wurde durch die Fähigkeit zur webbasierten, kollaborativen Entwicklung motiviert. Die Plattform bietet eine intuitive Möglichkeit zur Definition von Benutzerintentionen und Bot-Reaktionen durch Drag-and-Drop-Funktionalitäten [NLK19]. RASA NLU (Natural Language Understanding) wurde integriert, um eine solide Grundlage für die Konversation zu schaffen und gleichzeitig die Entwicklungserfahrung zu verbessern.

Das Team aus dem Bereich der Erziehungswissenschaften hat Beispiele für jede identifizierte Absicht und Entität erstellt. Für die NLU-Komponente wurden Modelle wie der Dual Intent and Entity Transformer (DIET) verwendet, eine hochmoderne Architektur zur gleichzeitigen Klassifizierung von Absichten und Entitäten [AAR22]. Dadurch kann der Chatbot relevante Informationen aus den Benutzerinteraktionen analysieren und herausfiltern. Auch der ursprüngliche Chatbot [Ne21] war bereits in die Webschnittstelle MWB integriert. Unser Ansatz zur vorlagengesteuerten natürlichen Sprachgenerierung (NLG) wies trotz dieser Fortschritte einige Einschränkungen und Herausforderungen auf. Insbesondere die begrenzte Flexibilität bei Abfragen, die über die vordefinierten Regeln hinausgehen, sowie das komplexe Konversationsmanagement aufgrund fehlenden Textbewusstseins und die Benutzerfrustration aufgrund einer weniger ansprechenden Benutzererfahrung stellten Herausforderungen dar.

4.1 LLM-basierter Prototyp

Der BiWi–AI-Tutor ist ein Chatbot-Prototyp auf Basis von LLM. Er wurde entwickelt, um eine skalierbare Lernunterstützung zu ermöglichen, indem er Wissen aus Vorlesungsfolien, Seminartexten und Organisationsmaterialien abruft. Der Chatbot nutzt das Modell `gpt-3.5-turbo` von OpenAI[5] und die LangChain[6]-Bibliothek, um nahtlos innerhalb der MWB eine reaktionsschnelle und kontextbewusste dialogische Interaktion zu bieten. Wir haben den Function Calling Agent in der LangChain-Bibliothek verwendet, eine leistungsstarke Funktion, die es großen Sprachmodellen (LLMs) ermöglicht, spezifische Funktionen oder Werkzeuge als Teil ihres Verarbeitungsworkflows aufzurufen. Diese Agenten können die Fähigkeiten von LLMs erweitern, indem sie Aufgaben ausführen, die externe Funktionen oder den Zugriff auf bestimmte Werkzeuge erfordern. Er wählt die am besten geeigneten Materialien für eine gegebene Anfrage aus. Durch die Kombination der generativen Leistung eines LLM mit Retrieval-Systemen findet das System effizient relevante Abschnitte der

4 `https://rasa.com/`

5 `https://platform.openai.com/docs/models/gpt-3-5-turbo`

6 `https://www.langchain.com/`

Lerninhalte, aus denen es Antworten generiert. Dies wird als Retrieval Augmented Generation (RAG) bezeichnet. Die Effektivität wird durch die Einbeziehung von direktem Feedback der Nutzenden, das durch ein Bewertungssystem (Daumen hoch oder Daumen runter) und Kommentare innerhalb des Chat-Interfaces gesammelt wird, erhöht. Die Interaktionsarchitektur des Chatbots ist sorgfältig aufgebaut, um ein benutzerfreundliches Erlebnis zu gewährleisten. Sie folgt einer logischen Abfolge, beginnend mit der Nachricht des Nutzers, dem Verstehen und Verarbeiten der Anfrage und der Auswahl des richtigen Werkzeugs, je nach Art der Frage. Die verwendeten Tools sind textlich beschrieben und dienen als Indizes für die relevantesten Inhalte, sodass der Chatbot relevante und präzise Informationen abrufen kann. Ein zentraler Aspekt der Funktionalität des Chatbots ist die Indizierung von Inhalten durch den innovativen Einsatz von Inhalteinbettungen. Textabschnitte werden mithilfe der Einbettungsmodelle von OpenAI zu Einbettungen verarbeitet und in einer Vektordatenbank gespeichert. Nach Erhalt einer Anfrage generiert der Chatbot eine semantisch reichhaltige Vektordarstellung der Frage. Dieser wird verwendet, um kontextuell relevante Texte aus den Inhaltsindizes durch semantische Ähnlichkeits- und Schlüsselwortabgleichsmaßnahmen abzurufen. Ein hybrider Ensemble-Retriever-Ansatz wird verwendet, indem ein semantischer Vektor-Retriever über eine Vektordatenbank und ein Schlüsselwort-Retriever über den BM25-Algorithmus [RZ09] eingesetzt wird. Wir haben die fünf besten Abschnitte vom Vektor-Retriever mit den fünf besten Abschnitten vom Schlüsselwort-Retriever abgerufen, um insgesamt zehn Textfragmente für die Abfrage zu erhalten, die als Kontext und Wissen für das LLM dienen. Nach dem Abruf des relevanten Kontexts konstruiert der Chatbot eine fundierte Antwort auf die Anfrage des Nutzenden.

4.2 Ergebnisse

Für die Evaluation des LLM-basierten Chatbots wurde ein Datensatz verwendet, der Fragen aus der Lehrveranstaltung, entsprechende richtige Antworten und die vom Chatbot generierten Antworten enthält. Der Datensatz wurde kuratiert, um die Vielfalt der Materialien widerzuspiegeln. Diese umfassen Vorlesungsfolien, Seminartexte und organisatorische Informationen, die in drei verschiedene Typen kategorisiert und verarbeitet wurden. Die Materialien umfassen Vorlesungsfolien, Seminartexte und organisatorische Informationen im PDF-Format. Sie wurden mithilfe des Moduls LlamaParse aus der Bibliothek LlamaIndex[7] geparst, gechunked und indiziert. Dabei wurde für jede Kategorie ein eigener Index erstellt. Zusätzlich wurde ein weiterer Index erstellt, der alle Kategorien enthält. Während des Indexierungsprozesses wurden die PDF-Dateien in Chunks von 1024 Token unterteilt, wobei eine Überlappung von 20 Token verwendet wurde. Anschließend wurden diese Chunks mithilfe des 'text-embedding-3-small'-Modells von OpenAI[8] in Einbettungen umgewandelt und in einer Vektordatenbank gespeichert, um sie abrufen zu können. Der Evaluierungsdatensatz bestand aus Fragen, die von den Lehrenden des bildungswissenschaftlichen Lehramtsmoduls auf Basis vorhandener Materialien entwickelt wurden, um

7 https://www.llamaindex.ai/

8 https://platform.openai.com/docs/guides/embeddings/embedding-models

die Fähigkeit des Chatbots, Antworten zu generieren, zu bewerten. Der Datensatz wurde mit dem QA-Evaluierungsprompt aus der LangChain-Bibliothek ausgewertet. Dabei wurde das gpt-4-Modell von OpenAI[9] verwendet, um die Korrektheit der Antworten des Chatbots zu beurteilen. Der Bewertungsprozess folgte einer strukturierten Eingabeaufforderung, die die sachliche Richtigkeit betonte und stilistische sowie formale Unterschiede zwischen der Ausgabe des Chatbots und der tatsächlichen Antwort außer Acht ließ. Der Bewertungsmechanismus orientierte sich an der folgenden Aufforderung an gpt-4:

> "You are a teacher grading a quiz. Grade the student answers based ONLY on their factual accuracy. Ignore differences in punctuation and phrasing between the student answer and true answer. It is OK if the student answer contains more information than the true answer, as long as it does not contain any conflicting statements. Begin!"

Die Leistung wurde für jede Materialkategorie separat bewertet. Es wurden zwei Ergebnisgruppen ermittelt: eine unter der Annahme, dass das richtige Lernmaterial zur Verfügung gestellt wurde, und ein weiterer, der die Fähigkeit des Chatbots bewertet, das richtige Lernmaterial auszuwählen, wenn alle Lernmaterialien für alle Kategorien als ein Index oder Tool gegeben sind. Die Ergebnisse zeigen, dass der Chatbot in der Lage ist, kontextuell relevante und sachlich korrekte Antworten zu generieren, wenn der passende Lernmaterialindex bereitgestellt wird. Sie verdeutlichen (Tab. 2), dass der Chatbot bei der Bereitstellung des entsprechenden Tools für die jeweilige Kategorie in 85-90% der Fälle korrekt antwortet.

Jedoch sinkt die Anzahl der richtigen Antworten, wenn alle Lernmaterialien in einem einzigen Index zusammengefasst werden. Die Wahl des richtigen Lernmaterials, wenn alle Materialien in einem Index gegeben sind stellt eine kritische Herausforderung für zukünftige Erweiterungen dar, um das kontextuelle Verständnis und die Entscheidungsfähigkeit des Chatbots zu verbessern. Diese Bewertung betont das Potenzial des Chatbots als akademisches Hilfsmittel und zeigt gleichzeitig Bereiche auf, in denen weitere Verbesserungen erforderlich sind, insbesondere bei der Analyse und Auswahl des richtigen und relevanten Lernmaterials, wenn alle Materialien in einem Index gegeben sind.

Tab. 1: Bewertungsergebnisse mit dem Anteil der richtigen Antworten bei ausschließlicher Bereitstellung des passenden Materials und bei Bereitstellung aller Materialien

Kategorie	**Anzahl Fragen**	**Passendes Material**	**Alle Materialien**
Seminar	20	17 (85%)	15 (75%)
Vorlesung	20	17 (85%)	17 (85%)
Organisatorisches	20	18 (90%)	17 (85%)

Tab. 2 zeigt jeweils ein Beispiel aus Vorlesung und Seminar. Beim Beispiel der Vorlesung wurde die Antwort des Chatbots im Vergleich mit der vorgegebenen Antwort als richtig bewertet, beim Beispiel des Seminars als falsch.

9 https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo

Tab. 2: Beispiele für Chatbot-Antworten

Beispiel	Details
Richtig	**Frage:** August Hermann Francke forderte, Kinder mit aller Freundlichkeit, Milde und unermüdlicher Geduld zur Liebe zur Wahrheit, Gehorsam und Fleiß anzuhalten. Wie stand Francke zu körperlichen Strafen? **Richtige Antwort:** Verstocktheit und Trotz dürfen mit körperlichen Strafen gebrochen werden. **Chatbot-Antwort:** August Hermann Francke war der Ansicht, dass Kinder mit aller Freundlichkeit, Milde und unermüdlicher Geduld zur Liebe zur Wahrheit, Gehorsam und Fleiß angehalten werden sollten. Er glaubte, dass nur auf diese Weise das kindliche Herz dem Erzieher in Liebe öffnet. Francke war auch der Meinung, dass Verstocktheit und Trotz mit körperlichen Strafen gebrochen werden dürfen.
Falsch	**Frage:** Wolfgang Klafki versucht den Widerstreit formaler und materialer Bildungstheorien zu überwinden. Welchen Namen gibt er diesem Bildungskonzept? **Richtige Antwort:** Kategoriale Bildung **Chatbot-Antwort:** Wolfgang Klafki gibt diesem Bildungskonzept den Namen "Allgemeinbildung".

5 Zusammenfassung

Zur Unterstützung der Studierenden haben wir im Rahmen der digitalen Arbeitsumgebung Mentoring Workbench (MWB) Chatbots entwickelt und erprobt. Wir verwendeten das Social Bot Framework und RASA, um regelbasierte Antworten zu generieren. Antworten auf häufig gestellte Fragen oder Kursinhalte waren jedoch fest vorgegeben. Durch den Einsatz von LLM bei der Chatbot-Entwicklung ist es nun möglich, von einer flexibleren und adaptiveren Antwortgenerierung zu profitieren, die eine tiefere Kontextualisierung ermöglicht. Dies führt zu einer präziseren und personalisierteren Nutzungserfahrung und ermöglicht es den Studierenden, komplexere Fragen zu stellen. Die nächsten Schritte umfassen eine umfassende Evaluierung mit den Nutzenden, auch unter Einbeziehung menschlicher Begutachtung des Chatbot-Outputs und eine entsprechende kontinuierliche Weiterentwicklung. Konzeptionell wird es in den nächsten Schritten u.a. darum gehen, LLM-gestützte Interaktionen des Chatbots zu ermöglichen, die sich an den Prinzipien der Online-Beratung orientieren [MP24].

Danksagung. Die Forschung, die zu diesen Ergebnissen geführt hat, wurde vom Bundesministerium für Bildung und Forschung (BMBF) im Rahmen des Projekts "Personalisierte Kompetenzentwicklung und hybrides KI-Mentoring" (tech4compKI) (Förderkennzeichen 16DHB2206, 16DHB2208, 16DHB2213) gefördert.

Literaturverzeichnis

[AAR22] Arevalillo-Herraez, M.; Arnau-Gonzalez, P.; Ramzan, N.: On Adapting the DIET Architecture and the Rasa Conversational Toolkit for the Sentiment Analysis Task. IEEE Access 10, S. 107477–107487, 2022, ISSN: 2169-3536.

[DRS18] Dyrna, J.; Riedel, J.; Schulze-Achatz, S.: Wann ist Lernen mit digitalen Medien (wirklich) selbstgesteuert? Ansätze zur Ermöglichung und Förderung von Selbststeuerung in technologieunterstützten Lernprozessen. 2018.

[ED15] Eby, L. T.; Dolan, E. L.: Mentoring in postsecondary education and organizational settings. 2015.

[Ki23] Kizilcec, R. F.: To advance AI use in education, focus on understanding educators. International Journal of Artificial Intelligence in Education, S. 1–8, 2023.

[MP24] Martin, A.; Pengel, N.: Beratung via Chatbot? Möglichkeiten und Anforderungen beim Einsatz Generativer KI in einem bildungswissenschaftlichen Lehramts-Modul. e-beratungsjournal.net, Zeitschrift für Onlineberatung und computervermittelte Kommunikation, 2024.

[Ne21] Neumann, A. T.; Arndt, T.; Köbis, L.; Meissner, R.; Martin, A.; de Lange, P.; Pengel, N.; Klamma, R.; Wollersheim, H.-W.: Chatbots as a Tool to Scale Mentoring Processes: Individually Supporting Self-Study in Higher Education. Frontiers in Artificial Intelligence 4, S. 64–71, 2021.

[Ng23] Nguyen, H. A.; Stec, H.; Hou, X.; Di, S.; McLaren, B. M.: Evaluating chatgpt's decimal skills and feedback generation in a digital learning game. In: European Conference on Technology Enhanced Learning. Springer, S. 278–293, 2023.

[NLK19] Neumann, A. T.; de Lange, P.; Klamma, R.: Collaborative Creation and Training of Social Bots in Learning Communities. In: 2019 IEEE 5th International Conference on Collaboration and Internet Computing (CIC). S. 11–19, 2019.

[Pe23] Pengel, N.; Soliman, H.; Neumann, A.; Kravcik, M.; Yin, Y.; Bez, M.; Stechert, L.; Haag, M.; Martin, A.; Köbis, L.; J., Z.; Jalilov, O.; Meissner, R.; Pögelt, A.; Köhler, T.; Wollersheim, H.-W.: Generative KI in der Hochschulbildung: Konzeption und Implementierung eines LLM-gestützten Chatbots. In: Tagungsband der Jahrestagung der GMW und CampusSource. 2023.

[RP23] Rüdian, S.; Pinkwart, N.: Auto-generated language learning online courses using generative AI models like ChatGPT. 2023.

[RZ09] Robertson, S.; Zaragoza, H.: The probabilistic relevance framework: BM25 and beyond. Foundations and Trends® in Information Retrieval 3 (4), S. 333–389, 2009.

[Se23] Seßler, K.; Xiang, T.; Bogenrieder, L.; Kasneci, E.: PEER: Empowering Writing with Large Language Models. In: European Conference on Technology Enhanced Learning. Springer, S. 755–761, 2023.

[Sh23] Sharples, M.: Towards social generative AI for education: theory, practices and ethics. Learning: Research and Practice 9 (2), S. 159–167, 2023.

[ZBJ23] Zawidzki, J.; Bez, M.; Jalilov, O.: Needs and requirements of teachers, learners and administrative staff for AI based mentoring tools in higher education. In: INTED2023 Proceedings. IATED, S. 4674–4679, 2023.

[Zi09] Ziegler, A.: Mentoring: konzeptuelle grundlagen und wirksamkeitsanalyse. Mentoring: Theoretische hintergründe, empirische befunde und praktische anwendungen, S. 7–29, 2009.

Enhanced Code Comprehension: Individualized Learning of Code Tracing with the Feedback Buddy

Nadine Nicole Koch [1], Ann-Kathrin Kapfenstein[2], Niklas Meißner [3], and Maria Wirzberger [4]

Abstract: The increasing importance of programming skills for various professions highlights the necessity of laying solid foundations for these skills in school. One critical step when learning programming is code tracing, i.e., the ability to analyze program code to predict the data changes when the code is executed. Considering the diverse levels of prior knowledge in computer science classes, it is essential to implement tailored teaching approaches, which can enhance the respective learning outcomes. To create personalized learning paths, we developed the tutoring system Feedback Buddy that teaches *if-else* branches, *for* loops, and combinations of them using tracing tables. Thereby, the Feedback Buddy adapts its feedback and task difficulty based on learners' affective, cognitive, and metacognitive states. Extending former systems, our approach focuses on school education and uses growth mindset feedback. Evaluation with 10th grade students in computer science classes yielded positive results, affirming the usability and usefullness of the Feedback Buddy in secondary school education.

Keywords: Tutoring Systems, Digital Tutors, Code Tracing, Computer Science Education, Feedback, Growth Mindset

1 Theoretical Background

The importance of computer programming extends across various fields such as science, technology, engineering, and mathematics (STEM) [Fe19]. Despite its importance, research indicates that even first-semester programming students often lack essential skills in comprehending program code [Li04; Mc01]. Given the necessity of programming skills across diverse fields and the lack of abilities in first-semester courses, it is imperative to increase the quality of programming education in schools.

A crucial skill for understanding code is *code tracing*, the systematic simulation of a program (ideally in written form) to predict the values of variables and program output [JM20;

1 University of Stuttgart, Department of Software Quality and Architecture & Department of Teaching and Learning with Intelligent Systems, Geschwister-Scholl-Straße 24D, 70174 Stuttgart, Germany, nadine.koch@iste.uni-stuttgart.de, https://orcid.org/0000-0002-3707-1668
2 University of Stuttgart, Department of Teaching and Learning with Intelligent Systems, Geschwister-Scholl-Straße 24D, 70174 Stuttgart, Germany,
3 University of Stuttgart, Department of Software Quality and Architecture, Universitätsstraße 38 , 70569 Stuttgart, Germany, niklas.meissner@iste.uni-stuttgart.de, https://orcid.org/0000-0001-9929-1220
4 University of Stuttgart, Department of Teaching and Learning with Intelligent Systems, Geschwister-Scholl-Straße 24D, 70174 Stuttgart, Germany, maria.wirzberger@ife.uni-stuttgart.de, https://orcid.org/0000-0003-3072-2875

doi:10.18420/delfi2024_16

NXK17]. Code tracing was shown to be related to other programming-related skills, such as explaining or writing program codes [Ku15; Lo08]. Hence, teaching code tracing can help to improve programming skills in general. This is also reflected in formal educational curricula in Germany, such as those in Baden-Wuerttemberg, where seventh-grade students are expected to dissect code segments step-by-step and interpret their effects [MJ17]. The curriculum also incorporates supplementary abilities associated with code tracing, such as analyzing computer system behavior and identifying errors in program code [MJ17; MJ18].

Teaching code tracing in schools faces a significant challenge due to the considerable diversity in prior knowledge and competencies among students in computer science classes [We07], alongside demographic, sociocultural, and affective (e.g., motivational) interindividual differences [SZ12]. To facilitate optimal learning outcomes for all students, it is necessary to adapt learning content to respective individual needs [SZ12], for example, by leveraging the advantages of tutoring systems. The demand for tutoring systems in higher education has, for example, been emphasized in a recent study with students and lecturers reflecting on the challenges of established learning management systems in the field of software engineering [Me24].

Characteristically, tutoring systems model the cognitive state of the learner, including their knowledge and abilities. Hence, they can support teachers by providing individualized instructions and learning paths [Ma14]. Although effect sizes vary, learning with tutoring systems has proven to positively affect learning outcomes [KF16; Wa23], i.e., in computer science education [MSB23; Ne14]. Tutoring systems even reveal comparable outcomes to human tutoring or small-group teaching [Ma14; Va11].

At the core of most tutoring systems, feedback plays a crucial role in the process of skill acquisition [WZH20]. This powerful mechanism provides information on the respective learner's performance [WZH20], enabling a more in-depth reflection on the already achieved skill level and potential pathways of improvement. Feedback has been shown to positively influence motivation [DKR99; HT07; WZH20], experience of competence, and interest [DKR99; Ha14; Ra19]. With regards to performance, while overall effects point in a positive direction, we can observe substantial variations in effect size and even some effects pointing in a negative direction [AK17; CH09; WZH20]. This can be explained by several moderators such as the timing [LF18; NH04; Sh08], valence [LF18; VK11] and complexity [NH04; Sh08] of the feedback. Additionally, influencing factors related to the individual learner are important, such as the experienced relevance of the feedback [Ki16; Sh08] and prior knowledge [NH04; Sh08]. Hence, feedback mechanisms have to be carefully chosen when implemented in tutoring systems.

The current landscape in learning-related research and applications has already brought forward a large number of tutoring systems that give automated feedback [KJH18]. Even more, Mason and Bruning [MB01] provide a decision tree to support an adequate choice of feedback approaches in computer-based instruction by respecting student achievement, task

difficulty, and prior knowledge. While this model hints at the way feedback should be given, it does not provide concrete formulations of feedback.

Inspecting the field of computer science from a broader lens, we observe that the dominant view regards success to be the result of innate abilities [Le15] primarily. Such a view on abilities or intelligence is described by Dweck and Leggett [DL88] as *fixed*. People with such a fixed mindset believe intelligence, in general or in a certain domain, is uncontrollable and unchangeable [DL88]. They avoid challenges and search for effortless successes, and when they fail, they feel helpless [DL88; Dw00]. By contrast, people with a *growth mindset* believe that effort can change their abilities and intelligence [DL88; Dw00]. Therefore, they react with perseverance and resilience to failure [Dw00]. Fixed and growth mindset beliefs are not clearly separated but form a continuum in both directions [Ad15].

In particular, growth mindset beliefs have a positive effect on learning as they lead to decreased cognitive load [RBD19; Xu21] as well as increased motivation [BTD07; Rh18], learning-related joy [AFG02], mental well-being [Bu23], self-efficacy [KN13], and interest in a career in computer science [Bu20]. Even though the reported effect sizes related to performance improvements are inconsistent and comparably small in size [Bu20; Bu23; MB23; Si18], they provide encouraging evidence. Hence, fostering a growth mindset in students holds educational value. Consequently, in subjects such as computer science, a change in the current way of thinking is desirable.

One way to foster a growth mindset in learners relates to feedback formulations and instructor behavior indicating growth mindset [MD98; RGD12; Tr17]. This further influences learners' affective response, sense of belonging to a class, interest, engagement, and performance [Mu20]. By contrast, according to Rattan et al., [RGD12], formulations that indicate a fixed mindset like "It's okay, you're just not a math person" can reduce motivation and self-efficacy expectation. Kroeper et al. [KFM22] assume that this effect occurs as learners derive the instructor's mindset from such formulations and behavior. To incorporate the positive effects of promoting a growth mindset, we developed a tutoring system that provides growth mindset feedback.

Most code-tracing tutoring systems focus on higher education [DK03; Hö19; Hu23; JM20; KO19; QF20; Ri21]. We could not observe systems specially tailored to intermediate-level grammar schools. Moreover, only some systems adapt their content or feedback based on the learner's behavior, and those that do typically focus only on cognitive aspects [Hö19; Hu23; Ku20]. In addition, most existing systems primarily indicate whether the answer is correct or not, sometimes also accompanied by providing the correct answer [Hö19; Hu23; JM20; KO19; Ku20; Ri21]. Only a few systems offer elaborative feedback [DK03; SK00].

Extending former systems, we developed the *Feedback Buddy*. This tutoring system adapts the content and feedback to the learner's cognitive, metacognitive, and affective states while simultaneously providing growth mindset feedback. As growth mindset feedback was shown to have supportive effects on learners' engagement and performance [VB23], we expect

these effects also to occur when using the Feedback Buddy. As a first step, we investigate if and to what extent the developed system is usable and useful to teach code tracing in an intermediate-level grammar school context.

2 The Feedback Buddy

Our Feedback Buddy is a program implemented in Processing [RF06] and exported to an executable binary for Windows. Aligned with established characteristics of a tutoring system [So16], the Feedback Buddy consists of four components and associated functionalities: (1) The skills, strategies, and knowledge of code tracing are represented in the *domain model*. (2) The *learner model* stores the cognitive, affective, and motivational state of the learner during learning. (3) Both the domain and learner models are used by the *pedagogical model* to predict a tutoring strategy, e.g., to present tasks with an appropriate difficulty. (4) For the communication between the human and the system, the *user interface* is used. It includes a pedagogical agent with the name *Bud* as an abbreviation for "Buddy".

While working with the Feedback Buddy, the interface conveys different states, as shown in Fig. 1. In the *initial* state, the pedagogical agent Bud welcomes the learner, introduces themselves and provides an overview of the system functionality. Upon selecting "Let's go", the learner starts a working session with a step-by-step example. Showing the example involves switching to the *activity* state. In this state, a program code is presented on the top left part of the screen, Bud gives an instruction on the bottom left part of the screen, and a trace table is shown on the right side of the screen. The chosen instruction depends on the state of the learner; for instance, if the learner displayed the solution in the last task quickly, the chosen statement indicates that quick success is not the aim. Code examples are presented row-by-row. Additionally, the activity state involves the presentation of exercises. They involve an empty task table, which can be filled by the learner, as shown in the left part of Fig. 2. By choosing the "Feedback" option, a learner can receive feedback for their entries. Progressing to the *feedback* state, the program code and trace table persist, with Bud providing feedback such as „That is not yet correct. It takes time to understand these things.“. If the task table is correct, the learner can quit the working session or proceed with the next exercise. If incorrect, the colors of the wrong entries in the table turn blue, giving the learner the opportunity to either retry the task or reveal the solution. In the *solution* state, the software mirrors the interface design shown upon correctly solving the trace table but with different feedback. For correct solutions, for example, „Great, you solved the task correctly and learned something new in the process!“ would be a potential feedback statement. Finally, the *summary* state, activated upon concluding working with the Feedback Buddy, presents a summary of the working session, including the number of examples shown, exercises initiated, and exercises successfully solved at each level. All feedback formulations are presented in supplementary materials S1, section 3.

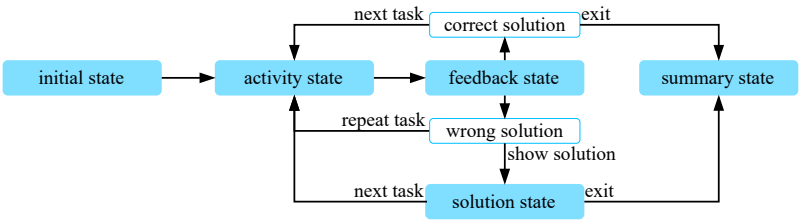


Fig. 1: The states the Feedback Buddy interface can take over the course of a working session.

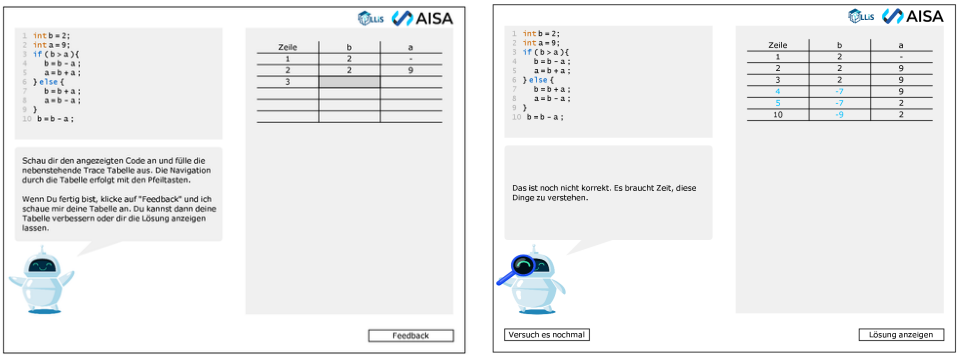


Fig. 2: Feedback Buddy: The left picture shows an instruction page (activity state), while the right one displays a feedback page (feedback state).

2.1 Domain Model – Code Tracing

The education plan defines implementing building blocks like *if-else* branches and *for* loops and analyzing given code for its functionality and its effects [MJ18, p. 37-38] as learning goals for grade nine. Further, as *if-else* branches and *for* loops are building blocks for nearly every program, we decided to cover those in our system. Concretely, *if-else* branchings represent the lowest level of difficulty, as they only involve evaluating a condition and choosing between paths. *For* loops represent the intermediate level of difficulty as they require an understanding of initialization, condition checking, and iteration. A combination of both constitutes the highest level of difficulty as combining both increases complexity and demands managing both conditional and iterative logic simultaneously. To provide a large number of different exercises [Th19], templates to generate randomized program codes were created for the different difficulty levels[5]. Java is used in the material provided by the Center for School Quality and Teacher Training Baden-Wuerttemberg for computer science and, hence, is often used in schools. Therefore, we decided to provide the code tracing tasks in Java.

5 see supplementary materials S1, section 1

2.2 Learner Model – State of the Learner

The learner's state is represented via locally stored variables during use, accessible solely within a running working session. Tying in with Arroyo et al. [Ar14], who infer the learner's cognitive (i.e., knowledge), metacognitive (i.e., engagement and problem-solving strategy), and affective states (i.e., motivation and interest) from the correctness of the exercise and effort of the learner (i.e., processing time, number of attempt and request of solution), we applied a similar approach. As we did not use hints like Arroyo et al. [Ar14], we adjusted the parameter according to whether the solution was shown before solving the example. Following Arroyo et al. [Ar14], the integration of those parameters leads to a detailed interpretation of a learner's behavior: When a learner, for example, gives fast and wrong answers, this should be interpreted as a lack of interest.

2.3 Pedagogical Model – Teaching Strategy

The Feedback Buddy combines a theoretically based choice of the code tracing strategy with feedback reflecting a growth mindset and an adaptive metric, which aligns content and feedback to the cognitive, meta-cognitive, and affective learner states.

2.3.1 Code Tracing Strategy

Existing research highlights several strategies for successfully teaching code tracing [Cu17]. For instance, Xie et al. [XNK18] showed that using an explicit code tracing strategy leads to improved learner performance, progress, and less excessive demands, especially for low-performing learners. The effectiveness of explicit code tracing strategies depends on the chosen technique [QF20]. One promising approach is the line technique, which includes a complete code tracing table and a tabular listing of the variable that records their change by the program code line by line [Cu17]. In the Feedback Buddy, tracing tables are first explained with examples, subsequently followed by exercises, as indicated by Xie et al. [XNK18]. Figure 2 shows an example exercise.

Importantly, the degree of support during exercises should be appropriate to prevent frustration [JM20; JM21]. To avoid unnecessary redundancy, learners first work on a problem without explanations [reduced-scaffholding; JM20]. The Feedback Buddy only provides extended support with step-by-step examples when the inferred learner state indicates that they are needed. The presentation of the table in the correct size and the names of the variables to track provides support for the learner [high-scaffholding; JM20]. Furthermore, we implemented the possibility that learners could request feedback and subsequently decide whether they want to keep on trying to solve a task or receive the solution. Such freedom of choice increases the experience of autonomy and prevents frustration [Sc18].

2.3.2 Growth Mindset Feedback

Assuming that our system targets learners who start to learn code tracing from scratch, we face an audience with only little prior knowledge. Hence, we decided to keep the level of the tasks at the beginner level, provide immediate feedback depending on the answer of the learner, and present the correct answer as part of this feedback [MB01].

With the goal to foster a growth mindset in learners, our feedback was oriented on formulations by Kroeper et al. [KFM22]. Thereby, the focus is set on characteristics of the learning process and not individual traits of the learning person [BH18]. Investing effort is praised, while fast successes are put in perspective [DL88]. In terms of failure, the individual learning curve is emphasized, highlighting the ability to master the task with more practice [Tr17]. We additionally provide the option to try again, which has proven a positive influence on enhanced performance [Sh08]. Such an approach also aligns with the concept of building a growth mindset[6], as this promotes effort and learning from mistakes.

2.3.3 Adaptive Metric

The adaptive metric for the analysis and choice of activities, implemented in the Feedback Buddy, follows the Effort-Based Tutoring (EBT) algorithm [Ar14]. Adapted to our use case, to choose a sequence of exercises, the range of expected values for the number of attempts and processing time is calculated and compared to the learner's actually achieved values. The processing time starts when a new task is started and ends as soon as the trace table has been filled in correctly or the learner decides to display the solution. Additionally, the information on whether the learner viewed the solution or not is taken into account.

To optimize the task difficulty, we define a range of behavior within which the task difficulty is deemed appropriate. This range, termed the expected behavior, is defined using a lower and upper bound. Arroyo et al. [Ar14] justify the choice of a lower and upper value of the distribution by noting that the behavior of learners is not normally distributed. Given the great heterogeneity in performance among computer science learners, we account for individual differences by following this approach. In our system, the boundaries are defined using the value of expected behavior and fractions of their standard deviations, denoted as δ-values. These fractions are determined by two fixed parameters, θ_{low} and θ_{high}, in the interval [0, 1]. The ranges of expected behavior[7] are computed as:

$$\delta_{attempt_i,low} = SD(attempt_i)\theta_{low} \qquad \delta_{time_i,low} = SD(time_i)\theta_{low}$$
$$\delta_{attempt_i,high} = SD(attempt_i)\theta_{high} \qquad \delta_{time_i,high} = SD(time_i)\theta_{high}$$

$$attempt_{low} = E(attempt_i) - \delta_{attempt_i,low}, \qquad time_{low} = E(time_i) - \delta_{time_i,low}$$
$$attempt_{high} = E(attempt_i) + \delta_{attempt_i,high}, \qquad time_{high} = E(time_i) + \delta_{time_i,high},$$

6 The exact formulations chosen can be found in the supplementary materials S1, section 3.
7 For the values of our sample and calculations, see supplementary materials S1, section 2.

Index i refers to the level of difficulty. By selecting $\theta_{low} = 1/4$ and $\theta_{high} = 1/2$, it is possible to set the lower limit (very high performance) closer to the value of expected behavior and the higher limit (very low performance) further from the value of expected behavior.

Depending on whether the behavior of the learner the state of the learner is inferred, and the difficulty of the next exercise and the feedback are adopted (see Table 1). If the learner solved a task on the highest difficulty level and the system suggests an increase in the difficulty level, the next task has the highest difficulty level again. If the learner has worked on a task on the lowest difficulty level and the system suggests a reduction in the difficulty level, an example on the lowest level is shown.

Tab. 1: The inference of the state of the learner and according adaptations based on a learner's behavior (i.e., number of attempts, time required, and tasks). We use *at* for attempt and *t* for time. In the column *solution*, 1 indicates that the solution was requested and 0 that the learner solved the task.

Task attempts	Time on task	Solution	Assumed learner state	Adaptation of difficulty	Adaptation of feedback
$< at_{low}$	$< t_{low}$	0	Abilities exceed task difficulty	Increase difficulty level	Emphasize learning progress
$< at_{low}$	$> t_{high}$	0	High motivation, appropriate abilities	Maintain difficulty level	Praise effort
$< at_{low}$	$< t_{low}$	1	Low motivation/ insufficient abilities/ least effort strategy	Reduce difficulty level	Relativize immediate success
$< at_{low}$	$> t_{high}$	1	High motivation, insufficient abilities	Example at same difficulty level	Praise effort
$> at_{high}$	$< t_{low}$	0/1	Low motivation/ least effort strategy	Reduce difficulty level	Relativize immediate success
$> at_{high}$	$> t_{high}$	0	High motivation, appropriate abilities	Example at same difficulty level	Praise effort
$> at_{high}$	$> t_{high}$	1	High motivation, insufficient abilities	Example at reduced difficulty level	Highlight the importance of effort
$> at_{low}$ & $< at_{high}$	$> t_{low}$ & $< t_{high}$	0/1	Appropriate abilities	Maintain difficulty level	Praise effort

2.4 User Interface – Pedagogical Agent

Our choice to enhance our user interface with a pedagogical agent emerged from their (small) positive effects on learning performance, especially in schools [Ca21; SAG13]. In general, pedagogical agents are virtual characters that guide learners through multimedia learning environments [HC11]. They facilitate learning by explaining and giving feedback or corrections [MM07; Si23].

Schroeder et al. [SAG13] show that non-human agents generally exert greater positive effects. However, especially in the case of robots, male learners have shown increased learning performance. By contrast, female learners displayed increased learning performance, in particular for gender-aligned agents [AOR22; Da22]. Building on these results, we chose a

gender-neutral robot to prevent strong and potentially confounding gender effects arising from the pedagogical agent [AOR22; Ba11; BC23]. Furthermore, existing evidence indicates pedagogical agents with facial expressions but no gestures support the learning best [BK09]. Hence, we equipped Bud with different facial expressions and different poses but decided to avoid gestures. The different poses further give orientation, as they indicate the different states of the program (see Figure 2). Tying in with considerably stronger effects of 2D agents compared to 3D agents [Ca21], we chose to present Bud as a 2D visualization.

Additional related research found that learners' emotional state, motivation, and performance improved with enthusiasm compared to a neutral pedagogical agent [LZS17] and that a positive emotional expression in instructors can significantly enhance learning gains [La1b]. Hence, Bud shows enthusiasm through feedback (see the right side of Figure 2). Further, for the user interface, we decided to keep a consistent structure. The interface in all states except the initial and summary state consists of the program code (top-left), the agent and feedback (bottom-left), and the trace table (right). Fig. 2 shows this structure. In the initials and summary, the state only the feedback Buddy and text are shown. In the feedback state, incorrect entries are marked blue, and a feedback text is shown.

3 Evaluation of User Experience and Visual Design

Tying in with our proposed research question if and to what extent our Feedback Buddy is usable and useful to teach code tracing in an intermediate-level grammar school context, we put the system to the test in an actual school environment. As we simultaneously investigated potential differences regarding the influence of growth mindset and corrective feedback on learning outcomes[8], participants either received growth mindset or corrective feedback. Even though not all participants received growth mindset feedback, and we disabled the adaptive metric for control purposes, we consider our evaluation to hold value as we focus on the user experience and visual design of the overall system.

The ethics committee of the University of Stuttgart approved our plan for participant recruitment, data collection, and analysis under file number 23-023. We collected informed consent for the students' participation in the evaluation from the respective parents or guardians.

3.1 Participants and Measures

A total of 79 students of 10th grade (23 female, 56 male) participated in the evaluation study. They were aged between 14 and 17 years ($M = 15.14, SD = 0.52$) and part of a class with the elective profile subject *computer science, mathematics, and physics* (IMP). Due to

8 This investigation is reported by Kapfenstein [Ka23]

incomplete data or missing personalized codes, we could not use several initially collected datasets, resulting in a final sample size of 79.

System evaluation started by indicating the overall impression of the system by positioning a slider (0 – *very bad* to 100 – *very good*) in response to the question *"How much do you like the Feedback Buddy overall?"*. Subsequently, we used an age-appropriate adaptation of the meCUE 2.0 questionnaire [St23; TM07], involving ten items to be rated on a Likert scale (1 – *I completely disagree* to 7 – *I completely agree*). The items are shown in Fig. 3

3.2 Procedure

After an introductory lecture on code tracing, participants interacted with the system for about 30 to 45 minutes. To ensure a controlled evaluation procedure, all participants thereby received identical code tracing tasks in the same order: three with *if-else* branches, two with a *for* loop, and one that combined the two control structures. Following their experiences with the Feedback Buddy, participants provided their responses to the introduced evaluation questions and items. Overall, they spent about 60 minutes in our study.

4 Results[9]

Taken together, participants rated the overall impression of the Feedback Buddy with 81.14 points (SD = 17.65). Only six participants awarded less than half of the points, whereas 15 participants awarded points in the highest range of 95 to 100.

Figure 3 shows the response frequencies of the individual test items. As displayed, participants' assessments of our system were predominantly positive. Particularly noteworthy are the high ratings for perceived usefulness and the prevalence of learning opportunities, as well as the intuitive and simple use. Overall, more than 90 % of the involved learners agreed that they could learn something with the Feedback Buddy and found it useful. At the same time, results indicate that there is still room for improvement in the visual and creative design of the Feedback Buddy.

5 Discussion

With the Feedback Buddy, we present a tutoring system that adapts its feedback and the difficulty of the task to the learner's cognitive, metacognitive, and affective states. Substantially extending the functionality of former systems, our system is tailored towards students in intermediate-level grammar schools; it is based on vested evidence from the learning sciences, combines an adaptive metric with feedback, and uses growth mindset

9 The analysis scripts and required files can be found in supplementary materials S2.

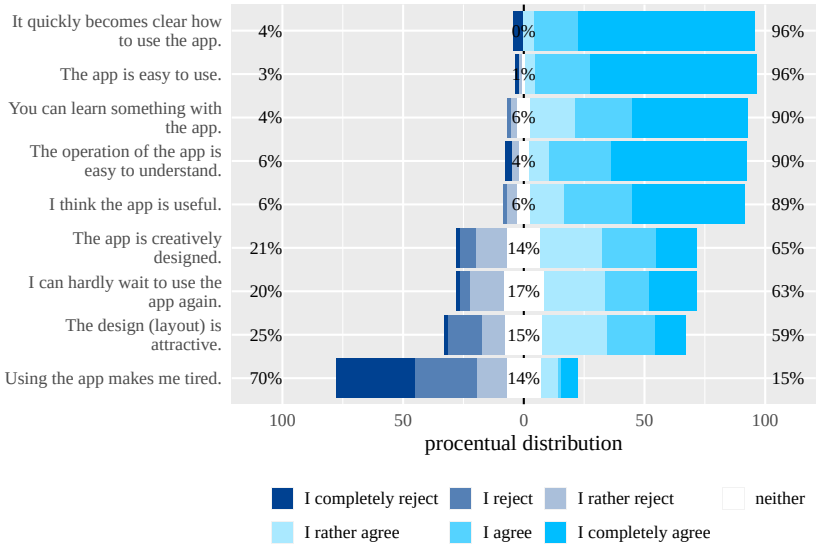


Fig. 3: Percentage distribution of the participants' ratings. The percentages on the left/right add up the numbers of negative/positive responses, respectively.

feedback formulations. As a first step, we investigated how well this approach to teaching code tracing was received. The evaluation showed an overall positive impression of the system and an especially positive evaluation of its usefulness, usability, and ability to provide learning opportunities. Simultaneously, we observe the potential for improvement related to the visual and creative design of our system.

With the EBT algorithm, the range of expected behavior may sometimes be very narrow and may not include any possible behavior. For instance, in our sample, the expected behavior for the *if-else* branching resides between 1.26 and 1.67 attempts[10]. A pattern like this indicates that learners who solve the task in one attempt are below the range of expected behavior, while those who solve it in two attempts exceed the expected number of attempts. Such a narrow range of expected behavior can cause the system to adjust the difficulty frequently. Such frequent adjustments can hinder the learning of the competencies intended to be developed at that difficulty level, as students are not given enough time to practice and master the necessary skills. However, the range of expected behavior can be adapted by increasing the δ_{high}-value, either for all difficulty levels or for the *if-else* branching only.

The currently existing prototype of the Feedback Buddy has some limitations. Although there are several schools teaching Python or other text-based programming languages, our system is restricted to Java code. Hence, an extension of the Feedback Buddy with syntax

10 see supplementary materials S1, section 2

trees [QF20] could enable a broader applicability of our system. Further, the current Feedback Buddy only teaches *if-else* branches, *for* loops, and combinations of them. Implementing further concepts such as *while* loops, *switch* statements, methods, or arrays can increase the usability of the system. Currently, the expected behavior for the adaptive metric is based on a rather limited data set. Hence, more data is needed that could be provided with long-term storage of log files on a server. This would also allow to create profiles to store individual data and make it available for formative assessment or further continuing based on saved data after terminating the software. Moreover, we chose to teach a tracing strategy explicitly [XNK18] by using trace tables [Ri21]. Other strategies, such as arrow flow techniques [Cu17], where the learner uses arrows to indicate changes in variables, may also be effective. Ideally, the choice of strategy is adapted to the respective student.

Future work should implement possibilities to customize the Feedback Buddy by teachers, for example, to choose a programming language or type of template and how many tasks per difficulty must be solved. An enhanced scope of functionalities could also include options to give learners a higher degree of autonomy by choosing a preferred design for the pedagogical agent, which has already demonstrated positive effects in prior research [Da22].

In conclusion, our Feedback Buddy represents a tutoring system that promises value for supporting school students with learning code tracing. With a design according to findings from research on code tracing, tutorial systems with adaptive learning metrics and pedagogical agents, as well as growth mindset interventions and feedback strategies, our system sets a new benchmark for the future development of tutorial systems. Given that promoting a growth mindset among learners is a rare endeavor, particularly in STEM fields, such an approach holds transformative value in programming education. Advanced opportunities to individually adjust difficulty levels in examples and tasks provide the opportunity to set personalized learning paths. Consequently, learners can receive tasks according to their current level of performance and effort, potentially giving low-performing learners more time and attempt to hone their competencies. Taken together, our Feedback Buddy could pave the way for STEM learners to develop enhanced academic performance, strong task motivation, and a constructive attitude toward challenges.

Acknowledgements

This research was supported by the Ministry of Science, Research, and the Arts (MWK) Baden-Württemberg, Germany, within the Artificial Intelligence Software Academy (AISA; reference number Az. 7542.2-9-47.10/36/2) and the funding scheme Künstliche Intelligenz & Gesellschaft: Reflecting Intelligent Systems for Diversity, Demography and Democracy (IRIS3D; reference number Az. 33-7533-9-19/54/5). We further acknowledge the support of the German Research Foundation (DFG; DFG reference number UP 31/1) for funding the Stuttgart Research Focus Interchange Forum for Reflecting on Intelligent Systems (IRIS).

References

[Ad15] Aditomo, A.: Students' Response to Academic Setback: 'Growth Mindset' as a Buffer Against Demotivation. International Journal of Educational Psychology 4/2, pp. 198–222, 2015, URL: http://doi.org/10.17583/ijep.2015.1482.

[AFG02] Aronson, J.; Fried, C. B.; Good, C.: Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence. Journal of Experimental Social Psychology 38/2, pp. 113–125, 2002, URL: https://doi.org/10.1006/jesp.2001.1491.

[AK17] Attali, Y.; van der Kleij, F.: Effects of Feedback Elaboration and Feedback Timing During Computer-Based Practice in Mathematics Problem Solving. Computers & Education 110/, pp. 154–169, 2017.

[AOR22] Armando, M.; Ochs, M.; Régner, I.: The Impact of Pedagogical Agents' Gender on Academic Learning: A Systematic Review. Frontiers in Artificial Intelligence 5/, 2022, URL: https://doi.org/10.3389/frai.2022.862997.

[Ar14] Arroyo, I.; Woolf, B. P.; Burelson, W.; Muldner, K.; Rai, D.; Tai, M.: A Multimedia Adaptive Tutoring System for Mathematics that Addresses Cognition, Metacognition and Affect. International Journal of Artificial Intelligence in Education 24/, pp. 387–426, 2014, URL: https://doi.org/10.1007/s40593-014-0023-y.

[Ba11] Baylor, A. L.: The design of motivational agents and avatars. Educational Technology Research and Development 59/, pp. 291–300, 2011, URL: https://doi.org/10.1007/s11423-011-9196-3.

[BC23] Beukman, M.; Chen, X.: Learner Perception of Pedagogical Agents. In (Wang, N.; Rebolledo-Mendez, G.; Dimitrova, V.; Matsuda, N.; Santos, O. C., eds.): Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky. Springer Nature Switzerland, Cham, pp. 587–592, 2023, ISBN: 978-3-031-36336-8.

[BH18] Brock, A.; Hundley, H.: In Other Words: Phrases for Growth Mindset: A Teacher's Guide to Empowering Students through Effective Praise and Feedback. Ulysses Press, 2018, ISBN: 978-1-61243-791-0.

[BK09] Baylor, A. L.; Kim, S.: Designing Nonverbal Communication for Pedagogical Agents: When Less is More. Computers in Human Behavior 25/2, Including the Special Issue: State of the Art Research into Cognitive Load Theory, pp. 450–457, 2009, URL: https://doi.org/10.1016/j.chb.2008.10.008.

[BTD07] Blackwell, L. S.; Trzesniewski, K. H.; Dweck, C. S.: Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention. Child Development 78/1, pp. 246–263, 2007, URL: https://doi.org/10.1111/j.1467-8624.2007.00995.x.

[Bu20] Burnette, J. L.; Hoyt, C. L.; Russell, V. M.; Lawson, B.; Dweck, C. S.; Finkel, E.: A Growth Mind-Set Intervention Improves Interest but Not Academic Performance in the Field of Computer Science. Social Psychological and Personality Science 11/1, pp. 107–116, 2020, URL: https://doi.org/10.1177/1948550619841631.

[Bu23] Burnette, J. L.; Billingsley, J.; Banks, G. C.; Knouse, L. E.; Hoyt, C. L.; Pollack, J. M.; Simon, S.: A Systematic Review and Meta-Analysis of Growth Mindset Interventions: For Whom, How, and Why Might Such Interventions Work? Psychological Bulletin 149/3-4, pp. 174–205, 2023, URL: https://doi.org/10.1037/bul0000368.

[Ca21] Castro-Alonso, J. C.; Wong, R. M.; Adesope, O.; Paas, F.: Effectiveness of multimedia Pedagogical Agents Predicted by Diverse Theories: A Meta-Analysis. Educational Psychology Review 33/, pp. 989–1015, 2021, URL: https://doi.org/10.1007/s10648-020-09587-1.

[CH09] Chase, J. A.; Houmanfar, R.: The Differential Effects of Elaborate Feedback and Basic Feedback on Student Performance in a Modified, Personalized System of Instruction Course. Journal of Behavioral Education 18/, pp. 245–265, 2009, URL: https://doi.org/10.1007/s10864-009-9089-2.

[Cu17] Cunningham, K.; Blanchard, S.; Ericson, B.; Guzdial, M.: Using Tracing and Sketching to Solve Programming Problems: Replicating and Extending an Analysis of What Students Draw. In: Proceedings of the 2017 ACM Conference on International Computing Education Research. ICER '17, Association for Computing Machinery, Tacoma, Washington, USA, pp. 164–172, 2017, ISBN: 9781450349680, URL: https://doi.org/10.1145/3105726.3106190.

[Da22] Dai, L.; Jung, M. M.; Postma, M.; Louwerse, M. M.: A Systematic Review of Pedagogical Agent Research: Similarities, Differences and Unexplored Aspects. Computers & Education 190/, p. 104607, 2022, ISSN: 0360-1315, URL: https://doi.org/10.1016/j.compedu.2022.104607.

[DK03] Dancik, G.; Kumar, A. N.: A Tutor for Counter-Controlled Loop Concepts and Its Evaluation. In: 33rd Annual Frontiers in Education, 2003. FIE 2003. Vol. 1, T3C–7, 2003, URL: https://doi.org/10.1109/FIE.2003.1263331.

[DKR99] Deci, E. L.; Koestner, R.; Ryan, R. M.: A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. Psychological bulletin 125/6, p. 627, 1999, URL: https://doi.org/10.1037/0033-2909.125.6.627.

[DL88] Dweck, C. S.; Leggett, E. L.: A Social-Cognitive Approach to Motivation and Personality. Psychological Review 95/, pp. 256–273, 1988, URL: https://doi.org/10.1037/0033-295X.95.2.256.

[Dw00] Dweck, C. S.: Self-theories: Their Role in Motivation, Personality, and Development. Psychology Press, 2000.

[Fe19] Fedorenko, E.; Ivanova, A.; Dhamala, R.; Bers, M. U.: The Language of Programming: A Cognitive Perspective. Trends in cognitive sciences 23/7, pp. 525–528, 2019, URL: https://doi.org/10.1016/j.tics.2019.04.010.

[Ha14] Harks, B.; Rakoczy, K.; Hattie, J.; Besser, M.; Klieme, E.: The Effects of Feedback on Achievement, Interest and Self-Evaluation: The Role of Feedback's Perceived Usefulness. Educational Psychology 34/3, pp. 269–290, 2014, URL: https://doi.org/10.1080/01443410.2013.785384.

[HC11] Heidig, S.; Clarebout, G.: Do Pedagogical Agents Make a Difference to Student Motivation and Learning? Educational Research Review 6/1, pp. 27–54, 2011, ISSN: 1747-938X, URL: https://doi.org/10.1016/j.edurev.2010.07.004.

[Hö19] Höppner, F.: Measuring Instruction Comprehension by Mining Memory Traces for Early Formative Feedback in Java Courses. In: Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education. ITiCSE '19, Association for Computing Machinery, Aberdeen, Scotland Uk, pp. 105–111, 2019, URL: https://doi.org/10.1145/3304221.3325529.

[HT07] Hattie, J.; Timperley, H.: The Power of Feedback. Review of Educational Research 77/1, pp. 81–112, 2007, URL: https://doi.org/10.3102/003465430298487.

[Hu23] Huang, Y.; Brusilovsky, P.; Guerra, J.; Koedinger, K.; Schunn, C.: Supporting Skill Integration in an Intelligent Tutoring System for Code Tracing. Journal of Computer Assisted Learning 39/2, pp. 477–500, 2023, URL: https://doi.org/10.1111/jcal.12757.

[JM20] Jennings, J.; Muldner, K.: When Does Scaffolding Provide Too Much Assistance? A Code-Tracing Tutor Investigation. International Journal of Artificial Intelligence in Education 31/4, pp. 784–819, 2020, URL: https://doi.org/10.1007/s40593-020-00217-z.

[JM21] Jennings, J.; Muldner, K.: Investigating Students' Reasoning in a Code-Tracing Tutor. In (Roll, I.; McNamara, D.; Sosnovsky, S.; Luckin, R.; Dimitrova, V., eds.): Artificial Intelligence in Education. Springer International Publishing, Cham, pp. 203–214, 2021.

[Ka23] Kapfenstein, A.-K.: Einfluss von Growth Mindset Feedback auf das Erlernen von Code-Tracing mit einem adaptiven tutoriellen System, unveröffentlichte Masterarbeit, 2023.

[KF16] Kulik, J. A.; Fletcher, J. D.: Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. Review of Educational Research 86/1, pp. 42–78, 2016, URL: https://doi.org/10.3102/0034654315581420.

[KFM22] Kroeper, K. M.; Fried, A. C.; Murphy, M. C.: Towards Fostering Growth Mindset Classrooms: Identifying Teaching Behaviors That Signal Instructors' Fixed and Growth Mindset Beliefs to Students. Social Psychology of Education 25/, pp. 371–398, 2022, URL: https://doi.org/10.1007/s11218-022-09689-4.

[Ki16] King, P. E.: When Do Students Benefit from Performance Feedback? A Test of Feedback Intervention Theory in Speaking Improvement. Communication Quarterly 64/1, pp. 1–15, 2016, URL: https://doi.org/10.1080/01463373.2015.1078827.

[KJH18] Keuning, H.; Jeuring, J.; Heeren, B.: A Systematic Literature Review of Automated Feedback Generation for Programming Exercises. ACM Trans. Comput. Educ. 19/1, 2018, URL: https://doi.org/10.1145/3231711.

[KN13] Komarraju, M.; Nadler, D.: Self-Efficacy and Academic Achievement: Why Do Implicit Beliefs, Goals, and Effort Regulation Matter? Learning and Individual Differences 25/, pp. 67–72, 2013, ISSN: 1041-6080, URL: https://doi.org/10.1016/j.lindif.2013.01.005.

[KO19] Kakeshita, T.; Ohta, K.: Student Log Analysis Functions for Web-based Programming Education Support Tool pgtracer. Journal of Information Processing 27/, pp. 456–468, 2019, URL: https://doi.org/10.2197/ipsjjip.27.456.

[Ku15] Kumar, A. N.: Solving Code-Tracing Problems and Its Effect on Code-Writing Skills Pertaining to Program Semantics. In: Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education. ITiCSE '15, Association for Computing Machinery, Vilnius, Lithuania, pp. 314–319, 2015, URL: https://doi.org/10.1145/2729094.2742587.

[Ku20] Kumar, A. N.: Long Term Retention of Programming Concepts Learned Using a Software Tutor. In (Kumar, V.; Troussas, C., eds.): Intelligent Tutoring Systems: 16th International Conference, ITS 2020, Athens, Greece, June 8–12, 2020, Proceedings. Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 382–387, 2020, URL: https://doi.org/10.1007/978-3-030-49663-0_46.

[La1b] Lawson, A. P.; Mayer, R. E.; Adamo-Villani, N.; Benes, B.; Lei, X.; Cheng, J.: The Positivity Principle: Do Positive Instructors Improve Learning from Video Lectures? Educational Technology Research and Development 69/6, pp. 3101–3129, 2021b, URL: https://doi.org/10.1007/s11423-021-10057-w.

[Le15] Leslie, S.-J.; Cimpian, A.; Meyer, M.; Freeland, E.: Expectations of Brilliance Underlie Gender Distributions Across Academic Disciplines. Science 347/6219, pp. 262–265, 2015, URL: https://doi.org/10.1126/science.1261375.

[LF18] Lechermeier, J.; Fassnacht, M.: How Do Performance Feedback Characteristics Influence Recipients' Reactions? A State-of-the-Art Review on Feedback Source, Timing, and Valence Effects. Management Review Quarterly 68/2, pp. 145–193, 2018, URL: https://doi.org/10.1007/s10864-009-9089-2.

[Li04] Lister, R.; Adams, E. S.; Fitzgerald, S.; Fone, W.; Hamer, J.; Lindholm, M.; McCartney, R.; Moström, J. E.; Sanders, K.; Seppälä, O.; Simon, B.; Thomas, L.: A Multi-National Study of Reading and Tracing Skills in Novice Programmers. In: Working Group Reports from ITiCSE on Innovation and Technology in Computer Science Education. ITiCSE-WGR '04, Association for Computing Machinery, Leeds, United Kingdom, pp. 119–150, 2004, URL: https://doi.org/10.1145/1044550.1041673.

[Lo08] Lopez, M.; Whalley, J.; Robbins, P.; Lister, R.: Relationships between Reading, Tracing and Writing Skills in Introductory Programming. In: Proceedings of the Fourth International Workshop on Computing Education Research. ICER '08, Association for Computing Machinery, Sydney, Australia, pp. 101–112, 2008, URL: https://doi.org/10.1145/1404520.1404531.

[LZS17] Liew, T. W.; Zin, N. A. M.; Sahari, N.: Exploring the Affective, Motivational, and Cognitive Effects of Pedagogical Agent Enthusiasm in a Multimedia Learning Environment. Human-centric Computing and Information Sciences 7/, p. 9, 2017, URL: https://doi.org/10.1186/s13673-017-0089-2.

[Ma14] Ma, W.; Adesope, O.; Nesbit, J. C.; Liu, Q.: Intelligent Tutoring Systems and Learning Outcomes: A Meta analysis. Journal of Educational Psychology 106/4, pp. 901–918, 2014, URL: https://doi.org/10.1037/a0037123.

[MB01] Mason, B.; Bruning, R.: Providing Feedback in Computer-based Instruction: What the Research Tells Us. Center for Instructional Innovation, University of Nebraska-Lincoln/, 2001.

[MB23] Macnamara, B. N.; Burgoyne, A. P.: Do Growth Mindset Interventions Impact Students' Academic Achievement? A Systematic Review and Meta-Analysis with Recommendations for Best Practices. Psychological Bulletin 149/3-4, p. 133, 2023, URL: https://doi.org/10.1037/bul0000352.

[Mc01] McCracken, M.; Almstrum, V.; Diaz, D.; Guzdial, M.; Hagan, D.; Kolikant, Y. B.-D.; Laxer, C.; Thomas, L.; Utting, I.; Wilusz, T.: A Multi-National, Multi-Institutional Study of Assessment of Programming Skills of First-Year CS Students. SIGCSE Bull. 33/4, pp. 125–140, 2001, ISSN: 0097-8418, URL: https://doi.org/10.1145/572139.572181.

[MD98] Mueller, C.; Dweck, C. S.: Praise for Intelligence Can Undermine Children's Motivation and Performance. Journal of Personality and Social Psychology 75/1, pp. 33–52, 1998, URL: https://doi.org/10.1037/0022-3514.75.1.33.

[Me24] Meißner, N.; Koch, N.; Speth, S.; Breitenbücher, U.; Becker, S.: Unveiling Hurdles in Software Engineering Education: The Role of Learning Management Systems. In: Proceedings of the 46th International Conference on Software Engineering: Software Engineering Education and Training. ICSE-SEET '24, Association for Computing Machinery, , Lisbon, Portugal, pp. 242–252, 2024, ISBN: 9798400704987, URL: https://doi.org/10.1145/3639474.3640060.

[MJ17] Ministerium für Kultus, Jugend und Sport Baden-Württemberg: Aufbaukurs Informatik. Neckar-Verlag GmbH, 2017.

[MJ18] Ministerium für Kultus, Jugend und Sport Baden-Württemberg: Informatik, Mathematik, Physik (IMP): Profilfach. Neckar-Verlag GmbH, 2018.

[MM07] Moreno, R.; Mayer, R. E.: Interactive multimodal learning environments. Educational Psychology Review 19/3, pp. 309–326, 2007, URL: https://doi.org/10.1007/s10648-007-9047-2.

[MSB23] Meißner, N.; Speth, S.; Breitenbücher, U.: An Intelligent Tutoring System Concept for a Gamified e-Learning Platform for Higher Computer Science Education. In: Software Engineering im Unterricht der Hochschulen (SEUH 2023). Gesellschaft für Informatik, Bonn, 2023.

[Mu20] Muenks, K.; Canning, E. A.; LaCosse, J.; Green, D. J.; Zirkel, S.; Garcia, J. A.; Murphy, M. C.: Does My Professor Think My Ability Can Change? Students' Perceptions of Their STEM Professors' Mindset Beliefs Predict Their Psychological Vulnerability, Engagement, and Performance in Class. Journal of Experimental Psychology: General 149/11, pp. 2119–2144, 2020, URL: https://doi.org/10.1037/xge0000763.

[Ne14] Nesbit, J. C.; Adesope, O. O.; Liu, Q.; Ma, W.: How Effective are Intelligent Tutoring Systems in Computer Science Education? In: 2014 IEEE 14th International Conference on Advanced Learning Technologies. Pp. 99–103, 2014, URL: https://doi.org/10.1109/ICALT.2014.38.

[NH04] Narciss, S.; Huth, K.: How to design informative tutoring feedback for multimedia learning. Instructional design for multimedia learning 181195/, 2004.

[NXK17] Nelson, G. L.; Xie, B.; Ko, A. J.: Comprehension First: Evaluating a Novel Pedagogy and Tutoring System for Program Tracing in CS1. In: Proceedings of the 2017 ACM Conference on International Computing Education Research. ICER '17, Association for Computing Machinery, Tacoma, Washington, USA, pp. 2–11, 2017, URL: https://doi.org/10.1145/3105726.3106178.

[QF20] Qi, R.; Fossati, D.: Unlimited Trace Tutor: Learning Code Tracing With Automatically Generated Programs. In. SIGCSE '20, Association for Computing Machinery, Portland, OR, USA, pp. 427–433, 2020, URL: https://doi.org/10.1145/3328778.3366939.

[Ra19] Rakoczy, K.; Pinger, P.; Hochweber, J.; Klieme, E.; Schütze, B.; Besser, M.: Formative Assessment in Mathematics: Mediated by Feedback's Perceived Usefulness and Students' Self-Efficacy. Learning and Instruction 60/, pp. 154–165, 2019, URL: https://doi.org/10.1016/j.learninstruc.2018.01.004.

[RBD19] Redifer, J. L.; Bae, C. L.; DeBusk-Lane, M.: Implicit Theories, Working Memory, and Cognitive Load: Impacts on Creative Thinking. Sage Open 9/1, 2019, URL: https://doi.org/10.1177/2158244019835919.

[RF06] Reas, C.; Fry, B.: Processing: Programming for the Media Arts. AI & Society 20/, pp. 526–538, 2006.

[RGD12] Rattan, A.; Good, C.; Dweck, C. S.: "It's Ok—Not Everyone Can Be Good at Math": Instructors with an Entity Theory Comfort (and Demotivate) Students. Journal of Experimental Social Psychology 48/3, pp. 731–737, 2012, ISSN: 0022-1031, URL: https://doi.org/10.1016/j.jesp.2011.12.012.

[Rh18] Rhew, E.; Piro, J. S.; Goolkasian, P.; Cosentino, P.: The Effects of a Growth Mindset on Self-Efficacy and Motivation. Cogent Education 5/1, ed. by Palikara, O., p. 1492337, 2018, URL: https://doi.org/10.1080/2331186X.2018.1492337.

[Ri21] Risha, Z.; Barria-Pineda, J.; Akhuseyinoglu, K.; Brusilovsky, P.: Stepwise Help and Scaffolding for Java Code Tracing Problems With an Interactive Trace Table. In: Proceedings of the 21st Koli Calling International Conference on Computing Education Research. Koli Calling '21, Association for Computing Machinery, Joensuu, Finland, 2021, URL: https://doi.org/10.1145/3488042.3490508.

[SAG13] Schroeder, N. L.; Adesope, O. O.; Gilbert, R. B.: How Effective are Pedagogical Agents for Learning? A Meta-Analytic Review. Journal of Educational Computing Research 49/1, pp. 1–39, 2013, URL: https://doi.org/10.2190/EC.49.1.a.

[Sc18] Schneider, S.; Nebel, S.; Beege, M.; Rey, G. D.: The autonomy-enhancing effects of choice on cognitive load, motivation and learning with digital media. Learning and Instruction 58/, pp. 161–172, 2018.

[Sh08] Shute, V. J.: Focus on Formative Feedback. Review of Educational Research 78/1, pp. 153–189, 2008, URL: https://doi.org/10.3102/0034654307313795.

[Si18] Sisk, V. F.; Burgoyne, A. P.; Sun, J.; Butler, J. L.; Macnamara, B. N.: To What Extent and Under Which Circumstances Are Growth Mind-Sets Important to Academic Achievement? Two Meta-Analyses. Psychological Science 29/4, pp. 549–571, 2018, URL: https://doi.org/10.1177/0956797617739704.

[Si23] Siegle, R. F.; Schroeder, N. L.; Lane, H. C.; Craig, S. D.: Twenty-Five Years of Learning with Pedagogical Agents: History, Barriers, and Opportunities. TechTrends 67/, pp. 851–864, 2023, URL: https://doi.org/10.1007/s11528-023-00869-3.

[SK00] Singhal, N.; Kumar, A. N.: Facilitating Problem-Solving on Nested Selection Statements Using C/C++. In: 30th Annual Frontiers in Education Conference. Building on A Century of Progress in Engineering Education. Conference Proceedings (IEEE Cat. No.00CH37135). Vol. 1, T4C/1–T4C/6 vol.1, 2000, URL: https://doi.org/10.1109/FIE.2000.897662.

[So16] Sottilare, R. A.; Graesser, A. C.; Hu, X.; Olney, A.; Nye, B.; Sinatra, A. M.: Design Recommendations for Intelligent Tutoring Systems: Volume 4 - Domain Modeling. US Army Research Laboratory, 2016.

[St23] Stock, A.; Stock, O.; Suren, M.; Koch, N. N.; Rey, G. D.; Wirzberger, M.: BeeLife: A Mobile Application to Foster Environmental Awareness in Classroom Settings. Frontiers in Computer Science – Digital Education 5/, demnächst veröffentlicht, 2023, URL: https://doi.org/10.3389/fcomp.2023.1298888.

[SZ12] Shute, V. J.; Zapata-Rivera, D.: Adaptive Educational Systems. In (Durlach, P. J.; Lesgold, A. M., eds.): Adaptive Technologies for Training and Education. Cambridge University Press, pp. 7–27, 2012.

[Th19] Thomas, A.; Stopera, T.; Frank-Bolton, P.; Simha, R.: Stochastic Tree-Based Generation of Program-Tracing Practice Questions. In: Proceedings of the 50th ACM Technical Symposium on Computer Science Education. SIGCSE '19, Association for Computing Machinery, Minneapolis, MN, USA, pp. 91–97, 2019, URL: https://doi.org/10.1145/3287324.3287492.

[TM07] Thüring, M.; Mahlke, S.: Usability, Aesthetics, and Emotions in Human-Technology Interaction. International Journal of Psychology 42/4, noch unveröffentlicht, pp. 253–264, 2007, URL: https://doi.org/10.1080/00207590701396674.

[Tr17] Truax, M. L.: The Impact of Teacher Language and Growth Mindset Feedback on Writing Motivation. Literacy Research and Instruction 57/2, pp. 135–157, 2017, URL: https://doi.org/10.1080/19388071.2017.1340529.

[Va11] VanLehn, K.: The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist 46/4, pp. 197–221, 2011, URL: https://doi.org/10.1080/00461520.2011.611369.

[VB23] Vestad, L.; Bru, E.: Teachers' Support for Growth Mindset and Its Links with Students' Growth Mindset, Academic Engagement, and Achievements in Lower Secondary School. Social Psychology of Education/, 2023, URL: https://doi.org/10.1007/s11218-023-09859-y.

[VK11] Van Dijk, D.; Kluger, A. N.: Task Type as a Moderator of Positive/Negative Feedback Effects on Motivation and Performance: A Regulatory Focus Perspective. Journal of Organizational Behavior 32/8, pp. 1084–1105, 2011, URL: https://doi.org/10.1002/job.725.

[Wa23] Wang, H.; Tlili, A.; Huang, R.; Cai, Z.; Li, M.; Cheng, Z.; Yang, D.; Li, M.; Zhu, X.; Fei, C.: Examining the Applications of Intelligent Tutoring Systems in Real Educational Contexts: A Systematic Literature Review from the Social Experiment Perspective. Education and Information Technologies 28/, pp. 9113–9148, 2023, URL: https://doi.org/10.1007/s10639-022-11555-x.

[We07] Weicker, N.: Zielorientierte Didaktik der Informatik – Kompetenzvermittlung bei engen Zeitvorgaben. In: Didaktik der Informatik in Theorie und Praxis:12. GI-Fachtagung „Informatik und Schule - INFOS 2007". Gesellschaft für Informatik e.V., pp. 337–348, 2007, URL: https://dl.gi.de/items/4475b383-d711-4d96-9a7e-cb65c530f7d0.

[WZH20] Wisniewski, B.; Zierer, K.; Hattie, J.: The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. Frontiers in Psychology 10/, 2020, URL: https://doi.org/10.3389/fpsyg.2019.03087.

[XNK18] Xie, B.; Nelson, G. L.; Ko, A. J.: An Explicit Strategy to Scaffold Novice Program Tracing. In: Proceedings of the 49th ACM Technical Symposium on Computer Science Education. SIGCSE '18, Association for Computing Machinery, Baltimore, Maryland, USA, pp. 344–349, 2018, URL: https://doi.org/10.1145/3159450.3159527.

[Xu21] Xu, K. M.; Koorn, P.; De Koning, B.; Skuballa, I. T.; Lin, L.; Henderikx, M.; Marsh, H. W.; Sweller, J.; Paas, F.: A Growth Mindset Lowers Perceived Cognitive Load and Improves Learning: Integrating Motivation to Cognitive Load. Journal of Educational Psychology 113/6, p. 1177, 2021, URL: https://doi.org/10.1037/edu0000631.

Entwicklung medien(fach-)didaktischer Kompetenzen bei Lehramtsstudierenden auf Basis curricular verankerter Medienprojekte

Alexander Aumann [1], Wolfgang Müller [2], Jörg Stratmann[3] und Holger Weitzel [1]

Abstract: Der vorliegende Beitrag beschreibt Ansätze, Konzepte und Prozesse im Rahmen der Integration verpflichtender curricularer Elemente zur Förderung digitaler Kompetenzen im Lehramtsstudium an einer Hochschule. Das Vorhaben wurde im Rahmen eines aktiven Change Prozesses umgesetzt, bei dem alle relevanten Stakeholder einbezogen wurden. Der Artikel beschreibt den hochschulischen Veränderungsprozess und geht insbesondere auf Unterstützungsangebote, den entwickelten Kompetenzrahmen sowie den entstandenen Makerspace ein.

Keywords: Lehramtsstudium, Digitale Kompetenzen, Change Management, Makerspaces

1 Einleitung

(Inter-)national wird dem Einsatz digitaler Technologien im Klassenzimmer, den medienbezogenen Fähigkeiten der Schüler:innen und der Förderung digitalisierungsbezogener Kompetenzen von (angehenden) Lehrkräften zunehmend Bedeutung beigemessen (z.B. [RP17]). Damit steigen auch die Anforderungen an die Kompetenzen der Lehrkräfte. Im Folgenden beschreiben wir ein Konzept zur verpflichtenden curricularen Verankerung von medien(fach-)didaktischen Kompetenzen im Lehramtsstudium für alle Fächer. Im Rahmen eines dreijährigen Change Prozesses wurde an der Weiterentwicklung der Hochschule gearbeitet.

2 Related Work

2.1 Förderung digitaler Kompetenzen von Lehrkräften

Die Notwendigkeit, dass Lehrkräfte über Kompetenzen zur effektiven Nutzung digitaler Medien im Unterricht und zu deren Vermittlung verfügen, wird seit Langem diskutiert. Für

1 Pädagogische Hochschule Weingarten, Biologiedidaktik, Kirchplatz 2, 88250 Weingarten, Deutschland, alexander.aumann@ph-weingarten.de, https://orcid.org/0000-0003-4746-5035; weitzel@ph-weingarten.de, https://orcid.org/0000-0001-6525-4341
2 Pädagogische Hochschule Weingarten, Mediendidaktik, Kirchplatz 2, 88250 Weingarten, Deutschland, muellerw@ph-weingarten.de, https://orcid.org/0000-0001-6474-3733
3 Pädagogische Hochschule Weingarten, Medienpädagogik, Kirchplatz 2, 88250 Weingarten, Deutschland, stratmann@ph-weingarten.de

doi:10.18420/delfi2024_17

diesen Diskurs ist das TPACK-Modell (Technological Pedagogical Content Knowledge, [MK06]) international eine wichtige Grundlage. Es differenziert medien(fach-)didaktische Kompetenzen für Lehrkräfte in sieben Teilbereiche aus, verbleibt jedoch aufgrund seines fächerübergreifenden Anspruchs auf einer abstrakten Ebene, die eine fach- oder domänenspezifische Konkretisierung erforderlich macht. Mit dem Ziel der Konkretisierung wurden in den letzten Jahren mehrere Rahmenmodelle entwickelt. Hierzu gehören insbesondere das UNESCO ICT Competency Framework for Teachers [UN18] und das European Framework for the Digital Competence of Educators (DigCompEdu, [RP17]). DigCompEdu wurde von der KMK [Ku21] als Orientierungsrahmen empfohlen, für den die Bundesländer spezifische Kompetenzrahmen entwickeln sollen. Ein solcher liegt für Bayern in Form von DigCompEdu Bavaria [me21] bereits vor. Zwar bieten diese Modelle und Rahmen eine Orientierung zur Entwicklung entsprechender Lernangebote, zur Operationalisierung hinsichtlich der für Lehrkräfte erforderlichen Kompetenzen bedarf es jedoch weiterer Konkretisierung und nationaler bzw. regionaler Spezifizierung. Solche Spezifizierungen wurden beispielsweise in Bezug auf die MINT-Fächer vorgeschlagen (z.B. [Be20], [Gh20]).

In Deutschland ist vor allem im Anschluss an die KMK-Strategie zur "Bildung in der digitalen Welt"[Ku17] und die Förderung der Digitalisierung an den Hochschulen im Rahmen der Qualitätsoffensive Lehrerbildung ein Digitalisierungsprozess in Gang gesetzt worden. Trotzdem ist die verpflichtende Integration von Inhalten zur Förderung digitalisierungsbezogener Kompetenzen in der Lehramtsbildung noch immer nicht gegeben.

2.2 Makerspaces in der Lehre

Makerspaces haben in den letzten Jahren zunehmend Eingang in verschiedene Bildungskontexte sowohl außerschulisch wie in Schule und Hochschule gefunden (z.B. [KKO20; KPZ21; Me21; RS21]). Ein Grund hierfür ist die Möglichkeit zur Verknüpfung von Hands-on Aktivitäten mit der Förderung von 21st Century Skills wie Kreativität, Kollaboration, Kommunikation und Problemlösen (vgl. [Va17; Va20]). Als Makerspaces werden Lernumgebungen bezeichnet, in denen sich Menschen zusammenfinden, um physische oder digitale Produkte zu entwickeln und in diesem Prozess ihr Wissen zu erweitern [Me21]. “Making” wird in einem solchen Kontext als intrinsisch motivierte Aktivität betrachtet und gilt aufgrund einer tiefergehenden Auseinandersetzung mit dem Lerngegenstand als hilfreich für erfolgreiches Lernen [DR85; Va04]. Entsprechend stellt “Making” eine vielversprechende Lernaktivität in konstruktivistisch orientierten Lernumgebungen dar [Be17; KKF14].

Pädagogische Makerspaces sind darauf ausgerichtet, Lehramtsstudierenden durch kreatives und kollaboratives Problemlösen sowohl medienbezogene Kompetenzen, 21st Century Skills, als auch Kompetenzen zur Planung, Durchführung und Reflexion von "Making-"Projekten zu vermitteln [MWL23]. Damit ist die Vorstellung verbunden, dass die Studierenden in die Lage versetzt werden, projektorientiertes und problemorientiertes Lernen in ihren späteren Schulalltag zu übertragen und somit einen Beitrag zur Modernisierung von Unterricht zu leisten. Damit pädagogische Makerspaces dieser Aufgabe nachkommen können, sollte die

Infrastruktur so reichhaltig sein, dass sie eine breite Palette an Produktionsmöglichkeiten ermöglicht (z.B. [Co17; JL19]). Zudem sollen die Projekte eng mit der späteren Praxis verbunden sein, um die Übertragbarkeit der Maker-Projekte ins Klassenzimmer zu erleichtern [Jo20; PDP18]. Damit folgt die Arbeitsweise dem pädagogischen Doppeldecker, nach dem selbst gemachte Lernerfahrungen leichter ins Klassenzimmer übertragen werden können [Jo20].

3 Umsetzung und Change-Prozess

Im Rahmen des hier beschriebenen Projekts wurde ein Change Prozess für das Lehramtsstudium an der Hochschule initiiert. Dieser Prozess betrifft damit beide Fakultäten der Hochschule mit ca. 130 Lehrenden sowie ca. 2.100 Studierenden.

Dieser Change Prozess wurde aktiv von der Projektleitung in enger Zusammenarbeit mit der Hochschulleitung und den Fakultätsleitungen begleitet. Zudem wurden jeweils Change Agents in den beiden Fakultäten installiert, die Fächer und Lehrende im Kontext notwendiger curricularer Umstellungen und Anpassungen begleiteten. Darüber hinaus unterstützten die Change Agents auch bei Einführung der für einige Fächer und Lehrende neuen Form des projektbasierten Lernens und bei der Integration medien(fach-)didaktischer Elemente, etwa durch gezielte Schulungsangebote.

3.1 Verbindliche Verankerung von zwei Medienprojekten im Lehramt

Gemeinsam mit allen relevanten Stakeholdern ((Pro)Rektorate, Fakultäten, Lehrende, Studierende, Verwaltung) wurde ein Konzept erarbeitet, welches die obligatorische Verankerung zweier medien(fach-)didaktischer Projekte im Verlauf des Lehramtsstudiums vorsieht. In diesem Prozess waren alle am Lehramtsstudium beteiligten Fächer dazu aufgefordert, medien(fach-)didaktische Kompetenzen in ihre Modulhandbücher zu integrieren. Da die Studierenden zwei Unterrichtsfächer sowie Bildungswissenschaften studieren, steht ihnen eine große Auswahl an Gelegenheiten und Themen zur Verfügung, um ihre beiden verpflichtenden Projekte zu bearbeiten. Dadurch erscheint die Möglichkeit zur intrinsisch motivierten Themenwahl ausreichend in der Hochschulstruktur abgebildet.

Die beiden Projekte ergänzen sich hinsichtlich des Kompetenzaufbaus. Im ersten Medienprojekt werden vorhandene Lehr-/Lernmedien für eine konkrete Lehr-/Lernsituation ausgewählt und aus einer medien(fach-)didaktischen Perspektive reflektiert. Im zweiten Medienprojekt entwickeln die Studierenden eigene digitale Lehr-/Lernmedien auf Basis einer medien(fach-)didaktischen Analyse. Hierzu benötigen sie nicht nur medien(fach-)didaktisches Wissen, sondern auch technisch-instrumentelle Fähigkeiten und Fertigkeiten. Dabei ist offen, was für ein mediales Artefakt entwickelt wird. Dies ist abhängig vom konkreten Fach- und Unterrichtskontext, in dem es eingesetzt werden soll.

3.2 Aufbau und Nutzung eines pädagogischen Makerspaces

Als zentraler Anlaufpunkt insbesondere für das Medienentwicklungsprojekt wurde ein pädagogischer Makerspace aufgebaut. Dieser setzt sich aus insgesamt fünf Räumen und zwei offenen Konzeptionsbereichen zusammen. Letztere geben den Studierenden Raum für kollaborative Ideenentwicklung, während die Räume eine Vielzahl an technischer Ausstattung für das anschließende Prototyping und die flexible Umsetzung von Projektideen bereithalten. Um große Gruppen von Studierenden aufzunehmen, wurden die Räume in unterschiedliche Arbeitsbereiche ausdifferenziert, sodass verschiedenste Projekte parallel bearbeitet werden können. Damit werden den Studierenden Technologien zur Erstellung von Erklärvideos sowie Podcasts, digitale Fertigungstechnologien wie 3D-Modellierung sowie -Druck, Lasercutting und -engraving, Textildruck und programmiertes Nähen, handgeführtes CNC-Fräsen, Mikrocontroller und Sensorik bis hin zu VR- und AR-Technologien zugänglich.

Auf Basis einer Bedarfsanalyse wurden die Räume auf eine maximale Kapazität von ca. 300 Studierenden pro Semester mit potenziell punktueller Nutzungshäufung ausgelegt. Um die Studierenden in der selbstgesteuerten Umsetzung ihrer Projekte zu unterstützen und somit die Gefahr vorzeitigen Abbruchs zu reduzieren, wurden redundante Unterstützungssysteme geschaffen, auf welche die Studierenden jederzeit im Verlauf ihrer Arbeit zurückgreifen können. Diese Unterstützungssysteme umfassen beispielsweise Selbstlernkurse, etwa zum Projektmanagement oder zum konkreten Einsatz bestimmter Technologien im Unterricht, sowie technische Bedienungsanleitungen in Form von Steckbriefen zu den einzelnen Technologien im Makerspace. Weitere Unterstützungssysteme stellen studentische Tutor:innen dar, welche (1) die Studierendengruppen in der technischen Umsetzung ihrer Projekte bedarfsorientiert begleiten und (2) Studierende wie Lehrende in mediendidaktischen Fragen beraten.

3.3 Kompetenzraster

Um den Lehrenden in den Fächern eine Grundlage für die Implementation digitalisierungsbezogener Kompetenzen an die Hand zu geben, wurde im Rahmen des Projekts ein Kompetenzraster zu digitalisierungsbezogenen Kompetenzen von Lehramtsstudierenden entwickelt, welches Studierenden und Lehrenden a) als konkreter Orientierungsrahmen zur Definition von Anforderungen und Elementen der Medienprojekte, b) als Grundlage für angebotene Lernaktivitäten im pädagogischen Makerspace, c) als Ausgangspunkt für die gezielte Entwicklung von Selbstlernmaterialien, die im pädagogischen Makerspace zur Verfügung gestellt werden, d) als Rahmen für die Entwicklung eines Self-Assessments für Studierende sowie e) als Grundlage zur Evaluation der Effektivität des Projektansatzes, dient.

Um eine internationale und nationale Anschlussfähigkeit zu garantieren wurde dabei auf den europäischen Rahmenplan DigCompEdu [RP17] aufgesetzt und für die sechs Bereiche

des Rahmenmodells jeweils ca. 10 Kompetenzen auf 3 verschiedenen Kompetenzniveaus (Kennen, Anwenden, Erzeugen) formuliert. Die Formulierungen der einzelnen Kompetenzen wurden in einer Fokusgruppe unter Einbeziehung von Fachdidaktiken vorgenommen und von ausgewählten Expert:innen überprüft. Bei der Auswahl und der Beschreibung der Kompetenzen wurde dabei auf geeignete Modelle und Kompetenzbeschreibungen existierender Modelle (z.B. [BW18]) zurückgegriffen. Tab. 1 zeigt ein Beispiel für die Formulierung der Kompetenzen.

(Bereich: Digitale Ressourcen) Ich kann meine digitalen Lehr-/Lernmaterialien sowie lizenzierten Open Educational Resources (OER) Materialien mit anderen Lehrenden teilen.		
Niveaustufe 1: Kennen	Niveaustufe 2: Anwenden	Niveaustufe 3: Erzeugen
Ich kann erklären, was OER sind, wie diese lizenziert werden (CC) und kann verschiedene OER-Plattformen benennen. Operatoren: Aufzählen, wiedergeben	Ich kann digitale Lehr-/Lernmaterialien aus verschiedenen OER-Plattformen zusammenstellen (auswählen) und für meinen Unterricht bzw. das eigene Lehrkonzept nutzen. Operatoren: Anwenden, übersetzen	Ich kann eigene OER-Materialien produzieren und lizenzieren (unter Berücksichtigung von Datenschutz und Copyright) und anderen online zur Verfügung stellen. Operatoren: Ausarbeiten

Tab. 1: Beispiel einer Formulierung im Kompetenzraster.

3.4 Umgang mit Hürden im Change Prozess

Ein so umfassender Veränderungsprozess des Studiums, der in verpflichtenden Vorgaben für die Änderung von Modulhandbüchern der Lehramtsstudiengänge mündet, läuft Gefahr auf Widerstände zu treffen, da hierfür potenziell fachliche Inhalte geopfert werden müssen. Zudem stellen die Integration von Inhalten zu digitalen Medien und Kompetenzen, aber auch die bislang geringen Erfahrungen mit Formen des projektbasierten Lernens für viele Lehrende Herausforderungen dar, die sich in der Praxis in Widerständen deutlich machen. Um diese Widerstände möglichst gering zu halten, wurden a) die Anforderungen an die Implementation der neuen Kompetenzen konsequent über eine so genannte Learner Journey aus der Perspektive der Studierenden heraus definiert, b) die Lehrenden der Fächer über Change Manager:innen durchgängig in die inhaltliche Ausgestaltung des Prozesses mit einbezogen und potentielle Lösungen für die Integration der Kompetenzen iterativ optimiert, c) Container in Form von ECTS geschaffen (in der Summe zwischen 5 und 8 ECTS für beide Projekte), in denen die Projekte ausgebracht werden können, d) ein gemeinsames Commitment der Hochschule für den Veränderungsprozess auch dadurch hergestellt, dass Rektorat und Fakultäten zeitlich und inhaltlich eng abgestimmt den Prozess begleiteten, e) mit dem zuständigen Kultusministerium eine Ausnahmeregelung entwickelt, die überhaupt erst die Erweiterung der Kompetenzen im Lehramtsstudium ermöglichte, f) Projekte als freieres hochschuldidaktisches Element in das Lehramtsstudium eingeführt und g) der

pädagogische Makerspace als zentraler Lernort der Hochschule mit personell unterfütterten Supportstrukturen nicht nur für Studierende, sondern auch für Lehrende ausgebaut.

4 Beispiele

Mittlerweile wurde bereits eine Vielzahl studentischer Projekte abgeschlossen. Produkte reichen von der Digitalisierung einzelner Präparate und Strukturmodelle in der Biologie mit hilfe von 3D-Scanning und Photogrammetrie, über zahlreiche Erklärvideos, Podcasts, Automatisierungen von Steuerungsprozessen bis hin zu vollständigen VR-Lehr-Lernumgebungen. Teilweise verbleiben die Projekte auf der Ebene von Prototypen, zum Teil sind die Produkte jedoch so ausgereift, dass sie beispielsweise Eingang in Lehrerbildungsplattformen finden oder auch in Unterrichtszeitschriften oder -reihen veröffentlicht werden können.

5 Zusammenfassung und Ausblick

Der in diesem Beitrag beschriebene Ansatz zur Integration verpflichtender curricularer Elemente zur Förderung digitaler Kompetenzen im Lehramtsstudium an einer Hochschule konnte erfolgreich mit den notwendigen curricularen Änderungen an der Hochschule umgesetzt werden. Trotz der vergleichsweise geringen Änderungen für einzelne Fächer in Form der Einführung eines Projektmoduls mussten Widerstände überwunden werden. Die aktive Begleitung des notwendigen Change Prozesses von Beginn an erwies sich hier als ein Schlüsselelement. Der aufgebaute pädagogische Makerspace wird inzwischen an der Hochschule als wertvolle Innovation wahrgenommen, die in Verbindung mit den entwickelten Unterstützungsangeboten für Studierende und Lehrende einen deutlichen Beitrag zur Digitalisierung der Hochschule leistet. Eine umfassende Evaluation, inwieweit der beschriebene Ansatz medien(fach-)didaktische Kompetenzen aller Lehramtsstudierenden nachhaltig stärken kann, konnte bislang nicht umgesetzt werden. Eine Evaluation ist ab dem Wintersemester 2024/2025 geplant. Ergebnisse kleinerer Lerngruppen während des Aufbaus der Strukturen sowie die umgesetzten Beispielprojekte sind jedoch vielversprechend. Mit der Umsetzung der Medienprojekte ist der Change Prozess an der Hochschule nicht abgeschlossen. Zunächst sollen einzelne Elemente der Unterstützungssysteme wie das Online-Self-Assessment oder die tutorielle Lernbegleitung evaluiert und iterativ weiterentwickelt werden. Geplant ist zudem, die Ergebnisse der Studierenden über maschinelles Lernen teilautomatisiert auszuwerten, um adaptives Feedback zu ermöglichen. Hieran wird derzeit in einem Folgeprojekt gearbeitet. Zudem ist die Übertragbarkeit der universitär erworbenen medien(fach-)didaktischen Kompetenzen in den Schulalltag bislang noch wenig erforscht. Eine entsprechende Untersuchung wurde im Rahmen des Projekts begonnen und soll zukünftig ausgebaut werden, um daraus wiederum Rückschlüsse auf Veränderungsbedarfe der Projekte zu generieren (z.B. [ASW24]). Darüber hinaus arbeiten wir daran, Erkenntnisse, die wir im Projekt über erfolgversprechende Ansätze der Vermittlung medien(fach-)didaktischer Kompetenzen erhalten haben, auch auf die folgenden Phasen der Lehrkräftebildung zu übertragen.

Literaturverzeichnis

[ASW24] Aumann, A.; Schnebel, S.; Weitzel, H.: Teaching Biology Lessons Using Digital Technology: A Contextualized Mixed-Methods Study on Pre-Service Biology Teachers' Enacted TPACK. Education Sciences 14 (5), S. 538, 2024, DOI: 10.3390/educsci14050538.

[Be17] Bevan, B.: The promise and the promises of Making in science education. Studies in Science Education 53 (1), S. 75–103, 2017, DOI: 10.1080/03057267.2016.1275380.

[Be20] Becker, S.; Bruckermann, T.; Finger, A.; Huwer, J.; Kremser, E.; Meier, M.; Thoms, L.-J.; Thyssen, C.; von Kotzebue, L.: Orientierungsrahmen Digitale Kompetenzen für das Lehramt in den Naturwissenschaften – DiKoLAN. Joachim Herz Stiftung, 2020.

[BW18] Brandhofer, G.; Wiesner, C.: Medienbildung im Kontext der Digitalisierung: Ein integratives Modell für digitale Kompetenzen. Online Journal for Research and Education 1 (10), 2018.

[Co17] Cohen, J.: Maker Principles and Technologies in Teacher Education: A National Survey. Journal of Technology and Teacher Education 25 (1), S. 5–30, 2017.

[DR85] Deci, E. L.; Ryan, R. M.: Intrinsic Motivation and Self-Determination in Human Behavior. Springer US, 1985.

[Gh20] Ghomi, M.; Dictus, C.; Pinkwart, N.; Tiemann, R.: DigCompEduMINT: Digitale Kompetenz von MINT-Lehrkräften. Kölner Online Journal für Lehrer*innenbildung 1 (1), S. 1–22, 2020, DOI: 10.18716/OJS/KON/2020.1.1.

[JL19] Jaatinen, J.; Lindfors, E.: Makerspaces for Pedagogical Innovation Processes: How Finnish Comprehensive Schools Create Space for Makers. Design and Technology Education: An International Journal 24 (2), S. 42–66, 2019.

[Jo20] Jones, W. M.; Caratachea, M.; Schad, M.; Cohen, J. D.: Examining K–12 teacher learning in a makerspace through the activity–identity–community framework. Journal of Research on Technology in Education 53 (3), S. 317–332, 2020, DOI: 10.1080/15391523.2020.1774824.

[KKF14] Kurti, R. S.; Kurti, D. L.; Fleming, L.: The philosophy of educational makerspaces part 1 of making an educational makerspace. Teacher Librarian 41 (5), S. 8–11, 2014.

[KKO20] Kajamaa, A.; Kumpulainen, K.; Olkinuora, H.: Teacher interventions in students' collaborative work in a technology-rich educational makerspace. British Journal of Educational Technology 51 (2), S. 371–386, 2020.

[KPZ21] Konstantinou, D.; Parmaxi, A.; Zaphiris, P.: Mapping research directions on makerspaces in education. Educational Media International 58 (3), S. 223–247, 2021, DOI: 10.1080/09523987.2021.1976826.

[Ku17] Kultusministerkonferenz (KMK): Bildung in der digitalen Welt - Strategie der Kultusministerkonferenz. Sekretariat der KMK, Berlin, 2017.

[Ku21] Kultusministerkonferenz (KMK): Lehren und Lernen in der digitalen Welt—Die ergänzende Empfehlung zur Strategie 'Bildung in der digitalen Welt'. Sekretariat der KMK, Berlin, 2021.

[Me21] Mersand, S.: The State of Makerspace Research: A Review of the Literature. TechTrends 65 (2), S. 174–186, 2021, DOI: 10.1007/s11528-020-00566-5.

[me21] mebis-Redaktion: DigCompEdu Bavaria – Digitale und medienbezogene Lehrkompetenzen, 2021, URL: https://mebis.bycs.de/beitrag/digcompedu-bavaria, Stand: 13. 01. 2024.

[MK06] Mishra, P.; Koehler, M. J.: Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. Teachers College Record 108 (6), S. 1017–1054, 2006.

[MWL23] Max, A.-L.; Weitzel, H.; Lukas, S.: Factors influencing the development of pre-service science teachers' technological pedagogical content knowledge in a pedagogical makerspace. Frontiers in Education 8, S. 1166018, 2023, DOI: 10.3389/feduc.2023.1166018.

[PDP18] Petko, D.; Döbeli Honegger, B.; Prasse, D.: Digitale Transformation in Bildung und Schule: Facetten, Entwicklungslinien und Herausforderungen für die Lehrerinnen- und Lehrerbildung. Beiträge zur Lehrerinnen- und Lehrerbildung 36 (2), S. 157–174, 2018.

[RP17] Redecker, C.; Punie, Y.: European Framework for the Digital Competence of Educators—DigCompEdu. Publications Office of the European Union, 2017.

[RS21] Rayna, T.; Striukova, L.: Fostering skills for the 21st century: The role of Fab labs and makerspaces. Technological Forecasting and Social Change 164, S. 120391, 2021, DOI: 10.1016/j.techfore.2020.120391.

[UN18] UNESCO: ICT Competency Framework for Teachers, 2018.

[Va04] Vansteenkiste, M.; Simons, J.; Lens, W.; Sheldon, K. M.; Deci, E. L.: Motivating Learning, Performance, and Persistence: The Synergistic Effects of Intrinsic Goal Contents and Autonomy-Supportive Contexts. Journal of Personality and Social Psychology 87 (2), S. 246–260, 2004, DOI: 10.1037/0022-3514.87.2.246.

[Va17] Van Laar, E.; Van Deursen, A. J. A. M.; Van Dijk, J. A. G. M.; De Haan, J.: The relation between 21st-century skills and digital skills: A systematic literature review. Computers in Human Behavior 72, S. 577–588, 2017, DOI: 10.1016/j.chb.2017.03.010.

[Va20] Valtonen, T.; Leppänen, U.; Hyypiä, M.; Sointu, E.; Smits, A.; Tondeur, J.: Fresh perspectives on TPACK: Pre-service teachers' own appraisal of their challenging and confident TPACK areas. Education and Information Technologies 25 (4), S. 2823–2842, 2020, DOI: 10.1007/s10639-019-10092-4.

VR/AR und immersives Lernen I

Phasen der Problemlösekompetenz bei der Korrektur elektrischer Schaltkreise in einer VR-Lernumgebung



Axel Wiepke [1], Alexander Hacke [2], Christian Manz [1] und Ulrike Lucke [1]

Abstract: Die Herausforderungen im alltäglichen Leben werden zunehmend durch mehr Komplexität und Dynamik bestimmt. Dadurch ist es für Lernende seltener möglich, routinierte Vorgehensweisen zur Überwindung auftretender kognitiver Barrieren zu nutzen. In solchen Fällen sind Problemlösekompetenzen gefragt, die meist auf Interaktion mit dem Problem beruhen, was Stift-und-Papier-Studien erschwert. Digitale Lernanwendungen ermöglichen es, Prozessdaten in interagierbaren Szenarien zu erfassen, können aber auf Kosten der ökologischen Validität gehen. Die immersive Eigenschaft von Virtual Reality (VR)-Lernanwendungen kann dabei ein möglicher Faktor zur Verbesserung der ökologischen Validität sein. Der vorliegende Beitrag analysiert eine VR-Lernanwendung hinsichtlich des Gefühls in der VR zu sein. In der Anwendung werden teilweise fehlerhafte, elektrische Schaltkreise präsentiert, die Lernende diagnostizieren und korrigieren sollen. Durch einen virtuellen Fragebogen werden Daten in den verschiedenen Phasen des Problemlöseprozesses erfasst. Die Ergebnisse zeigen ein kontinuierliches, befriedigendes Präsenzerleben.

Keywords: Virtual Reality, Problemlösekompetenz, Präsenz

1 VR als Lösung für Problemlösen

Die fortschreitende Digitalisierung prägt neben vielen anderen Bereichen des modernen Lebens zunehmend auch die Bildungslandschaft. So werden Aufgaben, die im gleichbleibenden Abarbeiten von Routinen bestehen, seltener, während häufiger neuartige oder komplexe Probleme bearbeitet und gelöst werden müssen [OE14]. Während die Bedeutsamkeit von Faktenwissen und Kompetenzen im automatisierbaren Ausführen von Algorithmen abnimmt, werden ausgeprägte Kompetenzen verlangt, die Schüler*innen dazu befähigen, Situationen zu bewältigen, für die nicht von vornherein ein eindeutiges Lösungsschema vorliegt [OE14]. Liegen dabei kognitive Barrieren vor, die Lernende an einer direkten Lösung durch routinierte Vorgehensweisen hindern, werden sogenannte Problemlösekompetenzen benötigt [Le12]. Dabei basiert das Problemlösen in realen Situationen meist auf Interaktion mit dem Problem [Ki20]. Die häufigen Veränderungen der Umgebungsvariablen, wie Positionen der Materialien, Zeitpunkte der Interaktionen und Interventionsmöglichkeiten der Lehrenden erschweren Studien, die auf Stift und Papier beruhen. Durch die Nutzung digitaler Lernanwendungen kann jedoch ein interaktives bzw. dynamisches Problemlösen

1 Universität Potsdam, Komplexe Multimediale Anwendungsarchitekturen, An der Bahn 2, 14476 Potsdam, Brandenburg, axel.wiepke@uni-potsdam.de, https://orcid.org/0000-0002-0555-4040; christian.manz@alumni.uni-potsdam.de, https://orcid.org/0009-0000-6257-3849; ulrike.lucke@uni-potsdam.de, https://orcid.org/0000-0003-4049-8088
2 Universität Potsdam, Didaktik der Informatik, An der Bahn 2, 14476 Potsdam, Brandenburg, alexander.hacke@uni-potsdam.de, https://orcid.org/0000-0002-2124-7956

doi:10.18420/delfi2024_18

für die Lernenden in einem realitätsnahen und interaktiven Problemszenario ermöglicht werden, bei gleichzeitiger Datenerhebung in Echtzeit (ebd.).

Doch nicht alle digitalen Lernanwendungen sind gleich gut geeignet zur Darstellung spezifischer Probleme. Probleme, die sich in der psychomotorischen oder affektiven Domäne befinden, werden z.B. in Desktopanwendungen oft nur unzureichend adressiert. Das Einüben von Bewegungsabläufen kann hier nicht authentisch geschehen, und es besteht eine gewisse emotionale und physische Distanz zur Situation. Diese Herausforderungen können von VR-Lernanwendungen, durch deren immersiven Charakter adressiert werden [Ar21]. So kann mit dem Eintauchen in das Medium einhergehen, dass Nutzende sich als Akteure in der virtuellen Umgebung wahrnehmen, mit der sie direkt interagieren können [Wi07]. Das Wahrnehmen dieser Interaktionsmöglichkeiten als potenzielle Handlungen im Prozess des Problemlösens ist bei dynamischen Problemen eine Vorbedingung des Problemlösens, da die benötigten Informationen hier oft erst durch Interaktion mit der Problemsituation verfügbar werden.

Solche und ähnliche Potenziale von VR werden seit längerer Zeit diskutiert [Dö19]. Die Möglichkeiten von VR als authentische Umgebung, um Lernprozesse zu beforschen, sind jedoch unzureichend beleuchtet. Dieser Beitrag stellt eine VR-Umgebung vor, die entwickelt wurde, um Phasen von Problemlöseprozessen untersuchbar zu machen. Dafür werden Arbeitsdefinitionen für Problemlösen, VR, und Präsenz bereitgestellt, und die entwickelte Anwendung „Spicy“ (Solve problems in circuits yourself [Ma23]) präsentiert. Danach wird eine Methodik beschrieben, mit der in Spicy Phasen des Problemlöseprozesses hinsichtlich der erfahrenen Präsenz untersucht wurden. Aus der zugrundeliegenden Studie werden dann Potenziale und Empfehlungen zur Untersuchung von Lernprozessen abgeleitet.

2 Phasen des Problemlösens

Unter einem *Problem* wird eine Situation verstanden, in der ein unerwünschter Anfangszustand unter der Überwindung einer kognitiven Barriere in einen Zielzustand überführt werden muss [Fl10]. Die kognitive Barriere besteht dabei in der Regel darin, dass Problemlösenden zu Beginn keine eindeutige und ihnen bekannte Lösungsroutine zur Verfügung steht. Traditionell wird Problemlösen spätestens seit Pólya (abud. [Pó45]) dabei als ein Prozess betrachtet, wobei in einer ersten Phase das Problem zunächst verstanden werden muss, um zu erkennen, welche Informationen bereits vorliegen und einzugrenzen, was gefunden oder gelöst werden muss. In der darauffolgenden Phase entwirft der bzw. die Problemlösende eine Strategie, um das Problem anzugehen. Diese kann beispielsweise beinhalten, dass Lösungsstrategien von ähnlich gearteten Problemen in Betracht gezogen bzw. adaptiert werden oder aber das Problem in leichter handhabbare Teilprobleme zerlegt wird. Im nachfolgenden Schritt wird die vorab überlegte Strategie ausgeführt und sich somit systematisch zum gewünschten Zielzustand vorgearbeitet. Ist das Ziel erreicht, prüft der oder die Problemlösende idealerweise in einem letzten Schritt die Lösung auf Korrektheit bzw. Effektivität. Diese normative Darstellung des Problemlöseprozesses veranschaulicht zwar,

welche Schritte das Problemlösen beinhaltet, lässt allerdings die nicht lineare, in Teilen auch iterative Art und Weise, wie Menschen Probleme lösen, außer Betracht. Spätere Modelle u.a. jenes von Wilson et al. [Fe94] berücksichtigen diesen Aspekt, indem sie Rückgriffe auf vorherige Phasen basierend auf dem Ergebnis oder beim Laufen in Sackgassen beinhalten. Ferner macht erfolgreiches Problemlösen nicht nur das Wissen und die erfolgreiche Anwendung von Strategien aus, sondern hängt auch stark vom eigenen Zutrauen ab und davon, wie gut der oder die Problemlösende in der Lage ist, das eigene Fortkommen zu überwachen und zu regulieren [Sc88].

Probleme lassen sich von Aufgaben dadurch abgrenzen, dass bei Problemen auf keine bekannten Lösungsroutinen zurückgegriffen werden kann, um den Zielzustand zu erreichen [Fu03]. Lösungsroutinen können unter anderem fachspezifisch erlernt werden, wodurch Problemlösende mit einem passenden Hintergrund eine Situation als Aufgabe wahrnehmen können, während andere es als Problem erkennen. Um in Untersuchungen diesen zusätzlichen Einflussfaktor auszuschließen, werden in vielen Studien vorwissensneutrale Probleme gewählt [GF13]. Dies kann jedoch mit Einschränkungen darin einhergehen, wie gut sich die Ergebnisse solcher Studien auf alltägliches Problemlösen übertragen lassen, da hier das Vorwissen oft eine größere Rolle spielt. Durch den zusätzlichen Fokus in der Forschung auf psychometrische Gütekriterien während des Lernens und der hohen Standardisierung der Situationen kann es in Studien zu geringerer ökologischer Validität kommen [GKL14].

Dabei hat Problemlösen als grundlegendes Konstrukt in verschiedenen Forschungsdisziplinen eine ausführlich betrachtete Tradition [Fu06]. Die Breite dieses Konstrukts zeigt sich dabei auch in der Vielfältigkeit der Perspektiven und Fragestellungen, unter denen Problemlösen untersucht wird. Die Entwicklungspsychologie befasst sich mit der Entwicklung problemlösenden Denkens im Lebensverlauf, die Sozialpsychologie mit Problemlöseprozessen in Gruppen, die Differenzialpsychologie mit den Unterschieden im Problemlösen verschiedener Personen, die klinische Psychologie mit pathologischen Störungen problemlösenden Denkens, die Arbeitspsychologie mit Problemlösen beispielsweise in technischen Arbeitsbereichen und die pädagogische Psychologie mit der Frage, ob und wie Problemlösen geschult oder trainiert werden kann [Fu03]. Auch die Konzeption einer digitalen Lernanwendung zu Problemlösen beruht auf der Annahme, dass es sich hier um eine trainierbare Kompetenz handelt.

2.1 So kann Virtual Reality Problemlösen unterstützen

Virtual Reality Lernanwendungen zeichnen sich dadurch aus, dass Prozessdaten aufgezeichnet und exakte Situationen repliziert werden können, wodurch eine sehr hohe experimentelle Kontrolle möglich ist [Ki14; Wi23]. Hohe experimentelle Kontrolle kommt häufig auf Kosten der ökologischen Validität [Ku16]. Es wird jedoch angenommen, dass durch eine realistische Darstellung der VR-Umgebung und das Gefühl sich darin zu befinden (Präsenz) die Transferierbarkeit gesteigert werden kann [Wi24].

Schubert [Sc03] teilt Präsenz in die drei Dimensionen Spatial Presence, Involvement und Realness auf. Hier beschreibt Spatial Presence das Gefühl, vollständig von der virtuellen Umwelt umgeben zu sein und direkt mit ihr interagieren zu können. Laut dem MeC-Model von Wirth et al. [Wi07] wird Spatial Presence in zwei Stufen gebildet. Im ersten Schritt wird von der Nutzerin bzw. dem Nutzer von VR ein mentales Modell der (virtuellen) Umgebung erstellt, wobei hierfür zunächst die Aufmerksamkeit auf der wahrgenommenen Umgebung liegen muss. In diesem Modellbildungsprozess werden die in der Situation wahrgenommenen Informationen verarbeitet, sodass am Ende ein Spatial-Situation-Model (SSM) vorliegt. Das SSM ist Voraussetzung für den zweiten Schritt, aus dem das Erleben von Spatial Presence hervorgehen kann. Hier werden Hypothesen, die aus dem SSM hervorgehen, getestet, wobei geprüft wird, ob die wahrgenommene Situation als der primäre egozentrische Referenzrahmen (engl. primary egocentric reference frame, kurz PERF) gedeutet werden kann. Sorgen die verschiedenen beteiligten personellen und medialen Faktoren dafür, dass eine virtuelle Umgebung als PERF akzeptiert wird, verorten sich Nutzende selbst in der virtuellen Umgebung und identifizieren meist auch verschiedene Möglichkeiten zur Interaktion. Auf diese Weise entsteht schlussendlich Spatial Presence [Wi07]. Involvement beschreibt das Gefühl, die Aufmerksamkeit ganz auf der virtuellen Umgebung anstatt der realen Umgebung fokussiert zu haben, während Realness umfasst, als wie realistisch die virtuelle Umgebung beurteilt wird [Sc03]. Auf diesen drei Dimensionen baut auch der Igroup Presence Questionnaire (IPQ) auf, der mit insgesamt 13 Items die drei beschriebenen Dimensionen und mit einem weiteren Item Präsenz als übergeordnetes Konstrukt erfasst [Sc03; SFR01].

Für den Einsatz von VR als Lernmedium werden unterschiedliche Potenziale und Erwartungen in der Literatur formuliert [ZW18]. Durch die vollständige Kontrolle der VR-Umgebung über das IT-System, können z. B. Just-in-Time Lernhilfen eingeblendet, Aufgabe angepasst oder Reflexionsunterstützungen in Form von Statistiken für Lernende angeboten werden. Der Simulationscharakter und der hohe Realismusgrad von VR-Lernanwendungen unterstützt den situativen Konstruktivismus, der davon ausgeht, dass Aktivitäten in der (simulierten) Situation spezifische Wissensstrukturen erzeugen [BCD89]). Durch die Isolation der Lernenden von der physischen Umgebung wird der wahrnehmbare Raum auf die Lernumgebung eingegrenzt. Diese Isolation kann gemeinsam mit der multimodalen Darstellung des Lerngegenstands zu geringerem aufgaben-unangemessenem cognitive Load führen, und damit das Verstehen und Erinnern für Lernende fördern [Ar21]. Weiterhin ist beim Einsatz von neuartigen Medien wie der VR ein gesteigertes Interesse und eine hohe Motivation der Lernenden zu erwarten [Ar21]. Auch die physische Interaktion mit virtuellen Lerngegenständen kann gemäß dem embodied Cognition Ansatz kognitive Prozesse fördern und das Lernen unterstützen [Wi24].

3 Problemlösen in der VR Lernanwendung Spicy

Im vorliegenden Beitrag wird eine Pilotversion einer VR-Lernanwendung auf wahrgenommene Präsenz während des Problemlöseprozesses untersucht. Präsenz dient dabei

sowohl als Qualitätskriterium der VR-Anwendung als auch als Indikator für die Übertragbarkeit von Problemlösekompetenzen in reale Situationen. In der Anwendung werden elektrische Schaltkreise dargestellt und verschiedene, funktionale Bauteile eines schulischen Steckbaukastens bereitgestellt. Der Steckbaukasten wurde als Kontext gegenüber echten Schaltungen bevorzugt, da er den Zugang zu Grundkonzepten in diesem Bereich erleichtert. Lernende erhalten teilweise fehlerhafte Schaltkreise, die sie diagnostizieren und korrigieren sollen. Um den immersiven Charakter des Mediums bewerten zu können, werden in den verschiedenen Phasen des Problemlöseprozesses Daten zur Spatial Presence erhoben. Da Problemlösen ein komplexes Konstrukt mit zahlreichen Modellen ist, kann nicht jede Perspektive in einer Anwendung berücksichtigt werden. Daher erscheint die Förderung der Problemlösekompetenz durch einen Fokus auf Teilkompetenzen als praktikabler Ansatz. Die Anwendung fokussiert sich auf einzelne Teilkompetenzen, wie das Aufstellen und Prüfen von Hypothesen.

Andere digitale Tools, die den gesamten Problemlöseprozess darstellen, beinhalten keine VR-Komponente (z.B. [PAD23]) oder untersuchen keine realweltlichen Probleme [Ar21]. Durch den Charakter des digitalen Tools müssen auch komplexe Prozesse gegebenenfalls weiter ausdifferenziert werden. Das Auswählen und Prüfen von Hypothesen lässt sich als Teilkompetenz von Problemlösen in die Phase der Strategieentwicklung einordnen.

Zur Implementierung von Spicy als prototypische VR-Lernanwendung für Problemlösekompetenz wurde die 3D-Engine Unity verwendet. Ein Überblick der entwickelten virtuellen Umgebung ist in Abbildung 1 dargestellt. Dort finden sich ein Arbeitstisch mit einem Steckbrett für elektrische Schaltkreise sowie drei Bildschirme, welche die Problemlösenden durch den Problemlöseprozess leiten. Durch den sparsamen Einsatz von Animationen, das Klassenzimmer als gewohnte Umgebung und das bekannte Steckbaukastenbeispiel, wurde der aufgaben-unangemessene cognitive Load auf ein Minimum beschränkt.

Durch Abarbeitung der Anweisungen auf dem Instruktionsbildschirm folgen Problemlösende dem in Abb. 2 skizzierten Problemlöseprozess. Dieser lässt sich mit dem von Pólya ([Pó45]) vergleichen, zieht aber die zyklische Natur von Problemlöseprozessen in Betracht, indem aus verschiedenen Phasen wieder zu einer der vorherigen zurückgesprungen werden kann, sofern gewünscht oder notwendig. Beim Auswählen und Prüfen von Hypothesen werden die Phasen nach Pólya *Entwickeln eines Plans* und *Plan durchführen* jeweils einmal unterteilt. So wurde das Entwickeln des Plans sequentialisiert, um Lernenden zu ermöglichen verschiedene Hypothesen zu generieren, jedoch nur einzeln zu überprüfen. In der Durchführung des Plans beinhaltet der Verlaufsplan von Spicy unterschiedliche Pfade für den Erfolg beim Problemlösen und den Misserfolg. Welches die jeweils aktuelle Phase im Problemlöseprozess ist, ergibt sich aus den bisherigen Interaktionen mit dem Hypothesen-Bildschirm. Wenn der bzw. die Problemlösende beispielsweise eine Hypothese ausgewählt hat, ein Urteil zu einer Hypothese abgibt oder auf den Knopf zur Bestätigung der aktuellen Problemlösung drückt, wird angenommen, dass er bzw. sie nun in die nächste Phase der Problemlösung übergeht. Je nach Phase aktualisieren sich daraufhin diese beiden Bildschirme, sodass auf dem Instruktionsbildschirm eine Anweisung für die neue Phase erscheint und auf dem



Abb. 1: Ausschnitt der virtuellen Umgebung in Spicy. In der Mitte ist ein Tisch zu sehen, an dem Problemlösende mit dem defekten Schaltkreis interagieren können. Auf den drei Bildschirmen werden das gewünschte Verhalten des Schaltkreises definiert, Instruktionen zum nächsten Schritt im Problemlöseprozess vermittelt und ein Interface für das Auswählen von Hypothesen bereitgestellt.

Hypothesen-Bildschirm andere Interaktionsmöglichkeiten aktiviert werden, beispielsweise das Bewerten einer Hypothese anstelle des Auswählens. Eine zusätzliche Schaltfläche ermöglicht es, aus späteren Phasen zur Vorauswahl der Hypothesen zurückzukehren. Die auf diese Weise im Programm möglichen Übergänge zwischen Phasen im Problemlöseprozess sind in Abbildung 2 dargestellt, in welcher erneut die iterative Natur des beschriebenen Problemlöseprozesses deutlich wird. Dadurch, dass Spicy durch die Phasen führt, wird den Problemlösenden eine Art von Scaffolding zuteil, das einerseits sehr unerfahrenen Teilnehmenden eine Art Guideline gibt und andererseits dazu beiträgt, den Fokus auf die spezifischen Phasen zu lenken.

Die Phasen werden dabei wie folgt beschrieben:

1. **Problem verstehen** Der bzw. die Problemlösende erfasst den Aufbau des defekten Schaltkreises sowie den Unterschied des derzeitigen Verhaltens zum gewünschten Verhalten.
2. **Vorauswahl relevanter Hypothesen** Der bzw. die Problemlösende wählt basierend auf seinem bzw. ihrem Vorwissen bis zu drei Optionen aus einer Sammlung relevanter Hypothesen, die mögliche Ursachen für den Defekt im Schaltkreis begründen.
3. **Hypothesen zur Prüfung auswählen** Der bzw. die Problemlösende wählt ausgehend von seinem bzw. ihrem Vorwissen eine Hypothese aus, um sie im Folgenden zu überprüfen. Die Auswahl erfolgt dabei aufgrund dessen, für wie wahrscheinlich das Zutreffen der Hypothese gesehen wird und als wie aufwendig die Überprüfung eingeschätzt wird.

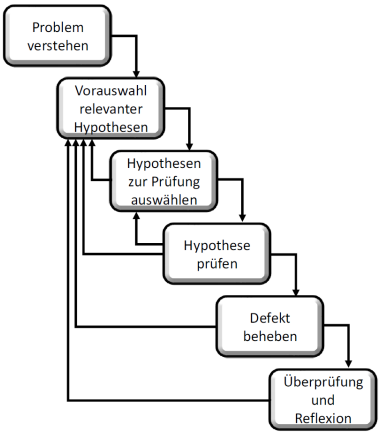


Abb. 2: Ablaufdiagramm zu von Spicy unterstützten Problemlöseprozessen, aufbauend auf Pólya [Pó45]. Die Übergänge erfolgen in der Regel durch die Betätigung von Schaltflächen am Hypothesen-Bildschirm.

4. **Hypothesen prüfen** Die gewählte Hypothese wird überprüft und anschließend dahingehend beurteilt, ob sie die Ursache für den Defekt im Schaltkreis darstellt oder nicht. Zur Überprüfung der Hypothese interagiert der bzw. die Problemlösende mit dem Schaltkreis, beispielsweise durch das Durchführen von Messungen oder das Umordnen und Austauschen von Bauteilen.
5. **Defekt beheben** Ausgehend von den zuvor identifizierten Ursachen behebt der bzw. die Problemlösende den Defekt im Schaltkreis, indem er bzw. sie diesen modifiziert.
6. **Überprüfung und Reflexion** Der bzw. die Problemlösende überprüft, ob die durchgeführte Behebung tatsächlich dafür gesorgt hat, dass der Schaltkreis nun das gewünschte Verhalten zeigt. Dabei wird evaluiert, ob die vorherigen Urteile zur Gültigkeit der überprüften Hypothesen weiterhin als zutreffend angesehen werden können.

Die Schritte der Anwendung korrespondieren mit den vier Phasen des Problemlöseprozesses nach Pólya. Der erste Schritt deckt sich mit dem Verstehen des Problems, während die Schritte zwei und drei das Entwickeln eines Plans darstellen. Die Schritte vier und fünf sind dem Ausführen des Plans zugeordnet. Abschließend spiegelt der sechste Schritt die Reflexion wider, was der letzten Phase in Pólyas Modell entspricht.

4 Demografie, Methodik und Verlauf der empirischen Studie

Mit einer empirischen Pilotstudie wurde die Angemessenheit der implementierten VR-Lernanwendung Spicy als Studienumgebung für das Problemlösen untersucht. Als Gütekriterium für die VR-Lernanwendung wurde Spatial Presence in vier der Phasen des

Problemlöseprozesses erhoben. Phase eins „Problem verstehen “ und zwei „Vorauswahl relevanter Hypothesen “ konnten nicht durch Interaktion mit der VR-Umgebung von den anderen Phasen abgegrenzt werden, wodurch kein Zeitpunkt identifiziert werden konnte, an dem eine Erhebung angebracht wäre. Die deutsche Version des IPQ wurde nach Abschluss der anderen Phase des Problemlöseprozesses als virtueller Fragebogen angezeigt, wobei Fragen durch Antippen auf einer 7-stufigen Likert-Skala beantwortet wurden (siehe Abbildung 3). Die Interaktion mit der virtuellen Umgebung wurde während der Erhebung auf die Beantwortung der virtuellen Fragebögen beschränkt, sodass ein Voranschreiten im Problemlösungsprozess erst nach der Datenerhebung möglich war. Da der in Abbildung 2 dargestellte Prozess durch die Hilfestellungen zum Problemlösen bereits in Spicy modelliert war, wurde die Annahme getroffen, dass sich Problemlösende tatsächlich kognitiv in der jeweiligen Phase des Problemlösens befinden, solange sie in Spicy zur Bearbeitung dieser Phase aufgefordert werden. So sollten die Spatial Presence Werte Aufschluss über die kognitiven Prozesse während der Korrektur der virtuellen, elektrischen Schaltkreise liefern.

Die Studie wurde über drei Tage hinweg mit insgesamt N = 14 Proband*innen im Institut für Informatik der Universität Potsdam durchgeführt. Von den 14 Proband*innen waren 13 im Alter von 18-27 Jahren und 13 Studierende unterschiedlicher Fachrichtungen. Die Proband*innen erhielten keine Aufwandsentschädigung. Spicy wurde auf einer Meta Quest 1 VR-Brille ausgeführt, die neben der Brille selbst über zwei Controller zur Interaktion mit der virtuellen Welt verfügt.

Die Studie begann mit der Erläuterung des Ablaufes. Nach dieser setzten die Proband*innen die VR-Brille auf, auf welcher Spicy bereits im Anfangszustand gestartet war. Dort wurde ein Einführungstext akustisch abgespielt, der die Proband*innen mit der Steuerung in Spicy vertraut machte. Anschließend bearbeiteten die Proband*innen bis zu fünf konsekutiv schwieriger werdende Probleme. Nach spätestens 35 Minuten wurden sie darauf hingewiesen, die erwartete Studiendauer überschritten zu haben. Die Proband*innen verbrachten in der Regel 30-35 Minuten in der VR-Lernanwendung. Proband*innen füllten während jedes Problems mehrere Male den virtuellen IPQ aus. Ein Datensatz musste wegen technischer Probleme während der Studie ausgeschlossen werden. Gegen Ende der Studie schauten viele Proband*innen seltener auf den Instruktionsbildschirm, wodurch sie die Definition des Problems nicht mehr gelesen haben. Stattdessen schienen sie anzunehmen, das Problem bestehe darin, die nicht leuchtenden Glühlampen im defekten Schaltkreis zum Leuchten zu bringen.

5 Ergebnisse der Pilotstudie

Die erhobenen Antworten der Likert-Skalen wurden mit Methoden von metrischen Skalen verarbeitet, wobei in der Interpretation der Ergebnisse auf die ordinale Natur der Daten geachtet wird (nach Empfehlung von [SL16]). Für die Auswertung der aufgenommenen Daten zu Spatial Presence wurden die Antworten zu den fünf Items für Spatial Presence jeweils gemittelt. Um Spatial Presence in den verschiedenen Phasen des Problemlöseprozesses

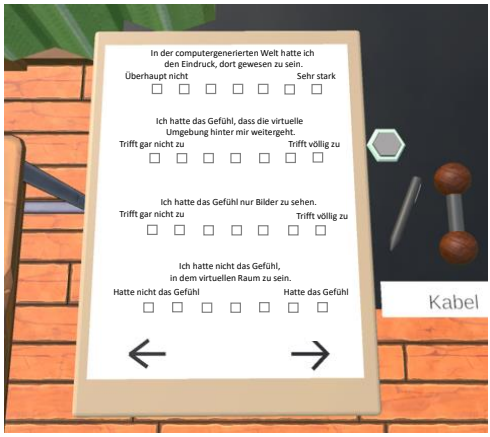


Abb. 3: Einbindung des IPQ [SFR01] als virtueller Fragebogen in Spicy. Die Items wurden auf mehrere Seiten verteilt, um eine bessere Lesbarkeit in VR zu gewährleisten.

auf statistisch signifikante Unterschiede zu überprüfen, wurden paarweise alle vier Phasen, in denen Spatial Presence gemessen wurde, mit einem zweiseitigen t-Test für abhängige Stichproben ausgewertet. Dieser Test wurde gewählt, da die gleichen Proband*innen zu unterschiedlichen Zeitpunkten befragt wurden und die erhobenen Daten optisch nahe einer Normalverteilung waren. Durch die geringe Anzahl an Teilnehmenden (N=14) hätte ein Studiendesign mit mehreren Gruppen eine noch geringere Aussagekraft gehabt. Die ermittelten Werte mit einem Minimum von 0 und einem Maximum von 6 für Spatial Presence in den jeweiligen Phasen können in Abbildung 4a) gesehen werden. Keiner der paarweisen t-Tests zwischen den Phasen zeigte ein Signifikanzniveau von $\alpha < 0.05$. Die für die einzelnen Phasen gemittelten Werte für Spatial Presence liegen zwischen 4,37 „Defekt beheben “ und 4,59 „Hypothesen zur Prüfung auswählen “. Zur Einordnung dieser Werte lässt sich die qualitative Skala für Spatial Presence von Melo et al. [Me23] verwenden, die in Abbildung 4b) dargestellt ist. Dort liegen die ermittelten Werte an der Grenze der Bewertungen *Satisfactory* und *Marginal*.

6 Diskussion

Auch wenn die ermittelten Werte von Spatial Presence insgesamt ein zufriedenstellendes Niveau erreichen konnten, konnten keine signifikanten Unterschiede zwischen den verschiedenen Phasen des Problemlöseprozesses ermittelt werden.

Mögliche Erklärungsansätze für diesen Befund bestehen darin, dass sich die unterschiedlichen kognitiven und handlungsbezogenen Anforderungen, die für die verschiedenen Phasen des skizzierten Problemlöseprozesses angenommen werden, nicht oder nur schwach auf das Erleben von Spatial Presence auswirken. Vor dem Hintergrund des MeC-Modells zur

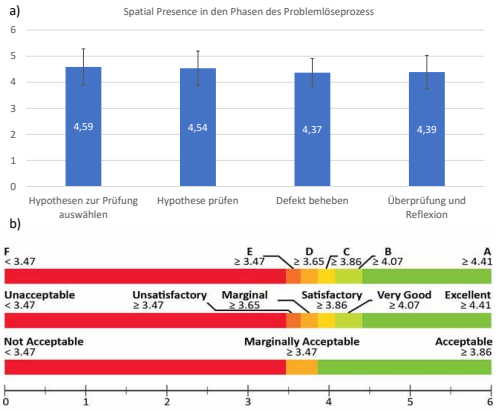


Abb. 4: Die Erhebung der Präsenz zu verschiedenen Zeitpunkten im Problemlöseprozesses resultiert in nicht signifikanten Unterschieden. Die Ergebnisse sind über alle Phasen hinweg befriedigend, gemäß [Me23].

Bildung von Spatial Presence kann zudem argumentiert werden, dass VR-Nutzende während der teilweise kurzen Dauer einer einzelnen Phase keine Neubewertung dessen vornehmen, ob sie die virtuelle Umgebung als primären egozentrische Referenzrahmen (PERF) deuten sollten. Falls die Bewertung der aus dem Spatial Simulation Model (SSM) hervorgehenden Hypothesen nicht erneut geprüft wird, würde dieses Modell somit nicht dafürsprechen, dass sich die erlebte Spatial Presence ändert.

Für die beschriebene Art der Operationalisierung lässt sich ein hohes Maß an Objektivität erwarten, da sie auf durch die Software festgelegten Regeln beruht und keine Urteile durch die Studienleitung benötigt. Sie könnte jedoch eine Schwachstelle bezüglich der Validität darstellen, falls sich Proband*innen nicht von den gegebenen Instruktionen im Problemlöseprozess leiten lassen, sondern die Problemlösung ohne kontinuierliche Interaktion mit dem Hypothesen-Bildschirm durchführen. Zudem konnte beobachtet werden, dass die Phase 6 „Überprüfung und Reflexion “ oft in zeitlich nur sehr geringem Umfang bearbeitet wurde. So finalisierten die Proband*innen die Bewertungen ihrer Hypothesen oft innerhalb weniger Sekunden. Zudem wurde das anschließende Feedback zu den Urteilen von vielen Teilnehmer*innen kaum zur Kenntnis genommen, wenn in dieser Phase ein Fragebogen ausgefüllt werden musste. In diesen Fällen schien die derzeitige Version von Spicy die Proband*innen nicht erfolgreich zu einer ausführlichen Reflexion anhand des Feedbacks anleiten zu können.

6.1 Potenziale für Forschung mit VR am Problemlösen

Basierend auf den vorliegenden Erkenntnissen wird deutlich, dass die Anwendung zur Messung der Spatial Presence mit dem gegenwärtig verwendeten normativen Problemlösemodell nur der Initialschritt auf dem Weg zu weiteren Untersuchungen ist. Die zweite Phase im Problemlöseprozess nach Pólya, in der die eigentliche Lösungsstrategie entwickelt wird [Pó45], wurde durch die Vorgabe von Hypothesen in Spicy nur teilweise repräsentiert. Obgleich ein lineares oder zyklisches Durchlaufen des Problemlöseprozesses dadurch nicht ausgeschlossen wird und die Problemlösenden nach Wahl einer Hypothese ihre konkrete Herangehensweise selbst organisieren müssen, nimmt die Vorgabe eine wichtige initiale Komponente der strategischen Denkphase vorweg.

Eine zweite Version der Anwendung müsste in der Lage sein, dieses Scaffolding nach und nach zu reduzieren, um Einblicke in die metakognitiven Fähigkeiten der Problemlösenden bei der Planung und Überlegung zu ermöglichen. In zukünftiger Forschung könnte man die Teilnehmenden ihr Vorgehen selbst bestimmen lassen, indem die Problemlöseprozesse anhand eines deskriptiven Modells wie dem von Rott [RSK21] aufzeichnet werden, anstatt sie vorzugeben. Es ließen sich dann Aussagen darüber machen, ob die Teilnehmenden bereits einen (Teil-)Lösungsweg im Hinterkopf haben, wie viel Zeit in den einzelnen Phasen verbracht wird und wie viele Wechsel zwischen unterschiedlichen Phasen vorgenommen werden. Hierfür wäre eine verbesserte Operationalisierung der Phasen notwendig, um sie deutlicher voneinander abzugrenzen und möglicherweise auch automatisiert erfassen zu können. In VR könnten beispielsweise gezielt Objekte, die phasenrelevant sind, im virtuellen Raum positioniert werden, um Bewegungen zwischen diesen Orten als Phasenwechsel zu dokumentieren. Am Beispiel von Spicy könnte etwa eine Glühlampe, die zum Leuchten gebracht werden soll, hinter den Lernenden platziert werden, damit durch ein Umdrehen erkennbar ist, dass gerade eine Ergebnisprüfung stattfindet. Weiterführende Untersuchungen könnten auch klären, ob sich Problemlöseprozesse in VR grundsätzlich von jenen mit Papier und Stift unterscheiden, beispielsweise in der Art der Fehler oder der Anzahl von Zyklen im Problemlösevorgang.

Letztlich könnte eine darauffolgende dritte Version der Anwendung, Problemlösende anhand gesammelter Forschungsdaten gezielt in ihrem individuellen Vorgehen unterstützen, indem beispielsweise durch die Erkennung von mehrfach auftretenden Zyklen ohne Fortschritt darauf hingewiesen wird, sich neu zu orientieren oder eine Neubewertung der aktuellen Lösungsstrategie vorzunehmen. Aus der Forschungsperspektive bietet das Problemlösen in VR den Forschenden die neuartige Möglichkeit, den Problemlöseprozess aus der Perspektive der Problemlösenden in authentischen (Problem-)Situationen zu beobachten und zu analysieren. Diese direkte Einsicht erlaubt es, präzise Daten über die Entscheidungsfindung und die Interaktion mit virtuellen Objekten zu sammeln, was zu tieferen Einsichten in kognitive Prozesse führen könnte.

Literaturverzeichnis

[Ar21] Araiza-Alba, P.; Keane, T.; Chen, W. S.; Kaufman, J.: Immersive virtual reality as a tool to learn problem-solving skills. en, Computers & Education 164, S. 104121, 2021, ISSN: 0360-1315, DOI: 10.1016/j.compedu.2020.104121, URL: https://www.sciencedirect.com/science/article/pii/S0360131520303195, Stand: 02. 08. 2023.

[BCD89] Brown, J. S.; Collins, A.; Duguid, P.: Situated Cognition and the Culture of Learning. en, Educational Researcher 18 (1), Publisher: American Educational Research Association, S. 32–42, 1989, ISSN: 0013-189X, DOI: 10.3102/0013189X018001032, URL: https://doi.org/10.3102/0013189X018001032, Stand: 10. 04. 2024.

[Dö19] Dörner, R.; Broll, W.; Grimm, P.; Jung, B., Hrsg.: Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität. Springer Berlin Heidelberg, Berlin, Heidelberg, 2019, ISBN: 978-3-662-58861-1.

[Fe94] Fernandez, M. L.; Hadaway, N.; Wilson, J. W.; Graeber, A. O.: Connecting Research to Teaching: Problem Solving: Managing It All. EN, The Mathematics Teacher 87 (3), Publisher: National Council of Teachers of Mathematics Section: The Mathematics Teacher, S. 195–199, 1994, ISSN: 0025-5769, 2330-0582, DOI: 10.5951/MT.87.3.0195, URL: https://pubs.nctm.org/view/journals/mt/87/3/article-p195.xml, Stand: 11. 04. 2024.

[Fl10] Fleischer, J.; Wirth, J.; Rumann, S.; Leutner, D.: Strukturen fächerübergreifender und fachlicher Problemlösekompetenz. Analyse von Aufgabenprofilen. Projekt Problemlösen. In: Zeitschrift für Pädagogik. Bd. Beiheft; 56, ISSN: 0514-2717, pedocs, 2010, ISBN: 978-3-407-41157-0, URL: http://nbn-resolving.de/urn:nbn:de:0111-opus-34329, Stand: 10. 04. 2024.

[Fu03] Funke, J.: Problemlösendes Denken. Google-Books-ID: bD94DwAAQBAJ, Kohlhammer Verlag, 2003, ISBN: 978-3-17-022830-6.

[Fu06] Funke, J.: Denken und Problemlösen. Google-Books-ID: DqCbEAAAQBAJ, Hogrefe Verlag GmbH & Company KG, 2006, ISBN: 978-3-8409-0527-8.

[GF13] Greiff, S.; Fischer, A.: Der Nutzen einer komplexen Problemlösekompetenz: Theoretische Überlegungen und empirische Befunde. Zeitschrift für Pädagogische Psychologie 27 (1-2), Publisher: Hogrefe AG, S. 27–39, 2013, ISSN: 1010-0652, DOI: 10.1024/1010-0652/a000086, URL: https://econtent.hogrefe.com/doi/10.1024/1010-0652/a000086, Stand: 10. 04. 2024.

[GKL14] Greiff, S.; Kretzschmar, A.; Leutner, D.: Problemlösen in der Pädagogischen Psychologie. Zeitschrift für Pädagogische Psychologie 28 (4), Publisher: Hogrefe AG, S. 161–166, 2014, ISSN: 1010-0652, DOI: 10.1024/1010-0652/a000140, URL: https://econtent.hogrefe.com/doi/10.1024/1010-0652/a000140, Stand: 10. 04. 2024.

[Ki14] Kinateder, M.; Ronchi, E.; Nilsson, D.; Kobes, M.; Müller, M.; Pauli, P.; Mühlberger, A.: Virtual Reality for Fire Evacuation Research. In. S. 313–321, 2014, DOI: 10.15439/2014F94, URL: https://fedcsis.org/proceedings/2014/drp/94.html, Stand: 23. 01. 2023.

[Ki20] Kipman, U.: Problemlösen bei PISA. In (Kipman, U., Hrsg.): Komplexes Problemlösen: Begriff – Einflussgrößen – Korrelate – Erkenntnisse am Beispiel der PISA-Studie. Springer Fachmedien, Wiesbaden, S. 9–31, 2020, ISBN: 978-3-658-30826-1, DOI: 10.1007/978-3-658-30826-1_3, URL: https://doi.org/10.1007/978-3-658-30826-1_3, Stand: 27. 03. 2024.

[Ku16] Kubbe, I.: Gütekriterien experimenteller Forschung. In (Kubbe, I., Hrsg.): Experimente in der Politikwissenschaft: Eine methodische Einführung. Springer Fachmedien, Wiesbaden, S. 61–76, 2016, ISBN: 978-3-658-09424-9, DOI: 10.1007/978-3-658-09424-9_4, URL: https://doi.org/10.1007/978-3-658-09424-9_4, Stand: 10. 04. 2024.

[Le12] Leutner, D.; Fleischer, J.; Wirth, J.; Greiff, S.; Funke, J.: Analytische und dynamische Problemlösekompetenz im Lichte internationaler Schulleistungsvergleichsstudien. Psychologische Rundschau 63 (1), Publisher: Hogrefe Verlag, S. 34–42, 2012, ISSN: 0033-3042, DOI: 10.1026/0033-3042/a000108, URL: https://econtent.hogrefe.com/doi/10.1026/0033-3042/a000108, Stand: 02. 08. 2023.

[Ma23] Manz, C., 2023, URL: https://github.com/DancingAsteroid/Spicy, Stand: 12. 06. 2024.

[Me23] Melo, M.; Gonçalves, G.; Vasconcelos-Raposo, j.; Bessa, M.: How Much Presence is Enough? Qualitative Scales for Interpreting the Igroup Presence Questionnaire Score. IEEE Access 11, Conference Name: IEEE Access, S. 24675–24685, 2023, ISSN: 2169-3536, DOI: 10.1109/ACCESS.2023.3254892, URL: https://ieeexplore.ieee.org/document/10064261, Stand: 09. 04. 2024.

[OE14] OECD: PISA 2012 Results: Creative Problem Solving (Volume V): Students' Skills in Tackling Real-Life Problems. OECD, 2014, ISBN: 978-92-64-20807-0.

[PAD23] Pechorina, Y.; Anderson, K.; Denny, P.: Metacodenition: Scaffolding the Problem-Solving Process for Novice Programmers. In: Proceedings of the 25th Australasian Computing Education Conference. ACE '23, Association for Computing Machinery, New York, NY, USA, S. 59–68, 2023, ISBN: 978-1-4503-9941-8, DOI: 10.1145/3576123.3576130, URL: https://dl.acm.org/doi/10.1145/3576123.3576130, Stand: 10. 04. 2024.

[Pó45] Pólya, G.: How to Solve It. Princeton University Press, 1945.

[RSK21] Rott, B.; Specht, B.; Knipping, C.: A descriptive phase model of problem-solving processes. en, ZDM – Mathematics Education 53 (4), S. 737–752, 2021, ISSN: 1863-9704, DOI: 10.1007/s11858-021-01244-3, URL: https://doi.org/10.1007/s11858-021-01244-3, Stand: 12. 04. 2024.

[Sc03] Schubert, T. W.: The sense of presence in virtual environments: Zeitschrift für Medienpsychologie 15 (2), Publisher: Hogrefe Verlag, S. 69–71, 2003, ISSN: 1617-6383, DOI: 10.1026//1617-6383.15.2.69, URL: https://econtent.hogrefe.com/doi/10.1026//1617-6383.15.2.69, Stand: 10. 04. 2024.

[Sc88] Schoenfeld, A.: Problem solving in context (s). In: The teaching and assessing of mathematical problem solving. S. 82–92, 1988.

[SFR01] Schubert, T.; Friedmann, F.; Regenbrecht, H.: The Experience of Presence: Factor Analytic Insights. Presence: Teleoperators and Virtual Environments 10 (3), S. 266–281, 2001, ISSN: 1054-7460, DOI: 10.1162/105474601300343603, URL: https://doi.org/10.1162/105474601300343603, Stand: 15. 08. 2022.

[SL16] Sauro, J.; Lewis, J. R.: Chapter 9 - Six enduring controversies in measurement and statistics. In (Sauro, J.; Lewis, J. R., Hrsg.): Quantifying the User Experience (Second Edition). Morgan Kaufmann, Boston, S. 249–276, 2016, ISBN: 978-0-12-802308-2, DOI: 10.1016/B978-0-12-802308-2.00009-6, URL: https://www.sciencedirect.com/science/article/pii/B9780128023082000096, Stand: 29. 05. 2024.

[Wi07] Wirth, W.; Hartmann, T.; Böcking, S.; Vorderer, P.; Klimmt, C.; Schramm, H.; Saari, T.; Laarni, J.; Ravaja, N.; Gouveia, F. R.; Biocca, F.; Sacau, A.; Jäncke, L.; Baumgartner, T.; Jäncke, P.: A Process Model of the Formation of Spatial Presence Experiences. Media Psychology 9 (3), Publisher: Routledge _eprint: https://doi.org/10.1080/15213260701283079, S. 493–525, 2007, ISSN: 1521-3269, DOI: 10.1080/15213260701283079, URL: https://doi.org/10.1080/15213260701283079, Stand: 10. 04. 2024.

[Wi23] Wiepke, A.: Virtual reality as experimental method in embodied cognition. In: Experimental Methods in Embodied Cognition. Num Pages: 13, Routledge, 2023, ISBN: 978-1-00-329069-8.

[Wi24] Wiepke, A.; Belli, F.; Fischer, M. H.; Miklashevsky, A.: Embodied Learning in Virtual Reality. In (Senkbeil, K.; Ahlers, T., Hrsg.): Virtual Reality in den Geisteswissenschaften. Konzepte, Methoden und interkulturelle Anwendungen. in press, Peter Lang Group AG, 2024.

[ZW18] Zender, R.; Weise, M.: Lehren und Lernen mit VR und AR – Was wird erwartet? Was funktioniert? In: Proceedings der Pre-Conference-Workshops der 16. E-Learning Fachtagung Informatik. Bonn, 2018.

Integration von Virtual Reality in die berufliche Aus- und Weiterbildung: Ein praxisorientierter Ansatz zur Konzeption immersiver Lernszenarien

Florian Nowotny [1], Chantal Berkhan [1] und Ulrike Lucke [1]

Abstract: Virtual Reality (VR) gewinnt aufgrund der Handlungsorientierung gerade in der beruflichen Aus- und Weiterbildung zunehmend an Relevanz. Um angestrebte Lernziele erreichen und überprüfen zu können, werden Lernmedien entlang didaktischer Strukturen entwickelt. Für diesen herausfordernden Prozess wird gleichermaßen fachliches, didaktisches und medientechnisches Wissen benötigt, was bei praxisnahen und zugleich technisch anspruchsvollen Settings besonderer Unterstützung bedarf. Der vorliegende Beitrag schlägt einen strukturierten Ansatz vor, der an der Schnittstelle zwischen fach- und mediendidaktischer Konzeption von Lernszenarien in immersiven Umgebungen wirkt. Mit Hilfe einer Methode zur Planung von gewerblich-technischem Unterricht wird zu vermittelndes Wissen nach inhaltlichen und lernpsychologischen Aspekten in ein Raster einsortiert, das als Rahmen für das Design einzelner Szenen einer Lernanwendung dient. Lehrenden und anderen Anbietenden dieser Lernanwendung kann das Raster anschließend bei der Integration in die individuelle Lehr-/Lernsituation helfen. Der Ansatz wird am Beispiel einer VR-Lernanwendung zur Vermittlung der fachgerechten Bedienung von Präzisionslandtechnik verdeutlicht.

Keywords: Virtual Reality, Trainingssimulation, Didaktisches Design, CTA, ArTWin

1 Virtual Reality in der Bildung

In der aktuellen Bildungslandschaft nimmt Virtual Reality (VR) eine zunehmend prominente Rolle ein. Dieser Trend spiegelt sich in der Bildungsforschung wider. Buchner und Mulders betonen die vielfältigen Einsatzmöglichkeiten von VR-Lernanwendungen als Trainings-, Konstruktions-, Explorations- und Experimentierwelten. Sie argumentieren für die Integration von VR in didaktisch fundierte Lernszenarien, um das Potenzial dieser Technologie vollständig zu erschließen [BM20]. In der Berufsbildung bieten besonders Trainingswelten das Potential, Fähigkeiten, die sich auf routinierte Handlungen im Arbeitsalltag beziehen, virtuell zu trainieren. Ein Beispiel dafür ist die VR-Lackierwerkstatt *HandleVR*, die didaktisch mit Hilfe des Vier-Komponenten-Instruktions-Modells (4C/ID) konzipiert wurde. Dieses Modell umfasst die Analyse notwendigen Wissens für kognitiv anspruchsvolle Handlungen und verweist, wenn nötig, auf eine kognitive Aufgabenanalyse (CTA) [Me20].

Der vorliegende Text skizziert ein Vorgehen, das die Nutzung einer CTA-Adaption als Teil der fach- und mediendidaktischen Konzeption von VR-Lernanwendungen für die

1 Universität Potsdam, Institut für Informatik und Computational Science, An der Bahn 2, 14476 Potsdam,
florian.nowotny@uni-potsdam.de, https://orcid.org/0000-0002-9027-3525;
chantal.berkhan@uni-potsdam.de, https://orcid.org/0009-0004-2802-7958;
ulrike.lucke@uni-potsdam.de, https://orcid.org/0000-0003-4049-8088

doi:10.18420/delfi2024_19

gewerblich-technische Berufsbildung vorschlägt. Das von Erlebach, Leske und Frank entwickelte Analyseraster für technische Wissensinhalte (ArTWin) [ELF20] ermöglicht es, benötigtes oder unterstützendes Wissen im fachlichen Kontext zu betrachten und Lernenden verschiedene, an der eigenen Lebenswelt orientierte Zugänge zu Lernaufgaben anzubieten. Gleichzeitig kann das ArTWin Lehrenden und Agierenden im Bildungswesen helfen, die Anwendung auf ihre Eignung für die eigene Lehre zu überprüfen und darin einzubinden. Der vorgestellte Ansatz soll lediglich eine von vielen Möglichkeiten darstellen, die fachdidaktische Aufbereitung des Domänenwissens mit Blick auf Binnendifferenzierung in eine mediendidaktische Herangehensweise zu überführen. Am Beispiel einer Szene aus einer VR-Lernanwendung soll verdeutlicht werden,

2 Analyse- und Entwurfsmethoden für VR-Bildungsanwendungen

Im Rahmen von Trainingssimulationen und Serious Games werden CTA-Methoden zur Analyse kognitiv anspruchsvoller Tätigkeiten und als Grundlage des Designs einzelner Szenen genutzt. [AS11] [Da16] [KGS18].

Antonova und Stefanov stützen sich in ihrer Arbeit auf eine *Applied Cognitive Task Analysis* (A-CTA), einer CTA-Adaption nach Militello und Hutton [MH98]. Dabei werden mithilfe von Interviews über die Formulierung von Teilaufgaben und damit verbundenem Domänenwissen kritische Momente identifiziert, welche als Ausgangspunkt für die szenische Gestaltung dienen. Antonova und Stefanov kommen zu dem Ergebnis, dass die A-CTA-Methode zur Konzeption von Lernanwendungen verwendet werden kann, wenngleich A-CTA einen Fokus auf implizites Wissen durch Erfahrungswerte aufweist. [AS11].

Erlebach, Leske und Frank entwickelten das ArTWin als eine Adaption des A-CTA Modells [MH98] hinsichtlich einer binnendifferenzierten – an der Lebenswelt von Lernenden orientierten – Unterrichtsplanung für Lehrende in der gewerblich-technischen Berufsbildung. Anstatt implizite Erfahrungswerte zu analysieren, soll das entstandene Analyseraster von Lehrenden genutzt werden, um das mit Lernaufgaben zusammenhängende Domänenwissen strukturiert aufzuarbeiten. [ELF20]. Des Weiteren empfehlen Erlebach und Frank das ArTWin-Modell als Grundlage zur Untersuchung des benötigten Vorwissens für das Verständnis technischer Repräsentationen [EF21].

Im Kontext von VR-Lernanwendungen offenbart die Forschungslandschaft noch eine erkennbare Lücke bezüglich der didaktischen Binnendifferenzierung innerhalb von Trainings-, Explorations- und Experimentierwelten. Diese Erkenntnislücke erschwert die Integration solcher Anwendungen in diverse Lehr- und Lernkontexte sowie die angemessene Adressierung heterogener Zielgruppen. Vor diesem Hintergrund schlägt das nachfolgend dargelegte Konzept einen innovativen Ansatz vor, der das Potential besitzt, diesen Herausforderungen zu begegnen, wenngleich seine praktische Wirksamkeit noch empirisch untersucht werden muss.

3 Konzeption einer tiefergehenden Entwurfsunterstützung

Das hier vorgestellte Vorgehen beschreibt einen möglichen Weg der Darbietung von Wissensbeständen in explorativen VR-Lernanwendungen oder explorativ gestalteten Szenen im Rahmen der technischen Berufs- und Weiterbildung.

Am Beispiel des 4C/ID [Me20] soll verdeutlicht werden, an welcher Stelle des Designprozesses das ArTWin eingesetzt werden kann und wofür es eher ungeeignet ist. 4C/ID umfasst die vier Komponenten: Lernaufgaben, unterstützende Informationen, prozedurale Informationen und Üben von Teilaufgaben. Nachdem Lernaufgaben definiert wurden folgt die Konzeption der unterstützenden Informationen für Nicht-Routineaspekte. Diese beinhaltet die Spezifizierung von Lehrinhalts- und mentalen Modellen, zu deren Analyse sich das ArTWin eignet. Die anschließend zu spezifizierenden systematischen Problemlöseansätze sowie kognitives Feedback lassen sich mit dem ArTWin nicht zwangsläufig hinreichend analysieren.

Das ArTWin dient somit – als formaler Rahmen – der Entscheidung welche Lehrinhalts- und mentalen Modellen für das Erreichen von Lernzielen relevant sind und wie diese vermittelt werden sollen – insbesondere, welche in VR zielführend adressiert werden können und welche im rahmenden Setting besser angesiedelt sind.

Mit gezielten Fragen, angelehnt an Kategorisierungen der Inhaltsbereiche aus dem Diskurs des Technischen Wissens [ELF20], wird zunächst das relevante Domänenwissen ermittelt. Im Anschluss wird das gesammelte Wissen in das ArTWin sortiert, wie in Abb. 1 zu sehen. Inhaltlich wird bei gesammeltem Wissen zwischen Sach- und Handlungswissen sowie den jeweiligen Bezug zum technischen Artefakt differenziert. Lernpsychologisch wird deklaratives und prozedurales Wissen jeweils dahingehend differenziert, ob es sich um konzeptionelles- bzw. Modell-Wissen handelt.

Das soll Lehrenden helfen, ausgehend vom individuellen Vorwissen, eigenen Erfahrungen sowie persönlichem Interesse, eine an der Lebenswelt der Lernenden orientierte Herangehensweise an Lernaufgaben zu ermöglichen bzw. individuelle Lernpfade zu identifizieren.

Erlebach und Frank heben hervor, dass auch im Kontext technischer Repräsentationen die Aktivierung des vorhandenen Vorwissens der Lernenden im Fokus steht. Das durch das ArTWin dokumentierte Wissen zielt darauf ab, ein profundes Verständnis eines technischen Artefakts zu fördern, ohne jedoch das implizite, für die Ausführung einer bestimmten Tätigkeit erforderliche Wissen zu analysieren [EF21]. Demnach kann und soll das ArTWin nicht die Nutzung von CTA-Methoden mit Interviews zur Analyse kognitiv anspruchsvoller Tätigkeiten ersetzen. Vielmehr geht es darum, Lehrenden über die horizontale Zuordnung die verschiedenen Zugänge zu Wissensbeständen zu verdeutlichen und dank der lernpsychologischen, vertikalen Sortierung mittels zugehöriger Operatoren fachdidaktisch fundiert mögliche Arbeitsaufträge direkt abzuleiten. Die inhaltliche Verknüpfung einzelner Wissensbestände ermöglicht eine nachvollziehbare Gestaltung von Lernpfaden [ELF20]. Über eine mediendidaktische Umsetzung trifft das Verfahren bislang keine Aussagen.

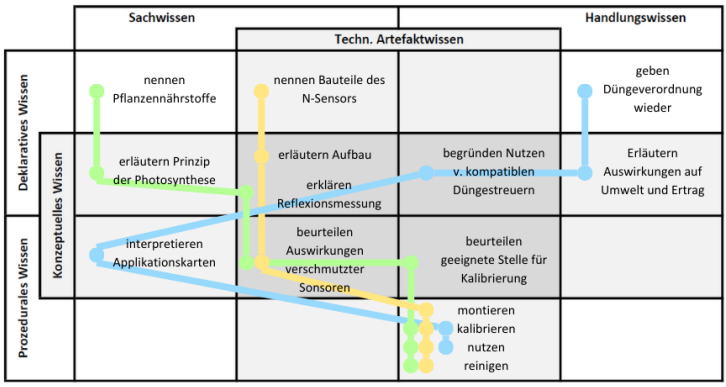


Abb. 1: ArTWin am Beispiel des Yara N-Sensor ALS 2

Im Hinblick auf explorative VR-Lernanwendungen sowie Lernaufgaben mit expositorischen Methoden nach Kerres [Ke18] kann das ArTWin helfen die szenische Gestaltung zu konkretisieren. Rein prozedurale Wissensbestände (vierte Zeile des ArTWin), die sich nicht auf Konzepte bzw. Modelle beziehen, haben zunächst einmal das Potential, vollständig in VR geübt und überprüft zu werden. Besonders von Bedeutung sind die Wissensinhalte des rein prozeduralen Handlungswissens mit Bezug zum technischen Artefakt – dem technischen Know-How –. Bei instruktional gestalteten Trainingswelten werden sich an dieser Stelle die konkreten Tätigkeiten der Lernaufgaben wiederfinden.

Konzeptuelles Wissen (zweite und dritte Zeile des ArTWin) entspricht den zu vermittelnden Lerhinhalts- und mentalen Modellen (gemäß 4C/ID). Notwendiges konzeptuelles Wissen kann in VR vermittelt werden. Eine Überprüfung zugehöriger Lernziele lässt sich innerhalb der VR-Anwendung aber nur dann umsetzen, wenn die inhaltlich verknüpfte Tätigkeit (vierte Zeile des ArTWin) nicht ohne vollstes Verständnis dieser Modelle durchführbar ist. Andernfalls wäre eine Überprüfung im rahmenden Setting angebracht.

Rein deklaratives Wissen (erste Zeile des ArTWin) lässt sich in VR vermitteln aber nicht sinnvoll überprüfen. Es handelt sich hierbei um grundlegendes Fakten-Wissen, dessen Überprüfung im Anforderungsbereich I – Reproduktion – verortet werden kann. Diese Wissensbestände stellen aber niedrigschwellige Anknüpfungspunkte bzw. kognitive Öffner dar.

Der vorgestellte Ansatz schlägt vor, innerhalb der betreffenden VR-Szenen diejenigen 3D-Modelle zu kuratieren, die mit diesen Fakten assoziiert werden können. Lernpsychologisch sollen diese Vorwissen aktivieren und visuell als Abrufhinweise für eine Konsolidierung neuer Informationen im Langzeitgedächtnis fungieren [BB17]. In Bezug auf die Szene soll es Lernenden ermöglicht werden, einen individuellen Lernpfad im Sinne eines Handlungsstrangs zu beschreiten. Abhängig von der Interaktion der Lernenden mit den zur

Verfügung gestellten Objekten erfolgt eine vertiefte Auseinandersetzung mit Lehrinhalts- sowie mentalen Modellen, welche eine Verknüpfung mit diesem Gegenstand innerhalb des ArTWin aufweisen. Somit müssen mehrere Handlungsstränge implementiert werden, die dem theoretischen Teil einer Trainings- oder Experimentierwelt den Charakter eines explorativen Lernangebots verleihen. Im Sinne des entdeckenden Lernens nach Kerres [Ke18] ist es dann möglich, sich jederzeit für einen anderen Pfad zu entscheiden, wenn das vermittelte Wissen keinen Bezug zum eigenen Interesse oder Vorwissen mehr aufweist. Zudem zielt die Existenz nicht beschrittener Pfade darauf ab, Lernende zu motivieren, zu diesen Punkten zurückzukehren und die unterstützenden Informationen mit anderem Fokus erneut abzurufen.

Zusammengefasst stellt sich das vorgeschlagene Vorgehen für den verzahnten fach- und mediendidaktischen Entwurf von VR-Lernanwendungen wie folgt dar:

1. Analyse des technischen Artefakts mit ArTWin
2. Identifikation von inhaltlich verbundenen Fakten / rein-deklarativem Wissen
3. Kuratierung assoziierter Objekte als 3D-Modelle in einer Szene
4. Konzipierung von Handlungssträngen zur Wissensvertiefung entlang inhaltlicher Pfade im ArTWin

4 Entwurf einer VR-Lernaufgabe zu Präzisionslandtechnik

Der vorgestellte Ansatz wird im Rahmen eines landwirtschaftlichen Forschungsprojekts genutzt, um innerhalb des Projekts generierte Erkenntnisse und Expertise für (angehende) LandwirtInnen, Landwirtschaftsdienstleitende, Lehrende sowie allgemein Interessierte zugänglich zu machen. Ziele des Projekttransfers sind Fähigkeiten zur fachgerechten Bedienung von Präzisionslandtechnik (PL) zu vermitteln und Verständnis für Potentiale der PL sowie ein Bewusstsein für übergeordnete Wirkungskreisläufe zu schaffen. Um die Folgen eigener Entscheidungen erlebbar zu machen und – über räumliche und zeitliche Skalen hinweg – in Bezug zu übergeordneten Aspekten zu setzen, erfolgt dies im Rahmen einer immersiven Experimentierwelt. Nutzende sollen die Interdependenz einzelner Technologien der PL erforschen und eigene Annahmen in Bezug auf Stickstoff-Düngung testen können, ohne negative Konsequenzen wie finanzielle Einbußen oder Umweltbelastung befürchten zu müssen.

Abb. 1 zeigt beispielhaft eine Anwendung des ArTWin für einen Stickstoff-Sensor. Inhaltliche Verknüpfungen sind farblich hervorgehoben. Für das *technische Know-How* lassen sich drei Handlungsstränge anhand verschiedener Zugänge ableiten: pflanzenphysiologisch (grün), instrumentell-technisch (gelb) und umsetzungsstrategisch (blau).

Das Fakten-Wissen in der ersten Zeile (Pflanzennährstoffe, Bauteile und Düngeverordnung) werden mit assoziierten Gegenständen (Dünger, N-Sensor und Düngeverordnung



Abb. 2: Mit dem N-Sensor assoziierte 3D-Objekte in der VR-Anwendung

als Dokument) – wie in Abb. 2 dargestellt – in einer Garage auf einem virtuellen Landwirtschaftsbetrieb dargeboten. Wird beispielsweise der Dünger in VR gegriffen, folgt eine inhaltliche Führung von Nährstoffen über den Prozess der Verstoffwechselung (Photosynthese) zur Messung des Chlorophyllanteils mittels Sensor (Reflexionsmessung). Dazu werden nach jeder Interaktion weitere Modelle und Inhalte freigeschaltet. Für die Photosynthese erscheint ein Blattquerschnitt, der auf Chloroplasten verweist. Die Reflexionsmessung wird anhand von – über-realistisch sichtbaren – auf den Blattquerschnitt treffenden Lichtstrahlen verdeutlicht. Die Erläuterung von Fehlerquellen (verschmutzte Sensoren) führt schließlich zur Ermittlung einer geeigneten Stelle (auf dem Feld) zur Kalibrierung des Sensors, welcher von ExpertInnen als kritischer Moment identifiziert wurde. Damit endet die Erarbeitung der Lehrinhalts- und mentalen Modelle für den pflanzenphysiologischen Handlungsstrang und es folgen die Übungen der damit verbundenen Prozesse im Sinne des 4C/ID inklusive systematischer Problemlöseansätze sowie kognitivem Feedback, auf deren Konkretisierung in diesem Beitrag aber nicht eingegangen wird.

5 Zusammenfassung und Ausblick

Das ArTWin bietet eine klare Struktur zur systematischen Erfassung technischer Wissensinhalte, die die szenische Gestaltung von VR-Lernanwendungen erleichtern soll. Des Weiteren soll es Lehrende bei der Integration der VR-Anwendung in den eigenen Lehrplan unterstützen. Die empirische Validierung des Konzepts steht jedoch noch aus. Zukünftige Untersuchungen sollten dessen Übertragbarkeit auf andere Disziplinen mittels angepasster Analyseraster und die mögliche Synthese mit bestehenden Methoden zur Reduktion des Entwicklungsaufwands erforschen. Eine Verbindung des Konzepts mit dem Ansatz von Tjiam et al. [Tj12] könnte die Integration in das 4C/ID-Modell vereinfachen, ebenso wie eine effektive Anknüpfung an On-Boarding-Prozesse denkbar ist.

Danksagung

Die hier beschriebenen Arbeiten werden teilweise vom Bundesministerium für Landwirtschaft und Ernährung unter Kennzeichen 28DE202C21 gefördert.

Literaturverzeichnis

[AS11] Antonova, A.; Stefanov, K.: Applied Cognitive Task Analysis in the Context of Serious Games Development. In (Dicheva, D.; Markov, Z.; Stefanova, E., Hrsg.): Third International Conference on Software, Services and Semantic Technologies S3T 2011. Springer, Berlin, Heidelberg, S. 175–182, 2011, ISBN: 978-3-642-23163-6.

[BB17] Buchner, A.; Brandt, M.: Gedächtniskonzeptionen und Wissensrepräsentationen. In (Müsseler, J.; Rieger, M., Hrsg.): Allgemeine Psychologie. Springer Berlin Heidelberg, Berlin, Heidelberg, S. 401–434, 2017, ISBN: 978-3-642-53897-1, DOI: 10.1007/978-3-642-53898-8_12.

[BM20] Buchner, J.; Mulders, M.: Lernen in immersiven virtuellen Welten aus der Perspektive der Mediendidaktik. Medienimpulse 58 (2), 23 Seiten–23 Seiten, 2020, DOI: 10.21243/mi-02-20-22.

[Da16] Dass, S.; Barnieu, J.; Cummings, P.; Cid, V.: A Cognitive Task Analysis for an Emergency Management Serious Game. The Interservice/Industry Training, Simulation, and Education Conference : I/ITSEC. Interservice/Industry Training, Simulation, and Education Conference 2016, 2016.

[EF21] Erlebach, R.; Frank, C.: Fachdidaktische Modellierung Technischen Wissens als Grundlage zur Analyse technischer Repräsentationen. Journal of Technical Education (JOTED) 9 (2), S. 59–86, 2021, DOI: 10.48513/JOTED.V9I2.205.

[ELF20] Erlebach, R.; Leske, P.; Frank, C.: Ein Analyseraster Technischer Wissensinhalte als Grundlage für eine lebenswelt - und ressourcenorientierte Unterrichtsplanung. Berufs- und Wirtschaftspädagogik - online (38), 2020, URL: https://www.bwpat.de/ausgabe38/erlebach_etal_bwpat38.pdf.

[Ke18] Kerres, M.: Mediendidaktik: Konzeption und Entwicklung digitaler Lernangebote. In: Mediendidaktik. De Gruyter Oldenbourg, 2018, ISBN: 978-3-11-045683-7, DOI: 10.1515/9783110456837.

[KGS18] Kaber, D. B.; Green, R. S.; Swangnetr, M.: An Integration of Cognitive Task Analysis Results for Situation Awareness-Focused Training Program Development. In (Chen, J., Hrsg.): Advances in Human Factors in Robots and Unmanned Systems. Springer International Publishing, Cham, S. 161–172, 2018, ISBN: 978-3-319-60384-1, DOI: 10.1007/978-3-319-60384-1_16.

[Me20] van Merriënboer, J. J. G.: Das Vier-Komponenten Instructional Design (4C/ID) Modell. In (Niegemann, H.; Weinberger, A., Hrsg.): Handbuch Bildungstechnologie: Konzeption und Einsatz digitaler Lernumgebungen. Springer, Berlin, Heidelberg, S. 153–170, 2020, ISBN: 978-3-662-54368-9, DOI: 10.1007/978-3-662-54368-9_8.

[MH98] MILITELLO, L. G.; HUTTON, R. J. B.: Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. Ergonomics 41 (11), S. 1618–1641, 1998.

[Tj12] Tjiam, I. M.; Schout, B. M.; Hendrikx, A. J.; Scherpbier, A. J.; Witjes, J. A.; Van Merriënboer, J. J.: Designing simulator-based training: An approach integrating cognitive task analysis and four-component instructional design. Medical Teacher 34 (10), e698–e707, 2012, DOI: 10.3109/0142159X.2012.687480.

Virtual Reality zur Professionalisierung von Sportlehrkräften

René Wilzeck [1] und Raphael Zender [2]



Abstract: Im Sportunterricht können vielfältige Situationen auftreten, die potenzielle Gefahren für Schüler*innen bergen und dies sowohl auf physischer als auch auf psychischer Ebene (die sogenannte doppelte Verletzbarkeit). Daher ist es sinnvoll, Sportlehrkräften Möglichkeiten zur Vorbereitung auf solche Szenarien zu bieten, ohne dabei sich selbst oder die Schüler*innen zu gefährden.

Die Integration von Virtual Reality (VR) in die Sportlehrkraftausbildung bietet effektive Ansätze für die Professionalisierung und Weiterentwicklung didaktischer Kompetenzen. Als Lernumgebung entsteht eine VR-Sporthalle als Nachbildung einer Sporthalle, in der Lehrkräfte die Möglichkeit erhalten, u. a. körperliche Präsenz und sportdidaktische Methoden in einem immersiven und physisch risikofreien, virtuellen Raum zu erproben.

Ein Konzept zu potenziellen Situationen doppelter Verletzbarkeit in einem Handstand-Übungsszenario wurde entwickelt und erfolgreich als Prototyp in einer ersten Iteration in die VR-Lernumgebung TeachR integriert. Dieser Beitrag beschreibt, wie darüber zu einem verstärkten Engagement und zur Reflexion über unterrichtsnahe Fähigkeiten angeregt wird, was letztendlich zu einer verbesserten Qualität des Sportunterrichts führen kann.

Keywords: Virtual Reality, Sportdidaktik, Simulation, Doppelte Verletzbarkeit, Digitale Lernorte, Professionalisierung

1 Motivation

Digitale Medien sind allgegenwärtig und unterstützen sowohl fachspezifische als auch übergreifende Kompetenzen [KM21]. Die digitale Transformation stellt nun auch Lehrkräfte vor neue Herausforderungen und Chancen, die eine spezialisierte Expertise und eine Anpassung ihrer Einstellung zu digitalen Medien verlangen [Lo22]. Mit Blick auf den Fachbereich Sport könnten digitale Angebote genutzt werden, um den Mangel an Lehrkräften [KM22] entgegenzukommen, Seiteneinsteiger zu qualifizieren und die unterschiedlichen Bedürfnisse von angehenden, sowie von etablierten Lehrkräften zu berücksichtigen [Lu20].

Zur Adressierung dieser Fortbildungsmöglichkeiten besteht der Fokus darauf, für den Sportkontext hochimmersive Medien wie virtuelle und erweiterte Realitäten sowie digitale Spielumgebungen als Teil der fachdidaktischen Konzeptbildung zu betrachten [NSR16]. Somit entstehen Konzepte, welche es Sportlehrkräften ermöglichen, verschiedene Unterrichtsszenarien in einer VR-Sporthalle zu durchlaufen und sie anschließend gemeinsam mit

1 Humboldt-Universität zu Berlin, Institut für Informatik, Rudower Chaussee 25, 12489 Berlin, Deutschland, rene.wilzeck@hu-berlin.de, https://orcid.org/0009-0004-0431-8689
2 Zeppelin Universität, Am Seemooser Horn 20, 88045 Friedrichshafen, Deutschland, raphael.zender@zu.de, https://orcid.org/0000-0001-9866-9455

doi:10.18420/delfi2024_20

ihren Kolleg*innen zu reflektieren. Die VR-Sporthalle kann natürlich keine Erfahrungen im echten Sportunterricht ersetzten, bietet im Vergleich zu diesem jedoch eine standardisierte Simulation von Unterrichtssituationen, die beliebig oft reproduzierbar sind, ohne dabei negative Konsequenzen zu befürchten. Außerdem werden die Unterrichtsszenarien so gestaltet, dass bestimmte Aspekte wie Mehrdimensionalität, Simultanität und Unmittelbarkeit gezielt beeinflusst werden können. Dadurch soll eine schrittweise Annäherung an die Komplexität und Dynamik des Unterrichts ermöglicht werden.

2 Grundlagen

2.1 Professionalisierung von Sportlehrkräften

Die Entwicklung von Lehr- und Lerneinheiten in der VR-Sporthalle stützt sich auf zwei sportdidaktische Untersuchungsfelder. Hierbei steht nicht die Schulung von Bewegungsformen im Fokus, sondern die Professionalisierung von Sportlehrkräften im Kontext des Klassenmanagements.

Das erste sportdidaktische Untersuchungsfeld legt den Fokus auf der körperlich-leiblichen Erfahrbarkeit von Nähe und Distanz. Diese Nähe-Distanz-Erfahrung kann sich einerseits auf das Verhältnis des eigenen realphysischen und gleichzeitig digital repräsentierten Zeichenkörpers in der VR gegenüber virtuellen Entitäten oder Avataren beziehen. Andererseits bietet die VR-Sporthalle auch einen Reflexionsraum, in dem mit Hilfe der Erfahrung einer fiktiven und simulierten Situation physische und soziale Grenzen und Grenzüberschreitungen im Sportunterricht thematisiert werden können. Ziel ist es, dieses Lernmodul mit Sportlehrkräften zu evaluieren, weiterzuentwickeln und als Fortbildungskonzept in der Lehrkräftebildung zu verankern.

Wahrnehmungsprozesse sind für das professionelle Handeln von Lehrkräften im Sportunterricht äußerst relevant, denn um situativ angemessene Entscheidungen für die Gestaltung von Unterricht treffen zu können, ist grundlegend die Fähigkeit bedeutsam, lernrelevante Unterrichtsereignisse erkennen und theoriebasiert interpretieren zu können [SBS10] [Me17]. In diesem zweiten sportdidaktischen Untersuchungsfeld werden inhaltlich die physische und psychische Verletzbarkeit von Schüler*innen (die sog. *doppelten Verletzbarkeit* [MK04]) sowie der kompetente Umgang damit thematisiert und in der VR-Sporthalle immersiv simuliert und erprobt. Ein besonderer Fokus liegt dabei auf dem Aspekt der Emotion Scham und dazugehörigen Präventionsansätzen. Zur Vermittlung dieser Inhalte wird eine Fortbildungskonzeption mit VR-Elementen für Sportlehrkräfte entwickelt, die ein Grundlagen- und ein Vertiefungsmodul in Präsenz mit einem E-Learning-Modul kombiniert. Die Konzeption soll unter anderem in Bezug auf die Förderung professioneller Wahrnehmung evaluiert werden. Das übergeordnete Ziel ist es, Wahrnehmungs-, Deutungs- und Handlungskompetenzen bei Sportlehrkräften mittels digitaler Medien sichtbar zu machen und zu fördern.

2.2 TeachR - das VR-Klassenzimmer

Im Rahmen des Open Source Systems TeachR[3] (ehem. VR-Klassenzimmer [Wi19]) wurde eine VR-Klassenraumsimulation entwickelt, die angehenden Lehrkräften ein realitätsnahes Training im Umgang mit Unterrichtsstörungen ermöglicht. Die Simulation erlaubt es, gezielt Störungen bei simulierten Schüler*innen auszulösen und das Klassenmanagement der Lehrkräfte zu evaluieren. Dabei konzentrierte sich das Training bisher auf ein traditionelles Klassenzimmer-Szenario [Wi19].

Die VR-Sporthalle basiert auf diesem etablierten System, welches speziell für dieses Vorhaben um den Sportbereich erweitert wurde und nun unter der GNU Affero General Public Lizenz[4] weiterentwickelt wird.

3 Methodik und Konzeption

Mit der Entwicklung der VR-Sporthalle sollen in einem konkreten Praxisfeld Lösungen für Bildungsprobleme gewonnen und wissenschaftliche Erkenntnisse in der Lehr-Lernforschung gesammelt werden. Eingesetzt in Seminaren an Universitäten und Workshops in Fortbildungen können somit Lernphänomene untersucht werden, wobei das Ziel einerseits darauf liegt, wichtige Probleme aus der Bildungspraxis zu lösen, während andererseits Ergebnisse generiert oder weiterentwickelt werden sollen, die zur Theoriebildung oder -entwicklung beitragen. Nach aktuellem Forschungsstand bietet das Design-Based Research Forschungsprinzip [Re05] diesbezüglich eine geeignete Methodik: 1. Vorprüfung, 2. Prototypenentwicklung, 3. Beurteilungsphase, 4. Lösung [SW21]. Dabei befindet sich die Entwicklung der VR-Sporthalle in den Phasen eins und zwei. Praktische Probleme werden unter Berücksichtigung theoretischer Ansätze analysiert, um darauf basierend Bildungsszenarien zu entwickeln. Anschließend werden Prototypen erstellt, die kontinuierlich anhand des Feedbacks aus der ersten Phase evaluiert und optimiert werden.

Die Implementierung der didaktischen Konzepte in die VR-Sporthalle folgt dem agilen Vorgehensmodell Scrum [Su20]. Nach diesem Framework erfolgt die Umsetzung neuer Anforderungen in sechswöchigen Entwicklungszyklen. Hierbei werden zu Beginn eines Zyklus die Anforderungen in Tickets organisiert und Verantwortlichkeiten zugewiesen. Im Laufe jedes Zyklus wird ein neuer Prototyp erstellt und am Ende von verschiedenen Kooperierenden getestet. Darauf folgend werden in einer Besprechung die umgesetzten Anforderungen mit allen Beteiligten reflektiert. Die Zyklusdauer von sechs Wochen wurde aufgrund der Teilzeitbeschäftigung des Entwicklungsteams strategisch gewählt. Diese Entscheidung ermöglichte eine ausgewogene Balance zwischen Agilität und verfügbarer Entwicklungszeit und förderte die Flexibilität des Teams bei der Reaktion auf neue Anforderungen.

3 GitLab Repository url: `https://gitup.uni-potsdam.de/mm_vr/vr-klassenzimmer`, Stand 11.04.2024

4 Free Software Foundation, Inc., 2007, url: `https://www.gnu.org/licenses/agpl-3.0.de.html`, Stand: 19.03.2024.

4 Entwicklung und Nutzung der VR-Sporthalle

Um die Herausforderungen des Sportunterrichts anzugehen, wurde eine Erweiterung von TeachR für den Bereich Sport angestrebt. Dabei wird folgend auf die Umsetzung eines ersten Lehr- und Lernkonzepts in der VR-Sporthalle eingegangen.

4.1 Erweiterung von TeachR um den Sportbereich

Im Rahmen einer Abschlussarbeit wurde das etablierte System TeachR um eine Sporthalle erweitert. Die Arbeit zielte darauf ab, die Integration von Sportinhalten in TeachR zu ermöglichen, den so entstandenen neuen Raum umfassend zu analysieren und zu verbessern. Sie identifizierte spezifische Herausforderungen des Sportunterrichts und entwickelte in Zusammenarbeit mit Sportlehrkräften Anforderungen für virtuelle Lehrumgebungen im Sportkontext.

Im Fokus steht nun die Weiterentwicklung dieser Sporthallen-Erweiterung von TeachR, um Anwendungen in hochimmersiven Medien (Virtual Reality) für Lehr- und Lernzwecke umzusetzen. Hierzu wurde die VR-Sporthalle bereits mit 3D-Modellen relevanter Sportgeräte erweitert und digitale Schüler*innen in Sportkleidung erstellt. Der Funktionsumfang der Schüler*innen wurde ebenso auf den Sportbereich ausgebaut. Dafür konnte das bestehende Kontrollsystem von TeachR visuell und funktionell angepasst werden, sodass eine Gruppenübung gestartet werden kann. Zuletzt wurden für die sportlichen Schüler*innen-Modelle passende Animationen erstellt (vgl. Abschnitt 4.2). Besonders herausfordernd war hierbei das korrekte Starten von mehreren Animationen, sodass diese synchron ablaufen.

Dieser Beitrag beleuchtet folgend die Entwicklung einer fachspezifischen Fortbildung zur situativ angemessenen Entscheidungsfindung bei der Unterrichtsgestaltung und die Professionalisierung des Klassenmanagements unter Berücksichtigung der doppelten Verletzbarkeit (vgl. Abschnitt 2.1).

4.2 Umsetzung eines Handstandkonzepts innerhalb der VR-Sporthalle

Die aktuelle Implementierung der VR-Sporthalle umfasst ein Lehr-, Lernkonzept, in welchem Lehrkräfte den virtuellen Schüler*innen eine sichere Übung des Handstands ermöglichen und dabei angehalten sind, auf verschiedene Fehlerquellen achtzugeben. Wie in den bisherigen Umsetzungen von TeachR und auf Abb. 1 dargestellt, wird eine Lehrkraft während der VR-Erfahrung von einer coachenden Person via PC geführt und angeleitet. Hierbei sind die Lehrkräfte innerhalb der VR-Anwendung während einer Lehreinheit dazu angehalten, ganz natürlich mit ihrer Umgebung zu interagieren, genau wie sie es auch in der realen Welt tun würden. Allerdings reagiert und antwortet das System nicht automatisch auf ihre Handlungen. Stattdessen kann die coachende Person außerhalb der VR

über ein Kontrollsystem das Szenario beeinflussen. Sie steuert das Verhalten der virtuellen Schüler*innen und reagiert auf die Ansagen der Lehrkraft. Sie überwacht die Szenarien und passt die Reaktionen der virtuellen Schüler*innen in Echtzeit an. Somit soll eine dynamische und kontrollierte Lernumgebung ermöglicht werden.



Abb. 1: Schematische Darstellung einer TeachR Lerneinheit. Die Lehrkraft erfährt die Handstand-Szene in der VR-Sporthalle (rechts) mittels VR-Headset und wird dabei von einer coachenden Person über das TeachR Kontrollsystem (links) am PC geleitet.

Die Szene findet auf zwei Mattenbahnen in der VR-Sporthalle statt, auf denen Schüler*innen zunächst einen Handstand ausführen und dabei von zwei weiteren Personen gestützt werden. Wie in einer realen Unterrichtssituation wird der Ablauf über ein Tablet aufgenommen, worüber die Ausführung der Übungen überprüft und beurteilt werden kann. Auf Abb. 2 ist eine Gegenüberstellung einer realen Szene zu der virtuellen Szene zu finden. Mit Start der Lehreinheit befindet sich die Lehrkraft auf der gelb markierten Position vor dem Tablet in der unteren Mitte des Bildes. Von hier aus kann mittels Teleportation auf sechs weitere, gelb markierte Positionen gesprungen oder per Joystick eine freie Bewegung ausgeführt werden.

Über das Kontrollsystem kann die coachende Person verschiedene Verhaltensweisen (VW) der virtuellen Schüler*innen aktivieren. Hierbei werden zwei rollenspezifische VW unterschieden: VW der Handstand-Person und VW der Hilfestellungs-Person. Zu den im Sportunterricht wünschenswerten VW, also der korrekten Ausführung der Übung, können auch falsche und unangebrachte VW (sog. Fehlerquellen) aktiviert werden. Bisher wurden folgende Fehlerquellen der Hilfestellungs-Personen implementiert, Fehlerquellen der Handstand-Personen folgen zu einem späteren Zeitpunkt:

1. Eine Person gibt keine Hilfestellung.
2. Eine Person führt die Hilfestellung nur mit einer Hand aus.
3. Die Handstand übende Person wird bei Ausführung der Übung unsittlich berührt.



Abb. 2: Vergleich einer realen Situation im Sportunterricht (oben) und einer digitalen Umsetzung (aktueller Arbeitsstand) eben dieser in Unity (unten) einer Übung des Handstands mit Hilfestellung. Im Bild unten markieren die gelben Punkte die möglichen Teleportationspunkte.

Die implementierte Szene erfordert eine große Anzahl von Animationen der virtuellen Schüler*innen. Diese wurden zunächst mithilfe von Blender händisch angefertigt. Sie dienen zu diesem Zeitpunkt ausschließlich zur Veranschaulichung und als Implementationshilfe und haben nicht den Anspruch einer flüssigen und natürlichen Bewegung. Parallel wird untersucht, wie mittels Motion-Trackings Animationen für die VR-Sporthalle erstellt werden können. Diese können dann zu einem späteren Zeitpunkt in die Szene übernommen werden.

5 Ausblick

In diesem Beitrag wurde beschrieben, wie ein VR-System zur Professionalisierung von Lehrkräften um den Sportbereich erweitert wurde, um in diesem aktuelle sportdidaktische Methoden untersuchen und als Lernmodule umsetzen zu können. In einem Verbundprojekt werden Lehr-, Lernmethoden konzeptioniert und in die VR-Sporthalle des Projekts TeachR implementiert. Es wurde ein Konzept zu potenziellen Situationen doppelter Verletzbarkeit in einem Handstand-Übungsszenario entwickelt und erfolgreich als Prototyp in einer ersten Iteration umgesetzt.

Das weitere Vorgehen wird zum einen das Handstand-Szenario iterativ weiterentwickeln. Des Weiteren werden neue Lernmodul-Konzepte in den beiden Forschungsfeldern 1. Körpererleben durch hochimmersive Medien von Lehrkräften (zum Thema: Nähe und Distanz), 2. Digitale Förderung der Wahrnehmungs- und Handlungskompetenzen im Sportunterricht (in Bezug auf: Doppelte Verletzbarkeit) entwickelt, implementiert und in Bezug auf die Förderung professioneller Kompetenz evaluiert.

Literaturverzeichnis

[KM21] KMK: Lehren und Lernen in der digitalen Welt: Ergänzung zur Strategie der Kultusministerkonferenz „Bildung in der digitalen Welt“. (Beschluss der Kultusministerkonferenz vom 09.12.2021). 2021, URL: https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2021/2021_12_09-Lehren-und-Lernen-Digi.pdf.

[KM22] KMK: Lehrkräfteeinstellungsbedarf und -angebot in der Bundesrepublik Deutschland 2021–2035: Zusammengefasste Modellrechnungen der Länder. (Beschluss der Kultusministerkonferenz vom 10.03.2022). 2022, URL: https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Dokumentationen/Dok_233_Bericht_LEB_LEA_2021.pdf.

[Lo22] Lorenz, R.; Yotyodying, S.; Eickelmann, B.; Endberg, M.: Schule digital–der Länderindikator 2021: Lehren und Lernen mit digitalen Medien in der Sekundarstufe I in Deutschland im Bundesländervergleich und im Trend seit 2017. Waxmann Verlag, 2022.

[Lu20] Lucksnat, C.; Richter, E.; Klusmann, U.; Kunter, M.; Richter, D.: Unterschiedliche Wege ins Lehramt–unterschiedliche Kompetenzen? Zeitschrift für pädagogische Psychologie, 2020.

[Me17] Meschede, N.; Fiebranz, A.; Möller, K.; Steffensky, M.: Teachers' professional vision, pedagogical content knowledge and beliefs: On its relation and differences between pre-service and in-service teachers. Teaching and teacher education 66, S. 158–170, 2017.

[MK04] Miethling, W.-D.; Krieger, C.: Schüler im Sportunterricht : die Rekonstruktion relevanter Themen und Situationen des Sportunterrichts aus Schülersicht (RETHESIS). Schorndorf: Hofmann (Verlag), 2004.

[NSR16] Nebel, S.; Schneider, S.; Rey, G. D.: Mining learning and crafting scientific experiments: a literature review on the use of minecraft in education and research. Journal of Educational Technology & Society 19 (2), S. 355–366, 2016.

[Re05] Reinmann, G.: Innovation ohne Forschung? Ein Plädoyer für den Design-Based Research-Ansatz in der Lehr-Lernforschung. Unterrichtswissenschaft 33 (1), S. 52–69, 2005.

[SBS10] Seidel, T.; Blomberg, G.; Stürmer, K.: „Observer “–Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE. 2010.

[Su20] Sutherland, J.: Das Scrum-Praxisbuch. Campus Verlag, 2020.

[SW21] Schmiedebach, M.; Wegner, C.: Design-Based Research als Ansatz zur Lösung praxisrelevanter Probleme in der fachdidaktischen Forschung. bildungsforschung (2), S. 1–10, 2021.

[Wi19] Wiepke, A.; Richter, E.; Zender, R.; Richter, D.: Einsatz von Virtual Reality zum Aufbau von Klassenmanagement-Kompetenzen im Lehramtsstudium, DELFI 2019, Bonn, 2019, DOI: 10.18420/delfi2019_319.

Immersive 360° Laborreise für Schüler*innen

Guido Klees[1], Kirsten Santelmann[2] und Markus Weber[3]

Abstract: Die Gestaltung interessefördernder Lernumgebungen ist eine zentrale Aufgabe des Merck – TU Darmstadt Lernlabors Biologie. In diesem Beitrag stellen wir ein 360°-basiertes Lernszenario vor, welches den Einflussbereich des Lernlabors erweitert und Schüler*innen direkt in den Schulen erreichen soll. Die Besonderheit des Lernszenarios ist dabei die Kombination der 360°-Technologie mit einem VR-Headset, wodurch das Labor explorativ und immersiv erlebt werden kann und den Schüler*innen die Möglichkeit eröffnet wird, schon in der Schule sehr nah an die Primärerfahrung im Labor zu gelangen. Designorientierte Forschungsansätze haben dabei ihre Eignung zur Implementierung von Lehrinnovationen in traditionellen Einrichtungen wie Schulen und Hochschulen bewiesen. Zur erfolgreichen didaktischen Integration des 360°- Labors in den Schulunterricht folgen wir dem designorientierten Entwicklungsansatz und berichten über erste praktische Erfahrungen und Ergebnisse, die im Rahmen des ersten Pilotprojekts gesammelt wurden.

Keywords: 360° Laborumgebung, didaktisches Konzept, 3D Vista, Immersion, interessefördernde Lernumgebung, designorientierte Forschungsansätze, Gestaltungsorientierung, Gamification

1 Einleitung

Lernlabore haben zum Ziel, die naturwissenschaftliche Grundbildung zu fördern und eine positive Einstellung zu den Naturwissenschaften bei Schüler*innen zu verankern [GP08]. Für den Biologieunterricht ist dies von besonderer Bedeutung, da Einblicke in moderne biologische Forschungsmethoden und -techniken geboten werden und das selbständige, praktische Arbeiten in situierter Lernumgebung ermöglicht wird [KL14]. Eine zentrale Herausforderung ist daher immer wieder die Gestaltung interessefördernder Lernumgebungen für verschiedene Altersstufen [BS18]. Aufgrund der immensen Nachfrage der Schulen nach experimentell orientiertem Unterricht und der Möglichkeit, Primärerfahrungen im Labor zu sammeln, wird die Durchführungskapazität des Lernlabors zunehmend zum Problem. Um den Einflussbereich und die Kapazität des Lernlabors dennoch zu steigern, nutzen wir erstmals die 360° Technologie zur Schaffung einer virtuellen Lehr/Lernumgebung. Durch die virtuelle Entsprechung des realen Raumes ist eine authentische Laborumgebung erzeugt worden, die in den Schulen selbst schon zum Einsatz kommen kann. Die so geschaffene Möglichkeit zur Einbindung von Laborbesuchen in den Unterricht hilft zudem das vorhandene Potenzial zur Lernwirksamkeit voll auszuschöpfen [GP08] [GPS06]. Die

1 Technische Universität Darmstadt, Fachbereich Biologie, Schnittspahnstrasse 10, 64287 Darmstadt, klees@bio.tu-darmstadt.de
2 Technische Universität Darmstadt, Fachbereich Biologie, Schnittspahnstrasse 10, 64287 Darmstadt, santelmann@bio.tu-darmstadt.de
3 Technische Universität Darmstadt, Hochschuldidaktische Arbeitstselle, E-Learning Arbeitsgruppe, Hochschulstraße 1, 64289 Darmstadt, markus.weber@tu-darmstadt.de, https://orcid.org/0000-0002-2888-6427

doi:10.18420/delfi2024_21

entstandene 360°-Laborreise, via VR-Headset, stellt dabei ein neuartiges, virtuelles und interaktives Lernszenario dar, in dem das Lernlabor Biologie virtuell begehbar wird und es den Schüler*innen ermöglicht die 360°- Laborumgebung, explorativ und mit hohem Immersionsgrad zu erleben.[He19] Somit gelingt schon in der Schule, eine Vorstufe der gewünschten Primärerfahrungen innerhalb der 360° Laborumgebung zu sammeln.[Ol18]. Zur erfolgreichen Integration des 360° Labors in den Schulunterricht, ist es notwendig, die komplexen Rahmenbedingungen in der jeweiligen konkreten Lernsituation grundlegend neu zu erfassen und zu berücksichtigen. Eine enge Zusammenarbeit mit den Lehrkräften und Schüler*innen sind daher absolute Voraussetzung. Designorientierte Ansätze, wie der Design-Based Research Ansatz (DBR) [Co03] [Re05] ermöglichen die notwendige enge Verknüpfung und identifizieren von Anfang an die grundlegenden Implementierungsmerkmale [Gr10] schon bei der Entwicklung des didaktischen Konzepts zur erfolgreichen Unterrichtseinbindung [KT15].

2 Ausgangslage und Zielsetzungen

Zielgruppe der 360° Laborreise sind Schüler*innen der 7. Jahrgangsstufe. Im Kerncurriculum Hessen der Sekundarstufe I – Gymnasium ist allgemein im Basiskonzept Struktur und Funktion das Mikroskopieren als unverzichtbarer Aspekt für den Unterricht im Kompetenzbereich der Erkenntnisgewinnung aufgeführt. Nach Abschluss der Jahrgangstufe 9/10 müssen die Schüler*innen in der Lage sein, ein Mikroskop zu bedienen und Zeichnungen von tierischen und pflanzlichen Zellen anzufertigen [4]. Darauf aufbauend sind in der E-Phase (früher 11. Klassenstufe), laut Kerncurriculum, verpflichtende Untersuchungen im Unterricht vorgegeben auf die die Schüler*innen der Jahrgangsstufe 7-10 entsprechend vorbereitet werden müssen. Molekularbiologische Verfahren, wie diese im Unterricht für die gymnasiale Oberstufe vorgeschrieben sind, können an den Schulen aufgrund der mangelnden Ausstattung und Vorerfahrung der Schüler*innen nur theoretisch vermittelt werden. Mit unseren Laborangeboten sind wir bestrebt, für die Naturwissenschaften zu begeistern und den Biologieunterricht an den Schulen zu unterstützen und zu erweitern und den Schulen so die Möglichkeiten zu bieten, die ansonsten nicht gegeben wären. In den Schulcurricularen unserer Kooperationsschulen ist die Einführung von Untersuchungsgeräten, zur Erkenntnisgewinnung in der Biologie, in der Jahrgangsstufe 7 verortet. Hier lässt sich thematisch die Erweiterung auf die Laborgeräte sehr gut in den Unterricht integrieren. Der naturwissenschaftliche Unterricht wird in dieser Jahrgangsstufe zudem epochal durchgeführt, wodurch die Klassengröße halbiert ist und praxisorientierter Unterricht begünstigt wird. Ziel beim gewählten designorientierten Forschungsansatz ist es zudem, die Möglichkeit auszuschöpfen, im konkreten praktischen Kontext die Lernumgebungen zu gestalten und gleichzeitig Lerntheorien im Anwendungskontext zu prüfen.

4 (https://kultus.hessen.de/unterricht/kerncurricula-und-lehrplaene/kerncurricula) (11.03.2024)

3 Designorientierter Forschungsansatz

Gestaltungsorientierte Entwicklungsansätze sind geeignet, um Innovationen für den Unterrichtsalltag im konkreten Praxisbezug hervorzubringen und eignen sich gerade auch zur Etablierung digitaler Medien in den Unterricht. Dabei gilt, dass die Wirkung digitaler Medien nicht in erster Linie durch die technischen Merkmale des Mediums selbst bestimmt werden, sondern vom zugrundeliegenden didaktischen Konzept abhängen, welches durch das technische und gestalterische Potential des Mediums unterstützt wird [Ke03]. Wesentlich ist also hier, dass vor lerntheoretischem Hintergrund didaktische Lösungen für konkrete „Praxisprobleme“ gefunden werden. Essentiell ist daher, dass die Rahmenbedingungen und Implementierungsmerkmale von Anfang an mitberücksichtigt werden [RV12] [KT15]. Sichergestellt wird dieses beim designorientierten Ansatz durch das zyklische, iterative Vorgehen, wobei Gestalter*innen und Anwender*innen wechselseitig beteiligt sind. In unserem Ansatz wiederholen sich die systematische Gestaltung, Durchführung, Überprüfung und Re-Design bei der Entwicklung der Lernumgebung. Das Forschungsdesign lässt sich dabei grob in drei Phasen einteilen, in Vorprüfung, Entwicklungsphase und Beurteilungsphase in denen qualitative und quantitative Evaluationsmethoden zum Einsatz kommen können [Pl13].

Vorprüfung: In Vorgesprächen mit Lehrkräften der Kooperationsschulen wurde zunächst das „Bildungsproblem“ erörtert. Von Schulseite aus sollten dauerhaft zugängliche, praxisnahe Möglichkeiten für die Schüler*innen geschaffen werden, die Einblicke in moderne Laborgeräte der biologischen Forschung ermöglichen. Die Schulausstattung kann so etwas nicht leisten. Beispielsweise würde ein einfacher Klassensatz an Mikropipetten Kosten von ungefähr 15 tausend Euro verursachen, was bei einem Jahresschuletat für die Biologie von wenigen hundert Euro utopisch ist. Als außerschulischer Lernort ist unser Lernlabor genau darauf ausgelegt, Schüler*innen einen Einblick in ein modern ausgerüstetes Forschungslabor zu geben und die Auseinandersetzung mit der moderneren Wissenschaft zu fördern und damit auch für die Naturwissenschaften zu begeistern. Das „Bildungsproblem“ von Seiten des Lernlabors aus, stellt jedoch die begrenzte Kapazität für Präsenzbesuche im Lernlabor dar, da die Nachfrage der Schulen unsere Möglichkeiten bei weitem übersteigt. An dieser Stelle sei erwähnt, dass die Einrichtung eines biologischen Forschungslabors in Klassengröße ungefähr eine Millionen Euro beträgt, sodass eine Erweiterung des bestehenden Laborangebotes ausgeschlossen ist. Die Lösung des „Problems“ stellt nun ein mit Hilfe der 360° Technologie erstelltes Dublikat des Lernlabors dar, welches schon in der Schule den Schüler*innen ermöglicht, erste Einblicke in die Laborumgebung und Gerätschaften zu erhalten. Der Einsatz des 360°- Labors sollte dann dauerhaft in den Unterricht integriert werden. In der Vorprüfung konnten zudem die Ausgangslage und die Rahmenbedingungen zur Etablierung eines praxistauglichen Unterrichtskonzepts erörtert werden. Demnach bietet sich thematisch die 7. Jahrgangstufe an. Der zeitliche Rahmen ist auf dreimal eine Doppelstunde im Unterricht begrenzt.

Entwicklungsphase: In der Entwicklungsphase wurde zunächst das 360°- Lernlabor gestaltet (vgl. Kapitel 4) und das Unterrichtskonzept entwickelt (vgl. Kapitel 5). Hierbei wurden

entwicklungsbegleitend die Zyklen der gestaltungsorientierten Entwicklung durchlaufen. Im ersten Zyklus wurde der Prototyp des 360°- Labors entwickelt und getestet. Im zweiten Zyklus wurde das Unterrichtskonzept unter Einsatz des 360°- Labors via VR-Headset mit ausgewählten Schüler*innen ausgearbeitet und überprüft. Im dritten Zyklus steht nun die Prüfung im Unterricht der 7. Klasse an, bevor es dann in den Regelbetrieb überführt werden kann.

Beurteilungsphase: Im Regelbetrieb erfolgt die Beurteilungsphase, in der mit mehreren Schulklassen überprüft werden kann, welchen Lernerfolg die 360°- Laborreise mit sich bringt und ob die Bedürfnisse und Erwartungen der Lehrkräfte und Schüler*innen erfüllt werden konnten. Die 360°- Laborreise kann darüber hinaus nicht nur den Einflussbereich und Kapazität des Lernlabors erweitern, sondern auch auf lerntheoretischer Basis [CS91] und durch die Einbindung in den Unterricht [GP08] die Wirkung des Lernlabors bei Präsenzbesuchen positiv beeinflussen. Von Interesse wären hier: ob oder wie stark die 360°- Laborreise in der Schule zu einer kognitiven extrinsischen Entlastung bei Präsenzbesuchen im Labor führt und ob, im konstruktivistischen Kontext, dieses letztendlich zu einer Förderung des Lernerfolgs im Labor bzw. zum besseren Verständnis der Themen in der Oberstufe beiträgt.

4 Technische Umsetzung

Technische Grundlage der immersiven 360° Laborreise bilden 360° Panoramabilder, welche durch Lernvideos und Quizzfragen, die einzelnen Laborgeräte vorstellen und deren Funktionsnweise beschreiben. Die Panoramabilder wurden mit einer 360° Kamera [5] produziert. Diese Kamera eignet sich aufgrund der kompakten Bauweise und einfachen Bedienung sehr gut für die Produktion von 360° Touren im Bildungskontext. Zwei 180° Linsen erzeugen das 360° Panoramabild. Ein Stitchen [6] ist nicht notwendig, dies erfolgt automatisiert durch die Kamerasoftware. Die technische Zusammenführung der produzierten Panoramabilder, erfolgte über die Software Virtual Tour Pro. Sie eignet sich aufgrund mehrerer Faktoren sehr gut zur Erstellung von gamifizierten 360° Touren. Die Bedienoberfläche ist einfach zu erfassen und sowohl der Import der Panoramabilder, als auch deren Verknüpfung untereinander, sowie die Erweiterung mit Interaktionselementen ohne Programmierkenntnisse möglich. Spielelemente, wie kleine Quizze, sind direkt in Virtual Tour Pro erstellbar. Die Anwendung sollte für den Schuleinsatz bewusst ohne Internetzugang über ein VR Headset, spiel- und erlebbar sein. Aus diesem Grund war es notwendig, Spielelemente, welche für eine browserbasierte Anwendung bspw. in H5P umgesetzt werden könnten, neu zu denken, und unter Berücksichtigung, der von Virtual Tour Pro möglichen technischen Umsetzungsvarianten, zu realisieren. Dies stellte auf produktionstechnischer, aber auch didaktischer Ebene, eine große Herausforderung dar. Hier waren der Umbau von Spielelementen (z.b. Drag an Drop, etc.) und eine didaktische Reduktion notwendig. Der Interaktionsgrad von gewünschten

5 Gopro Max 2

6 Zusammenfügen von Bildern zur Erstellung von segementierten Panoramabilder

Spielelementen stand den realen, technische Umsetzungsmöglichkeiten gegenüber. Die Abwägung darüber war Entscheidungsgebend für die finale technische Realisierung. Somit stand die Nutzung der Anwendung im VR Headset und die damit ermöglichte Immersion vor dem Gamifizierungs- und dem Interaktionsgrad der Anwendung. Neben den Panoramabildern stellen die verschiedenen, multimedialen Contentbausteine, welche die Laborgeräte erklären, eine zentrale Säule der Inhaltevermittlung dar. Dabei handelt es sich um Videos und Audiofiles, welche die Funktion der Laborgeräte in kurzer Form beschreiben und deren Einsatzzwecke erläutern. Für jedes Laborgerät wurde ein fachdidaktisch aufbereitetes Kurzvideo produziert. Die Videos haben eine Länge von ein bis zwei Minuten. Über Quizze wird das Verständnis zum Inhalt der Lernvideos spielerisch überprüft. Auch hier spielt der Immersionsmoment eine große Rolle. Das Video erscheint in einem VR Headset ähnlich groß, wie auf einer Kinoleinwand. Der Detailgrad der Beobachtung wird dadurch verstärkt, die Schüler*innen haben das Gefühl näher an den Objekten zu sein und können viel tiefer in die dargestellten Nutzungsmomente eintauchen und diese wahrnehmen. Virtual Tour Pro bietet unterschiedliche Veröffentlichungsformen der erstellten 360° Touren an. In unserem Fall erfolgte diese für die VR Headsets Quest II und III. Insgesamt sind acht Headsets im Einsatz. Das Aufspielen erfolgt manuell über eine kabelgebundene Verbindung zum PC. Die Nutzung des VR Headsets erfolgt kabelungebunden und frei im Raum.

5 Didaktische Umsetzung

Das Unterrichtskonzept für die 360° Laborreise beinhaltet drei aufeinander aufbauende Termine, die sich über je eine Schuldoppelstunde erstrecken. Am ersten Termin kommt das VR-Headset zum Einsatz und die Schülerinnen und Schüler spielen die 360° Laborreise. Am zweiten und dritten Termin kommt das VR-Headset nicht mehr zum Einsatz. Der zweite Termin dient der Ergebnissicherung und dazu, das erlernte Wissen vom ersten Termin zu rekapitulieren. Am dritten Termin erfolgt die Vertiefung indem Teile des erlernten, theoretischen Wissens zur praktischen Anwendung kommen. Als Leitfaden und für die Ergebnissicherung wurde ein Skript erstellt, welches die Schüler*innen durch die einzelnen Unterrichtsstunden leitet. Die erste Doppelstunde gliedert sich in drei inhaltlich aufeinander aufbauende Abschnitte: Onboarding, Tutorial zur Handhabung und die 360° Laborreise. Das Onboarding dient zur Einführung des VR-Headsets und zur Vorbereitung der Spielenden auf die neue Technik. Hierfür wurde speziell ein Onboarding-Video produziert. In diesem Video wird das VR-Headset erklärt und allgemeine Hinweise zum Umgang und zur Handhabung der VR-Brille erläutert. Die Durchführung der Laborreise erfolgt aufgrund der Rahmenbedingungen (acht VR-Headsets vorhanden, gegenseitige Hilfestellung) im Zweier-Team. Im Anschluss an das Onboarding Video bekommt jedes Schüler*innen Zweier-Team ein VR-Headset und Begleitskripte ausgehändigt. Das VR-Headset wird von den Schüler*innen nacheinander verwendet, sodass sich das Zweier-Team aufteilt. Eine Person nutzt das VR Headset, die zweite Person fungiert als Begleitperson. Die Person mit dem VR-Headset startet die 360° Laborreise und durchläuft zunächst ein Tutorial. Das Tutorial dient dazu, sich an die 360° Umgebung zu gewöhnen, sowie die Funktionen

der Hard- und Software kennenzulernen. Im Labor werden insgesamt 23 Geräte, deren Funktionsweisen und Einsatzzwecke, tierfergehend durch multimediale Inhalte (wie bspw. 2D-Videos und Audiofiles) erklärt. Spielelemente, wie Quizze, tragen zur Gamifizierung des Lernszenarios bei. Während die Person mit VR Headset sich im virtuellen Labor befindet, protokolliert die Begleitperson die Informationen zu den gefundenen Geräten im ausgehändigten Begleitskript. Am zweiten Termin erfolgt der Einstieg spielerisch in Form eines Ratespiels, welches der Wissensaktivierung und der Ergebnissicherung dient. Ziel des Spiels ist es, alle Laborgeräte benennen zu können. Im Anschluss erfolgen Kurzpräsentationen der einzelnen Teams zu ausgewählten Laborgeräten. Hierfür steht weiteres Unterrichtsmaterial zur Verfügung. Unbeantwortete Fragen können im Skript notiert und am dritten Termin den Expert*innen aus der Universität gestellt werden. Am dritten Termin kommt nun das theoretische Wissen zu den Laborgeräten zur praktischen Anwendung. Der dritte Termin kann dabei optional durch einen Besuch im Lernlabor ersetzt werden. Ist ein Besuch nicht möglich, werden verschiedene Kleingeräte aus dem Labor in die Schule mitgebracht. Die Schüler*innen haben somit die Möglichkeit, im Rahmen von kleineren Experimenten einige dieser Geräte praktisch einzusetzen und Primärerfahrungen zu sammeln.

6 Fazit und Ausblick

Im Rahmen der Entwicklungsphase wurde die 360° Laborreise erstmalig in der Schule mit einer Biologie-AG bestehend aus sieben Schüler*innen der Jahrgangstufen 5-7 durchgeführt. Durch die Pilotdurchführung konnten erste Erfahrungen zum Unterrichtskonzept und zu weiteren Rahmenbedingungen in den Schulen gewonnen werden. Die Erkenntnisse beruhen auf den Beobachtungen der Entwickler und der Lehrkräfte und auf Feedbackgesprächen mit den Schüler*innen. Die Vorbereitung der Schulräume war ein wichtiger Aspekt, um einen reibungslosen Ablauf der Unterrichtseinheiten zu gewährleisten. Der Einsatz der VR-Headsets benötigt im Klassenraum genügend Platz, wobei eine Fläche von 160x160 cm pro Zweier-Team ausreichend war. Die Anwendung sollte auf dem VR-Headset ohne Internetzugang nutzbar sein, da dieses in den Schulen meist nicht oder nur unter bestimmten Bedingungen verwendet werden kann. Die Wissensaktivierung und Ergebnissicherung am zweiten Termin verliefen reibungslos. Die Schüler*innenvorträge unter Verwendung des Zusatzmaterials funktionierten auch in dieser Altersklasse problemlos. Die Handhabung der Laborgeräte beim experimentieren am dritten Termin verlief sehr gut. Die Schüler*innen zeigten einen sicheren Umgang beim Einsatz der Geräte. Abschließend kann festgestellt werden, dass der Einsatz des 360° Labors eine Möglichkeit darstellt, um spielerisch und immersiv moderne Laborgeräte und deren Einsatzzwecke im Labor kennenzulernen. Die 360° Laborreise stellt damit eine innovative und besondere Art der Verknüpfung von außerschulischem Lernort Lernlabor und Schule dar. Als Ausblick bleibt spannend, welche Auswirkungen diese Art der Vorbereitung auf die praktischen Arbeiten und den Lernerfolg der Schüler*innen zeigt.

Literaturverzeichnis

[BS18] Blankenburg, Janet; Scheersoi, Annette: Interesse und Interessenentwicklung. Theorien in der naturwissenschaftsdidaktischen Forschung, S. 245–259, 2018.

[Co03] Collective, Design-Based Research: Design-based research: An emerging paradigm for educational inquiry. Educational researcher, 32(1):5–8, 2003.

[CS91] Chandler, Paul; Sweller, John: Cognitive load theory and the format of instruction. Cognition and instruction, 8(4):293–332, 1991.

[GP08] Guderian, Pascal; Priemer, Burkhard: Interessenförderung durch Schülerlaborbesuche-eine Zusammenfassung der Forschung in Deutschland. PhyDid A-Physik und Didaktik in Schule und Hochschule, 2(7):27–36, 2008.

[GPS06] Guderian, Pascal; Priemer, Burkhard; Schön, Lutz-Helmut: In den Unterricht eingebundene Schülerlaborbesuche und deren Einfluss auf das aktuelle Interesse an Physik. PhyDid A-Physik und Didaktik in Schule und Hochschule, 2(5):142–149, 2006.

[Gr10] Gräsel, Cornelia: Stichwort: transfer und Transferforschung im Bildungsbereich. Zeitschrift für Erziehungswissenschaft, 1(13):7–20, 2010.

[He19] Hebbel-Seeger, Andreas; Riehm, Philipp; Baranovskaa, Marianna et al.: LectureCast als 360°-Video. Welchen Einfluss haben Immersion und Präsenzerleben auf die Lernleistung? In: Teilhabe in der digitalen Bildungswelt. Münster: Waxmann. S. 118–127, 2019.

[Ke03] Kerres, Michael: Wirkungen und Wirksamkeit neuer Medien in der Bildung. In: Education Quality Forum. Wirkungen und Wirksamkeit neuer Medien. Münster: Waxmann. S. 31–44, 2003.

[KL14] Knogler, Maximilian; Lewalter, Doris: Design-Based Research im naturwissenschaftlichen Unterricht. Psychologie in Erziehung und Unterricht, 61:2–14, 2014.

[KT15] Klees, Guido; Tillmann, Alexander: Design-based research als Forschungsansatz in der Fachdidaktik Biologie. Entwicklung, Implementierung und Wirkung einer multimedialen Lernumgebung im Biologieunterricht zur Optimierung von Lernprozessen im Schülerlabor. Journal für Didaktik der Biowissenschaften, 6:91–110, 2015.

[Ol18] Olmos-Raya, Elena; Ferreira-Cavalcanti, Janaina; Contero, Manuel; Castellanos, M Concepción; Giglioli, Irene Alice Chicchi; Alcañiz, Mariano: Mobile virtual reality as an educational platform: A pilot study on the impact of immersion and positive emotion induction in the learning process. EURASIA Journal of Mathematics, Science and Technology Education, 14(6):2045–2057, 2018.

[Pl13] Plomp, Tjeerd et al.: Educational design research: An introduction. Educational design research, S. 11–50, 2013.

[Re05] Reinmann, Gabi: Innovation ohne Forschung? Ein Plädoyer für den Design-Based Research-Ansatz in der Lehr-Lernforschung. Unterrichtswissenschaft, 33(1):52–69, 2005.

[RV12] Reinmann, Gabi; Vohle, Frank: Entwicklungsorientierte Bildungsforschung: Diskussion wissenschaftlicher Standards anhand eines mediendidaktischen Beispiels. Zeitschrift für E-Learning–Lernkultur und Bildungstechnologien, 4:21–34, 2012.

Open Science und OER

Towards Open Science at the DELFI Conference



Natalie Kiesler [1], René Röpke [2], Daniel Schiffner [3], Sandra Schulz [4], Sven Strickroth [5], Matthias Ehlenz [6], Birte Heinemann [7], and Arno Wilhelm-Weidner [8]

Abstract: Despite the increasing awareness of Open Science within the educational technology community, conferences, such as DELFI, do not yet foster the publication of research data including software. To address this, we conducted a survey eliciting the community's needs, perspectives, and publication preferences. The analysis of 24 valid responses reveals a variety of research data formats used, and several uncertainties, e. g., regarding data ownership. Associated barriers comprise legal concerns and lacking resources to publish data. Nonetheless, researchers seem open for new publication formats. Moreover, we analyzed author's intentions to publish data related to their DELFI submissions in 2023 (n=66). Many researchers assume not to have data to share (n=28), or no intention to publish data in the future (n=16). Overall, the results imply a lack of awareness and recognition of data publications, so that further efforts and incentives are required to move toward Open Science practices in the DELFI community.

Keywords: Open Science, Open Data, Data publication, FAIR, DELFI, Educational Technology

1 Introduction

In educational technology research (EdTec), as in most scientific domains, the push towards Open Science has gained significant momentum, underscored by the increasing demands of funding institutions for greater transparency and accessibility. Initiatives such as the Open Science Framework [FD17] and Open Data mandates by major organizations [Eu23; Na13; PF07] highlight a global shift towards open research practices that are more inclusive, collaborative, and accessible to a broader audience. **F**indability, **A**ccessibility, **I**nteroperability, and **R**eusability [Wi16] have emerged as core principles (FAIR), providing a comprehensive guideline for enhancing the openness and usability of scientific data.

1 Nuremberg Tech, Computer Science, Keßlerplatz 12, 90489 Nuremberg, Germany, natalie.kiesler@th-nuernberg.de, https://orcid.org/0000-0002-6843-2729
2 TU Wien, Learning Technologies and eDidactics, Favoritenstr. 9-11, 1040 Vienna, Austria, rene.roepke@tuwien.ac.at, https://orcid.org/0000-0003-0250-8521
3 DIPF | Leibniz Institute for Research and Information in Education, Rostocker Str. 6, 60323 Frankfurt am Main, Germany, d.schiffner@dipf.de, https://orcid.org/0000-0002-0794-0359
4 University of Hamburg, Faculty of Education, Von-Melle-Park 8, 20146 Hamburg, Germany, sandra.schulz@uni-hamburg.de, https://orcid.org/0000-0002-2254-6579
5 LMU Munich, Institute for Informatics, Oettingenstraße 67, 80538 Munich, Germany, sven.strickroth@ifi.lmu.de, https://orcid.org/0000-0002-9647-300X
6 RWTH Aachen University, Lehrerbildungszentrum, MediaLab Lehramt, Kármánstr. 17-19, 52056 Aachen, Germany, ehlenz@lbz.rwth-aachen.de, https://orcid.org/0000-0001-6189-6056
7 RWTH Aachen University, Learning Technologies, Ahornstr. 55, 52074 Aachen, Germany, heinemann@cs.rwth-aachen.de, https://orcid.org/0000-0002-7568-0704
8 VDI/VDE-IT, Bildung und Wissenschaft, Steinplatz 1, 10623 Berlin, Germany, arno.wilhelm-weidner@vdivde-it.de, https://orcid.org/0000-0003-2604-3327

doi:10.18420/delfi2024_22

However, it is not common practice in the EdTec community to publish research data (including software) along with paper publications [KS22; KS23a]. Especially the DELFI conference by the German Informatics Society (GI) offers little room for the publication of research data, as its submission is not obligatory [Ge24b]. The limited availability of research data from DELFI papers published between 2018 and 2022 [MS23] confirms the lack of Open Data practices. We, as a community, thus need to critically assess and improve our practices to foster an alignment with Open Science.

This paper aims to provide insights into the current state of Open Science practices within the EdTec community, specifically focusing on the DELFI as a German EdTec conference. Based on the results of an online survey, we examine the community's perspective on publication practices, the characteristics of the research data, the barriers and challenges to open publication, and the preferences and expectations of researchers. Furthermore, we explore the intentions of authors with papers accepted at DELFI 2023 regarding the publication of their data, leveraging information gathered via the paper submission system. By providing these insights, we aim to start drafting a path forward towards Open Data practices at future DELFI conferences.

2 Background

Finding, accessing, understanding, and reusing data produced by others are challenges reported by many researchers. Likewise, replicating prior work is often difficult due to the lack of sufficient details [Ih15; Je22; Ki23; Pr23]. Suppose software is used as a method for research. In that case, it is hardly recognized within the community, rarely published, and thus not available, causing the reinvention of the wheel over and over again [KS22]. In particular, we observe the following challenges related to searching and reusing data: (1) Lack of (long-term) availability, e. g., loss of files, limitations through proprietary or deprecated software [GNT10; KGH21; KPC21]; (2) Lack of comprehensibility, and data maturity e. g., missing documentation due to required effort, time, and resources; (3) Lack of recognition for high quality, mature research data, e. g., publication is not recognized within the academic community [Mo11]; (4) Lack of obligatory reviews and guidelines, i. e., data submissions are rarely allowed/supported, and data review processes are not defined [KS23b]; and (5) Lack of data provenance, the data's origin, data preprocessing and cleaning steps, errors, or other aspects relevant for secondary research [CF10].

At the same time, publishing data can entail several obstacles for researchers. For example, Strickroth and Striewe [SS22] had to publish their research data on a personal website due to missing features on existing platforms and the lack of workflows for collaborative maintenance. Other limitations apply to research data centers, such as the German Centre for Higher Education Research and Science Studies [Ge23a]. Data publication processes are demanding and involve several time-consuming steps (over several months), people, and additional metadata and reports [Ki22a; Ki22b; SBH23]. Irwin [Ir13] adds ethical concerns due to the sensitivity of some (qualitative) data.

In an interview study with 13 researchers (from Germany, Peru, India, and China) in the context of learning analytics, a number of barriers preventing researchers from the publication of their data were identified. Among them are the following categories [BP21]:

- Legal concerns – barriers related to copyright or other legal considerations that may impact data publication;
- Fear of loosing control – barriers concerning the sovereignty over the data;
- Authority or practice considerations – established guidelines or research practices within a particular field affecting the acceptance of data publications;
- Technical / processing barriers – challenges associated with the publication process;
- Resource barriers – e. g., limited funding, human resources, or infrastructure hindering data publications.

Even though these concerns can vary across countries, the human and financial resources required for the preparation and publication of data, and the lack of infrastructure remain a common barrier, especially in countries with low incomes. Another result of the study is that most of the researchers have not published their data [BP21].

In a conservative estimate, a report on behalf of the European Commission calculates opportunity costs of 10.2 billion Euros due to the lack of FAIR research data [CRI18]. McGill [Mc19] concludes the lack of open research practices contributes to decreasing research standards. Therefore, it is crucial for EdTec researchers and the DELFI community to start focusing on research data and to foster data publications, including software.

3 Methodology

This paper aims to gather the EdTec community's perspective on the publication of research data so that we can start developing feasible formats for researchers to publish their data at the DELFI conference. Our research questions are as follows:

1. How can we characterize research data in the EdTec context?
2. What are the barriers and challenges associated with the publication of research data in the EdTec context?
3. What are the expectations and preferences regarding the publication formats of research data at the DELFI conference?
4. What were the author's intentions regarding the publication of their research data at the DELFI conference in 2023?

To address RQ1, RQ2, and RQ3, we adapted an available survey instrument [Bi23],which is aligned with the 2016 Open Data Survey framework [Ce17]. Moreover, we tailored it to the specifics of the DELFI community concerning the definition of research data formats and publication formats. The survey also incorporated common barriers for researchers attempting to publish their data. These barriers are categorized into five areas: (1) legal concerns, (2) loss of control of data, (3) technical constraints, (4) authority or practice

considerations, and (5) resource constraints [BP21]. In addition, the survey included questions about potential conference formats and how they accommodate the requirements for publishing research data, including software (see Appendix A). Survey questions Q1–Q8 address RQ1, questions Q9–Q14 address RQ2, and survey questions Q15–Q18 help answering RQ3. Most questions are multiple-choice questions, with few open input answering options (see also [Bi23]).

As part of the survey introduction, respondents were provided with a definition of research data based on the German Research Foundation (DFG), which includes, among others, "*measurement data, laboratory values, audiovisual information, texts, survey or observation data, methodological test procedures, and questionnaires. Compilations and simulations can likewise constitute a key outcome of academic research (. . .). The same applies if the software is required to create or process research data.* [Ge23b]". At the beginning of the survey, we asked participants to think of a recent representative educational research project in which research data had been created or gathered. The full survey will be available [upon acceptance]. Conducting the survey was approved by the Ethics Board of the DIPF | Leibniz Institute for Research and Information in Education.

The online survey was distributed between September and December 2023 within the German-speaking EdTec community following the purposeful sampling method [Pa02]. It should be noted that this community is considerably small. Specifically, we distributed our survey at the DELFI conference 2023, the GI's Doctoral colloquium workshop, and via the GI's educational technology group newsletter. This led to 58 responses (18 full, 40 partial). Due to the high number of partial responses, we had to define inclusion/exclusion criteria: (1) The processing time of the survey should be longer than 120 seconds, and (2) respondents had to fill out at least the initial five questions to remain in the sample. Responses that did not meet these requirements were ignored from the analysis. To analyze the remaining 24 response sequences (18 full, 6 partial), we use frequency and descriptive statistics for the quantitative results and thematic analysis [CB14] for open-ended, qualitative responses. Representative quotes highlight and contextualize the identified themes.

To determine the authors' intentions to publish their research data in the DELFI context (RQ4), an additional evaluation was conducted as part of the DELFI 2023 paper submission process. All authors were asked to specify their intentions by selecting one of the following options: (a) *I have already published my research data*, (b) *I will publish my research data*, (c) *I do not want to publish any research data* and (d) *I do not have any research data for publication*. Authors were also asked to provide links or references to their data. For the analysis, only the authors' responses from the 66 accepted papers (in the categories of long paper, short paper, demo, or poster) were considered. Moreover, if authors indicated the intention to publish their data or to have published their data already, we reviewed the DELFI 2023 proceedings, i. e., to examine whether links to external resources and data were provided.

4 Results

In this section, we summarize the findings for each research question, i. e., data characteristics (RQ1), associated barriers (RQ2), researchers' expectations and preferences (RQ3), and their intentions to publish data (RQ4). Illustrative figures highlight some of the results. The research data is published in an online repository [Ki24b].

4.1 Characterization of the Research Data (RQ1)

To address the first research question, we asked a number of questions related to the research data commonly gathered or used in EdTec projects. The goal of the respective survey questions was to characterize the research data (Q1–Q5), to identify data ownership and currently used publication formats (Q6–Q8).

First, the participants were asked what research data they used in one of their last EdTec projects (see Appendix A). Most of them reported using software developed by themselves (n_t=16) or by others (n_o=16), qualitative data collected by themselves (n_t=18) or by others (n_o=8), and quantitative data collected by themselves (n_t=17) or by others (n_o=7). For the answer option *other data*, the video and audio formats were selected by two participants.

In response to Q2, the following specific research formats were selected: plain text (n=16), structured text (n=18), general purpose formats (n=19), domain-specific formats (n=7), multimedia (n=17), compiled binary artifacts (n=8), and source code (n=14). As part of Q3, the participants chose the following research formats (see Fig. 1), which are important to be made available to others for the purpose of validation: plain text (n=12), structured text (n=16), general purpose formats (n=14), domain-specific formats (n=4), multimedia (n=13), compiled binary artifacts (n=5), and source code (n=11). Additionally, some participants indicated that none of these research data formats are important to be made available (n=2).

Regarding their last EdTec project, the participants described the volume of data they created and intended to share (Q4). Six participants reported having data between 10 and 400 megabytes. Other participants described a volume of 3 to 20 gigabytes (n=11). No one reported to have a volume of terabytes. Only one person did not know the volume of their

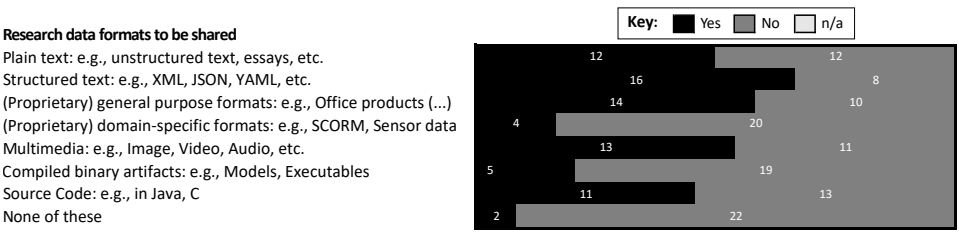


Fig. 1: Reported responses regarding research data formats to be shared (Q3)

data. Furthermore, researchers were asked for the approximate number of data files created in the last EdTec research project (Q5). Some participants did not know (n=8), and others (n=13) chose values between 5 and 5000.

When asked about data ownership (Q6), three respondents answered that they did not know who owned the data before or after publication. Most respondents answered that the data belonged to either *themselves* (n=12), *the project* (n=8) or *the institution* (n=5) before publication. After publication, it was assumed that the data belongs to *them* (n=8), *the project* (n=7), *the institution* (n=5) or others (n=5).

In response to Q7, most researchers stated they had previously published research data either in *a public software repository, e. g., GitHub, GitLab, DockerHub* (n=12), in *a publication's appendix* (n=7), in *a repository of their institution* (n=6) or as *a pure data publication* (n=5). The used publication platforms were *OSF* [Ce24], *Zenodo* [CE24], and *ArXiv* [Co24].

There were only a few responses to Q8, which asked for reasons why research data was not published. The most common responses are *too many barriers to overcome* (n=4), that *the research data was not sufficiently documented* (n=3), or that *the process of anonymization was not clear* (n=3). In the free text responses, it was noted that not all data could be anonymized, that the data was required for further processing, or that the study subjects had not given consent.

4.2 Associated Barriers (RQ2)

The second research question aims to identify barriers and challenges associated with the publication of research data in the EdTec context. In the survey, Q9–Q14 address well-known barriers, and specific challenges.

The barriers were prompted in a two-layer approach. First, participants could select up to five superordinate barriers (Q9) they faced in their last EdTec research project. Second, for each selected barrier area, subordinate, more specific barriers and concerns were surveyed (Q10–Q14). The respondents received those survey questions depending on their selection of barriers in Q9. Item Q9 was answered by 22 participants. Five of the participants did not experience any of the given superordinate barriers in their last research project. The others had experienced between one and five barriers (median: 2). The most frequently experienced barrier in the last project relates to *legal concerns* (n=12), followed by *resource constraints* (n=10), fears for *loss of control* (n=9), *technical/processing constraints* (n=5, e. g., data set too big, complicated), and *authority or practice considerations* (n=4).

In the following, interesting results of Q10–14 are highlighted (see Figure 2). Two participants discontinued the questionnaire after Q9, resulting in *n/a* answers for Q10–Q14. Participants experienced up to seven legal barriers or concerns (median: 2, Q10). The most frequently selected option in Q10 (legal concerns) was *It is difficult to balance privacy and openness* (n=8), followed by *I'm not sure how to deal with personal or sensitive data* (n=5), and

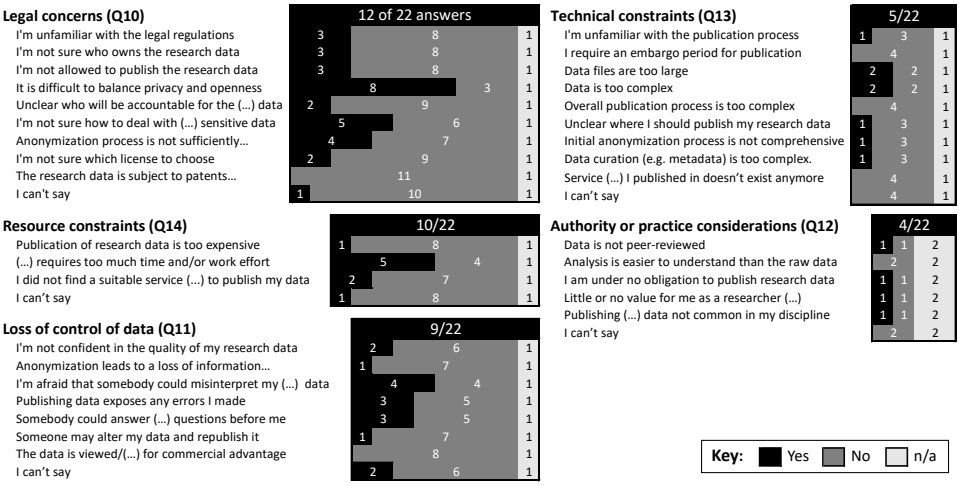


Fig. 2: Overview of the selected barriers and concerns. The number of researchers with issues in an area (based on Q9) is shown in each table header. The blocks below show the responses to Q10–Q14.

Anonymization process is not sufficiently secure (n=4). Three researchers explicitly selected *I'm not allowed to publish the data*. Nobody experienced problems with patents. Up to five barriers or concerns concerned the *Loss of control* (Q11) (median: 0). Two participants selected *no* for all options, and two selected *I can't say* as well as *no* for all other options. The most frequent reply was *I'm afraid that somebody could misinterpret my research data* (n=4). The following options were selected by three participants each: *somebody could misinterpret my research data*, and *Somebody could answer one of my research questions before me*. There were no reported concerns about research data being commercially exploited. All researchers (n=2) answering Q12 about *authority or practice considerations* selected *no* or *I can't say*, or *Analysis is easier to understand than the raw data*. The remaining replies are inconclusive (see Figure 2). Two of the five researchers who experienced barriers and concerns regarding *technical/processing constraints* did not select any of the provided options in Q13. Up to five barriers and concerns were experienced by the other three researchers (median: 0). The most frequently selected options were *Data files are too large* and *Data is too complex*, both selected twice. Four respondents each explicitly stated that the following aspects are not perceived as barriers: *I require an embargo period for publication*, and *Overall publication process is too complex*. Regarding the *resource constraint* (Q14), researchers often stated that the *Publication (. . .) requires too much time and/or work effort* (n=5). Two researchers selected *I did not find a suitable service* and according to one researcher, *Publication (. . .) is too expensive*. One researcher selected *I can't say*.

In the additional open-ended answer fields of Q10–Q14, the anonymization of audio recordings was mentioned as an issue. Furthermore, two respondents stated each that the additional effort is not recognized by the community: "on the contrary, you get punished if

you try". In addition, data tracks were identified to be missing, and the competencies of reviewers were questioned.

4.3 Expectations and Preferences for Data Publications at DELFI (RQ3)

Regarding EdTec researchers' expectations, another series of questions were asked, including past experiences with publishing, data formats and data uploads (Q15–Q17). Another question addressed how reviewers should handle these submissions (Q18).

When asked about the publication of the research project (not the data), 15 out of 20 respondents said that their research is or will be published in the context of an EdTec conference (Q15). Only 4 could not answer this question and chose *I can't say*. In the additional input field, two respondents added that the data could not be published or the project was not considered to be public.

Q16 addressed current conference formats and their appropriateness, and which new formats should be considered. We provided two options on existing formats (*Demo/Tools* and *Full paper tracks*) that could be extended with specifics of research software and data. Of 19 responses, 13 agreed to a more detailed *Demo/Tools track*, including additional materials and guidelines. Eleven selected the *Full paper track* and support the addition of data material and software. For the new tracks, a slight majority chose a six-page format for software within a new *Tools track* (11 out of 19) compared to the twelve-page variant. 13 participants selected a specialized *Data paper track* only for research data comprising six pages. In general, there is a positive trend towards the inclusion or extension of existing tracks with research software and research data (59 %).

When asked for supplementary material and the upload of metadata and data (Q17) along with a publication, ten participants out of 15 responded with yes, confirming its necessity. This includes the possibility of providing hyperlinks to external resources.

The last question (Q18) focused on researchers' expectations when reviewing software and research data. 14 participants answered the question and provided several common themes in their responses: Reviewers require guidelines how the artifacts should be evaluated, with FAIRness and reproducibility as driving factors. Quality is also considered important, but only on the metadata level. The participants included examples to illustrate what might help guide or incentivize the authors of research data, e. g., tutorials to setup the software, explanations how to run/read the files, or Badges.

4.4 Authors' Intentions to Publish Research Data at DELFI (RQ4)

The analysis of authors' intentions to publish their data reveals that for 28 of the 66 papers (42 %), authors stated to have no research data for publication. For 19 of the 66 papers, the

authors have no intention to publish their research data. Moreover, only 3 authors indicated they had published their research data already. Indeed, all of these three published papers contain hyperlinks to respective repositories. Two point towards source code, while one link points towards Zenodo [CE24].

Among the 16 (out of 66) papers for which authors indicated the intention to publish their research data, there are only 4 published papers with a link or reference to a data repository. These links point to code repositories (e. g., GitLab or GitHub). For the other 12 of these 16 papers, the authors did not seem to have published their data within the scope of the DELFI 2023 proceedings. However, they may or may not have published the data elsewhere.

In Table 1, we summarize the authors' intentions by paper type. Taking a look at the 6 full papers, which by the community's definition [Fa24] require the presentation of advanced or completed research work and therefore may contain empirical data or developed software, authors indicated the intention to publish the research data in only one of the six cases. The published paper, however, contains no reference or link to published data. For the five remaining full papers, the authors did not state the intention to publish any data (4) or do not have any data for publication (1). Similarly, for the 21 demo papers, most authors indicated they have either no data for publication (9) or do not want to publish it (5). For the remaining seven demo papers with authors' intentions to publish the data, only three papers contained links to published research data.

	full	short	demo	poster	$\sum$
No data for publication	1	8	9	10	28 (42 %)
No intention to publish	4	6	5	4	19 (29 %)
Intention to publish in the future	1	6	6	3	16 (24 %)
Already published the data	0	2	1	0	3 (5 %)

Tab. 1: Authors' intentions to publish the research data classified by paper type.

5 Discussion

The survey results reveal several points worth discussing with regard to our research questions. For brevity, we highlight and discuss the most interesting findings before pointing out the limitations of this work.

Regarding RQ1, survey results show that there is a high heterogeneity of data sources, types, and usages. One observation is the potential misunderstanding of Q1, where respondents may have had varying interpretations of the "data by others" answer option. It was selected by 16 respondents, which was an unexpectedly high number. Respondents may have referred to coauthors or doctoral students when choosing this option. Concerning data ownership (Q6), many respondents seem uncertain about who owns the data before and after publication. Results indicate varying assumptions, with some attributing ownership to themselves,

the project, or the institution. The term "the project" raises questions about its definition and seems to indicate false assumptions within the research community. While software repositories are widely used for sharing research outputs, the survey suggests that they may not be universally suitable for different types of data (Q7). This finding highlights the need for diverse and specialized data repositories tailored to accommodate the variety of data formats and research domains.

In response to RQ2, missing recognition and large efforts to prepare data remain key challenges. Interestingly, the question addressing barriers to data sharing itself emerged as a potential barrier, as indicated by a notable drop in the number of responses. Researchers may find the topic challenging, they might have been overoptimistic in their claimed data publication goal in the beginning, or barriers were not well-defined. Particularly, legal concerns on balancing privacy and openness are perceived to limit data publication practices. These findings underline the need for targeted efforts to reduce these barriers and streamline the data sharing process.

Looking at the expectations for data publication (RQ3), we see that the author's willingness to publish software or research data seems rather high in principle. However, the survey results show a demand of more appropriate formats that actually reflect the challenges related to creating software or processing research data. Especially defining individual tracks that focus on demos or data (n=13 more detailed demo, n=11 tools track, n=13 data track) might be worthwhile for upcoming venues. Evaluating the open responses to the evaluation criteria (Q18), a majority of respondents defines topics or tasks that avoid additional load on the reviewers. Some explicitly expressed this, making it obvious that good guidelines need to be created. Thus, not only reviewers but also authors should be supported when handing in research data or software.

The analysis of authors' intentions to publish their data in the DELFI 2023 context (RQ4) raised further questions, e. g., related to demo submissions. Assuming demos present developed research prototypes, it seems as if the majority of authors at DELFI 2023 do not categorize software as research data. However, for demos presenting only the innovative usage of existing software, the authors correctly stated they have no data to publish. Furthermore, the limited availability of references to published research data in all papers at DELFI 2023 underlines the missing recognition. It is tempting to assume the lack of recognition by the community negatively impacts authors' willingness to take on the extra work of preparing and publishing their research data. In the future, it should be evaluated how awarding open data badges for publications with published research data changes the perceived recognition and authors' intentions to publish their research data.

To sum up, researchers' have issues categorizing research data. They also assume and experience a variety of barriers while showing a general intent to publish their research data and software. These results underscore the need for training and guidance to support open science practices in the EdTec community.

A limitation of the presented study is the relatively low participation rate. This may partially be due to the small community, and the self-selection biases to participate or not. At the same time, it is another result of the survey, as Open Science and related practices do not seem to play a major role in the EdTec community, yet. Therefore, we consider it important to continue raising researcher's awareness, and offering support measures and incentives to engage with Open Data and Open Science practices.

6 Conclusions and Future Work

This paper presents the results of a survey study in the EdTec community that explores key areas of concern related to data ownership, sharing, and associated barriers of open science practices. We conclude that the respondents use software as well as qualitative and quantitative data by themselves and others in a broad range of formats, and recognize them as research data (RQ1). However, challenges arise when it comes to the ownership of the data. Several barriers were identified, some of a technical nature and some concerning how much data publication efforts are being recognized (RQ2). Preferred publication formats include more detailed demo and tools tracks, the enhancement of full paper tracks by the publication of the associated research data and the introduction of special data paper tracks (RQ3). Only a minority of authors published or intended to publish research data in the context of DELFI 2023 (RQ4), raising questions of missing recognition by the community and its negative impact for data publications.

Future work may entail a comparison of the results with a survey study conducted in the context of the CompEd conference [Ki23]. While the survey study was conducted in analogy, also based on [Bi23], the audience was computing education researchers worldwide [Ki24a]. The comparison may allow the identification of more barriers and issues regarding research data publications and help guide further research in support of increasing the awareness and recognition of open science practices. Furthermore, we aim to pursue the continuation of this research, e. g., by repeating the survey at the next DELFI conference, or by adding a qualitative study. For example, authors who did not publish research data along with their publications can be interviewed to elicit barriers, preferences, and intentions. Qualitative insights may also guide the design and dissemination of educational materials on open science practices for EdTec and related research communities.

Moreover, dedicated formats for the publication of research data and guiding authors during publication processes are needed as part of conferences. More networking and visibility of Open Data as a relevant topic in the EdTec community is required in the future. There is potential, for example, in the connection with other relevant actors, such as the German Research Foundation (DFG)-funded NFDIxCS [Ge24a] project, which works towards establishing standards, providing specialized data storage offers, and community support. The latter includes guidelines for publishing data as well as supporting reviewers.

A Survey Questions

Q1 **Categories of research data used**: Please select all the categories of research data that you used in one of your last educational technology research projects.

Q2 **Research format(s) used**: Please select the research data formats that you used or created as part of your last educational technology research project:

Q3 **Research data format(s) to be shared**: Which of the following research data formats you have indicated that you used or created as part of your last educational technology research project are important to be made available to others for the purpose of validation?

Q4 **File size**: Approximately, what is the volume of the data you created and intend to share as part of your last educational technology research project? If you have none, please insert zero.

Q5 **Number of data files**: Approximately how many files of data did you produce as part of your last educational technology research project? If you have none, please insert zero.

Q6 **Research data ownership**: Who do you believe 'owns' the research data that you have made or will make available to others as part of your last educational technology research project?

Q7 **Publishing research data**: Have you published the research data that you used or created as part of your last educational technology research project in any of the following ways?

Q8 **Why not publish research data**: Why haven't you published your research data?

Q9 **Barriers to publishing data**: Thinking about your educational technology project still, has any of the following limited your ability to publish research data?

Q10 **Barriers and concerns regarding the publication of research data**: What were the legal concerns that you had regarding publishing research data? (depends on Q9)

Q11 **Barriers and concerns regarding the publication of research data**: What were your concerns regarding loss of control when publishing research data? (depends on Q9)

Q12 **Barriers and concerns regarding the publication of research data**: Was one of these authority or practice considerations a reason for not publishing research data? (depends on Q9)

Q13 **Barriers and concerns regarding the publication of research data**: Did you face one of these technical/processing barriers when publishing research data? (depends on Q9)

Q14 **Barriers and concerns regarding the publication of research data**: Did you face one of these resource barriers when publishing research data? (depends on Q9)

Q15 **Research project**: The educational technology research project you reported in this study ...

Q16 **Publication formats**: What could be a possible way for publishing research data and/or research software at EdTec conferences (e. g., DELFI, LAK, EC-TEL, etc.)?

Q17 **Mandatory upload of research data and software**: Should each of these formats be associated with the mandatory delivery of a repository, resources, etc.?

Q18 **Review of research data and software**: What are your expectations for the review of delivered resources, repos, etc.?

References

[Bi23] Biernacka, K.; Mulligan, A.; Zimmermann, J.; Rudiak, R.: Research Data Sharing and Reuse 2020, Online, Mendeley Data, V1, 2023, DOI: 10.17632/nr9n75cpv2.1.

[BP21] Biernacka, K.; Pinkwart, N.: Opportunities for adopting open research data in Learning Analytics. In: Advancing the Power of Learning Analytics and Big Data in Education. IGI Global, Hershey, PA, pp. 29–60, 2021.

[CB14] Clarke, V.; Braun, V.: Thematic Analysis. In: Encyclopedia of Critical Psychology. Springer New York, New York, NY, pp. 1947–1952, 2014, DOI: 10.1007/978-1-4614-5583-7_311.

[Ce17] Centre for Science and Technology Studies, Elsevier and Leiden University: Open Data. The researcher perspective, tech. rep., Centre for Science and Technology Studies, Elsevier and Leiden University, 2017, URL: https://www.elsevier.com/open-science/research-data/open-data-report.

[CE24] CERN European Organization for Nuclear Research: Zenodo.org, Accessed on 02/21/2024, 2024, URL: https://www.zenodo.org.

[Ce24] Center for Open Science: osf.io, Accessed on 02/23/2024, 2024, URL: https://www.osf.io.

[CF10] Casadevall, A.; Fang, F. C.: Reproducible Science. Infection and Immunity 78 (12), pp. 4972–4975, 2010, DOI: 10.1128/IAI.00908-10.

[Co24] Cornell University: arXiv.org, Accessed on 02/23/2024, 2024, URL: https://www.arxiv.org.

[CRI18] Commission, E.; for Research, D.-G.; Innovation: Cost-benefit analysis for FAIR research data – Cost of not having FAIR research data. Publications Office, 2018.

[Eu23] European Union: European Open Science Cloud, 2023, URL: https://eosc-portal.eu/.

[Fa24] Fachgruppe Bildungstechnologien der Gesellschaft für Informatik e.V.: DELFI Tagungen: Themen, Formen der Forschung, Beitragsformate, Anforderungen und Einreichung, Accessed on 21.02.2024, 2024, URL: https://delfi-tagung.de/fileadmin/FG/WI-ELE/DELFI_Tagungen_Anforderungen-und-Formate.pdf.

[FD17] Foster, E. D.; Deardorff, A.: Open science framework (OSF). Journal of the Medical Library Association: JMLA 105 (2), p. 203, 2017.

[Ge23a] German Centre for Higher Education Research and Science Studies (DZHW): Find Higher Education and Science Research Data Packages, 2023, URL: https://metadata.fdz.dzhw.eu/en/start.

[Ge23b] German Research Foundation (DFG): Handling of Research Data, 2023, URL: https://www.dfg.de/en/research_funding/principles_dfg_funding/research_data/index.html.

[Ge24a] German Informatics Society (GI): National Research Data Infrastructure for and with Computer Science (NFDIxCS), Accessed on 03/11/2024, 2024, URL: https://nfdixcs.org/.

[Ge24b] Gesellschaft für Informatik e.V.: Call for Participation, 2024, URL: https://delfi-tagung.de/call-for-participation.

[GNT10] Goecks, J.; Nekrutenko, A.; Taylor, J.: Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome biology 11 (8), pp. 1–13, 2010.

[Ih15] Ihantola, P.; Vihavainen, A.; Ahadi, A.; Butler, M.; Börstler, J.; Edwards, S. H.; Isohanni, E.; Korhonen, A.; Petersen, A.; Rivers, K.; Rubio, M. Á.; Sheard, J.; Skupas, B.; Spacco, J.; Szabo, C.; Toll, D.: Educational Data Mining and Learning Analytics in Programming: Literature Review and Case Studies. In: Proceedings of the 2015 ITiCSE on Working Group Reports. ITICSE-WGR '15, ACM, New York, pp. 41–63, 2015.

[Ir13] Irwin, S.: Qualitative secondary data analysis: Ethics, epistemology and context. Progress in Development Studies 13 (4), pp. 295–306, 2013, DOI: 10.1177/1464993413490479.

[Je22] Jeuring, J.; Keuning, H.; Marwan, S.; Bouvier, D.; Izu, C.; Kiesler, N.; Lehtinen, T.; Lohr, D.; Peterson, A.; Sarsa, S.: Towards Giving Timely Formative Feedback and Hints to Novice Programmers. In: Proceedings of the 2022 Working Group Reports on Innovation and Technology in Computer Science Education. ITiCSE-WGR '22, ACM, Dublin, Ireland, pp. 95–115, 2022, DOI: 10.1145/3571785.3574124.

[KGH21] Katz, D. S.; Gruenpeter, M.; Honeyman, T.: Taking a fresh look at FAIR for research software. Patterns 2 (3), p. 100222, 2021, DOI: 10.1016/j.patter.2021.100222.

[Ki22a] Kiesler, N.: Dataset: Recursive problem solving in the online learning environment CodingBat by computer science students, Online, Datenerhebung: 2017. Version: 1.0.0. Datenpaketzugangsweg: Download-SUF. Hannover: FDZ-DZHW, 2022, DOI: 10.21249/DZHW:studentsteps:1.0.0.

[Ki22b] Kiesler, N.: Daten- und Methodenbericht Rekursive Problemlösung in der Online Lernumgebung CodingBat durch Informatik-Studierende, tech. rep., DZHW, 2022, URL: https://metadata.fdz.dzhw.eu/public/files/data-packages/stu-studentsteps$/attachments/studentsteps_Data_Methods_Report_de.pdf.

[Ki23] Kiesler, N.; Impagliazzo, J.; Biernacka, K.; Kapoor, A.; Kazmi, Z.; Ramagoni, S. G.; Sane, A.; Tran, K.; Taneja, S.; Wu, Z.: Where's the Data? Exploring Datasets in Computing Education. In: Proceedings of the ACM Conference on Global Computing Education Vol 2. CompEd 2023, ACM, Hyderabad, India, pp. 209–210, 2023, DOI: 10.1145/3617650.3624951.

[Ki24a] Kiesler, N.; Impagliazzo, J.; Biernacka, K.; Kapoor, A.; Kazmi, Z.; Ramagoni, S. G.; Sane, A.; Tran, K.; Taneja, S.; Wu, Z.: Where's the Data? Finding and Reusing Datasets in Computing Education. In: Proceedings of the Working Group Reports on Global Computing Education. CompEd 2023, ACM, Hyderabad, India, 2024.

[Ki24b] Kiesler, N.; Röpke, R.; Schiffner, D.; Schulz, S.; Strickroth, S.; Ehlenz, M.; Heinemann, B.; Wilhelm-Weidner, A.: Towards Open Science at the DELFI Conference, 2024, DOI: 10.17605/OSF.IO/VE5DB, URL: https://osf.io/ve5db.

[KPC21] Katz, D. S.; Psomopoulos, F. E.; Castro, L. J.: Working Towards Understanding the Role of FAIR for Machine Learning. In: 2nd Workshop on Data and Research Objects Management for Linked Open Science. ZB MED-Publikationsportal Lebenswissenschaften, Online, pp. 1–6, 2021, DOI: 10.4126/FRL01-006429415.

[KS22] Kiesler, N.; Schiffner, D.: On the Lack of Recognition of Software Artifacts and IT Infrastructure in Educational Technology Research. In (Henning, P. A.; Striewe, M.; Wölfel, M., eds.): 20. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., Bonn, pp. 201–206, 2022, DOI: 10.18420/delfi2022-034.

[KS23a] Kiesler, N.; Schiffner, D.: Open Science in den Bildungstechnologien: Zur Publikation und Begutachtung von Forschungsdaten inklusive Software im Rahmen der DELFI. In: Workshops der 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., Bonn, pp. 159–168, 2023, DOI: 10.18420/wsdelfi2023-40.

[KS23b] Kiesler, N.; Schiffner, D.: Why We Need Open Data in Computer Science Education Research. In: Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education Vol. 1. ITiCSE 2023, ACM, Turku, Finland, pp. 348–353, 2023, DOI: 10.1145/3587102.3588860.

[Mc19] McGill, M. M.: Discovering Empirically-Based Best Practices in Computing Education Through Replication, Reproducibility, and Meta-Analysis Studies. In: Proceedings of the 19th Koli Calling International Conference on Computing Education Research. Koli Calling '19, ACM, Koli, Finland, 2019, DOI: 10.1145/3364510.3364528.

[Mo11] Mons, B.; van Haagen, H.; Chichester, C.; Hoen, P.-B.; den Dunnen, J. T.; van Ommen, G.; van Mulligen, E.; Singh, B.; Hooft, R.; Roos, M., et al.: The value of data. Nature genetics 43 (4), pp. 281–283, 2011.

[MS23] Melkamu Jate, W.; Striewe, M.: Umgang der DELFI-Community mit Forschungsdaten und Softwareartefakten - Eine Erhebung auf Basis der Tagungsbände im Zeitraum 2018-2022. In: 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., Bonn, pp. 167–172, 2023, DOI: 10.18420/delfi2023-27.

[Na13] National Science Foundation: Open Data at NSF, 2013, URL: https://www.nsf.gov/data/.

[Pa02] Patton, M. Q.: Qualitative Research & Evaluation Methods. Sage, Thousand Oaks, 2002.

[PF07] Pilat, D.; Fukasaku, Y.: OECD principles and guidelines for access to research data from public funding. Data Science Journal 6, OD4–OD11, 2007.

[Pr23] Prather, J.; Denny, P.; Leinonen, J.; Becker, B. A.; Albluwi, I.; Craig, M.; Keuning, H.; Kiesler, N.; Kohn, T.; Luxton-Reilly, A.; MacNeil, S.; Petersen, A.; Pettit, R.; Reeves, B. N.; Savelka, J.: The Robots Are Here: Navigating the Generative AI Revolution in Computing Education. In: Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education. ITiCSE-WGR '23, ACM, Turku, Finland, pp. 108–159, 2023, DOI: 10.1145/3623762.3633499.

[SBH23] Schulz, S.; Berndt, S.; Hawlitschek, A.: Gruppenarbeit beim Programmieren lernen (GAPL). Datenerhebung: 2020. Version: 1.0.0. Datenpaketzugangsweg: SUF: Download, ed. by İkiz-Akıncı, D.; Niebuhr, J., Hannover, 2023, DOI: 10.21249/DZHW:dipit2020:1.0.0.

[SS22] Strickroth, S.; Striewe, M.: Building a Corpus of Task-based Grading and Feedback Systems for Learning and Teaching Programming. International Journal of Engineering Pedagogy (iJEP) 12 (5), pp. 26–41, 2022, DOI: 10.3991/ijep.v12i5.31283.

[Wi16] Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E., et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific data 3 (1), pp. 1–9, 2016.

FAIR Learning Technologies with Web Components and Packages

Frederic Salmen [1], Martin Breuer [2], Sergej Görzen [1], Malte Persike [2], and Ulrik Schroeder [1]

Abstract: Making the diverse software artifacts of the learning technologies community findable, accessible, interoperable, and reusable (FAIR) can be a technical challenge. We introduce a concept informed by our research involving packages and components to achieve FAIRness for web-based artifacts. This result is presented as a guideline to make FAIR technology choices when creating web-based learning technologies. The guideline compares classic choices with new paths afforded by technological innovation of the web platform. Supported by practical examples (learning analytics dashboards, e-assessment, and explorables) we discuss practical applications of our result.

Keywords: FAIR principles, Web Components, Packages, Reusable Software

1 Motivation

Current learning technologies like Learning Management Systems (LMS), multimodal learning resources, and e-assessment systems are used at many educational institutions such as universities or schools. Due to different requirements and contexts, there is a broad range of different platforms for individual needs. Recent discussions in the learning technology community question the current practices and lack of FAIRness related to developing software artifacts [KS22]. The authors favor fostering community collaboration, enhancing software quality and interoperability, and ensuring that software and IT infrastructure are acknowledged as critical scholarly research outputs, creating a more open and efficient scientific ecosystem. The FAIR principles were first proposed for research data in 2016 [Wi16]. Since that pivotal publication on this topic in the learning technologies community, progress has been made across other communities. In 2022, an adaptation of the FAIR model specifically for research software [Ch22] was proposed, which is the result of a three-year standardization process by an international and interdisciplinary working group of over 250 members, including researchers, developers, policymakers, and users. The model retains the FAIR acronym with four foundational principles, each supported by some guiding principles. The principles as stated in [Ch22] are:

1 RWTH Aachen University, Learning Technologies, Aachen, Germany,
salmen@cs.rwth-aachen.de, https://orcid.org/0009-0007-5206-1146;
goerzen@cs.rwth-aachen.de, https://orcid.org/0000-0003-3853-2435;
schroeder@cs.rwth-aachen.de, https://orcid.org/0000-0002-5178-8497
2 RWTH Aachen University, Learning Technologies, Aachen, Germany,
breuer@medien.rwth-aachen.de, https://orcid.org/0009-0008-0749-5110;
persike@cls.rwth-aachen.de, https://orcid.org/0000-0002-7825-089X

doi:10.18420/delfi2024_23

- F Software and its associated metadata are easy for humans and machines to find.
  - F_1 Software is assigned a globally unique and persistent identifier (ID).
  - F_2 Software is described with rich metadata.
  - F_3 Metadata clearly and explicitly include the identifier of the software described.
  - F_4 Metadata are FAIR, searchable and indexable.
- A Software and its metadata are retrievable via standardized protocols.
  - A_1 Software is retrievable by its identifier using a standardized comm. protocol.
  - A_2 Metadata are accessible, even when the software is no longer available.
- I Software interoperates with other software by exchanging data and/or meta-data and/or through interaction via application programming interfaces (APIs) described through standards.
  - I_1 Software reads/writes/exchanges data with domain-rel. community standards.
  - I_2 Software includes qualified references to other objects.
- R Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).
  - R_1 Software is described with a plurality of accurate and relevant attributes.
  - R_2 Software includes qualified references to other software.
  - R_3 Software meets domain-relevant community standards.

In this practical report, we outline a concept for applying these principles to a major subset of artifacts in the learning technologies research community: Web-based artifacts, which are implemented using web technologies, are prevalent in the community today (see next section). Their specific properties may be exploited to improve FAIRness. To this end, we pose the following research question:

RQ: How can the FAIRness (findability, accessibility, interoperability, and reusability) of web-based artifacts in learning technologies be achieved practically?

2 Related Work

In the 1990s and 2000s, learning objects (LO) emerged as an initial endeavor to address the challenge of blending educational materials [Wi02]. Through the establishment of standards, the aim was to make learning resources easily identifiable, reusable, and compatible, essentially to make them FAIR. Yet, a key drawback of LO surfaced when viewed through the lens of remix: “Learning objects can be aggregated but not adopted“ [Wi08].

A necessary first step is providing access to the source code of the artifact. Open sourcing makes the implementation of the artifact transparent and allows for fine-grained reuse of parts of the implementation. This total control over the granularity of the reuse is both an advantage and a disadvantage: Any part of the implementation, no matter how small, can be extracted. However, this is essentially looking inside the black box. Usually, there

is no standard abstraction to allow the reuse of something specific, forcing researchers to understand the whole implementation and to re-implement the part they need.

What learning technology artifacts have in common is that they are likely to be web-based, using the browser as a platform. The browser platform has seen many innovations in the last 10 years. Starting with the development of the HTML5 standard by the W3C and WHATWG, today's web marks a shift from the static, document-oriented web of the 1990s and 2000s towards a modern, app-oriented web. Previously, a lot of a web app's functionality had to be implemented on the server side (e.g., storage or routing) or integrated by plugins (e.g., Flash for multimedia). Today's web unites all these features in a single platform as a static document that can be called from everywhere (e.g., by using the browser on a smartphone). [Ka17]

Recently, the learning technologies research community has started adopting some of the new concepts of the modern web platform, such as components [SS23; St23]. Still, wide adoption is lagging behind compared to the software industry (React pioneered components in 2013 [Co15], and cross-browser support for Web Components was achieved in 2018 [WH24]) and is a promising path towards more FAIRness in research software.

3 FAIR Research Web Technology Flowchart

As an intermediary tool between the rather abstract FAIR-RS principles and the practical implementation of web-based learning technologies, we present a schema that can help make web-based artifacts more FAIR (compare Fig. 1), which we call the "FAIR Research Web Technology Flowchart". The FAIRness of the source code as research data, e.g. by forking an open source repository, is also encouraged but not addressed here.

Our flowchart (see Fig. 1) has two main paths, and each path leads to a different degree of FAIRness, as technology choices may or may not fulfill certain principles. We will go through two possible paths: the classic path (static file) and the FAIR path (modularized technology). In both paths, we start with a web application, achieving R_3 (domain-relevant community standards) by default since the web is heavily standardized.

Classic path: In the first case, we may produce a website as an artifact (e.g., an 'index.html' with related assets). Then, that website is served on any web hosting solution (for example, a self-hosted web server such as nginx). This covers both F_1 (globally unique, persistent ID) and A_1 (retrievable by ID) - the artifact can be retrieved using the URL as a globally unique identifier. It is worth noting that this identifier has limited persistency: If the site is restructured or the resource becomes otherwise unavailable, both principles no longer apply. Finally, the served website may be embedded into other web applications, for example, with an 'iframe.' While this path does not provide much FAIRness, it is very easy to follow: Since the software needs to be published as a whole anyway to be used normally, this almost enables reuse by default. It also provides security benefits as websites have strong and well-testing sandboxing protections.

FAIR path: In this path, a web application (R_3 applies) forks into two options for modularization: Component or Library. With Components, we mean encapsulated, reusable user interface elements (for example, a performance graph for learning analytics). Established libraries such as React and standards such as Web Components facilitate this common pattern to enable reuse across applications. Similarly, parts of the model layer or application logic can be extracted as an API or Library. With a Library, we mean related functionality intended for reuse (for example, jQuery [BRK15]). Almost all programming languages facilitate this pattern with a module system (e.g., ES Modules [Ec23]). Both options can fulfill multiple principles, namely I_1 (read/write/exchange data) with attributes, events, function calls, etc., I_2 (qualified references to objects) with the language's built-in reference system, R_1 (described with license and provenance) with associated docstrings and comments, and R_2 (qualified references to software) with composition of components or libraries through imports. Of course, components and modules are not exclusive: Any web application may have components and libraries worth modularizing. At this step, the researcher must deal with the fundamental challenges of modularization, for example, decoupling the component or library from other application parts or finding the right level of abstraction.

After that, two choices supplement each other to produce an artifact. An asset may be produced, such as a single file bundle using a bundle (e.g., Vite), so the user of the library or component needs only load a single file. This is optional since some libraries and components can be used directly (e.g., a small library in a single JS file). To achieve more FAIRness, the assets may be wrapped in a package. In some use cases, this may be a Node style 'package.json' or a container (e.g., Docker). Creating a package fulfills principles F_2 and F_3 as packages include the identifier and metadata such as the author, license, etc.

Next, there are two choices again. Assets may be served directly like a website, fulfilling F_1 and A_1 (see classic path). Alternatively, a package can be published in a registry such as NPM, Github, or Docker Hub (alongside version control to associate the code, the artifact, and their version). This fulfills F_1 (globally unique, persistent ID) with guaranteed immutable, versioned package names, F_4 (metadata are FAIR, searchable, and indexable) with supported metadata fields, A_1 (retrievable by ID) with package managers such as yarn, and A_2 (metadata available when software is not anymore) with deprecated packages remaining registered. While registered packages may also become unavailable like any web resource, dedicated registries such as NPM are both stable and easy to mirror, should the registry provider ever become defunct. Ultimately, this provides guarantees similar to the DOI system regarding stable, unique identifiers.

Finally, the package or served resource can be integrated into another web application. Compared to the classic path, reuse is more granular: The other application does not need to embed a whole website but can reference the URL of the served resource for a script or stylesheet or use a package manager to install a registered package. An even smaller granularity of reuse is possible by importing only the needed parts of a library. As a disadvantage, this kind of reuse requires more trust as code from the original web application is directly executed in the other web application.

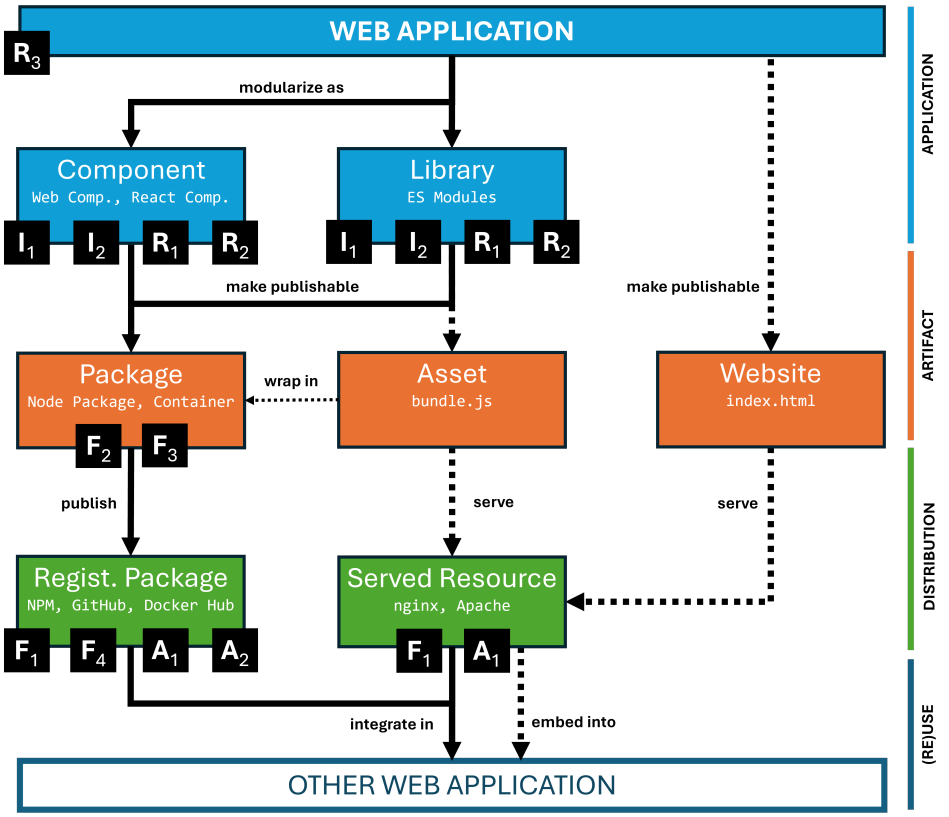


Fig. 1: FAIR Research Web Technology Flowchart. For R_1 etc. see FAIR-RS principles in Sect. 1

4 Practical experiences

One use case for the workflow introduced is an assessment analytics tool to support examiners in improving teaching and assessment practices [Br23]. Current efforts include the development of frontend components with a seamless integration in the assessment system Dynexite. Since the data analysis and result storage take place in a central learning analytics system [Br23], an additional goal is to decouple the development process from the assessment system, also enabling reuse in other systems. Further requirements include displaying multiple instances of the frontend components, such as an information pane for a course on an overview page. For example, that information pane was modularized as a standard web component using the Custom Element API without a library. Using GitLab CI pipelines, the web component and the service were bundled into a single bundle.js file, which was exposed via a Docker image and pushed onto the GitLab Docker Registry. The component can then be reused via the registry or by referencing the served resource.

The authoring tool WebWriter [SRS23][SS24] uses both components and packages. It is intended to help teachers create and remix digital content such as text, images, audio, video, and interactive content (e.g., simulations, interactive maps or charts), bringing them together in multimodal resources called explorables [Ho20; Vi11]. The interactive elements that can be added to explorables are called widgets. Developers may implement their widgets using web components and distribute them as Node packages via NPM. In short, web components and packages become the plugin system of the authoring tool. This encourages developers to create FAIR widgets to be compatible with WebWriter. Widgets are loosely coupled to WebWriter, meaning web applications can reuse them without changing the source code.

5 Conclusion

In this paper, we introduced the FAIR Research Web Technology Flowchart, a tool to apply the FAIR-RS principles to web-based learning technologies. This tool helps researchers by guiding them through the step-by-step implementation of a FAIR web application. While it is specific to modern frontend web technologies, the approach advocated can be used with almost any web-based framework and technology stack, as showcased by the diverse set of practical experiences reported. It also builds upon the best practices of the wider web development community.

While our approach aims to foster modularization, there are some limitations. In terms of scope, the concept is intended for modern single-page applications. Whether it applies to a server-focused architecture where most of the application logic is run server-side is unclear and worth investigating. Also, the flowchart intentionally only covers the reuse of software as an artifact, not the reuse of source code through a public, open-source repository. Regarding assuring software quality, following the "FAIR path" of the flowchart *can* lead to FAIR research software, quality control (including transparent version control) is still needed at every level. Another challenge lies in the validation of FAIRness, for example in the context of a conference where reviewers need to ensure the FAIRness of research software associated with submissions efficiently and reliably. Finally, institutional challenges, such as poor maintenance and a lack of collaboration, cannot be addressed through a technical concept.

For future work, standardized solutions such as Custom Elements Manifest (a JSON file describing Web Components) can be helpful to develop assistive tools for that validation. It may also be valuable to create a more permanent identifier for software artifacts (in case of changing registries), such as a DOI resolving to a registry name. Overall, we assume that our process can be refined further through more and different practical applications using different technology stacks (e.g., Learning Analytics components as Web Components or WebXR) and may also be aided by developing tools specific to the learning technologies community (for example integrating learning analytics with xAPI into learning technology web components). Further, through our flowchart, we plan to explore mechanisms to check the FAIRness of web projects automatically.

Acknowledgments We were supported by the research project NERD II (005-2201-0014), sponsored by the state of North Rhine-Westphalia.

References

[Br23] Breuer, M.; Brocker, A.; Persike, M.; Schroeder, U.: AxEL - Eine modulare Softwarekomponente für ein dediziertes E-Prüfungssystem zur Generierung von xAPI-Statements für Assessment Analytics. ISBN: 9783885797326 Publisher: Gesellschaft für Informatik e.V., 2023, DOI: 10.18420/DELFI2023-17.

[BRK15] Bibeault, B.; Rosa, A. D.; Katz, Y.: jQuery in Action. Simon and Schuster, 2015, ISBN: 978-1-63835-337-9.

[Ch22] Chue Hong, N. P.; Katz, D. S.; Barker, M.; Lamprecht, A.-L.; Martinez, C.; Psomopoulos, F. E.; Harrow, J.; Castro, L. J.; Gruenpeter, M.; Martinez, P. A.: FAIR Principles for Research Software (FAIR4RS Principles). Zenodo, 2022.

[Co15] Cory Gackenheimer: Introduction to React. 2015.

[Ec23] Ecma International: ECMA-262, 2023, visited on: 04/11/2024.

[Ho20] Hohman, F.; Conlen, M.; Heer, J.; Chau, D. H. (: Communicating with Interactive Articles. Distill 5 (9), e28, 2020, ISSN: 2476-0757, DOI: 10.23915/distill.00028.

[Ka17] Kaul, M.; Kless, A.; Bonne, T.; Rieke, A.: Game Changer for Online Learning Driven by Advances in Web Technology. 2017.

[KS22] Kiesler, N.; Schiffner, D.: On the Lack of Recognition of Software Artifacts and IT Infrastructure in Educational Technology Research. In: 20. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., pp. 201–206, 2022, ISBN: 978-3-88579-716-6.

[SRS23] Salmen, F.; Roepke, R.; Schroeder, U.: WebWriter: A System to Author and Remix Explorables–Requirements & First Prototype. 2023.

[SS23] Selmanagić, A.; Simbeck, K.: Breaking It Down: On the Presentation of Fine-Grained Learning Objects in Virtual Learning Environments. In: 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., pp. 155–160, 2023, ISBN: 978-3-88579-732-6.

[SS24] Salmen, F.; Schroeder, U.: Evaluating Authoring Tools with the Explorable Authoring Requirements, 2024, arXiv: 2403.17714 [cs], visited on: 04/09/2024.

[St23] Striewe, M.: Strukturformeln für Moleküle zeichnen und differenziertes Feedback erhalten: Eine integrierte Lösung im Rahmen des E-Assessment-Systems JACK. In: 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., pp. 273–274, 2023, ISBN: 978-3-88579-732-6.

[Vi11] Victor, B.: Explorable Explanations, http://worrydream.com/ExplorableExplanations/, 2011, visited on: 08/23/2022.

[WH24] WHATWG: HTML Living Standard, https://html.spec.whatwg.org/, 2024, visited on: 04/10/2024.

[Wi02] Wiley, D. A.: The Instructional Use of Learning Objects. Agency for instructional technology Bloomington, IN, 2002.

[Wi08] Wiley, D. A.: The Learning Objects Literature. In: Handbook of Research on Educational Communications and Technology. Routledge, pp. 345–353, 2008, visited on: 04/10/2024.

[Wi16] Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.: The FAIR Guiding Principles for Scientific Data Management and Stewardship. Scientific data 3 (1), pp. 1–9, 2016.

Implementation of flexible and open course material for computer programming courses [2]

Sebastian Speiser [1], Melanie Baur [1], and Johanna Sedlmair [1]

Abstract: We report on our practical experiences of teaching computer programming across several study courses in multiple disciplines. We focus on creating learning material in a way that improves upon the current state of using presentations slides. We address: i) collaboration, ii) interactive teaching, iii) self-study, iv) progress tracking and v) individualisation. We are deriving requirements for course material that improves in these categories. We describe three systems that we used in different study courses, assessing them against our specified requirements. Finally, we outline our vision of creating an open, modular and decentralized approach to learning materials.

Keywords: Programming education, Jupyter Notebook, Individualisation

1 Introduction

Learning the basics of computer programming is at the core of the computer science education and becomes more and more important as part of degree courses in other disciplines as well. However, failure rates in programming courses have been historically high - with no clear improvement levers found, but hints in the data, e.g. a higher pass rate for smaller classes (< 30 students) [WL14]. Smaller cohorts instead of large/massive courses are implemented e.g. in German universities of applied sciences[3]. Many schools, including ours, address this by offering several courses in parallel. These courses are mostly held by different lecturers and adapted to the target audience. Course material often consists to a large extent of slides that the lecturers share via email, shared storage or as file uploads in the LMS. While during and after the pandemic, the usage of video material and asynchronous formats has increased, in our experience these are still commonly based on slide presentations.

Slide presentations have established themselves as standard lecture materials and are suitable for frontal lectures; however, we see room for improvement, especially regarding:

1. Collaboration among lecturers (and between lecturers and students)
2. Interactivity and practicality of lecture
3. Suitability for self-study
4. Tracking progress
5. Individualisation of material

1 Hochschule für Technik Stuttgart, Schellingstr. 24, 70174 Stuttgart, Germany,
sebastian.speiser@hft-stuttgart.de, https://orcid.org/0009-0007-6587-1527;
melanie.baur@hft-stuttgart.de, https://orcid.org/0009-0005-9683-2446;
johanna.sedlmair@hft-stuttgart.de, https://orcid.org/0009-0007-4478-3245

2 Supported by Stifterverband and MWK Baden-Wuerttemberg

3 Referring to “Hochschule für angewandte Wissenschaften” (HAW)

doi:10.18420/delfi2024_24

As discussed, there are no clear winning strategies for improving passing rates. However, our proposed improvement areas are common themes in the literature and Vihavainen et al. [VAW14] found that about any change is often having a positive impact.

To bring about these improvements, we implemented lecture notes for computer programming courses using text-/code-based formats such as Jupyter Notebooks [Kl16] or Markdown. In this paper we describe our practical experiences and outline our vision of an open, modular and decentralized approach to learning materials.

In Section 2, we describe examples for the improvement opportunities listed above and deduce requirements for an improved concept. Section 3 reports on the approaches that we implemented for several lectures. We discuss our learnings and outline the next steps in Section 4.

2 Requirements for an improved concept

In this section, we further detail the improvement opportunities and derive requirements. Additionally, we collect non-functional, i.e. organizational and technical, requirements for a system. Requirements are listed as *R* and a number.

Better collaboration among lecturers and between lecturers and students: Collaboration is hard, especially for PowerPoint slides as illustrated by the following examples:

- Lecturer A replaces an outdated slide by a new one and sends the PowerPoint via email to the colleagues. Each colleague generates a PDF version and uploads it to their respective courses in the learning management system. If any of the colleagues has made any changes not reflected in the slideset used by A, trouble begins.
- Student X fixes some typos during the semester on her PDF-version and sends it to her lecturer who has to manually transfer those fixes and potentially merge with other updates during the timeframe.

R1: Text-/code-based formats that can be managed in a version control system such as git. Especially the usage of a platform such as GitHub or GitLab enables collaboration with features such as issue tracking, pull requests and pipelines that automatically update published material [Za15].

Addressing the examples above, we can ensure a clear lineage of versions and in text-based formats often automatic merging if changes do not contradict. In case of student collaboration, we can use a pull request that is reviewed by a lecturer before it is merged into the material.

More interactive and practical lecture style: Presentation slides usually contain static code examples and the output that is/would be shown when the program was executed. Student-centered and interactive methods can enhance students' skills [Ka21]. There is

great value in demonstrating varieties of the shown code or playing with, e.g. to i) answer student questions along the lines "could we also do it with a for-loop?", ii) test border cases or show limitations of the code by introducing and fixing bugs, iii) extend the code for additional functionality.

A common situation with slide-based material is the following: Lecturer B copies code from her IDE to a slide. Lecturer C wants to demonstrate the code during lecture - in best case has to leave presentation mode and copy paste code into his own IDE. Often enough, however, code is copied as screenshot to keep the formatting and/or split into several to better use horizontal space and/or shortened to focus on the core concepts.

R2: Executable code should be used for all examples and exercises. Meaning that code is integrated in plaintext in or linked from the lecture notes in a way that it can be executed and modified without the need for copying or conversion.

Suitability for self-study: Some students need to study the material on their own for various reasons, e.g. i) they were ill for a part of the lecture, ii) they have to repeat a lecture not fitting in their time schedule, or (iii) they need an individual pace due to previous knowledge or being fast/slow learners (see e.g. [Of17]). While slide presentations are suitable for frontal lecture, they are not ideal for self-study with no one explaining them. To get an overview of all topics relevant in a lecture, students have to download all slide presentations and open each to review the (hopefully existing) agenda slide. Additionally they need to download the exercise sheets. Often the slides and exercises are only made available week by week.

R3: Table of content (TOC) or a concept map should be available in the overall materials to give students an overview of all topics and direct links to the material they want to review. This should also serve as a single source of truth regarding which material is relevant.

R4: Textual content should contain all relevant information. Some students prefer video material due to the easier consumption compared to reading. To our knowledge, there is no clear evidence that video material is superior to textual content for learning programming [La16]. However, textual content is easier for reviewing the material or using it as a reference. Also it supports collaboration as discussed above. Of course, the content can be accompanied by video material that is linked in the textual content.

Tracking progress: Students should be able to see what is expected from them and what their progress in that respect is. For testing their abilities in a competence-oriented way, they need to be provided with suitable programming exercises.

R5: Progress tracking in the system should store which topics (see R3) have been covered by each student to enable them to review their progress either on their own or together with a lecturer. Different kind of trackings are thinkable: self-reporting, completion of exercises or official examinations.

R6: Linked exercises to the topics should be available either directly within the material or with links to a consistently used external system.

Individualized materials: More and more, we have students with different prior knowledge from school due to different chosen specialization paths. Additionally, the degrees are getting more specialized and instead of only CS students learning Java, we have e.g. students of AI learning Python, business informatics students learning ABAP and IoT students learning rust.

To address this, we need material that is tailored to cohorts or individuals based on prior knowledge and requirements of their study courses. We still want to collaborate and reuse materials leading to situations such as this: Lecturer D uses shared slides for explaining arrays but is replacing the code example with an example of matrix multiplication tailored to his audience of mathematics students. Every update involves a lot of manual copying of slides to the correct places.

R7: Modularized content that covers individual topics in a fine-granular way, potentially even in variants for different programming languages or business domains for the examples. A composition approach is required to combine modules into integrated material for a given lecture - comparable to playlists for music.

Technical and organizational requirements - not directly functional but necessary for a sustainable system:

R8: Low IT admin load to run the system, as this has often to be done by lecturers themselves, especially in smaller schools without a large IT department.

R9: Low individual technology that has to be developed and maintained. Instead it should be based on open source projects that are maintained by an active community.

R10: Accessibility not only at the school but also from home or mobile devices. Students are expected to spend more time with studying and exercising than what is in their timetable.

3 Implementation

In this section, we describe several approaches that we have implemented. The variance has historical reasons (what was known and technically feasible at that point in time) and a conscious decision for a try-analyze-improvement loop. For space reasons, the descriptions only focus on key aspects. Tab. 1 gives a complete overview of the requirement coverage.

Jupyter Books in Open Repository: For a course on data analytics using Python in a Master degree, we created lecture notes with integrated exercises using Jupyter Books. The book can be viewed in a web browser or downloaded as PDFs. The book and the corresponding git repository are publicly available. This fosters feedback from students and we received several issues regarding small mistakes and some pull requests for improvements. The

Approach	*R1: Text-/code-based formats*	*R2: Exec. code*	*R3: TOC*	*R4: Textual content*	*R5: Progress tracking*	*R6: Linked exercises*	*R7: Modular content*	*R8: Low IT admin load*	*R9: Low indiv. tech.*	*R10: Accessibility*
Jupyter Books	+	o	+	+	-	+	-	+	+	+
Jupyter Notebook self-hosted	+	+	o	+	-	+	-	-	-	-
Jupyter Notebook in GitHub Codespaces	+	+	o	+	-	+	-	o	-	+

Tab. 1: Table of Requirements

approach is rather low tech in the sense that after initializing a GitLab workflow the whole book is built and published automatically when an update is made. The Jupyter Notebooks underlying the book can be downloaded, allowing for direct modification and execution of the code. Some custom scripts had to be developed, e.g. to hide solutions to exercises in the published pages. The GitLab repository contains a *readme* file that links both all relevant book chapters for a course, as well as exercises and requirements for the portfolio-based project work.

Self-hosting Jupyter Notebook Java: For a course on data structures and algorithms, we created a set of Jupyter Notebooks that contain the lecture notes and instructions for exercises as well as placeholders where students can fill in their code. As setting up Jupyter Notebooks for Java is a bit of an effort and relies on libraries of unknown maintenance status, we decided to provide a centrally hosted server for the notebooks. As it was hosted in the internal network of the school, it was not easily accessible from home. There was significant own development required to facilitate the interaction between Java and the Jupyter Notebooks, e.g. for correct display of arrays or visualization of sorting algorithms. The same set of notebooks was used during lecture and based on interaction with the students new code was inserted or modified.

GitHub Codespaces: As the admin overhead for the Notebook server was quite high and it was not available outside of the school, we decided to set up GitHub Codespaces (an IDE in the web browser) preconfigured for the Java Jupyter Notebooks. This approach was used for an introductory programming course. The codespaces were in the beginning all the same as they are declaratively defined but configuration drift came quickly and every student was faced with an individual complex environment that was hard to fix if it broke. Especially the additional complexity of the tech layer confused beginners, e.g. that all code is stored in a git repository and instances of the IDE had to be launched and managed. We also had issues with the security settings of some browsers, especially in the labs of the school, that disabled the whole IDE. In case of a student not willing to use a commercial service like GitHub, we would have to provide an alternative solution.

4 Discussion and Future Work

All of the used systems were satisfactory for the lecturers and positively acknowledged by some students. However they were still i) overly complex for some other students and ii) not fully addressing all requirements.

To address the requirements *R8: Low IT admin load* and *R9: Low indiv. tech.*, we are currently evaluating LiaScript. We think it can lower our individual development efforts or channel it into an open source project, where they will be leveraged more effectively and maintained better. Also we expect lower admin load and easier authoring by more lecturers.

Also, LiaScript brings capabilities for storing the state of a user. We are reviewing how to leverage this for addressing *R5: Progress tracking*.

The major open topic is *R7: Modular content*. Here we are planning to adapt an approach of another (yet unpublished) system that we developed for managing programming exercises. If we generalize that system, we can use it to compose courses for specific cohorts or even individuals with base material developed e.g. in LiaScript. As a promising alternative, we review the recently presented system *WebWriter*, which has a similar goal [SRS23].

Finally, we are in the process of adapting our educational concepts to effectively teach using the developed systems across various study courses and with different lecturers.

References

[Ka21] Kadar, R.; Wahab, N. A.; Othman, J.; Shamsuddin, M.; Mahlan, S. B.: A Study of Difficulties in Teaching and Learning Programming: A Systematic Literature Review. Int. J. of Academic Research in Progr. Education and Development 10 (3), pp. 591–605, 2021.

[Kl16] Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; Ivanov, P.; Avila, D.; Abdalla, S.; Willing, C.; Jupyter development team: Jupyter Notebooks – a Publishing Format for Reproducible Computational Workflows. In: 20th Int. Conf. on Electronic Publishing. 2016.

[La16] Lang, G.: The Relative Efficacy of Video and Text Tutorials in Online Computing Education. Information Systems Education Journal 14 (5), p. 33, 2016.

[Of17] Offutt, J.; Ammann, P.; Dobolyi, K.; Kauffmann, C.; Lester, J.; Praphamontripong, U.; Rangwala, H.; Setia, S.; Wang, P.; White, L.: A Novel Self-Paced Model for Teaching Programming. In: ACM Conference on Learning @ Scale. 2017.

[SRS23] Salmen, F.; Roepke, R.; Schroeder, U.: WebWriter: A System to Author and Remix Explorables – Requirements & First Prototype. In: Workshops Der 21. Fachtagung Bildungstechnologien (DELFI). GI, 2023.

[VAW14] Vihavainen, A.; Airaksinen, J.; Watson, C.: A Systematic Review of Approaches for Teaching Introductory Programming and Their Influence on Success. In: 10th Annual Conference on Int. Computing Education Research. ACM, 2014.

[WL14] Watson, C.; Li, F. W.: Failure Rates in Introductory Programming Revisited. In: Conference on Innovation & Technology in Computer Science Education. ACM, 2014.

[Za15] Zagalsky, A.; Feliciano, J.; Storey, M.-A.; Zhao, Y.; Wang, W.: The Emergence of GitHub as a Collaborative Platform for Education. In: ACM Conference on Computer Supported Cooperative Work & Social Computing. 2015.

From Receiving to Characterizing: Improving Training Strategies for Research Data/Software Management by Another Domain

Jan Bernoth [1], Christian Riedel [1], Axel Wiepke [1] und Firas Al Laban [1]

Abstract: Constructing learning objectives for Research Data and Software Management (RDSM) primarily focuses on the cognitive domain of Bloom's taxonomy. However, surveys exploring why RDSM is not fully implemented across all research fields have revealed that the issue extends beyond a mere lack of knowledge. It appears that researchers also lack inner conviction and are not open to RDSM practices. To address this, we aim to initiate a discussion through this position paper on incorporating another dimension of Bloom's Taxonomy into training competencies: the affective domain. Considering emotional factors might help persuade individuals to receive and respond to information more effectively. By integrating the affective domain, we suggest that a shift in learning strategies and activities from purely cognitive objectives in RDSM could be beneficial.

Keywords: Research Data Management, Research Software Management, Learning Objectives, Bloom's Taxonomy

1 Introduction

In almost all scientific disciplines, digital methods are increasingly being used. Therefore, for the sake of traceability of scientific findings, the underlying data and – in the case of software is developed for the research purpose [Gr21] – the associated software is relevant for publications. The FAIR principles (Findable, Accessible, Interoperable, Reusable) for data [Wi16] and software [Ch22] express this requirement. Initiatives such as the National Research Data Infrastructure (NFDI)[2] or the European Open Science Cloud (EOSC)[3] are advancing the associated cultural change. Constructing procedures and infrastructure to support FAIR principles is primarily addressed by Research Data Management (RDM). However, as Katz et al.[Ka16] have shown, software is not data. Acknowledging that research data and software require different approaches for management, maintenance, and archiving, we use the term Research Data and Software Management (RDSM) for approaches that are predominantly associated with RDM, to highlight this distinction.

1 Universität Potsdam, Institut für Infromatik und Computational Science, An der Bahn 2, 14476 Potsdam, Deutschland, jan.bernoth@uni-potsdam.de, https://orcid.org/0000-0002-4127-0053;
christian.riedel@uni-potsdam.de, https://orcid.org/0000-0001-5154-4153;
axel.wiepke@uni-potsdam.de, https://orcid.org/0000-0002-0555-4040;
firas.al.laban@uni-potsdam.de, https://orcid.org/0000-0001-8072-9384

2 https://www.nfdi.de/

3 https://eosc.eu/

doi:10.18420/delfi2024_25

In this context, a task force comprised of members from various projects and organizations, including DINI[4], published a learning objective matrix to assist people in training RDM competencies in academia [Pe23]. To structure learning objectives, there are six different thematic clusters, which also helps to structure the field of RDM: 1. Basic and overarching concepts, 2. Working with Data, 3. Documentation and Metadata, 4. Long-term preservation, publication, reuse, 5. Law and ethics, 6. Supporting structures. To each thematic cluster are also subcategories in which multiple objectives are listed for different level of academic qualifications, like Bachelor, Master, PhD or Data Steward. Additionally, each learning objective is annotated with one or more competence area(s) and a learning level based on the revision of the cognitive domain of Bloom's Taxonomy [KA01].

The matrix presents a good framework for encapsulating content. However, it is important to note that in most disciplines, there is not yet a common understanding of which standard is suitable for the RDSM content aspect. In this regard, there are already many discussions concerning the usability, the necessity, and the time required to integrate such practices. In some cases, there is even a resistance to being open with research results, or there's a belief that the time required for effective RDSM is not a worthwhile investment. For example, some researchers or decision makers who have reservations about such processes could decelerate or halt their curricula integration. Additionally, researchers who are trained in RDSM might not incorporate the necessary competencies into their routines for various reasons, resulting in the persisting of barriers to a much-needed cultural shift toward Open Science. In Section 2, we present studies that have closely examined these reservations about RDSM.

In this position paper, we aim to spark a discussion on incorporating another dimension of Bloom's Taxonomy into training RDSM competencies: the affective domain. By constructing objectives to address the affective domain, we can create opportunities to design strategies and activities to further integrate Open Science practices among researchers. A further exploration of this domain and an example of how learning objectives can be constructed are discussed in Section 3.

Therefore, there is a need for a discussion on how to construct learning objectives within the affective domain. The DELFI community is an ideal forum for this discussion, given its extensive knowledge at the intersection of learning scenarios and software-based solution finding. These competencies help in comprehending the issues at hand and in developing a supportive framework for creating such learning objectives. Section 4 presents further steps that could be taken.

2 Barriers to produce FAIR Research Data and Research Software

Research has shown that implementing RDSM workflows across various scientific disciplines can bring significant benefits. These benefits include improved quality, reproducibility,

4 https://dini.de/

and long-term usability of scientific data and software, enhanced transfer of research methodologies and results, and strengthened scientific integrity [BCG20]. However, the adoption of these workflows is complex due to several factors. Firstly, the demands of implementing RDSM workflows vary significantly across disciplines due to differences in data acquisition, analysis, and methodological approaches [Fe20]. Secondly, the involvement of multiple stakeholders with divergent interests, including researchers, research institutions, publishers, and funding agencies, introduces further complexity. Despite the proven advantages of RDSM workflows and the promotion of Open Science practices, numerous barriers hinder their widespread implementation [Be17; Ho18].

Previous studies have identified technical, systematic, and individual-level obstacles that hinder the widespread adoption of open science practices. Technical issues such as limited interoperability, dependence on closed proprietary tools, management of large data sets, and computational challenges pose significant limitations [Fe20]. In addition, research institutions, publishers, and funding agencies often lack direct support and supportive regulatory frameworks, which is crucial since researchers are often guided by such principles when implementing RDSM workflows [Fe20; Go22b]. Moreover, the prevailing culture within scientific research does not sufficiently encourage the extensive sharing of diverse research outputs, and some research fields might lack a common practice for data sharing [AA22; Ho18].

At the individual researcher level, personal attitudes toward research outputs and knowledge barriers can significantly influence implementing open science practices [Go22b]. For example, the additional workload to prepare and share data or software for publication, concerns over intellectual property rights and potential data misuse, and complex legal constraints, with many researchers finding licenses overly restrictive or difficult to interpret, add to the perceived barriers to implementing RDSM practices. Additionally, concerns about the inappropriate use or misinterpretation of shared data discourage researchers from sharing their data and code [Be17; Go22a; Ho18].

Beyond systematic and technical issues and individual attitudes, the lack of formal education and training is a significant barrier to implementing RDSM workflows. This deficiency often leads to unstructured practices and perpetuates barriers to open science. Therefore, it is crucial to integrate training in RDSM as a core component of formal educational programs to ensure the sustainability and long-term utility of research data and software [Be17; Fe20; Ho18; Le14]. However, current training approaches tend to focus mainly on overcoming knowledge barriers and may not address other significant limitations.

3 Affectivitiy in RDSM Training

Teachers across all disciplines must believe in the value of their subject to ignite enthusiasm in their students [Rî20]. Viewing one's subject in this way serves as a tool for personal growth rather than an end goal. Additionally, theories of development conceptualize individuals as

multifaceted beings, incorporating aspects such as intellect, emotions, and morality, among others, highlighting the complexity of human evolution.

When it comes to affectivity, numerous theories aim to comprehend intentions to better forecast the likelihood of an individual engaging in a behavior. One of the initial promising approaches for predicting behavior was the Theory of Reasoned Action (TRA), which explains behavior based on the individual's preferences (attitude towards behavior) and the social pressure from significant others (subjective norm) [Br09]. This theory was later expanded into the Theory of Planned Behavior, the Decomposed Theory of Planned Behavior, and subsequently refined into models such as the Technology Acceptance Model (TAM and TAM2), among others [MJ17]. These theories have been developed across various scientific and social disciplines, each offering different levels of detail in their models. However, while these theories aim to aid in understanding or predicting individual intentions, they do not offer insights about how stable or volatile these intentions are. Bloom's taxonomy of the affective domain provides an approach for categorizing behavioral attitudes.

According to Bloom, the affective aspects of education aim to attain learning objectives that focus on altering interests, attitudes, and values [Rî20]. Therefore, teachers must delineate a set of values deemed desirable for the respective profession. Subsequently, a corresponding set of attitudes must be formulated and eventually demonstrated through behaviors consistent with these values and attitudes. It is crucial for all these aspects to be meticulously planned in advance, taking into account the cognitive-affective balance in education [Bu16]. Failure to do so may lead to resistance from learners, which can manifest in their efforts to achieve cognitive or evaluative equilibrium. This often results in unaddressed, modified, or even reversed learning outcomes, alongside diminished regard for the learners' instructors [Ma80].

As mentioned earlier, Petersen et al. introduced a comprehensive list of cognitive learning goals pertaining to six pertinent fields in RDM [Pe23]. To complement these goals, we advocate for addressing not only the cognitive aspects but also the affective domain as per Bloom's taxonomy. Among the most crucial objectives for learners is cultivating a value for learning itself, alongside fostering a positive attitude towards learning activities, novelty, creativity, and curiosity, all underpinned by joy and enthusiasm for the learning process [KBM64]. Hence, we briefly outline the steps of the affective domain, drawing values from the FAIR principles, and propose initial learning goals for an example field: "(research-) software & Coding," which falls within the broader category of "working with data." We chose this field to be best approachable by the DELFI community, and will focus on an untouched subfield of the learning goal matrix: FAIR Research Software development.

> *Receiving*: Being willing to listen and be aware to receive knowledge: Software designers are willing to follow debates about alternative solutions of the programming issues.

Responding: Actively participating and engaging to transfer knowledge: Software designers assist the community by sharing minimal working examples of their code.

Valuing: Finds value and worth in one's learning and is motivated to continue: Software designers demonstrate best-practice documentation even in minimal working examples.

Organization: Integrating and comparing values, ordering them according to priorities: Software designers contrast readability against efficiency of code on an abstract level according to the FAIR principles.

Characterization: Value that will control the outcome and behavior: Software designers justify their design decisions consistently with Open Science criteria.

4 The need to integrate this into the Open Science discussion

The process of concerning more openness in research has already started with the position paper of Kiesler and Schiffner [KS22]. In the process, it's important to address doubts, consider barriers, and support efforts with infrastructure and good practices, all while convincing people of the necessity. The matrix of learning objectives is an enormous starting point, but omitting the emotional parts could hinder the underlying goal of a cultural change. We sketched an example for software related affective learning objectives to give a better understanding of the affective domain, but to complement the given learning matrix there is a larger community process needed.

Therefore, it is important to consider that training should ideally impart knowledge while also fostering the internalization of values aligned with good scientific practice. This requires cultivating a broader cultural change within the scientific community, including researchers, students, and leaders in academia, publishing, or funding agencies who influence decision-making. The goal is to ensure that open science practices are understood as essential to scientific work. Accordingly, training initiatives should extend beyond the cognitive domain, where traditional learning objectives improve understanding but do not necessarily translate to value-based application in everyday scientific activities.

Literaturverzeichnis

[AA22] Adimoelja, A.; Athreya, A.: Reducing Barriers to Open Science by Standardizing Practices and Realigning Incentives. Special Issue on Open Science Policies as an Accelerator for Achieving the Sustainable Development Goals 21 (02), 2022, ISSN: 2372-2193, DOI: 10.38126/jspg210201, URL: http://dx.doi.org/10.38126/jspg210201.

[BCG20] Briney, K.; Coates, H.; Goben, A.: Foundational Practices of Research Data Management. Research Ideas and Outcomes 6, 2020, ISSN: 2367-7163, DOI: 10.3897/rio.6.e56508, URL: http://dx.doi.org/10.3897/rio.6.e56508.

[Be17] Beno, M.; Figl, K.; Umbrich, J.; Polleres, A.: Perception of Key Barriers in Using and Publishing Open Data. JeDEM - eJournal of eDemocracy and Open Government 9 (2), S. 134–165, 2017, ISSN: 2075-9517, DOI: 10.29379/jedem.v9i2.465, URL: http://dx.doi.org/10.29379/jedem.v9i2.465.

[Br09] Bradley, J.: The Technology Acceptance Model and Other User Acceptance Theories. In: Handbook of Research on Contemporary Theoretical Models in Information Systems. IGI Global, S. 277–294, 2009, ISBN: 978-1-60566-659-4, DOI: 10.4018/978-1-60566-659-4.ch015, URL: https://www.igi-global.com/chapter/technology-acceptance-model-other-user/www.igi-global.com/chapter/technology-acceptance-model-other-user/35835, Stand: 11. 04. 2024.

[Bu16] Buckland, R.: Affective and Cognitive - Designing Educational Experiences that Transform. In: Proceedings of the 2016 ACM Conference on International Computing Education Research. ICER '16, Association for Computing Machinery, New York, NY, USA, S. 1, 2016, ISBN: 978-1-4503-4449-4, DOI: 10.1145/2960310.2960311, URL: https://dl.acm.org/doi/10.1145/2960310.2960311, Stand: 11. 04. 2024.

[Ch22] Chue Hong, N. P.; Katz, D. S.; Barker, M.; Lamprecht, A.-L.; Martinez, C.; Psomopoulos, F. E.; Harrow, J.; Castro, L. J.; Gruenpeter, M.; Martinez, P. A.; Honeyman, T.: FAIR Principles for Research Software (FAIR4RS Principles), 2022, DOI: 10.15497/RDA00068.

[Fe20] Feger, S. S.; Wozniak, P. W.; Lischke, L.; Schmidt, A.: "Yes, I comply!": Motivations and Practices around Research Data Management and Reuse across Scientific Fields. Proceedings of the ACM on Human-Computer Interaction 4 (CSCW2), S. 1–26, 2020, ISSN: 2573-0142, DOI: 10.1145/3415212, URL: http://dx.doi.org/10.1145/3415212.

[Go22a] Gomes, D. G. E.; Pottier, P.; Crystal-Ornelas, R.; Hudgins, E. J.; Foroughirad, V.; Sánchez-Reyes, L. L.; Turba, R.; Martinez, P. A.; Moreau, D.; Bertram, M. G.; Smout, C. A.; Gaynor, K. M.: Why don't we share data and code? Perceived barriers and benefits to public archiving practices. Proceedings of the Royal Society B: Biological Sciences 289 (1987), 2022, ISSN: 1471-2954, DOI: 10.1098/rspb.2022.1113, URL: http://dx.doi.org/10.1098/rspb.2022.1113.

[Go22b] González-Teruel, A.; López-Borrull, A.; Santos-Hermosa, G.; Abad-García, F.; Ollé, C.; Serrano-Vicente, R.: Drivers and barriers in the transition to open science: the perspective of stakeholders in the Spanish scientific community. El Profesional de la información, 2022, ISSN: 1386-6710, DOI: 10.3145/epi.2022.may.05, URL: http://dx.doi.org/10.3145/epi.2022.may.05.

[Gr21] Gruenpeter, M.; Katz, D. S.; Lamprecht, A.-L.; Honeyman, T.; Garijo, D.; Struck, A.; Niehues, A.; Martinez, P. A.; Castro, L. J.; Rabemanantsoa, T.; Chue Hong, N. P.; Martinez-Ortiz, C.; Sesink, L.; Liffers, M.; Fouilloux, A. C.; Erdmann, C.; Peroni, S.; Martinez Lavanchy, P.; Todorov, I.; Sinha, M.: Defining Research Software: a controversial discussion, 2021, DOI: 10.5281/zenodo.5504016.

[Ho18] Houtkoop, B. L.; Chambers, C.; Macleod, M.; Bishop, D. V. M.; Nichols, T. E.; Wagenmakers, E.-J.: Data Sharing in Psychology: A Survey on Barriers and Preconditions. Advances in Methods and Practices in Psychological Science 1 (1), S. 70–85, 2018, ISSN: 2515-2467, DOI: 10.1177/2515245917751886, URL: http://dx.doi.org/10.1177/2515245917751886.

[KA01] Krathwohl, D. R.; Anderson, L. W.: A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman, New York, 2001.

[Ka16] Katz, D. S.; Niemeyer, K. E.; Smith, A. M.; Anderson, W. L.; Boettiger, C.; Hinsen, K.; Hooft, R.; Hucka, M.; Lee, A.; Löffler, F.; Pollard, T.; Rios, F.: Software vs. data in the context of citation, 2016, DOI: 10.7287/peerj.preprints.2630v1.

[KBM64] Krathwohl, D.; Bloom, B.; Masia, B.: Taxonomy of Educational Objectives: The Classification of Educational Goals. David McKay Company University of Mississippi School of Education, New York, 1964.

[KS22] Kiesler, N.; Schiffner, D.: On the Lack of Recognition of Software Artifacts and IT Infrastructure in Educational Technology Research, 2022, DOI: 10.18420/delfi2022-034.

[Le14] Lenhardt, W. C.; Ahalt, S.; Blanton, B.; Christopherson, L.; Idaszak, R.: Data Management Lifecycle and Software Lifecycle Management in the Context of Conducting Science. Journal of Open Research Software 2 (1), e15, 2014, ISSN: 2049-9647, DOI: 10.5334/jors.ax, URL: http://dx.doi.org/10.5334/JORS.AX.

[Ma80] Marshall, S.: Cognitive-Affective Dissonance in the Classroom. Teaching Political Science 8 (1), Publisher: Routledge _eprint: https://doi.org/10.1080/00922013.1980.11000222, S. 111–117, 1980, ISSN: 0092-2013, DOI: 10.1080/00922013.1980.11000222, URL: https://doi.org/10.1080/00922013.1980.11000222, Stand: 11. 04. 2024.

[MJ17] Momani, A. M.; Jamous, M. M.: The Evolution of Technology Acceptance Theories. en, International Journal of Contempory Computer Research (IJCCR) 1, S. 51–58, 2017, URL: https://ssrn.com/abstract=2971454.

[Pe23] Petersen, B.; Engelhardt, C.; Hörner, T.; Jacob, J.; Kvetnaya, T.; Mühlichen, A.; Schranzhofer, H.; Schulz, S.; Slowig, B.; Trautwein-Bruns, U.; Voigt, A.; Wiljes, C.: Lernzielmatrix zum Themenbereich Forschungsdatenmanagement (FDM) für die Zielgruppen Studierende, PhDs und Data Stewards, 2023, DOI: 10.5281/zenodo.8010617.

[Rî20] Rîşnoveanu, A.: How can the teacher shape attitudes? - A taxonomy of the affective field. Strategic changes in security and international relations, S. 157, 2020.

[Wi16] Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; Da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; 't Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1), S. 160018, 2016, ISSN: 2052-4463, DOI: 10.1038/sdata.2016.18, URL: https://www.nature.com/articles/sdata201618.

VR/AR und immersives Lernen II

Immersive Räume zur Kreativitätsunterstützung: Ein intelligenter Lehr- und Lernraum

Andreas Fuchs [1], Sven Appel [2] und Paul Grimm [3]

Abstract: Dieser Beitrag präsentiert einen neuartigen Ansatz zur Gestaltung immersiver Räume für die Hochschullehre, die basierend auf Verhalten, gesprochenem Wort und Stimmung eine Unterstützung für kollaborative Kreativitätsprozesse bieten. Ziel ist es, Lehrenden sowie Lernenden in einer interaktiven Virtual Reality-Umgebung durch KI-analysierte und -generierte Inhalte neue Gedankenanstöße zu geben. Durch die Integration von Natural language processing (NLP) und künstlicher Intelligenz wird die Mensch-Computer-Interaktion verbessert, um eine nahtlose Zusammenarbeit zu fördern. Das intelligente System verarbeitet Nutzerdaten und passt die Umgebung an die individuellen Bedürfnisse der Teilnehmenden an. Dies ermöglicht kollaboratives Arbeiten in einer geteilten und zugleich individualisierten Umgebung. Die Anwendung nutzt generative KI zur Erzeugung von Bildern, die auf der verarbeiteten Sprache bzw. den Gesprächsinhalten basieren und beeinflusst gestaltende Elemente wie Beleuchtung, Farbstimmung und Akustik. Der Beitrag erörtert technische Aspekte und potenzielle Anwendungen in Bildung, Unterhaltung und am Arbeitsplatz. Die Forschungsergebnisse deuten darauf hin, dass dieser Ansatz vielversprechend ist, um Kreativität zu fördern und das Wohlbefinden zu steigern.

Keywords: Virtual Reality, Artificial Intelligence, Large Language Models, Generative AI, Education

1 Einleitung

Kollaborative Ideenfindung ist in vielen Sektoren essenziell, aber herausfordernd hinsichtlich Motivation und Kreativität [AH18]. Die Gestaltung eines adaptiven Raums für diese Zwecke erfordert die Balance zwischen Anregung und Ablenkungsfreiheit [Ma13]. Einerseits muss die Umgebung anregend genug sein, um Kreativität, Engagement und Motivation der Nutzenden zu fördern. Andererseits sollte die Umgebung nicht so ablenkend sein, dass sie den kreativen Prozess behindert und den Ideenfluss unterbricht. Virtual Reality (VR) hat sich als förderlich für Zusammenarbeit und Kreativität erwiesen [Do22]. Jedoch fehlt es oft an Personalisierung und dynamischer Anpassungsfähigkeit [Go22]. Experteninterviews zeigen den Bedarf an einem flexiblen Raum, der Kreativität und Innovation unterstützt, da statische Umgebungen oft limitierend wirken.

Interviews mit Designer:innen und Wissenschaftler:innen sowie erste Tests [FAG23] haben den Bedarf an einem kollaborativen und kreativen Raum hervorgehoben, der als Basis für

1 Hochschule Darmstadt, Media, Max-Planck-Straße 2, 64807 Dieburg, Germany, andreas.fuchs@h-da.de, https://orcid.org/0000-0003-1477-9556
2 Hochschule Darmstadt, Informatik, Birkenweg 7, 64295 Darmstadt, Germany, sven.appel@h-da.de, https://orcid.org/0009-0003-7796-6970
3 Hochschule Darmstadt, Media, Max-Planck-Straße 2, 64807 Dieburg, Germany, paul.grimm@h-da.de, https://orcid.org/0000-0003-4189-2642

doi:10.18420/delfi2024_26

zukünftige Forschung dient. Die Befragten äußerten, dass traditionelle, statische Räume Kreativität und Anpassungsfähigkeit begrenzen, wobei besonders räumliche Anordnung und Architektur als einschränkend für Zusammenarbeit und kreatives Denken empfunden werden. Eine auf individuelle Bedürfnisse zugeschnittene und flexible Gestaltung wird als vorteilhaft für die Förderung von Kreativität und Ideenfindung angesehen.

Diese Arbeit präsentiert das *Imaginarium*, einen kollaborativen VR-Raum, der Künstliche Intelligenz (KI) einsetzt, um eine immersive, personalisierte Plattform für kreative Arbeitsprozesse zu bieten (Abb. 1). Das System analysiert Nutzerinteraktionen und passt den Raum dynamisch an, um kreative Prozesse zu unterstützen. Ziel ist die Bereitstellung eines interaktiven und ansprechenden Erlebnisses, das sich an die Bedürfnisse der Nutzenden in Echtzeit anpasst. Erste Tests mit diversen Gruppen haben stattgefunden.



Abb. 1: Das *Imaginarium* aus Sicht eines Nutzenden (links) sowie Darstellung eines kollaborativen Arbeitens in einem Multi-User-Szenario (rechts).

Die Arbeit ist wie folgt aufgebaut: Es wird zunächst ein Überblick über relevante Arbeiten im Bereich VR und kollaborative Umgebungen vermittelt. Dabei wird besonderer Wert auf wichtige Studien gelegt und Chancen identifiziert, die durch das vorgeschlagene Konzept weiterentwickelt werden können. Anschließend erfolgt eine detaillierte Beschreibung des Konzepts sowie der VR-Raum-Implementierung. Daraufhin werden die Ergebnisse von Benutzertests präsentiert und der Umfang sowie die Auswirkungen des Konzepts diskutiert. Abschließend werden die Ergebnisse zusammengefasst und Empfehlungen für zukünftige Forschung in diesem Bereich gegeben.

2 Verwandte Arbeiten

VR hat sich als wirksam erwiesen, um Kreativität und Ideengenerierung zu fördern. 2017 betonte Carruth die Vorteile von VR für Kreativität und Innovation, indem Nutzer in einer virtuellen Umgebung experimentieren können [Ca17]. 2020 hat He et al. CollaboVR eingeführt, ein rekonfigurierbares Framework für kreative Zusammenarbeit in VR, das es mehreren Nutzern ermöglicht, virtuelle Objekte gemeinsam zu erstellen [HDP20]. Bourgeois-Bougrine und Bonnardel zeigten 2022, dass immersive VR-Umgebungen die

persönliche und Gruppenkreativität durch verbesserte Aufmerksamkeit und reduzierte kognitive Belastung steigern [Bo22]. Hagedorn et al. untersuchten 2023, wie immersive Umgebungen Denkprozesse während der kreativen Ideenfindung beeinflusst, und fanden heraus, dass VR Ablenkungen reduziert und eine anregende Umgebung eine positive Wirkung erzielt [HDA23].

Die Cognitive-Affective Theory of Learning with Media (CATLM) von Moreno und Mayer und das darauf aufbauende CAMIL-Modell von Makransky und Petersen postulieren, dass effektives Lernen durch die Integration kognitiver und affektiver Prozesse erreicht wird [MM07; MP21]. Diese Theorien unterstreichen die Bedeutung von Präsenz, Emotion und Motivation im immersiven Lernen und die Notwendigkeit, Lernenden Handlungsspielraum und Autonomie zu gewähren.

Die vorliegende Arbeit erweitert diese Erkenntnisse durch die Einführung eines VR-Raums, in welchem umfangreiche Sprachmodelle und generative KI für eine immersive und personalisierte kollaborative Ideenfindung genutzt werden. Durch die Kombination von KI-Techniken mit den Prinzipien von CATLM und CAMIL zielt die Arbeit darauf ab, ein personalisiertes Erlebnis zu schaffen, das individuelle Präferenzen berücksichtigt und den Ideenfindungsprozess verbessert.

3 Konzept

Um einen ansprechenden, motivierenden und für den Nutzenden relevanten kreativen Raum zu schaffen, wurden die folgenden Punkte unter Beachtung der CATLM- und CAMIL-Grundsätze entwickelt:

1. Adaptive Systeme passen die Umgebung an die individuellen Bedürfnisse und Präferenzen der Nutzenden an. Mittels Sensoren, Kameras und maschinelles Lernen werden Beleuchtung, Akustik und visuelle Elemente modifiziert, um ein personalisiertes, immersives Erlebnis zu schaffen. Der Raum reagiert auf verschiedene Nutzeranforderungen wie Konzentration, Entspannung oder soziale Interaktion, um eine förderliche Atmosphäre zu schaffen.
2. Die Umwandlung gesprochener Worte in KI-generierte Bilder schafft unerwartete visuelle Darstellungen, die neue Ideen und Diskussionsansätze stimulieren. Diese Methode ermöglicht es Nutzenden, Konzepte auf neuartige Weise zu visualisieren, wodurch die kreativen Prozesse durch die Anregung unkonventioneller Denkwege bereichert werden. Die entstehenden Bilder dienen als Katalysator für die Generierung origineller Ideen und tragen dazu bei, komplexe Inhalte auf intuitive und greifbare Weise zu kommunizieren.
3. Das *Imaginarium* passt sich durch Farben, Klänge und Beleuchtung an die verbalen Äußerungen der Nutzenden an, um eine inspirierende und objektive Diskussionsumgebung zu fördern. Bei subjektiven Abweichungen lenkt es die Diskussion visuell

zurück zu Objektivität und Ausgewogenheit. Es ermöglicht individuell angepasste Erlebnisse in einer kollektiven VR-Umgebung, wobei jede Person eine auf ihre persönlichen Emotionen und Vorlieben abgestimmte Atmosphäre erleben kann, was eine zentrale Stärke des Systems darstellt.

4. Die Raumgestaltung und die Visualisierung gesprochener Worte in Bildern sollten ansprechend und anregend sein, ohne die Nutzenden in ihren Aktivitäten zu stören. Intelligente Systeme müssen unaufdringlich und nahtlos funktionieren, sich Verhalten und Stimmung der Nutzenden anpassen und deren Erlebnis verbessern, ohne die Konzentration zu beeinträchtigen.

3.1 Anwendungsarchitektur

Ein System, welches die Anforderungen umsetzen kann, muss gesprochene Worte mithilfe von Speech-to-Text-Technologie in Text umwandeln und NLP-Methoden zur Identifizierung von Schlüsselbegriffen einsetzen, die das Kernthema des Gesprächs erfassen. Für die Visualisierung der Gesprächsinhalte wird ein Verfahren benötigt, das eine Text-zu-Bild-Konvertierung ermöglicht. Die Ableitung der Stimmung des Nutzenden erfordert darüber hinaus eine Sentimentanalyse, die Rückschlüsse auf den emotionalen Status zulässt. Die folgenden Abschnitte beschreiben jeden Bereich im Detail und erläutern, wie die erforderlichen Unterfunktionen innerhalb der Anwendung bearbeitet werden:

Speech-to-Text bezieht sich auf den technologischen Prozess der Umwandlung von gesprochener Sprache in geschriebenen Text. Dabei wird der Ton über ein Mikrofon aufgenommen und ein Verfahren verwendet, um die gesprochenen Wörter zu erkennen und in ein Textformat umzuschreiben. Es ist wichtig zu beachten, dass dieser Prozess an die jeweilige Sprache angepasst werden muss, da verschiedene Sprachen unterschiedliche phonetische und grammatikalische Strukturen aufweisen.

Die **Text-to-Mood**-Verarbeitung erfolgt durch Sentimentanalyse, die mithilfe von maschinellem Lernen die Stimmung aus Gesprächsthemen extrahiert und Text als positiv, negativ oder neutral klassifiziert. Diese Technik ermöglicht die Schaffung einer immersiven Umgebung, die auf die emotionalen Zustände der Nutzer reagiert. Bei erkanntem Stress kann die Umgebung beispielsweise Beleuchtung, Musik und Geräusche dynamisch anpassen, um beruhigende, motivierende und unterstützende akustische Signale zu bieten. Zudem kann der adaptive Raum seine Form und Größe automatisch ändern, um Engegefühl zu reduzieren und Komfort sowie räumliche Wahrnehmung zu verbessern.

Die Transformation von **Text-to-Image** erfolgt mittels eines fortschrittlichen Ansatzes, der Diffusionsmodelle [Ro22] breitgefächert einsetzt. Diese Methode nutzt die Stärken diffusionsbasierter Strukturen, um aus Textprompts mithilfe generativer KI Bilder zu generieren. Der Vorgang ist rechenintensiv und benötigt fortgeschrittene Rechenleistungen, wie GPUs oder Cloud-Services, um hochwertige, kontextgetreue Bilder zu erstellen. Eine geringe Latenz ist essenziell, um die Übereinstimmung der Bilder mit dem laufenden Dialog

zu gewährleisten. Zur Berücksichtigung der Beiträge verschiedener Nutzer:innen und zur Wahrung der Aktualität wird ein Warteschlangensystem für Eingaben empfohlen, das neue hinzufügt und veraltete entfernt, sobald ein Limit erreicht ist, um die Balance zwischen Einbeziehung und Relevanz zu halten.

Adaptive Raumanpassung in virtuellen Umgebungen kann durch den synergetischen Einsatz von Speech-to-Text-, Text-to-Mood- und Text-to-Image-Technologien erreicht werden, um eine optimale Arbeitsatmosphäre zu schaffen: *Text-to-Image* liefert KI-generierte Bilder, die beispielsweise auf großen Bildschirmen im virtuellen Raum angezeigt werden und so neue Denkprozesse auslösen. Die Nutzenden können diese Bilder einfrieren, sodass sie sich nicht mehr von selbst verändern können. Auf diese Weise können Bildsets zusammengestellt werden, die weitere Arbeitsschritte optimal begleiten können. *Text-to-Mood* verwendet vordefinierte Einstellungen, um *entspannte*, *fesselnde* oder *anregende* Arbeitsumgebungen zu schaffen. Dazu wird zu Beginn ein gewünschtes Szenario ausgewählt, zum Beispiel *entspannen*. Unter Berücksichtigung der durch die Analyse der Speech-to-Text gewonnenen Informationen sowie des Verhaltens der Nutzenden werden entsprechende Anpassungen im Raum vorgenommen, sodass die aktuelle Stimmung der Teilnehmenden so manipuliert wird, dass sie sich in Richtung der zuvor definierten positiven Kommunikationsrichtlinien verschiebt. Dies kann unter anderem durch eine beruhigende Stimmungsbeleuchtung, gedämpfte, warme Farben und Raumelemente, die beruhigend wirken, erreicht werden.

3.2 Server Architektur

Die Analyse transkribierter Sprache nach Inhalt und Kontext mithilfe von NLP-Techniken erfordert umfangreiche Rechenressourcen. Schritte wie das Zusammenfassen von Texten zur Erfassung ihres Inhalts und das Erzeugen von Bildaufforderungen sind Teil dieser Prozesse. Diese Prompts dienen dazu, die zusammengefassten Texte in nachfolgenden Schritten durch KI in Bilder umzuwandeln. Die integrierte Hardware von VR-Geräten ist jedoch nicht in der Lage, all diese Vorgänge auszuführen. Um diese Einschränkung zu umgehen, wird ein Middleware-Server als Auslagerung für rechenintensive Aufgaben eingesetzt. Als zentraler Knotenpunkt sichert und verwaltet er Daten und ermöglicht so eine effiziente Analyse, die Verarbeitung von Spracheingaben und die gleichzeitige Bilderzeugung aus Text für mehrere Benutzende. Durch die Zusammenführung von Texttranskriptionen verschiedener Nutzenden wird das Verständnis der gesamten Konversation und der diskutierten Themen verbessert. Daher ist es unerlässlich, eine ressourceneffiziente Datenstromverarbeitung zu implementieren, um eine Entlastung der VR-Geräte bezüglich der Prozessverarbeitung zu erreichen. Dies gewährleistet ein hochwertiges Nutzer:innenerlebnis, da die VR-Geräte von rechenintensiven Aufgaben entlastet werden.

Außerdem können die auf diese Weise verarbeiteten und gespeicherten Daten effizienter über das Netz synchronisiert und an die verschiedenen Netzwerk-Clients verteilt werden, um alles auf dem gleichen Stand zu halten. Um die Datenübertragung zu optimieren, wird jedes erzeugte Bild durch einen Identifikator (Token) repräsentiert, der dann anstelle der

vollständigen Bilddaten an die Netzwerk-Clients verteilt wird, was zu einer erheblich geringeren Datennutzung im Netzwerk führt. Wie in Abb. 2 dargestellt, können die Netzwerk-Clients dann das bereitgestellte Token verwenden, um das entsprechende Bild selbst aus der Datenbank anzufordern, wobei sie bei Bedarf nur das Token bereitstellen. Dieser Ansatz reduziert die Latenzzeit und die Bandbreitennutzung, insbesondere in Fällen, in denen viele Clients wiederholt auf dieselben Bilder zugreifen müssen, was zu einer verbesserten Leistung und Bandbreite bei der Audioübertragung für Sprachchats führt.

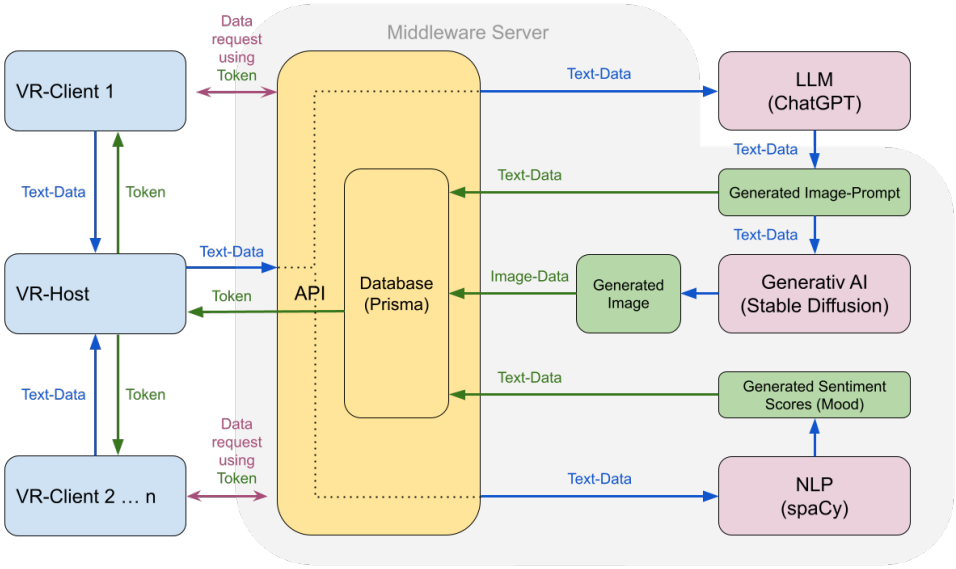


Abb. 2: Flussdiagramm eines beispielhaften Kommunikations- und Datenflusses zwischen den internen und externen Komponenten: 1. Speech-to-Text-Daten werden im Host-System gesammelt und gebündelt. 2. Die gebündelten Daten werden an den Middleware-Server (API) übertragen. 3. Der Middleware-Server steuert die folgenden Prozesse: 3.a LLM (*ChatGPT* als externer Dienst) wird verwendet, um Bildaufforderungen zu erzeugen. 3.b Generative KI (*Stable Diffusion*) erstellt Bilder auf der Grundlage dieser Aufforderungen. 3.c NLP (*spaCy*) generiert Stimmungswerte auf der Grundlage der ursprünglichen Texteingabe. 4. Die Ergebnisse der oben genannten Prozesse werden in der Datenbank (Prisma) gespeichert. 5. Die API antwortet, indem sie ein Token ausgibt. 6. Die Clients verwenden dieses Token, um die generierten Informationen anzufordern, wodurch der Datenverkehr zwischen den Hosts und den Clients reduziert wird.

Wie in Abb. 2 schematisch dargestellt, ist der Middleware-Server zentral für die Verarbeitung verschiedener Komponenten. Die Spracherkennung erfolgt direkt auf dem VR-Gerät, während Prozesse wie Textanalyse, Zusammenfassung, Bilderzeugung und Sentimentanalyse an den Server ausgelagert sind. Nach Übertragung der Textdaten vom VR-Host zum Server erfolgt eine NLP-basierte Vorverarbeitung zur Extraktion von Gesprächsinhalten und Sprecherstimmung, einschließlich des Entfernens von Füllwörtern und unerwünschten Inhalten. Die Stimmungsanalyse nutzt benutzerdefinierte Filterregeln und der Text-to-Image-

Prozess wandelt Text in visuelle Darstellungen um. Die resultierenden Daten – verarbeiteter Text, Inhaltsinformationen, Bilderzeugungsaufforderung und Stimmungsanalyse – werden sicher gespeichert, um zukünftige Analysen zu ermöglichen.

Im Ergebnis führt der Middleware-Server die Ergebnisse der Konversation effizient zusammen und bietet einen präzisen Überblick über die wichtigsten diskutierten Themen, die von den Nutzenden geäußerten Emotionen und andere relevante Erkenntnisse. Durch die Zentralisierung der Daten kann die Stimmungsanalyse auch Einblicke in die kollektive Stimmung einer Gruppe und nicht nur von Einzelpersonen geben, was sich bei Gruppendiskussionen und kollaborativen kreativen Prozessen als nützlich erweist. Letztendlich bietet der Middleware-Server eine skalierbare Lösung für die Analyse großer Datenmengen, was zu genaueren Ergebnissen und einer verbesserten Benutzer:innenfreundlichkeit der entwickelten Anwendung führt.

4 Implementierung

Es wurde eine VR-Anwendung erstellt, die die beschriebenen Technologien umsetzt. Um die Kontrolle über die Gestaltung der Umgebung zu haben und die Raum-, Licht- und Akustik gezielt beeinflussen zu können, wurde die VR-Technologie eingesetzt. Die flexible und ortsunabhängige Nutzung der Anwendung wurde am effizientesten durch Inside-Out-Tracking erreicht, wodurch keine zusätzliche Hardware benötigt wurde und die VR-Brille schnell in Betrieb genommen werden konnten, ohne besondere Anforderungen an den physischen Raum oder externe Sensoren. Die Meta Quest 3 wurde aufgrund ihrer breiten Verfügbarkeit und der Unterstützung durch bestehende Entwickler-Frameworks für die Entwicklung und das Testen verwendet.

Als Basis für die Entwicklung dient das Unity-Framework [Un23]. Um die Fernnutzung und kollaborative Funktionen zu ermöglichen, beinhaltet die Anwendung Netzwerkfunktionalitäten von Photon PUN [Ex23a] und Photon Voice [Ex23b]. Diese Dienste integrieren eine Server-Infrastruktur, die eine nahtlose Kommunikation und Interaktion zwischen mehreren Benutzenden in der virtuellen Umgebung ermöglicht. Für die Transkription von Sprache in Text nutzt die Anwendung den Dienst wit.ai von Meta [Me23]. Dieser Dienst wandelt gesprochene Worte in geschriebenen Text um und ermöglicht eine effiziente Verarbeitung und Analyse von Benutzer:inneneingaben. Für die Inhaltsanalyse und die Generierung von Prompts, die als Eingabe für die Bildgenerierung dienen, nutzt die Anwendung das Application Programming Interface (API) von ChatGPT [Op23a]. ChatGPT ist ein von OpenAI entwickeltes KI-Sprachmodell, das die Anwendung in die Lage versetzt, dynamische und kontextbezogene Prompts auf der Grundlage von Benutzer:inneninteraktionen zu erzeugen. Die korrekte Erstellung der Prompts ist ein entscheidender Aspekt von GPT-basierten Chat-Modellen. Der Prompt, der als Eingabe für das Modell dient, gibt die Richtung für die Antwort des Modells vor. Er liefert den notwendigen Kontext und spezifiziert die Aufgabe, die das Modell erfüllen soll. Daher hat die Qualität der in das Modell eingegebenen Prompts einen direkten Einfluss auf die Qualität und Relevanz der generierten Antworten.

Um die zuverlässige Verarbeitung des aus den Äußerungen der Nutzenden transkribierten Textes zu gewährleisten, wurde ein präziser Ablauf von Anweisungen für die KI entwickelt. Dieser Prozess führt das Modell durch eine sequenzielle Abfolge von drei Aufgaben: Die resultierende Anweisung, die jedem Satz beigefügt wird, lautet: *"Führe die folgenden Aufgaben aus: 1. Fasse den Text in englischer Sprache zusammen. 2. Erzeuge einen Bildprompt aus dieser Zusammenfassung. 3. Generiere eine Antwort basierend auf Schritt 2, ohne das Präfix 'Bildprompt'. Sollte eine Zusammenfassung nicht möglich sein, verwende Schlüsselwörter aus dem Text. Scheitert auch dies und kann keine Zusammenfassung erstellt werden, antworte mit der Fehlermeldung 'Fehler: keine Zusammenfassung', gefolgt von den transkribierten Worten der Nutzenden."*.

Dieser Prompt ist so konzipiert, dass eine Ausgabe erzeugt wird, die sich nahtlos in die Weiterverarbeitung des Systems einfügt und nachgelagerten Operationen oder Antworten erleichtert. So wird das Modell angewiesen, die Zwischenschritte auszulassen und stellt somit sicher, dass der Middelware Server nur das finale, relevante Ergebnis erhält. Wenn das Modell Schwierigkeiten hat, den Inhalt zusammenzufassen, wird es angeleitet, stattdessen Objekte aus dem Text zu verwenden. Wenn eine Zusammenfassung nicht möglich ist, wird das Modell angewiesen, "Fehler: keine Zusammenfassungäuszugeben. Diese letzte Anweisung stellt sicher, dass das Modell unter allen Umständen eine eindeutige Antwort an das System gibt, wodurch die Kommunikation aufrechterhalten und der Status der Aufgabe angezeigt wird. Diese Integration ermöglicht eine interaktive und adaptive Erfahrung, bei der die Umgebung anpassend reagiert und auf der Grundlage der Eingaben des Systems Änderung an den Elementen des virtuellen Raums vornimmt. Durch die Integration dieser Technologien und Dienste bietet die Anwendung ein umfassendes und immersives Erlebnis, das Echtzeitkommunikation, Speech-to-Text-Transkription, Inhaltsanalyse und Prompt kombiniert.

Zusätzlich zu den genannten Techniken spielt die Stimmungsanalyse eine wichtige Rolle bei der Ableitung relevanter Erkenntnisse aus gesprochenen Sätzen. Dabei werden die Fähigkeiten des *spaCy* NLP-Frameworks [Ex23c] in Kombination mit der Sentimentanalyse-Komponente *spaCyTextBlob* [Sa22] genutzt. Die Stimmungsanalyse extrahiert Eigenschaften wie Subjektivität und Polarität, um persönliche Meinungen und die Stimmungsausrichtung (positiv, negativ, neutral) in Sätzen zu bewerten.

Für das Hosting und den Betrieb des Middleware-Servers zur Vorverarbeitung der Sprachdaten sowie des Servers zur Bilderzeugung wird die Cloud-Computing-Plattform OpenStack [Op23b] verwendet. Das Front- und Backend des Middleware-Servers basiert auf node.js. Eine mit Prisma [Pr23a] abstrahierte MySQL-Datenbank wird als Datenbank für die Persistenz der Bild- und Sprachinformationen auf dem Middleware-Server verwendet. Die WebUI Automatic1111 [Au23], die Stable Diffusion mit dem Openjourney [Pr23c] Modell betreibt, wird für die Bilderzeugung verwendet. Damit die Bilderzeugung nahezu in Echtzeit abläuft, wird für die Berechnung eine Nvidia A100 Tensor Core GPU verwendet, die Bilder in wenigen Sekunden erzeugen kann.

5 Ergebnisse und Diskussion

Ziel dieser Forschung war es, eine immersive, interaktive Umgebung zu schaffen, die eine personalisierte Erfahrung bietet und sich nahtlos an die Handlungen der Nutzenden anpasst. Der vorgestellte VR-Raum, das *Imaginarium*, soll die Herausforderungen traditioneller Methoden der kollaborativen Ideenfindung überwinden, indem er sich den Anforderungen der Nutzenden in Echtzeit anpasst. Zur Bewertung des *Imaginariums* wurden verschiedene Evaluationsmethoden eingesetzt:

Mit 30 Personen, aufgeteilt in drei Gruppen von jeweils 10 Personen, wurden **Nutzertests und Befragungen** durchgeführt. Jede Gruppe interagierte gemeinsam in der VR-Anwendung und behandelte eine bestimmte thematische Problemstellung. Direkt nach der Nutzung wurden sie in einem leitfadengestützten Interview zu ihren Erfahrungen befragt. Zusätzlich wurde eine anonyme schriftliche Umfrage mittels Fragebogen durchgeführt, die von 13 der 30 Teilnehmenden ausgefüllt wurde. In den Kategorien „volle Zustimmung“, „Zustimmung“, „Ablehnung“ und „volle Ablehnung“ wurden verschiedene Fragen bewertet.

Die Umfrageergebnisse zeigten, dass 58,3% der Teilnehmenden der Meinung waren, dass die KI-unterstützte Bilderzeugung ihnen half, sich leichter in die Erzählungen ihrer Mitstudierenden hineinzuversetzen und ein besseres Verständnis für das Thema oder Problem zu entwickeln, während 41,7% dies ablehnten. Bei der Bewertung der virtuellen Lernumgebung in Bezug auf die Erreichung der Lernziele antworteten 41,7% mit „Ablehnung“, 41,7% mit „Zustimmung“ und 16,7% mit „voller Zustimmung“. Es zeigte sich, dass es den Lernenden schwerfiel, Fragen mit Bezug zum Lernziel zu beantworten, was vermutlich daran lag, dass der kreative Ansatz auf offene Prozesse ausgerichtet war und nicht auf ein bestimmtes Arbeitsergebnis abzielte. Drei Viertel der Teilnehmenden würden die Anwendung anderen Studierenden weiterempfehlen, insbesondere in den Bereichen Sprachenlernen, Ingenieurkonzepte, Visualisierung abstrakter Konzepte, Prototyping, Zusammenarbeit, Ideenfindung sowie in Meetings zur Visualisierung von Problemen, Ideensammlung und Inspiration.

Die Interviews bestätigten, dass die Erfahrung die Zusammenarbeit, Motivation und Kreativität förderte. Die Teilnehmenden berichteten, dass die individuell angepasste Umgebung ihnen half, mehr Ideen zu entwickeln und sich stärker in den Prozess einzubringen. Sie stellten fest, dass die immersive Erfahrung unerwartete Bilder und Anregungen hervorbrachte, die ihren Horizont bei den Brainstorming-Themen erweiterten. Die VR-Technologie erleichterte es ihnen, sich auf das kreative Arbeiten zu konzentrieren, ohne durch das physische Umfeld abgelenkt zu werden.

In einer **Beobachtungsstudie** im *Imaginarium* wurde untersucht, wie dynamische visuelle Umgebungskonfigurationen Einzel- und Gruppeninteraktionen beeinflussen, insbesondere während Diskussionen und Debatten. KI-generierte Bilder, die thematisch mit den Diskussionsthemen verknüpft waren, führten bei den Teilnehmenden zu Überraschungsmomenten, die neue Perspektiven und Ideen anregten. Diese visuellen Abweichungen intensivierten die Diskussionen und förderten Argumentationen sowie die Erkundung alternativer Stand-

punkte. Dieser Effekt kann auf die Variabilität und den stimulierenden Charakter der Bilder zurückgeführt werden.

Ein **Experteninterview** mit einem Spezialisten für Moderations-, Ideenfindungs- und Designmethoden im geschäftlichen Kontext offenbarte, dass für einen kontrollierten und zielgerichteten Arbeitsprozess möglicherweise mehrere Räume und Szenarien erforderlich sind. Ein Ansatz könnte darin bestehen, zunächst in individueller Arbeit mit vordefinierten Bildern zu beginnen und diese dann in einer Gruppenphase mit den Erkenntnissen und Ideen anderer Teilnehmender zu kombinieren. Dieses Vorgehen könnte es ermöglichen, einzelne Aspekte stärker zu gewichten, unabhängig von der anfänglichen Gruppendynamik. Die aktuelle Anwendungsentwicklung unterstützt die Integration und Erprobung solcher Prozesse und Workflows im Rahmen des bestehenden Multi-User-Szenarios, was die Implementierung und einheitliche Evaluation dieser Konzepte erleichtert. In Zusammenarbeit mit Fachexperten werden konkrete Szenarien entwickelt, die in einer umfassenden Nutzerstudie praktisch umgesetzt und evaluiert werden sollen.

Wie in Abb. 1 dargestellt, arbeitet eine Gruppe von Personen per Remoteverbindung gemeinsam und diskutiert über die von einer künstlichen Intelligenz generierten Bilder. Eine Stimmungsanalyse, implementiert auf einem Middleware-Server, erfasst die einzelnen Stimmungen und passt das Ambiente des *Imaginariums* entsprechend an. Jede Person in der Gruppe erlebt einen individuellen Raum, der sich dynamisch an den Gemütszustand der Teilnehmenden anpasst.



Abb. 3: Gegenüberstellung von vier Gestaltungsvarianten des virtuellen Kreativraums: Geteilte unterschiedliche Boden- und Wandtexturen (links) sowie auf der Stimmung basierende individuelle Lichtstimmungen (rechts).

Abb. 3 stellt gegenüber, wie das Raumambiente durch dynamische Anpassungen der Wand- und Bodentexturen modifiziert wird, um Ablenkungen zu minimieren und illustriert, wie sich die Lichtstimmung nahtlos an die emotionale Verfassung der Nutzenden anpasst, um die Stimmung zu verstärken oder positiv zu beeinflussen.

Durch die Entwicklung eines Middleware-Servers sind entscheidende Verbesserungen erzielt worden, die die Stabilität der Anwendung erhöhen und Abstürze verhindern. Dadurch ist das zuverlässige Synchronisieren der KI-generierten Bilder zwischen allen Teilnehmern

überhaupt erst möglich, und sowohl die Brillen als auch das Netzwerk werden entlastet. Im Rahmen eines iterativen Entwicklungsprozesses verbesserten sich sukzessive die Verbindungsqualität und die Tonqualität. Die Speicherung von begleitenden Dialogen und Stimmungsbewertungen, zusätzlich zu den Bildern, wird durch einen Datenbankserver ermöglicht. Dies schafft die Grundlage für zukünftige Entwicklungen, sichert die Reproduzierbarkeit von Ergebnissen und erlaubt eine umfassende Datenanalyse.

Vorangegangene Studien haben die Effektivität von VR bei der Steigerung der Kreativität und der Förderung von Zusammenarbeit nachgewiesen [HDA23; HDP20]. Aufbauend auf diesen Erkenntnissen integriert die aktuelle Arbeit NLP und KI-basierte Bildgenerierung, um eine hochgradig immersive VR-Welt zu schaffen. Der Einsatz fortschrittlicher KI-Technologien zielt darauf ab, den Nutzenden ein interaktives und personalisiertes Erlebnis zu bieten, das nicht nur die Kreativität fördert, sondern auch eine natürlichere und intuitive Interaktion mit der virtuellen Umgebung ermöglicht. Der Ansatz unterstreicht die Integration von kognitiven und affektiven Prozessen und steht im Einklang mit der Cognitive-Affective Theory of Learning with Media, um ein engagierendes und bedeutungsvolles Lernerlebnis zu schaffen. Dies geschieht nicht nur durch Anpassungen des Raumes [MM07], sondern auch durch personalisierte Licht- und Musikstimmungen. Das System ist darauf ausgerichtet, die emotionale Befindlichkeit der Nutzenden positiv zu beeinflussen oder bestehende Emotionen zu verstärken und hervorzuheben.

Darüber hinaus ist es bedeutsam, den Nutzenden das Gefühl der Selbstbestimmung und Kontrolle über ihre Umgebung zu geben. Während die Anpassungen derzeit automatisch erfolgen, wäre es von Vorteil, wenn die Nutzenden das erstellte Profil nach ihren Vorlieben ändern könnten, um so ihren virtuellen Raum individuell zu gestalten. Eine solche individuelle Profilanpassung fördert nicht nur das Engagement und die Motivation, sondern verringert auch Ablenkungen.

Insgesamt deuten die Ergebnisse darauf hin, dass das *Imaginarium* eine vielversprechende Plattform für die kollaborative Ideenfindung ist, da es einen immersiven und hochgradig personalisierten Raum bietet, der sich in Echtzeit an die Bedürfnisse der Nutzenden anpasst. Das Erreichen dieses Gleichgewichts förderte die Kreativität und Zusammenarbeit in immersiven Räumen, während Unterbrechungen des kreativen Prozesses vermieden wurden. Zukünftige Forschungen könnten sich auf die weitere Optimierung der Personalisierungen sowie den dynamischen Anpassungen des *Imaginarium* konzentrieren, um die Effektivität weiter zu steigern.

6 Zusammenfassung und Ausblick

In dieser Arbeit wird die Entwicklung einer Multi-User-VR-Umgebung vorgestellt, die Sprache-zu-Text-Technologie und generative KI nutzt, um auf Benutzerinteraktionen zu reagieren und Bilder zu generieren. Der generierte Text aus dem gesprochenen Wort wird mithilfe von Inhaltsprüfung vorverarbeitet, um Aufforderungen für ein KI-System zu

erstellen, das Bilder auf Basis fortlaufender Benutzerbeobachtungen generiert. Durch die Einbeziehung von Inhaltsprüfung und Benutzereingaben bietet die VR-Umgebung eine hochgradig personalisierte Erfahrung, die das Engagement und die Kreativität des Nutzers fördert. Die generierten Bilder werden den Nutzern subtil angezeigt, was eine immersive und dynamische Umgebung schafft.

Insgesamt deuten unsere Ergebnisse darauf hin, dass der VR-Raum mit Multi-User-Fähigkeit, Sprache-zu-Text-Technologie und Inhaltsfilterung das Potenzial hat, weitreichende Auswirkungen auf eine Vielzahl von Branchen zu haben, darunter Bildung, Gesundheitswesen und Spiele. Weitere Forschung ist notwendig, um das volle Potenzial dieser Technologie und ihrer Anwendungen in verschiedenen Bereichen zu erforschen.

Die künftige Arbeit wird sich auf die Verfeinerung der Verarbeitungsprozesse, die Einbeziehung von Nutzerfeedback und die Erforschung weiterer Anwendungen in verschiedenen Branchen konzentrieren. Dazu gehört auch die Gewährleistung von Datenschutz, Sicherheit und ethischen Überlegungen. Ein möglicher Bereich für künftige Arbeiten könnte die Verwendung von Sensoren und Kameras sein, um Echtzeitdaten über das Verhalten und die Stimmung der Nutzenden zu sammeln. Diese zusätzlichen Daten könnten die Fähigkeit des Systems weiter verbessern, die Umgebung an die Bedürfnisse und Vorlieben des Benutzers anzupassen.

Zu den Optimierungen des *Imaginariums* könnte eine Feedback-Schleife gehören, die es den Nutzenden ermöglicht, Rückmeldungen zu den dargebotenen Stimuli zu geben und diese auf der Grundlage ihrer Antworten anzupassen. Durch den Einsatz zusätzlicher Algorithmen des maschinellen Lernens kann die Anwendung nach und nach ein personalisiertes Profil der Benutzerpräferenzen erstellen und so die Umgebung im Laufe der Zeit immer genauer auf die individuellen Bedürfnisse der Benutzenden abstimmen.

Der Middleware-Server und das daraus resultierende Framework für virtuelle Lernumgebungen sollen in Zukunft sowohl als Bundle als auch als Einzelkomponenten zur Verfügung gestellt werden. Damit soll die Erstellung von virtuellen Lernumgebungen vereinfacht und beschleunigt werden. Entsprechend werden strukturierte und gut dokumentierte Assets für die zukünftige Nutzung entwickelt.

7 Danksagung

Diese Arbeit wird gefördert durch das Hessische Ministerium für Wissenschaft und Kunst (HMWK). Wir bedanken uns bei unseren Kollegen und Partnern aus dem Projekt *fuels - Future Learning Spaces* [Pr23b] für ihre Zeit und wertvollen Kommentare.

Literaturverzeichnis

[AH18] Al-Samarraie, H.; Hurmuzan, S.: A review of brainstorming techniques in higher education. Thinking Skills and Creativity 27, S. 78–91, 2018, ISSN: 1871-1871, DOI: https://doi.org/10.1016/j.tsc.2017.12.002, URL: https://www.sciencedirect.com/science/article/pii/S1871187117302729.

[Au23] Automatic1111: Stable Diffusion web UI, accessed May 11, 2024, 2023, URL: https://github.com/AUTOMATIC1111/stable-diffusion-webui.

[Bo22] Bourgeois-Bougrine, S.; Bonnardel, N.; Burkhardt, J. M.; Thornhill-Miller, B.; Pahlavan, F.; Buisine, S.; Guegan, J.; Pichot, N.; Lubart, T.: Immersive Virtual Environments' Impact on Individual and Collective Creativity. European Psychologist 27 (3), S. 237–253, 2022, DOI: 10.1027/1016-9040/a000481.

[Ca17] Carruth, D. W.: Virtual reality for education and workforce training. In: 2017 15th International Conference on Emerging eLearning Technologies and Applications (ICETA). S. 1–6, 2017, DOI: 10.1109/ICETA.2017.8102472.

[Do22] Doerner, R.; Broll, W.; Grimm, P.; Jung, B.: Virtual and augmented reality (VR/AR): Foundations and methods of extended realities (XR). Springer Nature, 2022.

[Ex23a] Exit Games GmbH: PUN The Ease-of-use of Unity's Networking with the Performance & Reliability of Photon Realtime, accessed Mar. 22, 2023, 2023, URL: https://www.photonengine.com/pun.

[Ex23b] Exit Games GmbH: VOICE Crystal-Clear 3D Voice Chat for Unity Games, VR/AR/MR, accessed Mar. 22, 2023, 2023, URL: https://www.photonengine.com/voice.

[Ex23c] ExplosionAI GmbH Inc.: Industrial-Strength Natural Language Processing, accessed Jul. 12, 2023, 2023, URL: https://spacy.io.

[FAG23] Fuchs, A.; Appel, S.; Grimm, P.: Immersive Spaces for Creativity: Smart Working Environments. In: 2023 International Electronics Symposium (IES). 2023, DOI: 10.1109/IES59143.2023.10242458.

[Go22] Gong, Z.; Lee, L. H.; Soomro, S. A.; Nanjappan, V.; Georgiev, G. V.: A systematic review of virtual brainstorming from the perspective of creativity: affordances, framework, and outlook. Digital Creativity 33 (2), S. 96–127, 2022, DOI: 10.1080/14626268.2022.2064879.

[HDA23] Hagedorn, L. J.; De Rooij, A.; Alimardani, M.: Virtual Reality and Creativity: How Do Immersive Environments Stimulate the Brain during Creative Idea Generation? In: Proceedings of the 2023 ACM Conference on Virtual Reality Software and Technology. New York, NY, USA, 2023, DOI: 10.1145/3544549.3585848.

[HDP20] He, Z.; Du, R.; Perlin, K.: CollaboVR: A Reconfigurable Framework for Creative Collaboration in Virtual Reality. In: 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Porto de Galinhas, Brazil, S. 542–554, 2020, DOI: 10.1109/ISMAR50242.2020.00082.

[Ma13] Mandal, S.: Brief introduction of virtual reality & its challenges. International Journal of Scientific & Engineering Research 4 (4), S. 304–309, 2013.

[Me23] Meta: Build Natural Language Experiences, accessed Apr. 18, 2023, 2023, URL: https://wit.ai.

[MM07] Moreno, R.; Mayer, R. E.: Cognitive-affective theory of learning with media. In (Mayer, R. E., Hrsg.): The Cambridge Handbook of Multimedia Learning. Cambridge University Press, S. 161–177, 2007.

[MP21] Makransky, G.; Petersen, G. B.: The Cognitive Affective Model of Immersive Learning (CAMIL): a Theoretical Research-Based Model of Learning in Immersive Virtual Reality. Educational Psychology Review 33, S. 937–958, 2021, DOI: 10.1007/s10648-020-09586-2.

[Op23a] OpenAI Inc.: Introducing ChatGPT, accessed Feb. 13, 2023, 2023, URL: https://openai.com/blog/chatgpt.

[Op23b] OpenInfra Foundation: The Most Widely Deployed Open Source Cloud Software in the World, accessed Feb. 12, 2023, 2023, URL: https://www.openstack.org/.

[Pr23a] Prisma Data Inc.: Next-generation Node.js and TypeScript ORM, accessed Mar. 18, 2023, 2023, URL: https://www.prisma.io/.

[Pr23b] Projekt fuels: Future Learning Spaces, accessed Mai 11, 2024, 2023, URL: https://futurelearning.space/.

[Pr23c] Prompthero: openjourney, accessed Feb. 13, 2023, 2023, URL: https://huggingface.co/prompthero/openjourney.

[Ro22] Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B.: High-Resolution Image Synthesis With Latent Diffusion Models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). S. 10684–10695, 2022.

[Sa22] Sam, E.: A TextBlob sentiment analysis pipeline component for spaCy, accessed June 6, 2023, 2022, URL: https://github.com/SamEdwardes/spaCyTextBlob.

[Un23] Unity Technologies: Unity Technologies, accessed Jan. 21, 2023, 2023, URL: https://unity.com/.

360°-Perspektiven: von der physischen zur virtuellen Ausstellung im Hochschulkontext

Ulrike Mascher[1] und David Fernes[1]

Abstract: Das Paper gibt Einblick in die Entwicklung, Umsetzung und Evaluation einer 360°-Ausstellung im Hochschulkontext. Nach einer kurzen Einführung zum Lernen mit Hilfe von 360°-Technologie wird zunächst der Umsetzungsprozess von der physischen Ausstellung zur 360°-Ausstellung dargestellt. Schließlich werden die Ergebnisse einer Nutzendenbefragung – mittels System Usability Scale sowie Freitexteingaben – vorgestellt. Basierend auf dem Feedback der Nutzer*innen wurde die 360°-Ausstellung in einem nächsten Schritt weiterentwickelt. Die identifizierten Stärken, wie der 360°-Panoramablick, der räumliche Eindruck, selbstgesteuerte, explorative Wissenserschließung, sowie ein hoher Interaktivitäts- und Individualisierungsgrad zeigen das Potenzial der Anwendung. Die dargestellten Erfahrungen und Erkenntnisse können als Best-Practice-Beispiel Orientierung für ähnliche Projekte bieten.

Keywords: 360°-Tour, digitale Ausstellung, Provenienzforschung, Universität

1 Einleitung

Bei 360°-Umgebungen handelt es sich um Foto- oder Videoaufnahmen, die einen Rund- bzw. Panoramablick ermöglichen. Das heißt, dass die Betrachterin ihren Blick in alle Richtungen schweifen lassen kann. Zur Erstellung von 360°-Umgebungen werden meist spezielle 360°-Kameras verwendet, zusätzlich bieten sie die Möglichkeit, multimediale Inhalte mittels sogenannter Hotspots einzubinden und so räumlich zu verknüpfen: „[I]nformation-rich virtual environments embed symbolic information within a realistic 3D environment. [...] In this way, symbolic and perceptual information are integrated in a single environment“ [Bo99](S. 3 f.).

360°-Umgebungen lassen sich nicht nur als Desktopanwendungen nutzen, sondern können über ein Head-Mounted Display auch in Virtual Reality (VR) übertragen werden. Allerdings sind die Interaktionsmöglichkeiten hier beschränkter als in einer programmierten VR-Umgebung. Ein Vorteil von 360°-Umgebungen sind dagegen der im Gegensatz zu VR relativ geringe Produktionsaufwand und die geringen Kosten bei der Erstellung.

In Lehr-/Lernkontexten lassen sich mit Hilfe von 360°-Umgebungen zum Beispiel Orientierungsszenarien umsetzen, etwa für Labore [BKM22] oder zur Vor- oder Nachbereitung von Exkursionen. Auch lassen sich schwer oder nicht (mehr) zugängliche Orte digital erkunden – explorativ oder anhand eines festgelegten Lernpfades. Außerdem können Immersionseffekte

1 Goethe-Universität Frankfurt, studiumdigitale, Eschersheimer Landstr. 155-157, 60323 Frankfurt am Main, Deutschland, mascher@sd.uni-frankfurt.de;
fernes@sd.uni-frankfurt.de, https://orcid.org/0000-0001-7063-6221

doi:10.18420/delfi2024_27

von realitätsnahen 360°-Umgebungen genutzt werden, um beispielsweise Lernende für bestimmte Sachverhalte zu sensibilisieren oder Perspektivwechsel zu unterstützen [Mo20].

Konstruktivistische Ansätze begreifen Lernen als kreativen und individuellen Konstruktionsprozess (s. z.B. [Ke18; Pe20]). Einen hohen Stellenwert nimmt dabei das selbstregulierte bzw. selbstgesteuerte Lernen ein: Lernende erarbeiten weitgehend selbstständig Problemlösungs- und Wissensstrukturen, d.h. sie bestimmen ihren Lernweg in weiten Teilen selbst. Ebenso können Nutzer*innen selbstgesteuert durch Wissensräume in 360°-Umgebungen navigieren. Zahlreiche Studien zeigen positive Effekte von 360°-Anwendungen auf die Motivation und das Interesse von Lernenden (für einen Überblick s. [FS]). Im Museumskontext sehen Hayes & Yoo [HY18] besonders für das Erreichen jüngerer Zielgruppen in der 360°-Technologie großes Potenzial.

2 Entwicklung

2.1 Von der physischen Ausstellung...

Die Ausstellung „StolperSeiten – NS-Raubgut in der Universitätsbibliothek Frankfurt am Main“ war vom 20. Mai bis 28. August 2022 im Schopenhauer-Studio der Universitätsbibliothek Johann Christian Senckenberg zu sehen. Das Schopenhauser-Studio versteht sich als universitären „Kommunikations- und Ausstellungsraum“ [2] , in dem regelmäßig Forschungs- und Lehrprojekte der Goethe-Universität präsentiert werden. Entstanden ist die Ausstellung im Rahmen eines Provenienzforschungsprojekts, in dem die Bibliothek mit Förderung des Deutschen Zentrums Kulturgutverluste ihren Bestand nach während der NS-Zeit geraubten Büchern durchsucht. [3]

In acht Themeninseln nimmt die Ausstellung „StolperSeiten“ die wissenschaftlichen Bibliotheken Frankfurts in der Zeit vom Beginn der NS-Herrschaft 1933 bis zur Nachkriegszeit in den Blick. Dabei ‚stolpert‘ man in der Ausstellung über eine Vielzahl ermittelter Einzelschicksale, zusätzlich werden Arbeitsweisen und Werkzeuge, aber auch Hindernisse der Provenienzforschung thematisiert.

2.2 ... zur digitalen Ausstellung

Besucher*innen der digitalen Ausstellung sollte die Möglichkeit geboten werden, das Erlebnis der ‚echten‘ Ausstellung nachzuerleben. Dazu sollten sie sich frei durch die Ausstellungsräume bewegen und einzelne Ausstellungsstücke genauer betrachten können. Die Möglichkeit einer Erklärung zu einzelnen Teilen der Ausstellung sollte, wie bei einer Führung, auch gegeben sein.

2 https://www.ub.uni-frankfurt.de/ausstellung/raum.html
3 Mehr Informationen dazu: https://www.ub.uni-frankfurt.de/ausstellung/stolperseiten.html

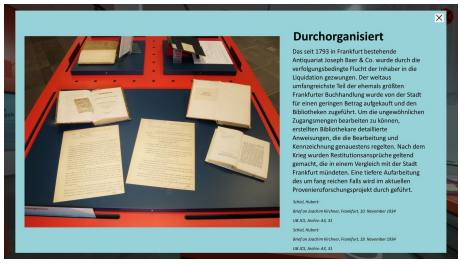
Neben der zeit- und ortsunabhängigen Verfügbarkeit und der digitalen Verstetigung der Ausstellung war außerdem die Erhöhung der Reichweite durch die Digitalisierung ein Anliegen des Projekts. Nachfolgend wird die Umsetzung der digitalen Ausstellung „StolperSeiten 360°“ nachgezeichnet. [4]

2.3 Aufnahmen

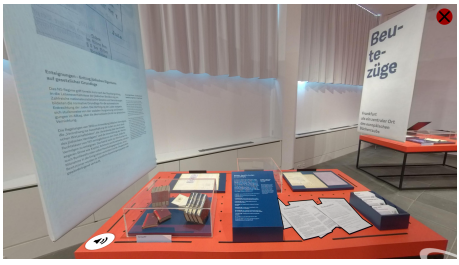
Die Panoramen wurden mit einer Insta360 Pro 2 Kamera aufgenommen. Zusätzlich wurden Nahaufnahmen der Ausstellungsobjekte gemacht. Weitere Bilder sowie bereits digitalisierte Dokumente wurden von den Kurator*innen zur Verfügung gestellt, außerdem wurden Audioaufnahmen zur Erklärung der einzelnen Stationen von den Kurator*innen aufgenommen.



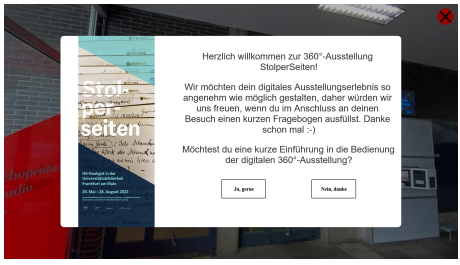
(a) Punkte zum Navigieren der Ausstellung



(b) Infotafeln zu Objekten



(c) Anklickbare Objekte



(d) Optionales Tutorial

Abb. 1: Screenshots der digitalen Tour

2.4 Umsetzung

Aus den 360°-Aufnahmen wurde eine digitale 360°-Tour mit Hilfe der Software 3DVista erstellt, die es ermöglicht die Aufnahmen mit weiteren Interakionsmöglichkeiten anzureichern. In dieser Tour können sich Benutzer*innen von Panorama zu Panorama durch die Ausstellungsräume bewegen, indem sie auf Punkte auf dem Boden klicken (s. Abbildung 1a).

4 http://stolperseiten.studiumdigitale.uni-frankfurt.de/

Ausstellungsstücke können detailliert betrachtet werden, indem man sie anklickt (s. Abbildung 1b). Wenn die Maus über ein Objekt bewegt wird, welches genauer betrachtet werden kann, wird ein Rahmen um dieses Objekt angezeigt (s. Abbildung 1c). Wenn ein Exponat angeklickt wird, öffnet sich ein vergrößertes Bild dieses Exponats. Bei einigen Objekten wird zusätzlich noch ein Erklärungstext angezeigt. Vereinzelt können auch PDF-Scans von Büchern oder Dokumenten betrachtet und durchblättert werden. Dies war in der Ausstellung vor Ort nicht gegeben, so dass die digitale Ausstellung hier zusätzliche Interaktionsmöglichkeiten bietet. Auch ein Video, das weitere Informationen zum Hintergrund der Ausstellung bietet, kann wiedergegeben werden.

Zu jeder der Themeninseln kann optional eine kurze Erklärung der Kurator*innen angehört werden (ähnlich einem Audioguide), indem ein virtueller Knopf am jeweiligen Tisch gedrückt wird. Diese Erklärungen können nach dem Start auch wieder pausiert oder abgebrochen werden. Während die Audiodatei abgespielt wird, kann die Ausstellung weiter frei erkundet werden. Zu Beginn der Tour wird optional eine kurze Einführung in die Bedienung und die Interaktionsmöglichkeiten gegeben, da nicht davon auszugehen ist, dass alle Nutzer*innen mit 360°-Touren vertraut sind (s. Abbildung 1d). In diesem kurzen Tutorial werden Möglichkeiten zur Fortbewegung und zur Auswahl von Exponaten vorgeführt. Erfahrenere Nutzer*innen können diese Erklärung jedoch auch überspringen. Während der Tour kann eine kurze Version dieser Erklärung jederzeit nochmals angezeigt werden.

3 Evaluation

Um Feedback zur Verbesserung von „StolperSeiten 360°“ sowie zukünftiger digitaler 360°-Ausstellungen zu sammeln, wurde eine kurze optionale Befragung mit Besucher*innern der Webseite durchgeführt.

3.1 Stichprobe

An der Befragung nahmen 17 Personen teil. Von diesen Teilnehmenden identifizierten sich vier als männlich, zwölf als weiblich und eine Person machte keine Angabe zu ihrem Geschlecht. Das durchschnittliche Alter der Teilnehmenden betrug 48,21 Jahre (sd = 11.61), wobei drei Personen ihr Alter nicht nannten.

3.2 Instrumente

In der Befragung wurde die Usability der digitalen Ausstellung mit Hilfe der deutschen Version der System Usability Scale (SUS) [Br96] erhoben. Weiterhin wurde offenes Feedback zu der Ausstellung mit den folgenden drei Fragen gesammelt:

- Was hat dir an der digitalen Version der Ausstellung „StolperSeiten“ besonders gut gefallen?
- Was hat dich an der digitalen Version der Ausstellung „StolperSeiten“ gestört?
- Hier ist Platz für deine Anmerkungen:

3.3 Durchführung

Besucher*innen der digitalen Ausstellung wurden beim Start der Tour darauf hingewiesen, dass sie am Ende an einer optionalen Befragung teilnehmen können. Danach konnten sie nach Belieben die Ausstellung erkunden. Beim Schließen der Tour wurden sie dann gefragt, ob sie an der Befragung teilnehmen möchten. Wenn sie sich dafür entschieden, wurden sie zu einem Onlinefragebogen weitergeleitet.

3.4 Ergebnisse

Im Mittelwert erhielt die Ausstellung einen SUS-Wert von 79,12 (sd = 22,08). Nach Bangor et al. [BKM09] entspricht dieser SUS-Wert einer „guten“ Usability.

Auf die Frage, was den Teilnehmenden besonders gut gefallen habe, wurde die räumliche und zeitliche Unabhängigkeit der digitalen Ausstellung genannt. Auch dass durch die Kombination von Bildern und Tonaufnahmen mehrere Sinne angesprochen werden, wurde als positiver Punkt genannt.

Auf die Frage, was sie gestört habe, nannten die Teilnehmer*innen teilweise unscharfe Bilder sowie zu kleine Texte, was das Betrachten von Objekten und das Lesen von Erklärungen erschwert habe. Auch wurde kritisiert, dass die Audio-Erklärungen sich zwar pausieren ließen, aber sonst keine Möglichkeit zur Steuerung der Wiedergabe – beispielsweise in Bezug auf die Abspielgeschwindigkeit – vorhanden war.

Als weitere Anmerkung wurde der Wunsch geäußert, noch mehr Objekte – beispielsweise ganze Bücher – als Digitalisate anzubieten, um diese im Detail betrachten zu können. Eine Person betonte, dass digitale Ausstellungen keine Ausstellungen vor Ort ersetzen, sondern diese vielmehr ergänzen sollten. Obwohl die Usability bereits gut ausfiel, zeigten die Rückmeldungen, dass es dennoch Potenzial gibt, die digitale Ausstellung noch weiter zu verbessern.

3.5 Einschränkungen

Da die Befragung am Ende der Ausstellung durchgeführt wurde, konnten nur Daten von Nutzer*innen erhoben werden, die sich bereits dafür entschieden hatten, die digitale

Ausstellung zu besuchen. Dadurch konnten keine Information dazu gewonnen werden, welche Änderungen nötig wären, um ein breiteres Publikum anzusprechen. Da die Umfrage nur aufgerufen werden konnte, wenn die Ausstellung über den eingebauten „Beenden"-Knopf verlassen wurde, besteht die Möglichkeit, dass einige Nutzer*innen nicht an der Umfrage teilnehmen konnten, wenn sie stattdessen direkt ihr Browserfenster geschlossen haben. Diese Einschränkungen führen dazu, dass nur 17 Besucher*innen der digitalen Ausstellung diese auch bewerteten. Diese niedrige Anzahl macht es schwer, objektive Rückschlüsse aus der Befragung zu ziehen, das Feedback konnte trotzdem verwendet werden, um Verbesserungspotentiale der digitalen Ausstellung zu identifizieren.

4 Fazit

Diese Verbesserungen wurden abschließend, soweit technisch möglich, in eine zweite Version der Ausstellung eingebaut. Dazu wurden etwa die Erklärungstexte so angepasst, dass sie sich dynamisch an die Bildschirmgröße adaptieren und damit auf unterschiedlichen Geräten, von Smartphones bis Desktop-PCs, gut lesbar sind. An einigen Stellen wurden schärfere Bilder eingefügt. Neben den aus dem Feedback der Nutzer*innen gewonnenen Änderungsvorschlägen wurden auch noch weitere Anpassungen vorgenommen. So wurde beispielsweise ein einfaches Gamification-Element eingebaut mit dem Ziel, anhand von drei Dokumenten verschiedene Provenienzmerkmale zu finden. Außerdem wurden, wie von den Nutzer*innen gewünscht, weitere Digitalisate von Dokumenten eingefügt.

Durch diese Anpassungen konnten die von den Nutzenden identifizierten Mehrwerte der Ausstellung noch weiter ausgeschöpft werden. Diese beinhalten vor allem den 360°-Rundumblick sowie den räumlichen Eindruck, also sozusagen die ‚Begehbarkeit' der digitalen Ausstellung. Im Sinne konstruktivistischer Auffassungen wurde die selbstgesteuerte, explorative Wissenserschließung positiv wahrgenommen und User*innen erlebten somit ihren ganz individuellen Ausstellungsbesuch.

Weitere von den Nutzenden wahrgenommene Vorteile waren der zeit- und ortsunabhängige Zugang, die Interaktionsmöglichkeiten mit den Objekten sowie die zeitliche Unbegrenztheit des virtuellen Ausstellungsbesuchs.

Zusätzliche Vorteile, die zu nennen sind, sind die digitale Verstetigung über die Dauer der physischen Ausstellung hinaus sowie eine höhere Reichweite. Mit Blick auf weitere bereits in den digitalen Raum übertragene Ausstellungen der Universitätsbibliothek Frankfurt fällt auf, dass es sich dabei primär um Digitalisate von Textdokumenten handelt, die meist linear angeordnet sind.[5] Es bleibt festzuhalten, dass die 360°-Ausstellung im Vergleich dazu einen räumlichen Eindruck, einen höheren Interaktivitätsgrad sowie eine höhere Individualisierung bietet.

5 https://www.ub.uni-frankfurt.de/ausstellung/uebersicht.html#virtuell

Literaturverzeichnis

[BKM09] Bangor, A.; Kortum, P.; Miller, J.: Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of usability studies 4 (3), S. 114–123, 2009.

[BKM22] Braun, C.; Kayali, F.; Moser, T.: Einsatz von virtuellen Rundgängen in der Distanz-Laborlehre: Praxisbeispiele aus dem Projekt Lab4home. MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung 47, S. 196–219, 2022, DOI: 10.21240/mpaed/47/2022.04.10.X.

[Bo99] Bowman, D. A.; Hodges, L. F.; Allison, D.; Wineman, J.: The educational value of an information-rich virtual environment. Presence: Teleoperators & Virtual Environments 8 (3), S. 317–331, 1999, DOI: 10.1162/105474699566251.

[Br96] Brooke, J. et al.: SUS-A quick and dirty usability scale. Usability evaluation in industry 189 (194), S. 4–7, 1996.

[FS] Funk, J.; Schmidt, L.: Vergleichende Evaluation von Explorationsarten in interaktiven 3D-360°-Anwendungen. Einsatz von explorativem Lernen zur Vorbereitung von Handwerkern auf Vor-Ort-Termine beim Kunden. DOI: 10.21240/mpaed/51/2023.01.22.X.

[HY18] Hayes, J.; Yoo, K.: Virtual reality interactivity in a museum environment. In: Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology. S. 1–2, 2018, DOI: 10.1145/3281505.3281620.

[Ke18] Kerres, M.: Mediendidaktik: Konzeption und Entwicklung digitaler Lernangebote. 5. Auflage. Berlin, 2018, DOI: https://doi.org/10.1515/9783110456837.

[Mo20] Morganti, F.; Palena, N.; Savoldelli, P.; Greco, A.: Take the first-person perspective to become dementia-friendly: the use of 360 video for experiencing everyday-life challenges with cognitive decline. Frontiers in Psychology 11, 2020, DOI: 10.3389/fpsyg.2020.01117.

[Pe20] Petko, D.: Einführung in die Mediendidaktik: Lehren und Lernen mit digitalen Medien. 2. Beltz, 2020.

Konzeption eines adaptiven Systems für Augmented Reality Lehr-Lernsysteme mit individuellen Lernszenarien

Dennis Wohlers [1], Roland Schröder-Kroll[1], Svetlana Golovanev [1], Michael Geis [2], Maximilian Mushoff [3], Valerie Varney [3], Joachim Metternich[2], Rüdiger Weißbach [1] und Axel Wagenitz [1]

Abstract: Adaptives Lernen, ein personalisierter Lehransatz der Lerninhalte an die individuellen Bedürfnisse der Lernenden anpasst, führt bei der Integration in Augmented und Virtual Reality (AR, VR) Lernumgebungen zu vielversprechenden Ergebnissen im Kontext lernförderlicher Faktoren. Im Bereich von AR Lehr-Lernsystemen, die es Lehrenden erlauben, unabhängig vom Lehrgegenstand selbstständig Lernsimulationen zu erstellen, konnten in der Literatur keine bestehenden Ansätze für adaptive Modelle identifiziert werden. Für das gegenwärtige Forschungsprojekt „easyARguide" erscheint die Integration von Adaptivität als vielversprechender Weg, um den Lerntransfer zu verbessern. In dem Paper wird ein adaptives AR-Lehr-Lernsystem mit anpassbaren Lehrinhalten entwickelt. Das System adressiert eine flexible Lerngestaltung, um die individuelle Präferenz von Lehrenden bei der Lernumgebungserstellung zu gewährleisten und gleichzeitig eine unabhängige Anwendung auf diverse Lerngebiete sicherzustellen. Dieser Ansatz ist als konzeptionelle Ausarbeitung zu verstehen, dessen Wirksamkeit noch nicht empirisch untersucht worden ist. Das entwickelte Modell soll als Grundlage für weitere Forschungsarbeiten fungieren.

Keywords: Adaptives Lernen, Augmented Reality, Lehr-Lernsystem

Disclaimer: Die in der vorliegenden Arbeit verwendeten Personenbezeichnungen beziehen sich gleichermaßen auf weibliche, männliche und diverse Personen. Auf eine Doppelnennung und gegenderte Bezeichnungen wird zugunsten einer besseren Lesbarkeit verzichtet.

1 Einleitung

Der demografische Wandel führt in Handwerk und produzierenden Unternehmen zum Verlust ausgebildeter Fachkräfte und deren Fachwissen [MH22]. Das Augmented Reality

1 Hochschule für Angewandte Wissenschaften Hamburg, Fakultät Wirtschaft und Soziales, Berliner Tor 5, 20099 Hamburg, Deutschland, dennis.wohlers@haw-hamburg.de, https://orcid.org/0009-0005-2233-7629; roland.schroeder-Kroll@haw-hamburg.de; svetlana.golovanev@haw-hamburg.de, https://orcid.org/0009-0003-3094-3851; ruediger.weissbach@haw-hamburg.de, https://orcid.org/0000-0002-7435-2135; axel.wagenitz@haw-hamburg.de, https://orcid.org/0009-0002-9612-9257

2 Technische Universität Darmstadt, Institut für Produktionsmanagement, Technologie und Werkzeugmaschinen (PTW), 64287 Darmstadt, Deutschland, m.geis@ptw.tu-darmstadt.de, https://orcid.org/0009-0000-6805-6340; j.metternich@ptw.tu-darmstadt.de

3 Technische Hochschule Köln , Department, 50678 Köln, Deutschland, maximilian.mushoff@th-koeln.de, https://orcid.org/0009-0007-2260-5755; valerie.varney@th-koeln.de, https://orcid.org/0000-0002-6252-7217

doi:10.18420/delfi2024_28

(AR) Forschungsprojekt easyARguide[4] zielt unter anderem darauf ab, diesem Verlust durch Sicherung des Wissens entgegenzuwirken. Lehrende können mit der im Projekt entwickelten Anwendung ohne Programmierkenntnisse domänenunabhängig Lerninhalte mit einem AR Head-Mounted-Display umsetzen und Lernenden zur Verfügung stellen.
Der Einsatz von AR kann in Aus- und Weiterbildungsprozessen eine Reduzierung von Anlernzeiten und Prozessfehlern bewirken [Ka21][WM98][LQB16][HWT15]. Ein Anstieg der Lerneffizienz kann sowohl während des als auch nach dem Training beobachtet werden [Ka21][WM98]. Dabei sind die Kosten für Entwicklung, Einrichtung und Implementierung von konventionellen AR-Technologien im Aus- und Weiterbildungsprozess eine der größten Herausforderungen [Fe17]. Durch die in easyARguide entwickelte Anwendung wird Fachpersonal in die Erstellung von AR-Trainingsszenarien einbezogen, was den monetären Aufwand reduziert und einen niederschwelligen Zugang zur Integration von AR in Lehr-Lernprozesse fördert. In dem Zusammenhang gewinnt auch die lernförderliche Gestaltung von AR-Szenarien an Bedeutung, z.B. durch die Umsetzung adaptiver Lernsysteme in AR- und VR-Lernumgebungen, welche in vorangegangen Studien untersucht wurden [Hu21][Si16]. Hierbei lässt sich Adaptivität als System definieren, das sich beim Lernen an Vorwissen, Lernfortschritt und Verhalten der Lernenden anpasst [Ke69]. Insgesamt besteht das Ziel darin, durch die Adaption der Lernumgebung an den Lernenden das Lernumfeld zu verbessern und den Lernerfolg positiv zu beeinflussen. Aktuell existieren Gestaltungsrichtlinien für adaptive VR- und AR-Lernsysteme [ZA20][Hu21], jedoch nicht für Anwendungen –wie im Rahmen des Projekts easyARguide entwickelt– mit frei erstellbaren Inhalten. Die vorliegende Arbeit erstellt auf Basis relevanter Studienergebnisse und lernpsychologischer Grundlagen ein Konzept für ein solches adaptives Lernsystem, welches die Anforderungen aus der Anwendung im Rahmen von easyARguide erfüllt.

2 Lern-adaptive und lernpsychologische Grundlagen

Zahabi et al. präsentieren ein Framework für adaptive VR-Lernszenarien mit den Aspekten Performanzmessung, adaptive Logik und adaptive Variablen [ZA20]. Für den Spezialfall des Trainings an Maschinen ist AdapTutAR ein adaptives AR Lehr-Lernsystem mit fünf Leveln von Hinweisen [Hu21]. Nach einer Zeit ohne Interaktion oder dem Manipulieren des falschen Objektes wird das Level erhöht, nach Abschluss eines Schrittes verringert [Hu21]. Verbunden mit diesen Beispielen wird in dieser Arbeit in lernpsychologischer Hinsicht die Theorie der kognitiven Belastung (KB) nach Sweller und Chandler als Grundlage angewendet [SC91]. Sie gliedert die Belastung des Arbeitsgedächtnis beim Wissenserwerb in drei Kategorien. Die intrinsische kognitive Belastung (IKB), welche die Belastung durch Komplexität und Zusammenhang des Lernmaterials beschreibt [Sw10], die extrinsische kognitive Belastung (EKB), welche durch die Darstellung der Lerninhalte beeinflusst wird [Ja18][Sw10] und die lernbezogene kognitive Belastung (LKB). Die LKB kann hierbei sowohl durch ein Zunehmen der IKB als auch der EKB ansteigen, wobei eine Zunahme durch

4 Ein vom Bundesministerium für Wirtschaft und Klimaschutz (BMWK) gefördertes Projekt als Teil des Mittelstand-Digital Netzwerks

EKB das Lernen reduziert und somit im Sinne eines möglichst effektiven Lernfortschritts zu vermeiden ist [Sw10][Ma09][HY13].
Daher ist es zielführend, die Lerninhalte auf das Kompetenzniveau des Lernenden zu adaptieren, um ein anregendes Niveau zu erreichen, welches weder über- noch unterfordert. Für die Adaption wird exemplarisch auf ein dreistufiges Beurteilungskonzept, das in "Kenner", "Könner" und "Experte" einstuft, zurückgegriffen [NRS13]. In dieser Arbeit wird die Bezeichnung der Kompetenzstufen verändert und der "Kenner" zur besseren Abgrenzung zum "Könner" als "Anfänger" bezeichnet.

3 Aufbau des adaptiven Systems

Der grundsätzliche Aufbau des hier entwickelten adaptiven Systems ist dem Framework zu adaptiven VR-Lernsystemen von Zahabi und Razak nachempfunden und gliedert sich in die Teilbereiche Performanzmessung, adaptive Variablen und adaptive Logik [ZA20].

Performanzmessung
Um die Anwendungsfreundlichkeit im beruflichen Kontext zu gewährleisten, wird hier auf Hardware zur sensorischen Erfassung des Lernempfindens der Lernenden verzichtet, was die Arten der möglichen Performanzmessung limitiert. Da die Anwendung auf verschiedenen Endgeräten mit abweichender Sensorik zum Einsatz kommen soll, muss eine universell zugängliche Messgröße eingesetzt werden. Daher wird für die Performanzmessung die Zeit zur Durchführung eines Lernschrittes als Leistungsindikator gewählt.

Adaptive Variablen
Adaptive Variablen sind Lernelemente, die in der Anwendung des easyARguide verwendet werden. Jedes Lernelement kann verschiedene Arten von übergeordneten Informationen vermitteln, wie beispielsweise Orientierung und Positionierung, die nachfolgend als Informationstypen bezeichnet und analog zu [La22] verwendet werden.
Die Definition der adaptiven Variablen des Lehr-Lernsystems dient dazu, den Funktionsumfang der Trainingserstellung einzuschränken und den vermittelten Informationsgehalt eines Lernelements zur Anwendung im Lernsystem zu konkretisieren und vergleichen zu können. Arbeitsschritte, aus denen sich Trainings zusammensetzen, sind definiert als atomare und in sich schlüssige Schritte (an [Ga19] angelehnt). Sie setzen sich aus mehreren Lernelementen zusammen. In der Tabelle 1 sind die im Projekt easyARguide entwickelten Lernelemente, sowie deren Definition und inhärente Informationstypen dargestellt.

Adaptive Logik
Die adaptive Logik basiert auf der fortlaufenden Erfassung der benötigten Zeit für jeden Arbeitsschritt und der darauf basierenden Einteilung der Lernenden in eine Kompetenzgruppe. Entsprechend dieser werden angemessene Lernelemente eingeblendet, um die Herausforderung des Lernprozesses und somit die kognitive Belastung auf einem für die Lernenden anregenden Niveau zu halten. Die Logik teilt hierbei Lernende auf Basis des dreistufigen Beurteilungskonzeptes in "Anfänger", "Könner" und "Experte" ein.

Lernelemente	Definition	Informationstypen
3D Modelle	3D Objekte, die zur Veranschaulichung von Handlungsschritten eingesetzt werden. Im easyARguide-Projekt werden Modelle von Werkzeugen und Arbeitsteilen bereitgestellt, die auch animiert werden können.	H, O, P, I
Arbeitsschritttitel	Bezeichnung des Arbeitsschrittes.	Arbeitsschritttitel
Auxiliary Modelle	Einfache 3D-Modelle, die das Lernen durch visuelle Hinweise unterstützen (z.B. Pfeile für räumliche Orientierung oder Bewegungen eines Objektes) und animiert werden können.	O, P
Bilder	2D-Grafiken, die durch das AR-System eingebunden werden, z.B. Handgriffe, Werkzeuge und Anwendungsbeispiele.	H, O, P, I
Handbewegung	Handbewegungen von 3D Handmodellen, die vom Lehrenden aufgenommen werden können.	H, O, P
Sicherheitshinweis	Optionale Sicherheits- oder Warnhinweise, um auf Besonderheiten im Arbeitsschritt aufmerksam zu machen.	Sicherheitshinweis
Sprachaufnahmen	Auditive Texte, die den Prozess erklären, mit der Möglichkeit, sie als Text auszugeben, um die Barrierefreiheit zu erhöhen.	Beschreibung des Arbeitsschritts
Text	2D-Texte, die ergänzende Handlungsinformationen in kurzen Stichworten oder Sätzen liefern.	Anweisung
Video mit Ton	2D-Videos mit Ton, die durch das AR-System aufgenommen werden. Sie ermöglichen Erläuterungen durch den Lehrenden.	Beschreibung des Arbeitsschritts H, O, P, I
Video ohne Ton	2D-Videos ohne Ton, die durch das AR-System aufgenommen werden und Handgriffe, Werkstücke und Anwendungsbeispiele zeigen.	Handhaltung H, O, P, I

Tab. 1: Definition der Lernelemente und inhärente Informationstypen (abgekürzt wie folgt: (H[andhabung], O[rientierung], P[ositionierung], I[dentität])

"Anfänger" verfügen über Grundkenntnisse, die ihnen das erfolgreiche Bearbeiten der Lernsituation ermöglichen, haben aber keine konkreten Kenntnisse über die Arbeitsschritte und bedürfen beim Erlernen einer engmaschigen Anleitung mit hoher Informationsdichte. "Könner" haben bereits erste Grundkenntnisse über das Lernszenario und benötigen für die Bearbeitung der Aufgabe ergänzende Informationen, um ihr Verständnis zu vertiefen.
"Experten" sind mit dem Lernszenario vertraut, sodass Lernaufgaben mit nur minimalen Kontextinformationen eigenständig zu bewältigen sind.
Die für einen Arbeitsschritt benötigte Zeit wird durch den Lehrenden nach der Erstellung des Trainings durch Durchführung der Lernsimulation eingemessen und als Idealwert für die Kompetenzgruppe "Experte" festgelegt. Abhängig von der durch einen Lernenden benötigten Zeit werden beim Überschreiten eines Schwellenwerts zusätzliche Kompetenzgruppen mit den zugehörigen adaptiven Variablen eingeblendet oder reduziert. Die Einstufung bleibt bestehen, solange die zeitliche Bearbeitung der Aufgabenschritte in den Schwellwerten liegt. Diese Art der Messung baut auf [Hu21] auf. Die Lernenden können vor Beginn des Trainings ihre Erfahrung in einem Lernszenario angeben. Dies dient als Ausgangspunkt für das adaptive System. Die genauen zeitlichen Schwellenwerte zum Übergang zwischen den

Kompetenzgruppen werden in zukünftigen Untersuchungen experimentell ermittelt.
Da die easyARguide Anwendung für eine lernaufgabenunabhängige Erstellung von Lernszenarien gedacht ist, muss das adaptive System so entwickelt werden, dass eine möglichst flexible Anwendung von Lernelementen nach individueller Präferenz der Lehrenden und Kontext möglich ist. Nach der vorgenommene Definition der Kompetenzgruppen benötigen „Experten“ Lernende am wenigsten Informationen zur Erfüllung der Lernschritte. Arbeitsschritttitel sollten als Lernelement ausreichend Kontextinformationen vermitteln. Der Arbeitsschritttitel muss zur Lernerstellung verwendet werden. Optional können Sicherheitshinweise auch erfahrenen Lernenden relevante Inhalte vermitteln.
Für die Kompetenzgruppe „Könner“ wurden die Informationstypen “Positionierung”, “Orientierung”, “Handhabung” und “Anweisung” als notwendig identifiziert. Um den Lehrenden eine hohe Flexibilität bei der Erstellung von Lernszenarien zu ermöglichen, ist es erforderlich, mindestens eine der Variablen "Text", "Handbewegung" und "Hilfsmodelle" zu nutzen. Dies ermöglicht, auch die Elemente der Lernsituation anzupassen und zu kombinieren. Zur Lernunterstützung wurden die informativen Merkmale „Identität“ und „Beschreibung des Arbeitsschritts“ als zusätzliche benötigte Informationen identifiziert.
Der Kompetenzgruppe „Anfänger“ wurden zwei Möglichkeiten der Trainingserstellung gegeben. „Video mit Erklärung“ entspricht dabei einem Informationsniveau, die einer Kombination aus „Sprachaufnahme“ und einem der Merkmale „3D Modelle“, „Bilder“ und „Video ohne Erklärung“ entspricht. Die resultierende Einteilung von adaptiven Variablen in den einzelnen Kompetenzgruppen ist in Abbildung 1 dargestellt.

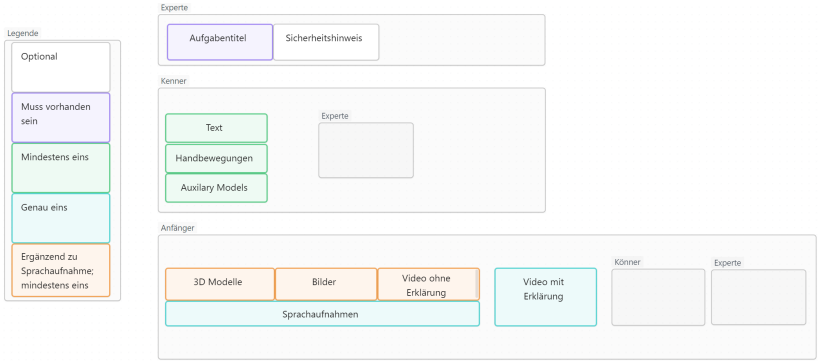


Abb. 1: Zuordnung der Lernelemente nach Kompetenzgruppen

4 Diskussion

In dieser Arbeit wurde ein adaptives Lernystem zur Anwendung in frei erstellbaren AR Lehr-Lernszenarien entwickelt. Im Vergleich zu dem adaptiven Modell von Huang et al. [Hu21] unterscheidet sich unser Modell durch eine Einteilung von Lernenden und Lernelementen in Kompetenzgruppen und die dementsprechende Adaption. Diese Entscheidung basierte auf einem Kompromiss zwischen einer möglichst hohen Flexibilität und einem möglichst

geringen Aufwand in der Erstellung von Lerninhalten. Lehr-Lernsysteme mit durch Lehrende anpassbaren Lernelementen sind von vielen Randbedingungen gekennzeichnete Systeme. Einerseits müssen sie die Funktionalität haben, Lernende in ihrem Fortschritt bestmöglich zu unterstützen. Gleichzeitig muss ihre Nutzung flexibel sein, um Lehrenden eine der Aufgaben und persönlichen Präferenzen entsprechende Lerngestaltung zu ermöglichen. Das von uns vorgestellte Modell adaptiven Lernens mit einer beispielhaften Einteilung in drei Kompetenzgruppen kann hier ein erster Ansatz sein, um einen Kompromiss aus diesen Faktoren herzustellen. Ob hierbei eine Lernunterstützung bestmöglich gewährleistet wird, soll im Laufe des Projekts easyARguide evaluiert werden. Eine Ergänzung der Kompetenzgruppen um eine weitere Gruppe wäre eine Möglichkeit, um eine kleinteiligere Anpassung der Lernsimulation vorzunehmen und die Unterstützung der Lernenden verbessern zu können. Auch könnte Optimierungsbedarf in der Anordnung der Lernelemente innerhalb der Kompetenzgruppen liegen. Nach der derzeitigen Aufteilung können teilweise Dopplungen von Informationstypen durch verschiedene Lernelemente entstehen. So besitzen beispielsweise beide Lernelemente Bilder und Videos die Informationstypen Orientierung und Positionierung. Dies kann auf Lernende ablenkend wirken oder verschiedene Lerntypen besser unterstützen.
Zur Performanzmessung wurde die für eine Aufgabe benötigte Bearbeitungszeit ausgewählt. Es kann sein, dass Rückschlüsse alleine basierend auf diesem Merkmal nicht für alle Lernenden ein optimales Performanzkriterium darstellt. Aufgrund der Rahmenbedingungen des Projektes ist es nicht möglich, das System durch eine sensorische Erfassung physiologischer Merkmale von Lernenden zu erweitern, um die Akzeptanz durch möglichst geringe Kosten und einer einfachen Anwendbarkeit in der Praxis zu gewährleisten.
Die Wirkung dieses adaptiven AR Lehr-Lernsystems ist noch nicht in seiner Effektivität untersucht. Eine Untersuchung im Anschluss an die Umsetzung des Projektes easyARguide wird angestrebt. Im Kontext von AR sind Forschungsarbeiten zu adaptiven Systemen wenig vertreten und nicht hinreichend in ihrer Effektivität bestätigt.

5 Fazit

Diese Arbeit präsentiert ein neuartiges adaptives Lehr-Lernmodell für frei erstellbare AR-Lernsimulationen, das eine variable Anpassung von Lerninhalten ermöglicht. Das vorgestellte adaptive Modell ist so konstruiert, dass es aufgabenstellungsunabhängig angewendet werden kann. Die Übertragbarkeit auf bestehende AR-Lernsimulationen eröffnet für Unternehmen und Bildungseinrichtungen die Möglichkeit, die Lernerfahrungen durch die Implementierung des adaptiven Systems zu verbessern. Viele Elemente des Modells basieren auf einer Kombination aus diversen theoretischen Vorarbeiten. Etwaige Synergien der Elemente sowie die Effektivität der Elemente in Kombination sind weitgehend unerforscht. Es besteht daher der Bedarf einer empirischen Untersuchung zur Wirksamkeit, um den Einfluss des Systems auf den Lerneffekt zu ermitteln. Als erste wissenschaftliche Arbeit, die ein solches adaptives Lehr-Lern-Modell entwickelt hat, bietet sie einen wichtigen Beitrag zur Weiterentwicklung im Einsatz von adaptiven AR-Systemen in der Bildungstechnologie.

Literaturverzeichnis

[Fe17] Fehling, C. D.: Erweiterte Lernwelten Für Die Berufliche Bildung. Augmented Reality Als Perspektive. In (Thissen, F., Hrsg.): Lernen in Virtuellen Räumen. De Gruyter, S. 125–142, 2017, ISBN: 978-3-11-050113-1, DOI: 10.1515/9783110501131-009, Stand: 11. 04. 2024.

[Ga19] Gattullo, M.; Scurati, G. W.; Fiorentino, M.; Uva, A. E.; Ferrise, F.; Bordegoni, M.: Towards Augmented Reality Manuals for Industry 4.0: A Methodology. Robotics and Computer-Integrated Manufacturing 56, S. 276–286, 2019, ISSN: 0736-5845, DOI: 10.1016/j.rcim.2018.10.001, Stand: 19. 03. 2024.

[Hu21] Huang, G.; Qian, X.; Wang, T.; Patel, F.; Sreeram, M.; Cao, Y.; Ramani, K.; Quinn, A. J.: AdapTutAR: An Adaptive Tutoring System for Machine Tasks in Augmented Reality. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. ACM, Yokohama Japan, S. 1–15, 2021, ISBN: 978-1-4503-8096-6, DOI: 10.1145/3411764.3445283, Stand: 11. 04. 2024.

[HWT15] Hou, L.; Wang, X.; Truijens, M.: Using Augmented Reality to Facilitate Piping Assembly: An Experiment-Based Evaluation. Journal of Computing in Civil Engineering 29 (1), S. 05014007, 2015, ISSN: 0887-3801, 1943-5487, DOI: 10.1061/(ASCE)CP.1943-5487.0000344, Stand: 11. 04. 2024.

[HY13] Hattie, J.; Yates, G. C. R.: Visible Learning and the Science of How We Learn. Routledge, 2013, ISBN: 978-1-134-64311-0.

[Ja18] Jadin, T.: Psychologische Aspekte Des Lernens Mit Digitalen Medien. Zeitschrift für Psychologie / Journal of Psychology 38 (5), S. 368–374, 2018, ISSN: 2151-2604.

[Ka21] Kaplan, A. D.; Cruit, J.; Endsley, M.; Beers, S. M.; Sawyer, B. D.; Hancock, P. A.: The Effects of Virtual Reality, Augmented Reality, and Mixed Reality as Training Enhancement Methods: A Meta-Analysis. Human Factors: The Journal of the Human Factors and Ergonomics Society 63 (4), S. 706–726, 2021, ISSN: 0018-7208, 1547-8181, DOI: 10.1177/0018720820904229, Stand: 11. 04. 2024.

[Ke69] Kelley, C. R.: What Is Adaptive Training? Human Factors: The Journal of the Human Factors and Ergonomics Society 11 (6), S. 547–556, 1969, ISSN: 0018-7208, 1547-8181, DOI: 10.1177/001872086901100602, Stand: 11. 04. 2024.

[La22] Laviola, E.; Gattullo, M.; Manghisi, V. M.; Fiorentino, M.; Uva, A. E.: Minimal AR: Visual Asset Optimization for the Authoring of Augmented Reality Work Instructions in Manufacturing. The International Journal of Advanced Manufacturing Technology 119 (3), S. 1769–1784, 2022, ISSN: 1433-3015, DOI: 10.1007/s00170-021-08449-6, Stand: 19. 03. 2024.

[LQB16] Loch, F.; Quint, F.; Brishtel, I.: Comparing Video and Augmented Reality Assistance in Manual Assembly. In: 2016 12th International Conference on Intelligent Environments (IE). IEEE, London, United Kingdom, S. 147–150, 2016, ISBN: 978-1-5090-4056-8, DOI: 10.1109/IE.2016.31, Stand: 11. 04. 2024.

[Ma09] Mayer, R. E.: Multimedia Learning. Cambridge University Press, 2009, ISBN: 978-0-521-51412-5 978-0-521-73535-3 978-0-511-81167-8.

[MH22] Malin, L.; Hickmann, H.: Fachkräftemangel Und Ausbildung Im Handwerk. Kompetenzzentrum Fachkräftesicherung (KOFA), 2022.

[NRS13] North, K.; Reinhardt, K.; Sieber-Suter, B.: Kompetenzmanagement in der Praxis: Mitarbeiterkompetenzen systematisch identifizieren, nutzen und entwickeln ; mit vielen Fallbeispielen. Springer Gabler, Wiesbaden, 2013, ISBN: 978-3-8349-2815-3.

[SC91] Sweller, J.; Chandler, P.: Evidence for Cognitive Load Theory. Cognition and Instruction 8 (4), S. 351–362, 1991, ISSN: 07370008, 1532690X, JSTOR: 3233599, Stand: 11. 04. 2024.

[Si16] Siu, K.-C.; Best, B. J.; Kim, J. W.; Oleynikov, D.; Ritter, F. E.: Adaptive Virtual Reality Training to Optimize Military Medical Skills Acquisition and Retention. Military Medicine 181 (5S), S. 214–220, 2016, ISSN: 0026-4075, 1930-613X, DOI: 10.7205/MILMED-D-15-00164, Stand: 11. 04. 2024.

[Sw10] Sweller, J.: Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. Educational Psychology Review 22 (2), S. 123–138, 2010, ISSN: 1040-726X, 1573-336X, DOI: 10.1007/s10648-010-9128-5, Stand: 11. 04. 2024.

[WM98] Waller, D.; Miller, J.: A Desktop Virtual Environment Trainer Provides Superior Retention of a Spatial Assembly Skill, 1998.

[ZA20] Zahabi, M.; Abdul Razak, A. M.: Adaptive Virtual Reality-Based Training: A Systematic Literature Review and Framework. Virtual Reality 24 (4), S. 725–752, 2020, ISSN: 1359-4338, 1434-9957, DOI: 10.1007/s10055-020-00434-w, Stand: 11. 04. 2024.

Schleif-R: Handwerkliches Schleifen in VR



Marc Schlicker [1] und Raphael Zender [2]

Abstract: Die Ausbildung an Schleifgeräten gestaltet sich oft als herausfordernd und ist aus verschiedenen Gründen begrenzt, wie aufgrund des Mangels an qualifizierten Lehrkräften, zeitlichen Einschränkungen, Kosten und dem potenziellen Risiko von Verletzungen und Gesundheitsproblemen. Die korrekte Bedienung von Schleifgeräten ist für viele Berufe jedoch von großer Bedeutung. Dieser Beitrag zielt darauf ab, das Potenzial von Virtual Reality (VR) in der beruflichen Bildung zur Vermittlung der Bedienung eines Exzenterschleifers in Diskussion zu stellen. Er präsentiert eine virtuelle Umgebung für die Schleifausbildung. Diese bietet eine Simulation des Schleifprozesses und benutzt einen speziell entwickelten VR-Controller, der das taktile Erlebnis eines tatsächlichen Schleifers nachahmt.

Keywords: VR, Simulation, Berufliche Bildung, Schleifen, VR-Learning

1 Motivation

Virtual Reality (VR) hat vielfältige Potenziale im Einsatz in der beruflichen Aus- und Weiterbildung [Th18]. Dieser Beitrag widmet sich der Analyse des VR-Einsatzes in der handwerklichen Ausbildung, insbesondere der Schulung im Umgang mit einem Exzenterschleifer. Zunächst ist es angebracht, die aktuellen Lehrmethoden in der handwerklichen Ausbildung zu betrachten. In der (Tischler-)Ausbildung werden Auszubildende derzeit geschult, indem ein erfahrener Mitarbeiter die Anwendung vorführt und die Auszubildenden sie nachahmen. Es besteht jedoch die Gefahr, dass entweder die Erklärungen unzureichend sind oder dass die Auszubildenden aus Angst vor Bloßstellung vorgeben, das Konzept verstanden zu haben, obwohl dies nicht der Fall ist. Dies gilt auch in der Schulung im Umgang mit einem Exzenterschleifer

Basierend auf den im Rahmen dieses Projektes durchgeführten Interviews ergibt sich die Problematik, dass Fachkräfte am Ende ihrer Ausbildung Schwierigkeiten im Umgang mit einem so grundlegenden Werkzeug haben, das für nahezu alle Tätigkeiten im Tischlerhandwerk und vielen anderen Gewerken von Bedeutung ist. Diese Lücken in der Ausbildung können sich negativ auf die Qualität der Arbeit und die Sicherheit am Arbeitsplatz auswirken und möglicherweise die Produktivität des Betriebes sowie das Handwerksgewerbes insgesamt beeinträchtigen. Es ist daher von großer Bedeutung, diese Herausforderungen anzugehen und Lösungen zu finden, um eine hohe Qualität der beruflichen Ausbildung an Exzenterschleifern zu gewährleisten.

1 Humboldt Universität zu Berlin, Institut für Informatik, Unter den Linden 6, 10117 Berlin, Deutschland, schlickm@hu-berlin.de, https://orcid.org/0009-0001-5417-380X
2 Zeppelin Universität, Am Seemooser Horn 20, 88045 Friedrichshafen, Deutschland, raphael.zender@zu.de, https://orcid.org/0000-0001-9866-9455

doi:10.18420/delfi2024_29

2 Potenzial von VR im Schleiftraining

Das Schleifen mit einem Exzenterschleifer findet in zahlreichen Bereichen Anwendung, wobei häufig eine Verbesserung der Oberflächenqualität angestrebt wird. Ein typisches Beispiel ist das Schleifen vor dem Lackieren von Werkstücken, bei dem eine gleichmäßige 8-förmige Bewegung mit konstantem Druck und Geschwindigkeit über das Werkstück geführt wird. Dabei ist es wichtig, den Absaugschlauch und das Stromkabel hinter dem Rücken oder über der Schulter zu halten, um eine Beeinträchtigung der Arbeit zu vermeiden. In der handwerklichen Ausbildung gibt es im Bereich des Schleifens mit dem Exzenterschleifer verschiedene Herausforderungen (Abb. 1).

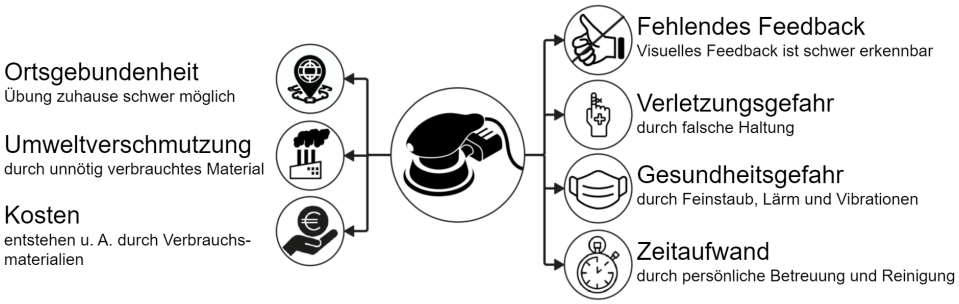


Abb. 1: Herausforderungen des Exzenterschleifens

Das Üben des Schleifvorgangs birgt Gesundheitsrisiken durch Feinstaub, Vibrationen [BG15; Ka07], und Lärm, während auch Verletzungsgefahren durch falsches Halten bestehen [Fi21]. Darüber hinaus mangelt es oft an Feedback und Betreuung und es stellt eine Herausforderung für die Umwelt und die Finanzen der Ausbildungsstätten dar. Auch der Zeitaufwand und die Ortsgebundenheit sind zu berücksichtigen.
VR bietet Möglichkeiten, um Bildungsanwendungen zu unterstützen, indem sie die Lernumgebung beeinflusst und Lerninhalte erlebbar macht [Ze18]. Im handwerklichen Bereich können zusätzliche Probleme auftreten, die durch VR gelöst werden können. In einer Schleifwerkstatt können die in Abbildung 1 dargestellten Herausforderungen reduziert oder vollständig beseitigt werden.
Ein vergleichbarer Ansatz wird im Projekt HandLeVR [Mu23] verfolgt, bei dem VR als zentrale Technologie genutzt wird, um Auszubildende beim Erlernen des Lackierens in der Fahrzeuglackierung zu unterstützen. Dies geschieht durch die Nutzung eines virtuellen Lackiersimulators und eines Tangibles in Form einer Lackierpistole. Ein wesentlicher Unterschied zum Schleifen liegt beim Lackieren darin, dass, abgesehen von der Lackierpistole, kein haptisches Feedback von den 3D-Werkstücken benötigt wird oder gewollt ist, da dies in der realen Welt das gerade lackierte Werkstück beschädigen könnte [Ze19].

3 VR-Trainingssystem Schleif-R

3.1 Konzeption

Angelehnt an klassische Systementwicklungsprozessen wurde ein Design-based Research-Ansatz [CJW14] genutzt. Dieser Ansatz umfasste zunächst eine gründliche Auseinandersetzung mit den theoretischen Grundlagen, während parallel ein erster Prototyp entstand.

Der Prototyp ermöglicht das Einsetzen eines VR-Controllers vom Typ Meta Quest 2 zur genauen Positionsbestimmung. Am unteren Teil des 3D-Drucks befindet sich ein Stift, der in Verbindung mit Schaumstoff den Druck auf das oben liegende Objekt misst. Zur Konzeption des Systems wurde eine erste Anforderungsanalyse unter Verwendung des ersten Prototyps durchgeführt. Die Datenerhebung in der Anforderungsanalyse erfolgte mittels halb-strukturierter Experteninterviews mit zwei ausgewählten Personen. Eine dieser Personen ist ein Tischler-Meister und Werkstattmeister an einer Berliner Universität, der zudem ein Informatikstudent ist. Die zweite Person ist eine Lehrkraft für Holzarbeiten und ausgebildeter Tischler-Geselle. Beide Personen sind männlich.

Die Interviews wurden anhand eines Kodierleitfadens ausgewertet und die Anforderungen anschließend in Anlehnung an das MoSCoW-Priorisierung [Kh15] klassifiziert. Dabei erfolgte eine Differenzierung zwischen Soft- und Hardwareanforderungen. Zu den wichtigsten Hardwareanforderungen gehörten die Druckmessung, die Positionsbestimmung und die Annäherung an die Form eines echten Exzenterschleifers. Die Software sollte Feedback bereitstellen, mindestens einen Schleifgang unterstützen und verschiedene Werkstücke bieten können. Die detaillierte Anforderungsliste ist im Git-Repository des Projekts[3] zu finden.

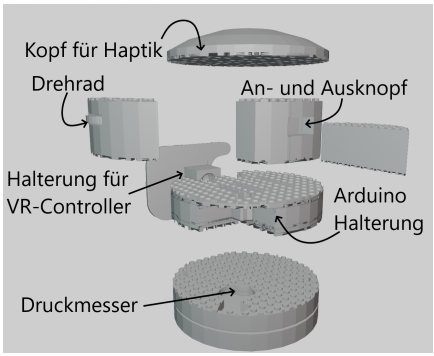
3.2 Implementierung

Das System besteht aus einer Hardwarekomponente, die ein Tangible und ein VR-System erfordert, sowie einer Software, die eine VR-Werkstatt darstellt.

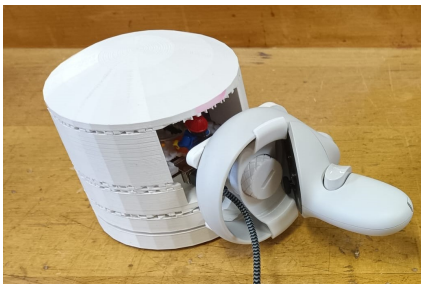
3.2.1 Tangible und VR-Hardware

Die Hardware wurde in mehreren Schritten größtenteils unter Verwendung der Software Blender[4] entwickelt. Die aktuelle Version des Tangible weist verschiedene Merkmale auf:

3 Schleif-R Git: `https://github.com/MarcSchlicker/Schleif-R`
4 Blender: `https://www.blender.org`



(a) Explosionszeichnung



(b) Gedrucktes Tangible

Abb. 2: Schleif-R Tangible

- Die Befestigung des Controllers erfolgt an einer Schraubbefestigung außerhalb des 3D-Drucks, die größtenteils von HandLeVR übernommen wurde. Dies ermöglicht einerseits eine Gewichtsverlagerung des Tangibles, wie mit einem Griff, und andererseits eine verbesserte Erkennung des Controllers durch die VR-Brille.
- Die Druckmessung erfolgt über einen Drucksensor, was eine präzise Druckmessung ermöglicht und gleichzeitig die Kompatibilität mit anderen VR-Systemen gewährleistet.
- Durch die Integration eines Arduino-Boards wurde das Tangible um einen Ein- und Ausschalter sowie ein Potentiometer als Drehzahlregler erweitert.
- Die Bauteile des Tangibles dienen auch als Klemmbausteine, wodurch Teile einfach ausgetauscht oder das Tangible um weitere Teile ergänzt werden kann. Dies ist besonders vorteilhaft für verschiedene Anwendungszwecke, da beispielsweise der Adapter für den VR-Controller je nach VR-System leicht angepasst werden kann. Zudem ermöglicht es, dass der Ein- und Ausschalter durch ein Ab- und Zusammenstecken, beispielsweise durch einen Kippschalter, ausgetauscht werden kann.

Eine Anleitung zum Zusammenbau des Tangibles sowie seine Schaltkreise und sein technischer Aufbau, als auch die Software, sind im Git-Repository des Projekts[3] veröffentlicht.

3.2.2 Trainingsanwendung

Die Software für das Projekt wurde auf der Unity-Engine[5] entwickelt. Das gesamte Programm läuft auf einem PC, wobei die Ausgabe über ein Datenkabel an das VR-Headset (HMD)

5 Unity: `https://unity.com/de`

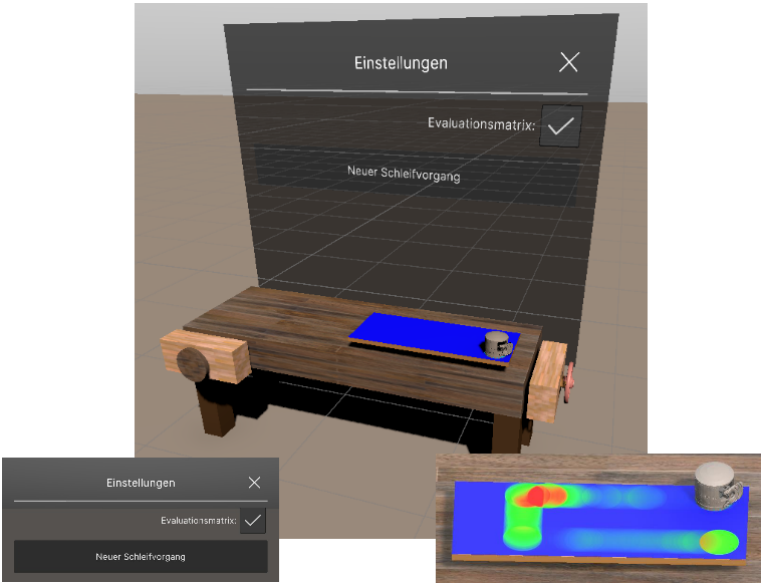


Abb. 3: Bestandteile der Trainingsanwendung

erfolgt. Zur Interaktion mit dem Arduino-Board des Tangibles nutzt das Unity-Programm eine serielle Verbindung in Verbindung mit Ardity[6], einem Unity-Arduino-Plugin.

Der Arbeitsbereich der Trainingsanwendung umfasst eine Hobelbank, ein Werkstück, eine Evaluationsmatrix und ein Kontrollpanel (Abb. 3). Das Werkstück ist auf der Hobelbank platziert und kann durch Knopfdruck auf die Höhe eines realen Tisches kalibriert werden, wodurch ein authentisches Schleifen simuliert wird. Dies ermöglicht beispielsweise das Kippen des Schleifers an der Tischkante. Das Kontrollpanel erlaubt die Überlagerung einer anfänglich blauen Matrix über das Werkstück, auf der Feedback angezeigt wird. In Abb. 3 ist zu erkennen, dass die Matrix sich färbt, wenn der Schleifer oft oder mit großem Druck über das Werkstück geführt wird. Grüne Farbe deutet dabei auf ein gutes Ergebnis hin, während eine rote Matrix darauf hinweist, dass zu viel abgeschliffen wurde. Das Feedback kann jederzeit ein- und ausgeschaltet werden, sodass ein authentisches Schleifen möglich ist und das Ergebnis am Ende betrachtet werden kann.

3.3 Evaluation

Die Implementierung wurde einer Zwischenbewertung unterzogen. Für eine umfassende Evaluation steht jedoch noch eine Fallstudie mit Auszubildenden aus.

6 Ardity: `https://ardity.dwilches.com`

Für die Zwischenbewertung des Projekts wurden drei Experten - alle Werkstattmeister einer Berliner Universität, davon 2 männlich und 1 weiblich - sowie drei Laien, davon 2 männlich und 1 weiblich, bei ihrer ersten Nutzung der Umgebung beobachtet. Ein Werkstattmeister, der auch an der Anforderungsanalyse beteiligt war, füllte zusätzlich einen strukturierten Interviewfragebogen aus.

Es wurde festgestellt, dass das Tangible-Gefühl dem eines echten Schleifers sehr nahe kommt. „Es fühlt sich an, als würde man wirklich etwas vom Material abtragen, was in diesem Fall durch das Filz auf der Schleiferunterseite und die leicht raue Holzoberfläche simuliert wurde“ (Interview 3).

Bei den Beobachtungen war zu erkennen, dass die Probanden die Umgebung intuitiv nutzen konnten, ohne Rückfragen zu stellen. Es gab jedoch Einschränkungen, wie den Verlust der Verbindungen zum Controller durch Bewegungen des Arms der Probanden und Probleme bei der Druckmessung aufgrund ungenauer Kalibrierung des Drucksensors. Zusammenfassend lässt sich sagen, dass alle "must-have” Anforderungen erfüllt wurden, während an den weiteren Anforderungen weitergearbeitet werden muss.

4 Zusammenfassung und Ausblick

Dieser Beitrag stellt die Entwicklung eines prototypischen, VR-basierten Schleifsystems vor. Ziel war es, effektive und benutzerfreundliche Anforderungen zu identifizieren und zu überprüfen, ob ein solcher Prototyp einen praxisbezogenen Mehrwert bietet. Die Anforderungsanalyse, unterstützt durch Experten aus dem Handwerksbereich, lieferte wertvolle Einblicke. Die Evaluierung des auf dieser Basis entwickelten Prototyps führte zu einer Vielzahl wertvoller Rückmeldungen von Experten bezüglich Benutzungsfreundlichkeit und Realitätsnähe.

Die Ergebnisse zeigen, dass eine VR-basierte Lernumgebung für das Schleifen mit einem Exzenterschleifer realisierbar und in der Praxis nützlich sein kann. Dieser Beitrag hebt jedoch auch Bereiche hervor, die für den tatsächlichen Einsatz als Lernumgebung weiterer Forschung und Verbesserung bedürfen, darunter die Anpassung von Hardware und Software an spezifische Nutzeranforderungen. Ferner kann die Lernumgebung nur als Ergänzung zum Üben am realen Exzenterschleifer gesehen werden. Die Entwicklung entsprechender didaktischer Szenarien steht noch aus.

Die Arbeit legt den Grundstein für zukünftige Forschung und Entwicklung im Bereich der VR-basierten handwerklichen Ausbildung. Zukünftige Arbeiten sollten beispielsweise die Wirksamkeit der Software mit Auszubildenden validieren, die Heatmap überarbeiten, Hardware optimieren, didaktische Ansätze integrieren und eine Schnittstelle mit HandleVR schaffen. Diese Optimierungen und Erweiterungen könnten die Bildungsrelevanz und den Mehrwert der VR-Lernumgebung für Auszubildende im Handwerk weiter steigern.

Literatur

[BG15] BGHM: Druckluft Exzenterschleifer in der Holzbearbeitung. FI/51, S. 1–4, 2015.

[CJW14] Christopher Gess; Julia Rueß; Wolfgang Deicke: Design-based Research zur Verbesserung der Lehre an Hochschulen: Einführung und Praxisbeispiel. 2014.

[Fi21] Fitzner: Handverletzungen – Die häufigste Art der Arbeitsunfälle | fitzner.de, 2021, URL: https://www.fitzner.de/blog/PROFI-raTgeber/Handverletzungen-Die-haeufigste-Art-der-Arbeitsunfaelle.html.

[Ka07] Kaulbars, U.: Erstmals Grenzwerte für die Vibrationsbelastung: Die neue Lärm- und Vibrations-Arbeitsschutzverordnung. TÜ/9, S. 17–23, 2007, URL: https://www.deutsche-gesetzliche-unfallversicherung.de/medien/ifa/de/pub/grl/pdf/2007_117.pdf.

[Kh15] Khan, J. A.; Rehman, I. U.; Khan, Y. H.; Khan, I. J.; Rashid, S.: Comparison of requirement prioritization techniques to find best prioritization technique. International Journal of Modern Education and Computer Science 7/11, S. 53, 2015.

[Mu23] Mulders, M.; Weise, M.; Schmitz, A.; Zender, R.; Kerres, M.; Lucke, U.: Handwerkliches Lackieren mit Virtual Reality (HandLeVR). MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung 51/, S. 214–245, 2023.

[Th18] Thomas, O., Hrsg.: Digitalisierung in der Aus- und Weiterbildung: Virtual und Augmented Reality für Industrie 4.0. Springer Gabler, Berlin, 2018, ISBN: 9783662565506.

[Ze18] Zender, R.; Weise, M.; Heyde, M.; Söbke, H.: Lehren und Lernen mit VR und AR – Was wird erwartet? Was funktioniert? In (CEUR Workshop Proceedings, Hrsg.): Proceedings of DeLFI Workshops 2018 co-located with 16th e-Learning Conference of the German Computer Society. 2018.

[Ze19] Zender, R.; Sander, P.; Weise, M.; Mulders, M.; Lucke, U.; Kerres, M.: HandLeVR: Action-Oriented Learning in a VR Painting Simulator. In: Proceedings of the 4th International Symposium on Emerging Technologies for Education. S. 46–51, 2019.

KI Einsatz in der Hochschule II

Automatische Generierung von Aufgaben für die konzeptuelle Modellierung

Paul L. Christ [1], Joerg M. Haake [1] und Torsten Munkelt[2]

Abstract: Die konzeptuelle Modellierung ist eine wichtige Kompetenz für verschiedene Disziplinen. Das Erlangen dieser Kompetenz erfordert Übung und die Auseinandersetzung mit Aufgaben, die verschiedene kognitive Prozessdimensionen adressieren. Viele solcher Aufgaben sind nötig, um eine große Anzahl an Lernenden mit unterschiedlichen Anforderungen zu bedienen. Der Bedarf einer großen Menge an diversen Aufgaben kann nicht manuell gedeckt werden. Derzeitigen automatischen Methoden, wie der Automatic Item Generation (AIG), mangelt es entweder an Skalierbarkeit oder der Möglichkeit, höhere kognitive Prozessdimensionen zu adressieren. Um diesen Mängeln entgegenzuwirken, wird ein generalisierter AIG-Prozess (gAIG) vorgeschlagen. Schritt 1 dieses Prozesses umfasst die Erzeugung einer Item-Spezifikation, welche eine Schablone für einen Aufgabentyp darstellt, und aus Templates einer Aufgabenbeschreibung, einem Input für Lernende, einer erwarteten Ausgabe (Lösungsvorschlag) und einem Antwort-Format besteht. Schritt 2 dieses Prozesses nimmt die Spezifikation eines parametrisierbaren Generators vor, welcher zur Item-Spezifikation passende Items erzeugt. Eine Fallstudie zeigt die Anwendung der gAIG zur Generierung von Aufgaben zur konzeptuellen Modellierung, welche die kognitiven Prozessdimensionen *Anwenden* und *Analysieren* adressiert, was bisherige Methoden nicht skalierbar vermögen.

Keywords: AIG, Automatic Item Generation, konzeptuelle Modellierung, kompetenzorientierte Aufgaben, E-Assessment, E-Learning, Bloom's Taxonomie, Graph-Rewriting

1 Einleitung

Die konzeptuelle Modellierung bildet ein reales oder geplantes System auf ein abstraktes Modell in einer gewählten Modellierungssprache ab [De18; GGM19; He07; Ro08]. Die Modalität des konzeptuellen Modells (z. B. textuell oder grafisch) hängt vom Typ der Modellierungssprache ab. Konzeptuelle Modellierung nimmt einen zentralen Platz im Curriculum der Wirtschaftsinformatik und verwandter Disziplinen ein [So22; St21]. Berufsverbände wie die Association for Computing Machinery (ACM) und die Gesellschaft für Informatik (GI) empfehlen die Einbeziehung der konzeptuellen Modellierung in die Curricula verschiedener IT-Disziplinen [AC21; GI17]. In IT-Disziplinen wird die konzeptuelle Modellierung z. B. für das Erstellen von Diagrammen zum Datenbankdesign und für die Beschreibung geplanter oder existierender Software oder Geschäftsprozesse in einer Organisation verwendet.

1 FernUniversität in Hagen, Fakultät Mathemathik und Informatik, Universitätstraße 1, 58084 Hagen, Deutschland, paul.christ@fernuni-hagen.de, https://orcid.org/0000-0002-6096-7403; joerg.haake@fernuni-hagen.de, https://orcid.org/0000-0001-9720-3100

2 HTW Dresden, Fakultät Informatik und Mathemathik, Friedrich-List-Platz 1, 01069 Dresden, Deutschland, torsten.munkelt@htw-dresden.de

doi:10.18420/delfi2024_30

Für das Lehren, Erlernen und zur Prüfung der Kompetenz zur konzeptuellen Modellierung werden summative und formative Assessments benötigt [UH23]. Ein Assessment besteht aus einem oder mehreren Assessment-Items (Items). Ein Item repräsentiert eine Aussage, Frage, Übung oder Aufgabe, für die Lernende entweder eine Antwort auswählen oder konstruieren, oder die Übung bzw. Aufgabe durchführen müssen [AA14].

Assessments können verschiedene kognitive Prozessdimensionen adressieren. Wegen ihrer häufigen Verwendung für die Kategorisierung von Modellierungsaufgaben [Bo19] verwenden wir die revidierte Bloom'sche Taxonomie [Lo01].

Striewe et al. beleuchten die bestehende Forderung nach studierendenzentrierter Lehre [St21]. Es sollen die Problemlöse- und Selbstlernkompetenzen der Studierenden gestärkt werden. Die Stärkung scheitert bislang insbesondere in Lehrveranstaltungen mit hohen Teilnehmendenzahlen aufgrund nicht ausreichender personeller Ressourcen u. a. zur Erstellung kompetenzorientierter Aufgaben und insbesondere von Aufgaben, die höhere kognitive Prozessdimensionen adressieren und über Multiple-Choice-Wissensabfragen oder andere geschlossene Aufgabenformate nicht zufriedenstellend abgebildet werden können. Die konzeptuelle Modellierung ist davon besonders betroffen, weil die praktische Anwendung der Modellierungsmethoden wichtig ist.

Die Konstruktion von Aufgaben zur konzeptuellen Modellierung ist anspruchsvoll. Das gilt besonders für Modellierungsaufgaben, die höhere kognitive Prozessdimensionen adressieren. Solche Aufgaben benötigen eine sinnvolle Problem- und Aufgabenbeschreibung und zumindest eine Musterlösung. Aufgrund des hohen Aufwands für die manuelle Erstellung stehen Lernenden gemeinhin nur wenige Aufgaben zur Verfügung, was die Gelegenheit zum Üben einschränkt und das Lernen behindert.

Es konnten drei Generierungsansätze identifiziert werden:

1. Traditionelles Item Development [Ru19] durch manuelle Erstellung von Aufgaben zur konzeptuellen Modellierung. Aufgrund begrenzter Kapazität der Lehrkräfte kann so keine große Anzahl an Aufgaben erzeugt werden.
2. Automatic Item Generation (AIG) [GLT21], welche von Lehrkräften erzeugte Aufgaben-Templates für die softwaregestützte Erzeugung vieler Aufgaben gemäß des Templates nutzt. AIG eignet sich nur für geschlossene Aufgabenformate und adressiert keine höheren kognitive Prozessdimensionen.
3. Fachbezogene Methoden [SA24] nutzen u. a. Transformationsregeln aus der Metamodellierung, um z. B. eine Modellrepräsentation (z. B. Text) in eine andere Form zu überführen (z. B. eine Zielmodellierungssprache). Diese Methoden erfordern von den Lehrkräften für jede Aufgabe die Erstellung einer eigenen Modellrepräsentation und skalieren daher nicht.

Es ist daher eine offene Forschungsfrage, wie eine große Anzahl von Aufgaben zur konzeptuellen Modellierung erzeugt werden kann, die höhere kognitive Prozessdimensionen

nach der revidierten Bloom'schen Taxonomie adressieren. Zur Beantwortung dieser Frage betrachtet der nächste Abschnitt Aufgaben zur konzeptuellen Modellierung und, wie diese höhere kognitive Prozessdimensionen adressieren. Nach einer kurzen Darstellung verwandter Arbeiten und ihrer Defizite stellen wir unseren Ansatz zur automatischen Generierung von Aufgaben für die konzeptuelle Modellierung vor und validieren ihn anhand einer Fallstudie. Ein Fazit, offene Fragen und ein Ausblick schließen den Beitrag ab.

2 Aufgaben zur konzeptuellen Modellierung für höhere kognitive Prozessdimensionen

[Bo19] bildet die Dimensionen kognitiver Prozesse auf potenzielle Typen von Modellierungsaufgaben ab. Wir definieren konkrete Aufgaben zur Modellierung durch eine Item-Spezifikation für die jeweilige Dimension. Ähnlich zu Gierl's Definition eines kognitiven Modells [GLT21] legt eine Item-Spezifikation die Konzepte, Annahmen und Logik für die Erstellung eines Items und die Annahme darüber fest, wie Lernende eine Aufgabe lösen.

Tab. 1: Item-Spezifikationen zweier Aufgaben zur konzeptuellen Modellierung.

Instruktions-Template	**Eingabe-Platzhalter**	**Ausgabe-Platzhalter**	**Antwort-Format**
Transformiere das in der Notation <SOURCE MODELING LANGUAGE> angezeigte Modell in ein Modell der Notation <TARGET MODELING LANGUAGE>.	⟨MODEL⟩	⟨TRANSFORMED MODEL⟩	Constructed-Response
Finde und markiere syntaktische Fehler in dem gezeigten Modell.	⟨ERRONEOUS MODEL⟩	⟨MARKED MODEL⟩	Constructed-Response

Tabelle 1 zeigt für die höheren kognitiven Prozessdimensionen *Anwenden* und *Analysieren* (Ebene 3 und 4 der revidierten Bloom'schen Taxonomie), welche Elemente die Item-Spezifikation für eine entsprechende Aufgabe zur konzeptuellen Modellierung beinhalten muss. Platzhalter, die in einer konkreten Aufgabe ersetzt werden müssen, werden in spitzen Klammern dargestellt. Eine Item-Spezifikation für eine Aufgabe zur konzeptuellen Modellierung besteht aus vier Teilen: (1) dem Instruktions-Template, welches spezifiziert, was Lernende ausführen sollen, (2) dem Eingabe-Platzhalter, welcher die Inhalte umfasst, die zur Durchführung der Aufgabe nötig sind, (3) dem Ausgabe-Platzhalter, welcher die erwartete korrekte Lösung beinhaltet, und (4) dem Antwort-Format, das festlegt, wie die Antwort des Lernenden erfasst werden soll.

Die erste Zeile der Tabelle 1 stellt eine Item-Spezifikation für Aufgaben zur konzeptuellen Modellierung in der kognitiven Prozessdimension *Anwenden* dar. Das Instruktions-Template dient Lernenden als Aufforderung, ein in der Notation einer Modellierungssprache <SOURCE MODELING LANGUAGE> angezeigtes Modell in ein äquivalentes Modell in der Notation einer anderen Modellierungssprache <TARGET MODELING LANGUAGE>

zu transformieren. Den Lernenden wird der Inhalt des Eingabeplatzhalters <MODEL> angezeigt, zu dem sie das transformierte Modell <TRANSFORMED MODEL> in der Zielsprache erstellen sollen. Die von Lernenden erstellten Modelle werden zur Bewertung mit der Musterlösung im Ausgabeplatzhalter verglichen. Wie Lernende das Modell erstellen (z. B. auf Papier oder mit einem Modellierungswerkzeug), ist nicht vorgegeben, jedoch wird ein offenes Antwortformat erwartet. Die zweite Zeile der Tabelle 1 stellt eine Item-Spezifikation für Aufgaben zur konzeptuellen Modellierung in der Dimension *Analysieren* dar. Lernende sollen Fehler in einem fehlerhaften Modell <ERRONEOUS MODEL> finden und markieren und so ein mit Fehlern markiertes Modell <MARKED MODEL> erstellen.

Für eine konkrete Aufgabe müssen alle Platzhalter der entsprechenden Item-Spezifikation mit passenden Modellen oder Modellierungssprachen befüllt werden, was manuell oder automatisch geschehen kann. Für das Erstellen einer großen Zahl solcher Aufgaben erscheint eine automatische Generierung notwendig, die semantisch passende Kombinationen von Instruktions-Templates und Eingabe- und Ausgabe-Platzhaltern erzeugt.

3 Verwandte Arbeiten zur Generierung von Aufgaben zur konzeptuellen Modellierung

3.1 Traditionelles Item Development

Bei dem traditionellen Item-Development erstellen Fachexperten Items manuell [GLT21]. Bestenfalls wechseln sie sich beim Erstellen, Editieren, Reviewen und Revidieren eines jeden Items ab, bis die gewünschte Item-Qualität erreicht worden ist [HS15]. Traditionelles Item-Development ist zeitaufwändig und teuer, da jede Aufgabe individuell erstellt und idealerweise auch editiert, reviewed und revidiert werden muss [GLT21; SCD06]. Deshalb sind viele Fachexperten nötig, um eine große Anzahl von Aufgaben zu erstellen. Es erscheint deshalb unrealistisch, viele Aufgaben zur konzeptuellen Modellierung mittels traditionellem Item-Development zu erstellen.

3.2 Automatic Item Generation

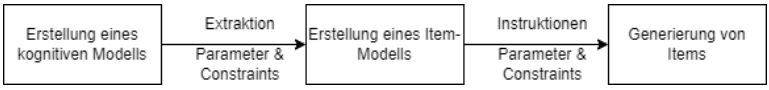


Abb. 1: Dreischrittiger AIG-Prozess nach [GLT21] (übersetzt)

Automatic Item Generation (AIG) ist ein Prozess, der Items mittels Modellen und Computertechnik erzeugt [GLT21]. Gierl et al. beschreiben AIG als einen dreischrittigen Prozess zur Item-Erzeugung (siehe Abbildung 1). Der erste Schritt erstellt den Inhalt für die Item-Erzeugung in Form eines kognitiven Modells, welches das zur Bearbeitung der

Aufgabe notwendige Wissen und die notwendigen Fähigkeiten und Problemlösungsprozesse enthält [GLT21]. Der zweite Schritt erzeugt ein Item-Model, das die Bestandteile und Inhalte spezifiziert, die zur Erzeugung neuer Items manipuliert werden können [La86]. Bestandteile beinhalten den Aufgabenstamm, die Optionen und die Zusatzinformationen, um Selected-Response Items zu generieren (z. B. Multiple-Choice-Aufgaben) [GLT21]. Der dritte Schritt bildet mithilfe computerbasierter Systeme alle vom Item-Modell erlaubten Kombinationen, um eine große Zahl an individuellen Items zu erzeugen [GLT21; Ku23].

Ein großer Nachteil existierender AIG-Methoden ist jedoch die Einfachheit der generierten Fragen, die meist nur niedrige kognitive Prozessdimensionen (*Erinnern* und *Verstehen*) adressieren [Ku20] und geschlossene Antwortformate verwenden, wie z. B. Multiple-Choice oder Lückentexte [Fa23; Ku20; PSS20]. Existierende AIG-Methoden sind zwar grundsätzlich in der Lage, Aufgaben zu erzeugen, die höhere kognitive Dimensionen adressieren, [Pu20; St23]. Die Verwendung geschlossener Antwortformate zur Prüfung der Kompetenzstufe *Anwenden* in der konzeptuellen Modellierung ist nicht möglich. Für die Prüfung der Kompetenzstufe *Analysieren* verbleibt die grundlegende Problematik geschlossener Antwortformate, dass die Anwendung der zu prüfenden Kompetenz bei der Lösung der Aufgabe zwar vermutet und durch die Formulierung der Aufgabenstellung aufgedrängt werden kann, diese Vermutung aufgrund des Antwortformats jedoch grundsätzlich nicht überprüfbar ist.

3.3 Fachspezifische Methoden

Fachspezifische Methoden zur Konstruktion der in Abschnitt 2 beschriebenen und für Aufgaben zur konzeptuellen Modellierung benötigten Eingaben und Ausgaben können die Beschränkungen der vorher genannten Ansätze überwinden. Für verschiedene Fachgebiete existieren Methoden zur Transformation einer eventuell vagen Modell-Repräsentation, z. B. einer textuellen Beschreibung in natürlicher Sprache, in potenziell viele valide Repräsentationen in einer vorgegebenen formalen Modellierungssprache. Beispiele umfassen die Transformation von Geschäftsprozessbeschreibungen in Modelle einer grafischen Modellierungssprache [SA24], die Erzeugung von Entity-Relationship Diagrammen aus einem Satz textueller Anforderungen in natürlicher Sprache [GB18] und das Erstellen von Modellinstanzen mittels handgeschriebener Generatoren, Modellanalysesoftware und SAT-Solvern [KSV19; SV20].

Defizite der fachspezifischen Methoden liegen in ihrer fehlenden Übertragbarkeit auf andere Fachgebiete und dem Umstand, für jede Aufgabe eine eigene Repräsentation als Eingabe für verschiedene Item-Varianten vorgeben zu müssen. Fachspezifische Methoden sind deshalb ebenfalls nicht geeignet, eine große Anzahl fachübergreifender Aufgaben zur konzeptuellen Modellierung zu erzeugen.

4 Lösungsansatz zur Generierung von Aufgaben zur konzeptuellen Modellierung

Für die Generierung vieler Varianten einer konzeptuellen Modellierungsaufgabe stellen wir eine Generalisierung von AIG in einem zweischrittigen Prozess vor: Im ersten Schritt erzeugen Item-Entwickler eine Item-Spezifikation gemäß den in Tabelle 1 definierten vier Bestandteilen: (1) dem Instruktions-Template, (2) dem Eingabe-Platzhalter, (3) dem Ausgabe-Platzhalter und (4) dem Antwort-Format. Im zweiten Schritt spezifizieren Item-Entwickler einen Generator für die Erzeugung von Item-Instanzen für die Ein- und Ausgabe-Platzhalter einer Item-Spezifikation. Der Generator sei eine Funktion $G(I) = O$, der eine Eingabe I auf eine Ausgabe O abbildet. Ein Generator kann selbst wieder andere Generatoren aufrufen.

Der zweischrittige Prozess ist eine Verallgemeinerung der AIG (siehe Abschnitt 3.2) und wird im weiteren Verlauf als generalisierter AIG (gAIG) bezeichnet. Der gAIG fasst den in Abbildung 1 beschriebenen AIG-Prozess als eine mögliche Generatorspezifikation auf, um z. B. Single-Choice-Aufgaben zu generieren (Schritt 2 im gAIG-Prozess), welche in der Item-Spezifikation in Tabelle 2 spezifiert werden (Schritt 1 im gAIG-Prozess).

Tab. 2: Item-Spezifikation für eine generische Single-Choice Aufgabe

Instruktions-Template	**Eingabe-Platzhalter**	**Ausgabe-Platzhalter**	**Antwort-Format**
Wähle die korrekte Antwort für die angegebene Aussage.	⟨STEM⟩ ⟨OPTIONS⟩	⟨CORRECT OPTION⟩	Selected-Response

Für eine generische Single-Choice Frage muss der Generator einen <STEM> erzeugen, der eine Frage formuliert, die zu den Antwortmöglichkeiten <OPTIONS> passt, die aus mehreren Distraktoren und einer korrekten Antwort <CORRECT OPTION> bestehen.

Wie bei Gierl et al. beschrieben, erzeugt AIG einen <STEM>, der den Inhalt oder die Frage enthält, die Lernende beantworten müssen, und die <OPTIONS>, eine Menge alternativer Antworten, von denen eine korrekte ist und die anderen falsch sind [GLT21]. Herkömmliche AIG kann deshalb als Generator im zweiten Schritt der gAIG zur Erzeugung notwendiger Werte für die Ein- und Ausgabe-Platzhalter verwendet werden, wie sie Tabelle 2 beschreibt. Wie in Abschnitt 2 diskutiert, kann AIG aber nicht zur Generierung von Aufgaben zur konzeptuellen Modellierung verwendet werden, die kognitive Dimensionen oberhalb der Dimension *Anwenden* adressieren, weil solche Aufgaben eine andere Art von Werten für Ein- und Ausgabe-Platzhalter benötigen. Um Werte für Platzhalter solcher Aufgaben zu generieren, sind andere Generatoren nötig.

Im Folgenden führen wir einen Generator ein, der auf einem Graphersetzungssystem (GES) basiert. Die Vorteile von GESen bestehen darin, dass sie 1.) die nötigen Ausgabe-Modalitäten (speziell grafische Modelle) erzeugen können, 2.) unabhängig von der Fachdomäne sind, und 3.) ihr Verhalten bei der Erzeugung der Ausgabe-Modalitäten konfigurierbar ist.

Die generate-Methode des GES nimmt einen GESINPUT und produziert als Ausgabe einen GRAPHEN, der aus einer Menge von Knoten, einer Menge von Kanten und einer Menge von Subgraphen besteht. Der GESINPUT besteht aus:

- einem Arbeitsgraphen g, auf dem Ersetzungsoperationen ausgeführt werden [Ro97],
- einer Menge benannter Graph-Muster, deren Elemente jeweils aus einem Muster und dessen Namen bestehen. Wenn ein Muster angewendet wird, erzeugt es, unter Berücksichtigung von Eingabe-Parametern und potenziellen Constraints, eine zufällige Instanz eines Graphen, welcher der Struktur des Musters entspricht.
- einer Menge von Ersetzungsregeln. Jede Ersetzungsregel $e(g, L, R)$ enthält einen Mustergraphen L und einen Ersetzungsgraphen R. Bei der Anwendung von e wird eine Instanz von L, die in g enthalten ist, durch eine Instanz von R in g ersetzt. Eine Ersetzungsregel wird nur angewendet, wenn g mindestens eine Instanz von L enthält.
- einer Menge von Parametern, welche die Anwendung der Ersetzungsregeln steuern:
  - wie oft jede Regel angewendet werden muss,
  - einer Ausführungsreihenfolge für die Regeln, und
  - Filter-Optionen für die Auswahl potenzieller Instanzen von L, die in g gefunden worden sind, von denen zufällig eine für das Ersetzen ausgewählt wird.

Das spezifizierte GES kann nun als Generator für Aufgaben im gAIG-Prozess verwendet werden. Die Konfiguration des Generators muss ein Item-Entwickler bereitstellen, ebenso wie die Item-Spezifikation. Das Einsetzen der Generator-Ausgaben in die finale Aufgabeninstanz erfolgt gemäß der Spezifikation des Item-Entwicklers.

Durch die Konfiguration existierender Generatoren und das Bereitstellen weiterer konfigurierbarer Generatoren kann der vorgestellte Lösungsansatz um andere Aufgabentypen erweitert werden, wodurch auch andere Fachbereiche jenseits von Aufgaben zur konzeptuellen Modellierung abgedeckt werden. Der Lösungsansatz reduziert den manuellen Aufwand für die Qualitätssicherung von Aufgaben drastisch, weil Verbesserungen nun auf Ebene der Generatorkonfiguration und nicht mehr auf Ebene der individuellen Aufgabe stattfinden, wie von Stark et al. [SCD06] angemerkt.

5 Fallbeispiel zur Generierung von Aufgaben zur konzeptuellen Modellierung

5.1 Hinführung zum Fallbeispiel

Das folgende Fallbeispiel veranschaulicht den Lösungsansatz aus dem vorangegangenen Abschnitt 4. Es behandelt die Generierung von Aufgabeninstanzen der Aufgabentypen zur konzeptuellen Modellierung aus Tabelle 1. Das Fallbeispiel bedient sich des Bereichs

der Geschäftsprozessmodellierung und der Modellierungssprache der Ereignisgesteuerten Prozessketten (EPKen) [KSN92]. Es werden EPKen gewählt, da sie einer einfachen und bündigen Notation folgen. Um die nachfolgend beschrieben Generierung syntaktisch korrekter EPKen nachvollziehen zu können, ist die Kenntnis der Notation von EPKen erforderlich. Es folgt ein kurzer Überblick über die grundlegende Notation der EPKen. Die EPKen bestehen aus den folgenden Symbolen: Ereignissen, Funktionen, Prozesswegweisern, logischen Konnektoren (AND, OR, XOR) und aus Verbindungspfeilen. Rollen, Verantwortlichkeiten und Daten werden der Einfachheit halber weggelassen. Jede EPK beginnt und endet mit einem Ereignis. Ereignisse und Funktionen müssen sich abwechseln. Die Symbole müssen durch gerichtete Verbindungspfeile verbunden sein. Öffnende Konnektoren müssen genau einen eingehenden und mindestens zwei ausgehende Pfeile haben. Schließende Konnektoren müssen mindestens zwei eingehende Pfeile und genau einen ausgehenden Pfeil haben. Öffnende OR- und XOR-Konnektoren dürfen nicht auf ein Ereignis folgen. Zur Vereinfachung ist die beschriebene Notation restriktiver als die originale Notation.

5.2 Spezifikation eines GESs zur Generierung von EPKen

Nachfolgend werden die Eingaben für das GES spezifiziert, das aus einem Ausgangsgraphen den für eine Aufgabe benötigten Zielgraphen erzeugt.

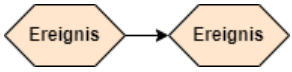


Abb. 2: Initialer Arbeitsgraph zur Generierung von EPK

Initialer Arbeitsgraph g sei der Graph aus Abbildung 2, der lediglich aus zwei verbundenen Ereignissen besteht und noch nicht mit der Notation von EPKen konform geht.

Die in Abbildung 3 dargestellten Muster bilden die Basis für die Ersetzungsregeln zur Generierung von EPKen. Jedes Muster wird als Prozesswegweiser dargestellt. Das Sequenz-Muster enthält eine variable Anzahl alternierender Ereignisse und Funktionen, welche mit Verbindungspfeilen verknüpft sind. Die zum Sequenz-Muster gehörende Konstruktionsfunktion sei $P_s(E_p, E_c, N)$, wobei E_p eine Funktion zur Ermittlung des benötigten Element-Typs zum Start der Sequenz ist, E_c eine Funktion zur Ermittlung des benötigten Element-Typs zum Abschluss der Sequenz und N die Anzahl der Knoten, welche zwischen Start- und End-Element der Sequenz generiert werden sollen. Die Funktion E_p erhält als Eingabe den Knoten, unter bzw. hinter dem die Musterinstanz eingefügt werden soll und traversiert ausgehend von diesem Eingabe-Knoten die Kette der Elternknoten, bis ein Element vom Typ „Ereignis“ oder „Funktion“ gefunden worden ist, und gibt dann entweder „Funktion“ oder „Ereignis“ zurück. Der Startknotentyp (entweder „Ereignis“ oder „Funktion“) für die Sequenz ist somit ermittelt worden. Die Funktion E_c arbeitet analog, aber erhält zusätzlich das Resultat von E_p und traversiert die Kette der Kindknoten. Der Parameter N bestimmt die Gesamtlänge der zu generierenden Sequenz. Das Rhombus-Muster enthält eine variable Anzahl von Zweigen zwischen den Konnektoren

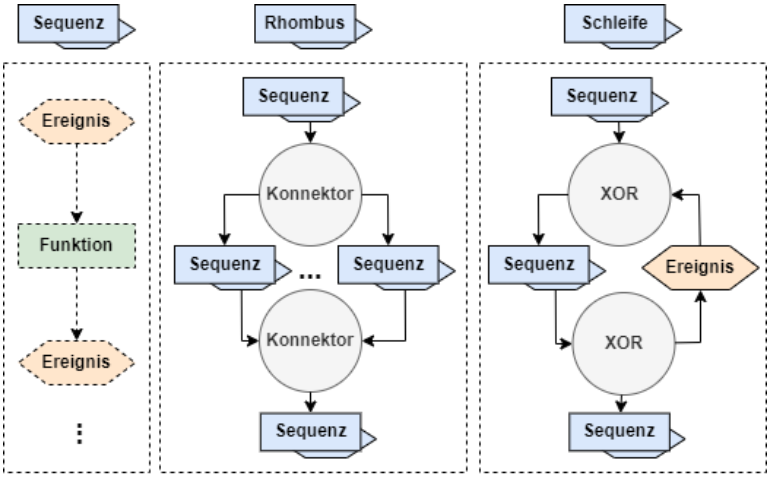


Abb. 3: Drei mögliche Muster zur Generierung von EPKen

und einen variablen Konnektortyp. Die Konstruktionsfunktion des Rhombus-Musters kann als $P_r(N,T)$ beschrieben werden, wobei N die Anzahl der zu erzeugenden Zweige und T der Konnektor-Typ ist. Das Schleifen-Muster besteht aus genau einem Vorwärtszweig und genau einem Rückwärtszweig. Der Rückwärtszweig enthält genau ein Ereignis zwischen den Konnektoren, und der Konnektortyp ist vom Typ „XOR“.

Mit den drei Mustern können Ersetzungsregeln formuliert werden, welche syntaktisch korrekte EPKen erzeugen. Eine Ersetzungsregel besteht aus einem Mustergraph L und einem Ersetzungsgraph R. Im Arbeitsgraph g werden alle Instanzen von L gesucht, und eine zufällig ausgewählte Instanz von L wird anschließend mit einer Instanz von R ersetzt. Der resultierende Graph g' ist der modifizierte Arbeitsgraph, welcher zum neuen Arbeitsgraph g für die nächste Regelanwendung wird.

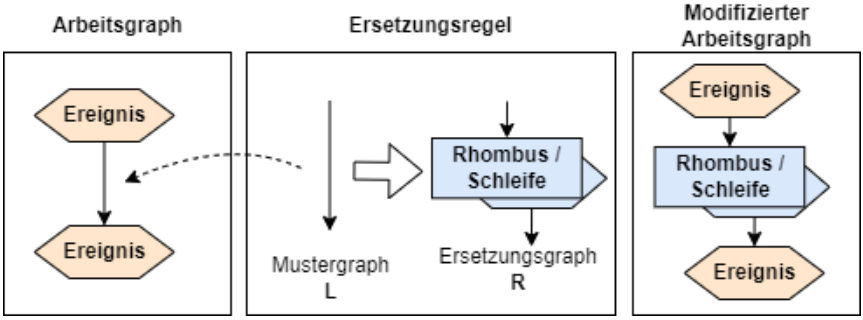


Abb. 4: Ersetzungsregel 1

Die Ersetzungsregel 1 aus Abbildung 4, zeigt, wie komplexe Muster in den Arbeitsgraphen

eingefügt werden. Die Regel selektiert einen Mustergraphen in der Form eines einzelnen Verbindungspfeils und ersetzt ihn durch eine Instanz eines Rhombus- oder eines Schleifen-Musters und einen eingehenden und ausgehendem Verbindungspfeil. Vereinfachend wählt die Regel zufällig ein Rhombus- oder Schleifenmuster aus.

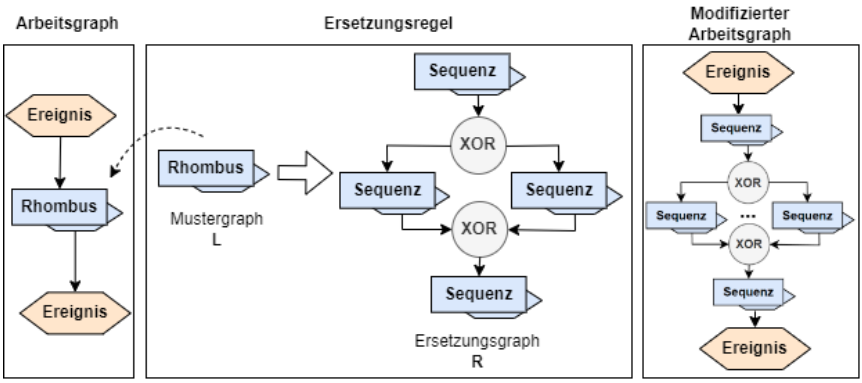


Abb. 5: Ersetzungsregel 2

Die Ersetzungsregel 2 aus Abbildung 5 zeigt, wie ein Ersetzungsmuster instanziiert wird, in dem ein zufällig ausgewähltes Rhombus-Muster (Mustergraph) im Arbeitsgraphen g durch eine Instanz eines Rhombus-Musters (Ersetzungsgraph) ersetzt wird. Vereinfachend wird angenommen, dass ein zufälliges Rhombus-Muster (Ersetzungsgraph) instanziiert wird.

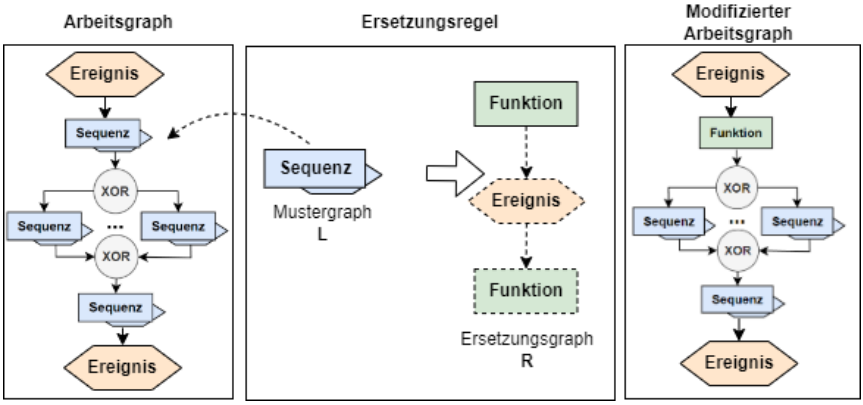


Abb. 6: Ersetzungsregel 3

Die Ersetzungsregel 3 aus Abbildung 6 zeigt, wie eine Sequenz instanziiert wird, indem ein zufälliges Sequenz-Muster (Prozesswegweiser) im Arbeitsgraphen g durch eine Instanz eines Sequenz-Musters (Prozess) ersetzt wird. Die benötigten Knotentypen für den Start und das Ende der Sequenz werden durch die zuvor definierten Funktionen E_c und E_p bestimmt. Zur Vereinfachung wird angenommen, dass der Parameter $N = 0$ ist, sodass nur eine Funktion eingefügt wird, die gleichzeitig Start- und Endknoten der Sequenz repräsentiert.

Bereits durch die zufällige Anwendung der drei Ersetzungsregeln könnte man eine große Anzahl sich stark voneinander unterscheidender EPKen generieren. Um die Generierung besser zu steuern, können zusätzliche Parameter eingeführt werden, wie z. B. die Häufigkeit der Anwendung der jeweiligen Regel oder eine Ausführungsreihenfolge der Regeln.

Die oben beschriebenen Muster und Ersetzungsregeln sind lediglich eine valide Möglichkeit, EPKen mittels eines GESs zu generieren und deckt nicht den gesamten Raum an möglichen, und syntaktisch korrekten EPKen ab. Das ist insofern nützlich, als dass bereits durch die Konstruktionsregeln der Raum an EPKen so beschränkt werden kann, dass lediglich Aufgaben generierbar sind, welche die Lehrenden als relevant erachten. Aufgrund der generierten EPKen lassen sich nun durch die Anwendung der in Abschnitt 3 beschriebenen domänenspezifischen Methoden verschiedene Aufgabenstellungen erzeugen, wie z. B. das Finden von Fehlern in einem Modell oder die Transformation eines Modells, das als EPK generiert worden ist, in eine andere Modellierungssprache.

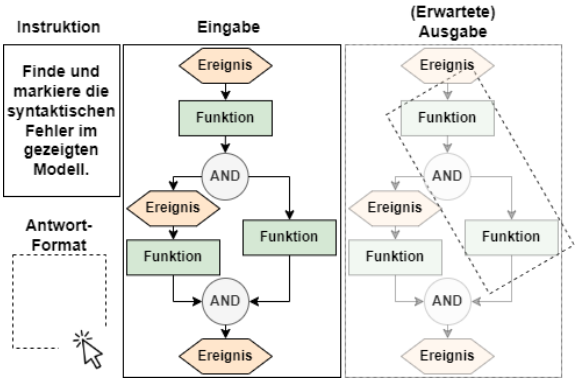


Abb. 7: Aufgabe zum Finden syntaktischer Fehler

Um Aufgaben zum Finden syntaktischer Fehler zu erzeugen, können häufig auftretende Fehlermuster mittels Graphersetzung in vorab syntaktisch korrekt generierte EPKen eingesetzt werden. Sammlungen von Fehlermustern liegen bereits vor [Sz11]. Eine mögliche Aufgabe zum Finden syntaktischer Fehler in EPKen stellt Abbildung 7 dar. Die Aufgabe enthält als Eingabe für Lernende eine fehlerbehaftete EPK, in welcher Lernende Fehler durch Einrahmen markieren sollen. Da die generierten Fehler bekannt sind, kann mittels der erwarteten Ausgabe geprüft werden, ob die Fehler korrekt markiert worden sind.

Aufgaben zu Modelltransformationen können anhand von in Metamodellen spezifizierten Transformationsregeln ebenfalls mittels Graphersetzung erzeugt werden. GESe können EPKen z. B. in UML-Aktivitätsdiagramme transformieren, die dann als erwartete Lösung dienen. Transformationsregeln liegen bereits vor [El23; YC22]. Eine mögliche Aufgabe zur Modelltransformation stellt Abbildung 8 dar. Die Aufgabe enthält eine EPK als Eingabe für Lernende, die sie in ein UML-Aktivitätsdiagramm transformieren sollen.

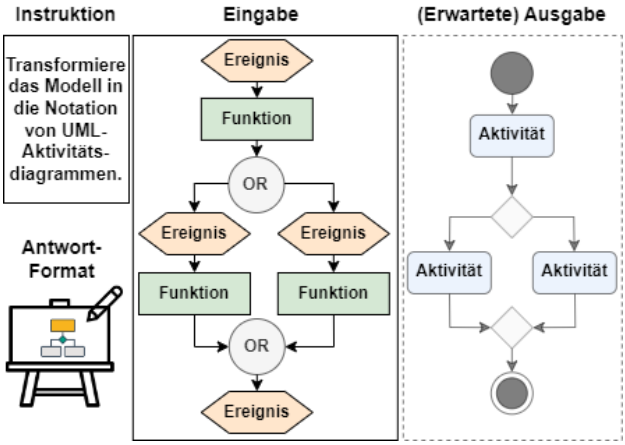


Abb. 8: Aufgabe zur Modelltransformation

6 Fazit

Konzeptuelle Modellierung ist eine wichtige Kompetenz in verschiedenen Disziplinen. Sie zu erlernen, erfordert Übung und das wiederholte Lösen entsprechender Aufgaben, die verschiedene kognitive Prozessdimensionen adressieren. Das manuelle Erstellen vieler solcher Aufgaben ist sehr aufwendig, und derzeitige Methoden zur automatischen Aufgabengenerierung sind aufgrund mangelnder Skalierbarkeit oder fehlender Möglichkeit, höhere kognitive Prozessdimensionen zu adressieren, dazu nicht in der Lage.

Unser Beitrag stellt einen generalisierten AIG-Ansatz und neue Methoden zu Generierung von Aufgaben vor. Der Ansatz erlaubt es, Aufgaben zur konzeptuellen Modellierung in großer Zahl zu generieren, welche die Dimensionen *Anwenden* und *Analysieren* adressieren. Der Ansatz erlaubt sowohl die kontrollierte Generierung von Aufgaben unterschiedlicher Komplexität als auch das *Einfrieren* von Parametern, welche die Komplexität der Aufgaben beeinflussen. Besagtes *Einfrieren* stellt den gleichen Schwierigkeitsgrad der generierten Aufgaben sicher, was den Einsatz der generierten Aufgaben in Prüfungen ermöglicht, und reduziert durch unterschiedliche Aufgaben Täuschungsmöglichkeiten. Der Ansatz ist erweiterbar und kann um weitere Generatoren, wie z. B. große Sprachmodelle/KI, ergänzt werden, die ggf. Aufgaben generieren, welche sogar die höchsten Dimensionen kognitiver Prozesse, *Evaluieren* und *Erschaffen*, adressieren und das Problem *semantikfreier* Aufgaben lösen, wie [CMM24] zeigt. Unser Ansatz trifft keine Annahmen über die Präsentations- und Interaktionsschicht zur Bearbeitung der generierten Aufgaben. Da er jedoch vielfach auf offene Antwortformate abzielt, sollten zukünftige Arbeiten die Verbindung von gAIG und elektronischen Assessment-Umgebungen betrachten, welche die Lernendenantwort in einem computerlesbaren Format erfassen. Computerlesbare Antworten könnten zukünftig automatisches individuelles Feedback und Learning Analytics ermöglichen.

Literaturverzeichnis

[AA14] American Educational Research Association; American Psychological Association: Standards for Educational and Psychological Testing, National Council on Measurement in Education (2014). American Educational Research Association, Washington, DC, 2014, ISBN: 0935302352.

[AC21] ACM, Association for Computing Machinery: Curricula recommendations, 2021, URL: https://www.acm.org/education/curricula-recommendations.

[Bo19] Bork, D.: A Framework for Teaching Conceptual Modeling and Metamodeling Based on Bloom's Revised Taxonomy of Educational Objectives. In: Hawaii International Conference on System Sciences. 2019, URL: https://api.semanticscholar.org/CorpusID:102353632.

[CMM24] Christ., P.; Munkelt., T.; M. Haake., J.: Generalized Automatic Item Generation for Graphical Conceptual Modeling Tasks. In: Proceedings of the 16th International Conference on Computer Supported Education - Volume 1: AIG. INSTICC, SciTePress, S. 807–818, 2024, ISBN: 978-989-758-697-2, DOI: 10.5220/0012753200003693.

[De18] Delcambre, L. M. L.; Liddle, S. W.; Pastor, O.; Storey, V. C.: A Reference Framework for Conceptual Modeling. In (Trujillo, J. C.; Davis, K. C.; Du, X.; Li, Z.; Ling, T. W.; Li, G.; Lee, M. L., Hrsg.): Conceptual Modeling. Springer International Publishing, Cham, S. 27–42, 2018, ISBN: 978-3-030-00847-5.

[El23] Elstermann, M.; Fleischmann, A.; Moser, C.; Oppl, S.; Schmidt, W.; Stary, C.: Modellierungssprachen für Geschäftsprozesse. In: Ganzheitliche Digitalisierung von Prozessen: Perspektivenwechsel - Design Thinking - Wertegeleitete Interaktion. Springer Fachmedien Wiesbaden, Wiesbaden, S. 71–133, 2023, DOI: 10.1007/978-3-658-41777-2_3.

[Fa23] Falcão, F.; Pereira, D. M.; Gonçalves, N.; De Champlain, A.; Costa, P.; Pêgo, J. M.: A suggestive approach for assessing item quality, usability and validity of Automatic Item Generation. Advances in Health Sciences Education 28 (5), S. 1441–1465, 2023.

[GB18] Ghosh, S.; Bashar, R.: Automated Generation of E-R Diagram from a Given Text in Natural Language. 2018 International Conference on Machine Learning and Data Engineering (iCMLDE), S. 91–96, 2018, URL: https://api.semanticscholar.org/CorpusID:58673089.

[GGM19] Guarino, N.; Guizzardi, G.; Mylopoulos, J.: On the Philosophical Foundations of Conceptual Models. In: European-Japanese Conference on Information Modelling and Knowledge Bases. 2019, URL: https://api.semanticscholar.org/CorpusID:208192387.

[GI17] GI, Gesellschaft für Informatik e.V.: Rahmenempfehlung für die Ausbildung in Wirtschaftsinformatik an Hochschulen, 2017, URL: https://dl.gi.de/collections/c7b8fe29-d5f5-410f-81cb-54eaf8edcbcb.

[GLT21] Gierl, M. J.; Lai, H.; Tanygin, V.: Advanced Methods in Automatic Item Generation, en, 2021, URL: https://www.routledge.com/Advanced-Methods-in-Automatic-Item-Generation/Gierl-Lai-Tanygin/p/book/9780367458324, Stand: 21. 02. 2024.

[He07] He, X.; Ma, Z.; Shao, W.; Li, G.: Metamodel for the notation of graphical modeling languages. In. Bd. 19, S. 219–224, 2007, DOI: 10.1109/COMPSAC.2007.27.

[HS15] Haladyna Mark R. Raymond, T. M.; Suzanne, L., Hrsg.: Handbook of Test Development. Routledge, New York, 2015.

[KSN92] Keller, G.; Scheer, A.-W.; Nüttgens, M.: Semantische Prozeßmodellierung auf der GrundlageËreignisgesteuerter Prozeßketten (EPK)". Inst. für Wirtschaftsinformatik, 1992.

[KSV19] Kafa, V.; Siegburg, M.; Voigtländer, J.: Exercise Task Generation for UML Class/-Object Diagrams, via Alloy Model Instance Finding. In. 2019, URL: https://api.semanticscholar.org/CorpusID:214372454.

[Ku20] Kurdi, G.; Leo, J.; Parsia, B.; Sattler, U.; Al-Emari, S.: A Systematic Review of Automatic Question Generation for Educational Purposes. International Journal of Artificial Intelligence in Education 30 (1), S. 121–204, 2020.

[Ku23] Kucharski, S.; Damnik, G.; Stahr, F.; Braun, I.: Revision of the AIG Software Toolkit: A Contribute to More User Friendliness and Algorithmic Efficiency. In. S. 410–417, 2023, DOI: 10.5220/0011982000003470.

[La86] Laduca, A.; Staples, W. I.; Templeton, B.; Holzman, G. B.: Item modelling procedure for constructing content-equivalent multiple choice questions. Medical Education 20 (1), S. 53–56, 1986, DOI: https://doi.org/10.1111/j.1365-2923.1986.tb01042.x.

[Lo01] Lorin W., A.; David R., K.; Peter W., A.; Kathleen A., C.; Mayer, R.; Paul R., P.; Raths, J.; Merlin C., W.: A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. 2001, ISBN: ISBN: 080131903X.

[PSS20] Ponto, R.; Schüler, T.; Striewe, M.: Ansätze zur automatischen Generierung von Aufgaben zum Modellverstehen am Beispiel von UML-Sequenzdiagrammen (Approaches for Automatic Generation of Tasks for Model Comprehension using UML Sequence Diagrams as an Example). In: Modellierung. 2020, URL: https://api.semanticscholar.org/CorpusID:211829645.

[Pu20] Pugh, D.; De Champlain, A.; Gierl, M.; Lai, H.; Touchie, C.: Can automated item generation be used to develop high quality MCQs that assess application of knowledge? Research and Practice in Technology Enhanced Learning 15 (1), S. 12, 2020.

[Ro08] Robinson, S.: Conceptual modelling for simulation Part I: Definition and requirements. Journal of the Operational Research Society 59, S. 278–290, 2008, DOI: 10.1057/palgrave.jors.2602368.

[Ro97] Rozenberg, G.: Handbook of Graph Grammars and Computing by Graph Transformation. 1997.

[Ru19] Rudolph, M. J.; Daugherty, K. K.; Ray, M. E.; Shuford, V. P.; Lebovitz, L.; DiVall, M. V.: Best Practices Related to Examination Item Construction and Post-hoc Review. American Journal of Pharmaceutical Education 83 (7), S. 7204, 2019, ISSN: 0002-9459, DOI: 10.5688/ajpe7204, URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6788158/, Stand: 22. 02. 2024.

[SA24] Schüler, S.; Alpers, S.: State of the Art: Automatic Generation of Business Process Models. In. S. 161–173, 2024, ISBN: 978-3-031-50973-5, DOI: 10.1007/978-3-031-50974-2_13.

[SCD06] Stark, S.; Chernyshenko, O. S.; Drasgow, F.: Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. Journal of Applied Psychology 91 (6), Place: US Publisher: American Psychological Association, S. 1292–1306, 2006, ISSN: 1939-1854, DOI: 10.1037/0021-9010.91.6.1292.

[So22] Soyka, C.; Schaper, N.; Bender, E.; Striewe, M.; Ullrich, M.: Toward a Competence Model for Graphical Modeling. ACM Trans. Comput. Educ. 23 (1), 2022, DOI: 10.1145/3567598, URL: https://doi.org/10.1145/3567598.

[St21] Striewe, M.; Forell, M.; Houy, C.; Pfeiffer, P.; Schiefer, G.; Schüler, S.; Soyka, C.; Stottrop, T.; Ullrich, M.; Fettke, P.; Loos, P.; Oberweis, A.; Schaper, N.: Kompetenzorientiertes E-Assessment für die grafische, konzeptuelle Modellierung. HMD Praxis der Wirtschaftsinformatik 58 (6), S. 1350–1363, 2021, ISSN: 2198-2775, DOI: 10.1365/s40702-021-00797-x, URL: https://doi.org/10.1365/s40702-021-00797-x.

[St23] Striewe, M.: Domain-Specific Automatic Item Generation for Higher Competence Levels: A Comparative Study on Three Cases. In (Milrad, M.; Otero, N.; Sánchez-Gómez, M. C.; Mena, J. J.; Durães, D.; Sciarrone, F.; Alvarez-Gómez, C.; Rodrigues, M.; Vittorini, P.; Gennari, R.; Di Mascio, T.; Temperini, M.; De la Prieta, F., Hrsg.): Methodologies and Intelligent Systems for Technology Enhanced Learning, 13th International Conference. Springer Nature Switzerland, Cham, S. 152–159, 2023, ISBN: 978-3-031-41226-4.

[SV20] Siegburg, M.; Voigtländer, J.: Generating Diverse Exercise Tasks on UML Class and Object Diagrams, Using Formalisations in Alloy. In: Modellierung. 2020, URL: https://api.semanticscholar.org/CorpusID:211251258.

[Sz11] Szczęśniak, B.: Syntax errors in flat EPC diagrams made by persons learning the methodology. EN, Zeszyty Naukowe / Akademia Morska w Szczecinie nr 27 (99) z. 2, S. 75–79, 2011, ISSN: 1733-8670, URL: http://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-article-BWM6-0007-0032, Stand: 27. 02. 2024.

[UH23] Ullrich, M.; Houy, C.: Automated Assessment of Conceptual Models in Education. Enterprise Modelling and Information Systems Architectures (EMISAJ) - International Journal of Conceptual Modeling 15, Nr. 2, S. 1–15, 2023, DOI: 10.18417/emisa.18.2.

[YC22] Yasmina, R.; Chaoui, A.: Automatic Bridge between BPMN Models and UML Activity Diagrams based on Graph Transformation. Computer Science 23, 2022, DOI: 10.7494/csci.2022.23.3.4356.

KI-basierte Analyse von E-Portfolios



Alexander Gantikow [1], Sara Durski [1], Andreas Isking [1],
Paul Libbrecht [1,2], Wolfgang Müller [1], Simon Ostermann [3] und Sandra Rebholz [1,4]

Abstract: E-Portfolios stellen eine effektive Form des Assessments im Kontext des kompetenzbasierten Lernens dar. Die Sichtung und Bewertung solcher Lerntagebücher bringt allerdings bislang einen hohen Zeitaufwand mit sich. Die natürliche Sprachverarbeitung (NLP) liefert vielversprechende Ergebnisse in der Analyse von Texten und eröffnet auch in den Bereichen des formativen und summativen Assessments sowie Feedback von E-Portfolios neue Möglichkeiten, Lehrende bei der Bewertung, aber auch Studierende bei der Komposition von E-Portfolios zu unterstützen. Im Rahmen des hier vorgestellten Projekts wurde eine webbasierte Infrastruktur entwickelt, die E-Portfolios von Studierenden analysiert und für das Assessment relevante Indikatoren und Aspekte in Form von Dashboards visualisiert, um so neue Zugänge zu deren Inhalten zu schaffen. Verfolgte Ansätze und Methoden werden beschrieben und erste Evaluationsergebnisse präsentiert, die die Machbarkeit des Ansatzes demonstrieren, sowie erste Erkenntnisse über den potenziellen Mehrwert des Systems und dafür notwendige Weiterentwicklungen abgeleitet.

Keywords: Learning Analytics, Natural Language Processing, Artificial Intelligence, E-Portfolios, Dashboard, Visualisierung

1 E-Portfolios als Instrument zur Kompetenzbewertung

E-Portfolios sind eine bewährte Form, um individualisiertes Lernen zu unterstützen [Ra05]. Die Erstellung von E-Portfolios fördert Studierende in ihrer Kompetenzentwicklung, indem sie dieser durch selbst verfasste Texte und ausgewählte multimediale Lernartefakte Ausdruck verleihen. Auf Seiten der Lehrenden ermöglichen E-Portfolios einen Einblick in individuelle Lernprozesse und versetzen sie in die Lage, den Studierenden gezieltes Feedback zu geben. Werden E-Portfolios als Instrument zur Leistungsbewertung eingesetzt, müssen die Lehrkräfte die verschiedenen in einem E-Portfolio enthaltenen Artefakte anhand geeigneter Kompetenzraster bewerten [BC15].

E-Portfolios sind digitale Sammelmappen, in denen Lernende ihr Wissen und ihre Lernprozesse dokumentieren. Durch die individuelle Zusammenstellung digitaler Artefakte (Texte,

1 Pädagogische Hochschule Weingarten, Mediendidaktik und Visualisierung, Kirchplatz 2, 88250 Weingarten, Deutschland, gantikow@ph-weingarten.de, https://orcid.org/0000-0002-6571-4683;
sara.durski@ph-weingarten.de, https://orcid.org/0000-0003-0744-7983;
andreas.isking@ph-weingarten.de, https://orcid.org/0000-0003-2531-376X;
muellerw@ph-weingarten.de, https://orcid.org/0000-0001-6474-3733
2 IU Internationale Hochschule, IT & Technologie, Juri-Gagarin-Ring 152, 99084 Erfurt, Deutschland,
3 Deutsches Forschungszentrum für Künstliche Intelligenz, Multilingualität und Sprachtechnologie, Stuhlsatzenhaus 3, 66123 Saarbrücken, Deutschland,
simon.ostermann@dfki.de, https://orcid.org/0000-0002-0899-0657
4 Ostbayerische Technische Hochschule Amberg-Weiden, Elektrotechnik, Medien und Informatik, Kaiser-Wilhelm-Ring 23, 92224 Amberg, Deutschland,

doi:10.18420/delfi2024_31

Grafiken, Videos, ...) können sich Lernende mit einem Vokabular ausdrücken, das sowohl für sie selbst als auch für ihr Zielpublikum zugänglich ist. Gängige E-Portfolio-Systeme sind webbasiert und erlauben sowohl die Präsentation eigens erstellter Artefakte als auch die Einbettung und Verlinkung existierender Artefakte (z.B. Videos oder relevante Lernressourcen). Die dadurch entstehende (Re-)Kombination und Vernetzung von Wissen stellt den Lernprozess und seine Ergebnisse auf sehr individuelle Weise dar. Die Umsetzung kann weit über eine bloße Reproduktion des Lehrstoffs hinausgehen: Das Wissen wird in eigenen Worten wiedergegeben, grafisch aufgearbeitet (z.B. in Form von Concept Maps) oder im Rahmen von Projekten angewendet. Individuelle Schwerpunktsetzungen und Vertiefungen erlauben den Studierenden, den Lernstoff gemäß ihren eigenen Interessen und Kompetenzen zu erarbeiten. Reflexionen des Lernprozesses fordern von den Studierenden zudem eine bewusste Auseinandersetzung mit den eigenen Stärken und Schwächen und helfen dabei, den Lernprozess gezielt zu steuern und anzupassen. Im Gegensatz zu vielen anderen Lernmethoden nutzen E-Portfolios offene, digitale Dokumentenformate, die eine Verknüpfung mit frei verfügbaren Online-Ressourcen erlauben. Diese als Teil der Informationskompetenz geltende Fähigkeit, eigenes Wissen mit bestehenden Wissensquellen zu vernetzen, auf einen bestimmten Kontext anzuwenden und das Ergebnis darzustellen, ist eine Schlüsselkompetenz, die im modernen Berufsleben unabdingbar ist [Eh20].

Für das Assessment der E-Portfolios wurden detaillierte Bewertungsraster (Rubrics) entwickelt, erprobt und optimiert [MRL17]. Mit Hilfe dieser Rubrics bewerten die Lehrenden die erworbenen Kompetenzen und das Wissen der Studierenden anhand von Kriterien wie thematische Vollständigkeit, fachspezifische Tiefe, Verwendung relevanter Quellenmaterialien, Originalität der präsentierten Artefakte, Qualität der Erklärungen, formalen Aspekten und der Reflexion über die eigenen Lernprozesse sowie selbst hergestellte, fächerübergreifende Verknüpfungen der Kursthemen. Beim Lesen der E-Portfolios beurteilen die Lehrenden, wie gut diese Kriterien erfüllt sind, und erhalten einen Einblick, welche Themen (mehr oder weniger ausführlich) behandelt wurden. Diese Bewertung erfolgte bisher größtenteils manuell und ist in bestimmten Kohorten aufgrund der hohen Anzahl von E-Portfolios und des Zeitaufwands für die Bewertung der einzelnen Kriterien sehr aufwändig und zeitraubend. Um eine Vorstellung des insgesamt notwendigen Zeitaufwands zu vermitteln, sei hier außerdem auf ein im Grundstudium zu absolvierendes Projektmodul zur Förderung der digitalen Kompetenzen von Lehramtsstudierenden verwiesen: Rund 600 Teilnehmende dokumentieren jedes Jahr ihre Lernprozesse mit E-Portfolios, wobei die Bewertung eines E-Portfolios in der Regel ein bis zwei Stunden in Anspruch nimmt. An dieser Stelle setzt das Projekt AISOP an und untersucht, wie KI-basierte Methoden und Werkzeuge eingesetzt werden können, um Lehrende beim Assessment (formativ und summativ) und Studierende bei der Erstellung von E-Portfolios zu unterstützen. Speziell im Bereich der automatisierten Textanalyse kann mit einer Themenerkennung schnell und effizient überprüft werden, ob alle geforderten Themenbereiche abgedeckt sind und in welcher Tiefe diese behandelt wurden. Dies entlastet die Lehrenden und ermöglicht es ihnen, sich stärker auf die qualitative Bewertung der Inhalte zu konzentrieren.

2 Relevanz der Textanalyse

Die Analyse geschriebener Sprache ist ein breites Feld der künstlichen Intelligenz und befasst sich mit der Verarbeitung von Informationen in Textform und der Ableitung nützlicher Dienste. Die Textanalyse hat in den letzten Jahren mehrere große Entwicklungsschritte durchlaufen: Nach den semantischen Bereicherungen von Wortähnlichkeiten auf Basis von Ontologien oder distributionellen, zählbasierten Methoden, wurden vektorbasierte Modelle entwickelt, die Wörter als kontinuierliche Vektoren darstellen, die mithilfe von neuronalen Netzen aus dem Satzkontext berechnet werden und den Sinn von geschriebener Sprache deutlich besser erfassen können. Aktuelle neuronale Sprachmodelle sind der Höhepunkt dieser Entwicklung. Sie ermöglichen eine Modellierung von Sprache durch unüberwachtes Training auf allgemeinen Sprachkorpora (etwa Texte aus Enzyklopädien oder Nachrichten). Durch dieses sog. "Pre-training" lernt das Sprachmodell Grundlagen, die für spezielle Anwendungszwecke optimiert ("fine-tuning") werden können. BERT [De19b] oder RoBERTa [Li19] sind Beispiele für solche Modelle. Seit der Etablierung der Technologie sind Sprachmodelle in Größe und Datenhunger immens gewachsen, was mit einer spürbaren Performanzverbesserung einherging.

Mit Hilfe von automatischer Textanalyse lassen sich verschiedene Aufgaben durchführen, darunter die Themenerkennung (engl. Topic Recognition). Sie erlaubt es, Wörter und benachbarte Wörter in Texten zu analysieren, um das zugrundeliegende Thema bzw. darin behandelte Konzept zu erkennen und zu extrahieren. Im Supervised Topic Modeling werden neuronale Netze oder maschinelle Lernmodelle, wie zum Beispiel die oben erwähnten Sprachmodelle, dahingehend trainiert (im Falle von Sprachmodellen optimiert), solche Themen automatisch erkennen zu können. Dafür müssen händisch Trainingsdaten erzeugt werden, indem Textfragmente mit Hilfe einer vorgegebenen Sammlung an Konzepten annotiert und als Korpus gebündelt werden.

Eine solche Themenerkennung wird beispielsweise in der Software "PapagAI" [So23] eingesetzt. Dort werden anhand mehrerer Erkennungssysteme verschiedene Dimensionen studentischer Reflexionen analysiert und als Feedback in Form eines Netzdiagramms den Studierenden präsentiert. [Wu23] wenden ebenfalls NLP-Methoden basierend auf BERT an, um schriftliche Reflexionen zu analysieren und qualitative Indikatoren zu deren Bewertung zu extrahieren. Erste Ergebnisse zeigen, dass Indikatoren wie die Anzahl an behandelten fachspezifischen Konzepten, die Wortanzahl und die Textkohärenz Hinweise auf die Textqualität geben und dies als Ausgangspunkt zur Entwicklung von automatisierten Bewertungsmethoden genutzt werden kann. In [Mu23] werden mit Themenerkennung und anderen Analysen wissenschaftliche Texte im Hinblick auf deren Qualität analysiert. Die Rückmeldung wird als Formulierungsvorschlag im Text ausgegeben, was den weiteren Schreibprozess erleichtern kann. [SR19] entwickelten mit Hilfe von Textanalyse eine Methode, um anhand von Lernmaterialien Ähnlichkeiten zwischen Lehrveranstaltungen festzustellen, um es Studierenden wie Lehrenden zu erleichtern, Relationen zwischen Lehrveranstaltungen zu entdecken. [MRL17] haben ein automatisiertes Werkzeug zur Analyse von E-Portfolios entwickelt, das auf semantischen und deskriptiven Analysen

beruht. Um diesen Ansatz zu erweitern, wird im Folgenden eine ganzheitliche, webbasierte Infrastruktur zur KI-basierten Themenerkennung in E-Portfolios vorgeschlagen.

3 KI-gestützte Analyse von E-Portfolios

Das Projekt AISOP hat das Ziel, Lehrende beim Assessment und Studierende bei der Erstellung von E-Portfolios zu unterstützen. Eine Themenerkennung in E-Portfolios soll speziell auf die Fachkonzepte der in der Lehrveranstaltung behandelten Themengebiete ausgerichtet werden, um so die in einem E-Portfolio enthaltenen Themen in einer interaktiven Concept Map [SAH07] hervorzuheben und die Themenabdeckung in einem Learning-Dashboard zu visualisieren. Neben der inhaltlichen Breite hat sich auch die inhaltliche Tiefe als wichtige Bewertungsgrundlage erwiesen. Hierfür gibt es verschiedene Indikatoren, für die Daten erhoben und auf einem Dashboard visualisiert werden können. Beispielsweise die Häufigkeiten von Konzepten in einem Portfolio, durchschnittliche oder absolute Wortanzahl pro Konzept, die lexikalische Dichte oder die Verwendung von Fachbegriffen und Quellenmaterialien. Hinzu kommen formelle Kriterien wie die Gesamtwortanzahl des Portfolios oder die Anzahl und Kategorisierung eingesetzter Medien. Außerdem soll das Dashboard klassische Elemente aus dem Bereich Learning Analytics beinhalten, wie z.B. eine Zeitleiste aus der das Nutzungsverhalten von Studierenden ersichtlich wird. Ein Vergleich verschiedener Messgrößen zwischen Studierenden ist für die Bewertung von großem Interesse und soll ebenfalls Teil des Dashboards sein. Im Gegensatz zu den gängigen Essay-Scoring-Algorithmen, wie etwa in [Sc23] oder [KB20] bietet AISOP eine visuelle und interaktive Unterstützung, damit Dozierende ein vertrauenswürdiges und fundierteres Assessment durchführen können.

Die automatisierte Unterstützung beim Assessment von E-Portfolios basiert auf den folgenden Schritten:

- Transformation des bewährten Kompetenzrasters: Damit sich die Kriterien auch beim automatisierten Assessment von E-Portfolios quantifizieren lassen, wurden diese als Indikatoren in objektive Messgrößen überführt. So sollen z.B. die thematische Vollständigkeit über die An- oder Abwesenheit behandelter Themen und die Qualität der Erklärungen mit Textmetriken gemessen werden.
- Erfassen und Darstellen erwarteter Inhaltskonzepte: Um die erwarteten Konzepte zu ermitteln, wurde eine SKOS-Ontologie [MPA07] erstellt, die das zugrundeliegende Fachwissen der Kurse konzeptualisiert. Das Ergebnis dieses Schrittes ist eine Concept Map, die als Expertenmodell dient.
- Erstellung eines domänenspezifischen Korpus zum Training von Machine-Learning-Modellen (ML-Modelle) zur automatischen Textanalyse: Verschiedene Arten von Lernressourcen (z.B. Folien, Webquellen, Videotranskriptionen) sowie Portfolios aus vergangenen Kursen bildeten die Grundlage für ein domänenspezifisches Korpus, das mit Annotationen angereichert wurde, die auf die entsprechenden Konzepte des

Expertenmodells verweisen. Das Korpus wurde für das Training diverser ML-Modelle zur Analyse derInhalte von E-Portfolios genutzt.

- Wissensextraktion aus E-Portfolios: Der Inhalt eines E-Portfolios wird als Input für das trainierte ML-Modell verwendet. Darüber hinaus werden Textmetriken sowie der Aufbau des HTML-Dokuments erhoben. Das identifizierte Wissen wird im E-Portfolio über HTML-Attribute kenntlich gemacht.
- Vergleich des extrahierten Wissens mit den erwarteten Lernergebnissen: Die semantische Ähnlichkeit zwischen dem extrahierten Wissen und dem Expertenmodell wird berechnet und in einer kombinierten Concept Map visuell dargestellt: Vorhandene, überlappende und fehlende Konzepte werden hervorgehoben und geben Aufschluss über Wissenslücken, spezifische Expertise und individuelle Interessensgebiete.
- Vergleich des extrahierten Wissens mit dem aggregierten Lernerwissen: Alle E-Portfolios, die im Rahmen einer Gruppenanalyse ausgewertet wurden, bilden das aggregierte Wissen aller Lernenden. Auf diese Weise lässt sich das in einem E-Portfolio enthaltene Wissen mit der Grundgesamtheit vergleichen.
- Visuelle Navigation innerhalb eines E-Portfolios: Die kombinierte Concept Map aus Studierenden- und Expertenwissen eignet sich auch als Werkzeug für die visuelle Navigation [SAH07] innerhalb eines E-Portfolios: Zum einen können die Knotenpunkte der Concept Map als interaktive Links zu den entsprechenden Abschnitten im E-Portfolio genutzt werden, zum anderen werden die Konzepte im aktuell sichtbaren Abschnitt des Portfolios erkannt und in der Concept Map hervorgehoben.
- Kontinuierliche Erweiterung des Korpus: Das zugrundeliegende Korpus für das Training des ML-Modells wird kontinuierlich durch neu erstellte und annotierte E-Portfolios erweitert.

3.1 Training eines Modells zur Textklassifizierung

Von besonderem Interesse für AISOP ist es herauszufinden, welche Konzepte die Studierenden in ihren E-Portfolios vollständig, teilweise oder gar nicht abgedeckt haben. Für das Kriterium der thematischen Vollständigkeit wurde dementsprechend die quantitative Abdeckung der Seminarthemen als einer der Indikatoren definiert. Zu diesem Zweck wurde zunächst eine Concept Map des Seminars “Fundamentale Ideen der Informatik” erstellt und deren Konzepte als Label-Set (“geschlossenes Label-Vokabular”) für das Training eines seminarspezifischen ML-Modells verwendet. Damit das ML-Modell auch auf reflexive Inhalte sowie fächerübergreifende Verknüpfungen trainiert werden konnte, wurde das Label-Set um die übergeordneten Konzepte “Reflexion” und “Link” ergänzt.

Datensatz Für die Erstellung eines Trainingsdatensatzes wurden neben den Foliensätzen die E-Portfolios von 18 Studierenden mit deren Einverständnis zunächst anonymisiert und in

insgesamt 4787 Paragraphen zerlegt. Die Zerlegung wurde von geschulten Mitarbeitenden durchgeführt, die mit einem Browserwerkzeug ausschließlich Portfolioabschnitte mit einem einzelnen Thema selektierten. Die extrahierten Textabschnitte sind im Schnitt 400 Zeichen lang. Alle identifizierten Abschnitte wurden anschließend mit Hilfe der Software "Prodigy" [Ex23a] und dem Label-Set in einer Mehrfachauswahl annotiert. Drei Annotatoren konnten so insgesamt 4396 Datensätze annotieren. Fälle, die von den Annotatoren als unklar markiert wurden, wurden von einer Expertin manuell verbessert. Eine Säuberung der Daten (Löschen von Duplikaten und nicht auflösbaren Fällen) resultierte in insgesamt 3618 Einzelinstanzen, die in 3 Teil-Datensätze aufgesplittet wurden: Trainingsset (2894 Instanzen), Validierungsset (362 Instanzen) und Testset (362 Instanzen).

Abbildung 1 zeigt die Verteilung annotierter Klassen auf den finalen Daten. Die Kategorie "Rest" beinhaltet alle Klassen mit weniger als 50 Instanzen (*Virtualisation*: 44, *Computer Science*: 35, *IT Security*: 30, *Input-Process-Output Model*: 3, *Von Neumann Architecture*: 2). Abschnitte, die als *Algorithmization* gelabelt wurden, stellen mit über einem Viertel der Gesamtdaten die größte Klasse dar. Über 50% der Daten können einer von 4 Klassen zugeordnet werden: *Algorithmization, Representation, Encryption,* und *Operation System.*

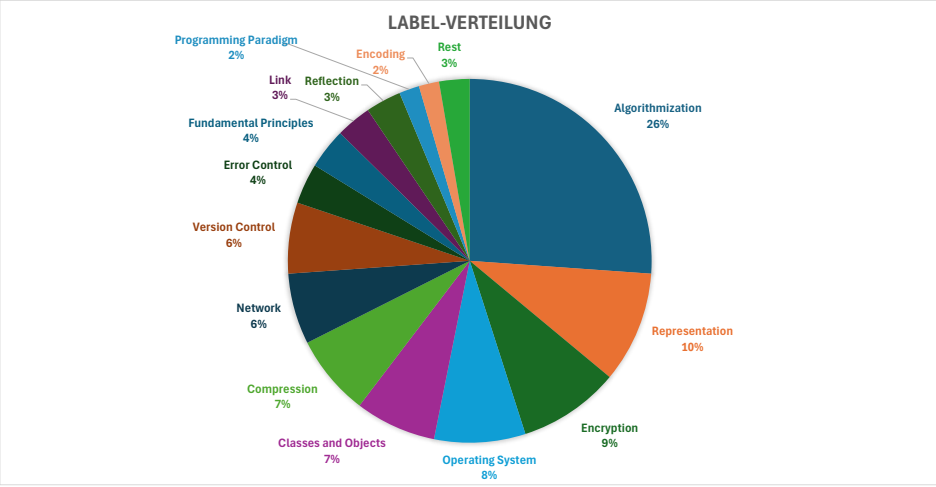


Abb. 1: Verteilung der annotierten Klassen auf den bereinigten Daten im Kurs "Fundamentale Ideen der Informatik"

Klassifikationsmodell Auf Grundlage des annotierten Korpus wurden verschiedene Machine-Learning-Modelle (ML-Modelle) zur Textklassifikation trainiert:

- Wir benutzen ein "spaCy"-basiertes Pipeline-Modell zur Textklassifizierung. Wir experimentieren mit einer Standard-Preprocessing-Pipeline und einem tok2vec-Modell [Ex23b], welches mit einem Multiclass-Head auf den annotierten Trainingsdaten

trainiert wird. Die Vorhersage der Labels erfolgt im Ensemble mit einem 1-Gramm-basierten linearen Bag-of-Words-Modell.

- Wir benutzen mono- und multilinguale ML-Modelle basierend auf vortrainierten mittelgroßen Sprachmodellen (ein deutsches BERT-Modell [Ch23] und ein multilinguales XLM-R-Modell [Co19], was in einem ersten Schritt auf deutschen ConLL-Daten [TKSDM03] optimiert wurde). Für beide Modelle erfolgte eine weitere Optimierung.

Die Genauigkeit der ML-Modelle beträgt 90, 1% (spaCy), 92% (BERT) und 91, 7% (XLM-R). Die Performanz-Unterschiede der benutzten Modelle fallen nur minimal aus. Die generell hohe Performanz aller Modelle zeigt, dass die Textklassifikation der Abschnitte ein wenig komplexes Problem darstellt. Aufgrund der geringen Performanz-Unterschiede und der einfachen Integrierbarkeit fand das spaCy-basierte ML-Modell in der durchgeführten Evaluation (s.u.) Verwendung.

Im Rahmen einer genaueren Untersuchung der Ergebnisse des besten Modells mit einer Konfusions-Matrix konnten wir keine problematischen Paarungen zwischen Klassen finden, die besonders häufig verwechselt würden; die Anzahl der Fehlklassifizierungen ist nahezu uniform über alle Klassen verteilt. Das absolut am häufigsten falsch klassifizierte Label ist *Algorithmization*, was aber zugleich auch die größte Klasse darstellt, mit lediglich 7 falsch klassifizierten Instanzen.

3.2 Die AISOP-Web-Application

Die AISOP-Web-Application wird als eigenständige Node.js-Anwendung neben dem E-Portfolio-System Mahara [Ca22] betrieben und über dessen Schnittstellen angebunden. Über eine Autorisierung mittels OAuth 1.0a können Dozierende und Studierende die Dienste der Webanwendung auf die Portfolios von Mahara anwenden. Nach einer Synchronisierung der Portfolio-Metadaten können beide Zielgruppen zunächst ihre eigenen E-Portfolios analysieren. Da auch Mahara-Gruppen in der Web-Application abgebildet werden, können deren Administratoren (i.d.R. Lehrende) darüber hinaus die E-Portfolios ihrer Gruppenmitglieder zunächst thematisch kategorisieren und diese gebündelt analysieren.

Im Rahmen der automatisierten Analyse werden die von Mahara bereitgestellten Webservice-Schnittstellen genutzt: Die zu analysierenden E-Portfolios werden für die weitere Bearbeitung zunächst auf Mahara gesperrt, deren "HTML Lite Export" angestoßen sowie heruntergeladen und anschließend wieder entsperrt. Das empfangene ZIP-Archiv enthält eine HTML-Datei, die mit geeigneten Python-Analyse-Skripten ausgewertet wird. Mit der Analyse gängiger Textmetriken sowie einer HTML-Analyse liegen zwei generische Skripte vor, die sich auch anwendungsunabhängig nutzen lassen. Ergänzt werden diese durch eine Textanalyse, die auf das zuvor mit spaCy trainierte ML-Modell zur Textkategorisierung zurückgreift. Es erkennt die vom Studierenden beschriebenen Konzepte und reichert die entsprechenden

Paragraphen um das ermittelte Konzept sowie den zugehörigen Wahrscheinlichkeitswert (Score) an. Auf diese Weise werden nicht nur die Treffer gezählt, sondern auch deren Fundstellen im E-Portfolio markiert. Wurde die Analyse auf Gruppenebene angestoßen, so werden die erhobenen Daten entlang der zuvor gebündelten E-Portfolios aggregiert.

Die Auswertungen werden für Lehrende und Studierende in zielgruppenspezifischen Learning-Dashboards dargestellt. Während für Studierende nur die Ergebnisse der eigenen E-Portfolios einsehbar sind, können Dozierende die Analyseergebnisse aller Studierenden einer Kohorte aufrufen. Für jedes analysierte E-Portfolio steht eine Übersicht bereit, die die Ergebnisse zusammenfassend visualisiert. Einen tieferen Einblick erhalten die Nutzenden mit der “Portfolioansicht”. Diese präsentiert das zuvor angereicherte E-Portfolio und erweitert dieses um zusätzliche Interaktions- und Navigationsmöglichkeiten. Eine Visualisierung der kursspezifischen Concept Map bildet hierbei das Herzstück in Abb. 2: Durch Anwählen eines Konzepts können die Nutzenden zu den entsprechenden Fundstellen im E-Portfolio springen. Gleichzeitig wird die Scrollbar um eine Visualisierung ergänzt, die die Verteilung des Konzepts über das E-Portfolio hinweg abbildet. Die Concept Map macht außerdem nicht erwähnte Konzepte sowie abgeleitete Informationen sichtbar, wie beispielsweise die vom Studierenden in Relation verwendeten Konzepte. Darüber hinaus erlauben auch die Ergebnisse der HTML-Analyse (z.B. Abbildungen, Tabellen, Hyperlinks, Tags) sowie ein automatisch erzeugtes Inhaltsverzeichnis einen Sprung zu den jeweiligen Fundstellen. Die Concept-Map-basierte Navigation schafft so neue Zugänge zu den Portfolioinhalten und verknüpft diese mit den für das Assessment erhobenen Daten.

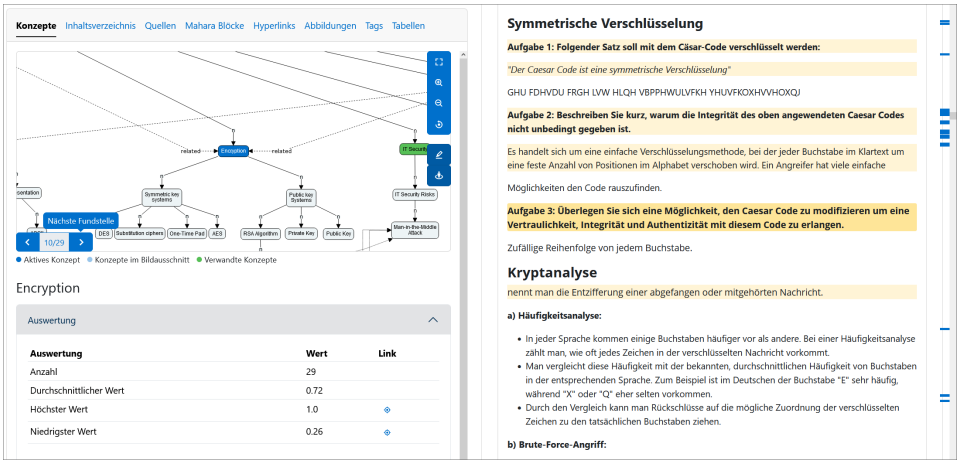


Abb. 2: Die Portfolioansicht besteht aus einer Visualisierung der kursspezifischen Concept Map (links) und dem E-Portfolio (rechts), das angereichert wurde. Durch Anwählen und Weiterblättern eines Konzepts kann ein Nutzer zu den entsprechenden Fundstellen im E-Portfolio springen.

4 Erprobung

Um eine erste Erprobung der webbasierten Infrastruktur zu ermöglichen, wurde die Portfolioprüfung im Wintersemester 2023/2024 so gestaltet, dass sie wenig von der traditionellen Prüfungsform abweicht: Die Studierenden haben ihr E-Portfolio weiterhin im Mahara der Hochschule erstellt, zum Assessment freigegeben und anschließend eine mündliche Prüfung absolviert. Darüber hinaus sollten sie einen Export ihres E-Portfolios einreichen, der in das projekteigene Mahara importiert wurde. Diese Mahara-Instanz wurde eingerichtet, um die technische Machbarkeit des vorgestellten Assessment- und Analyseworkflows zu erproben. Sie entspricht der Version an der Hochschule und dient als Testumgebung, bis das Produktiv-Mahara der Hochschule entsprechend konfiguriert ist. Der hier beschriebene Evaluationsdurchgang beschränkt sich auf den Einsatz des Dashboards zur Analyse und Bewertung von E-Portfolios durch Lehrende. Eine umfassende Evaluation mit Studierenden kann erst durchgeführt werden, wenn die Web-Application an das Hochschul-Mahara-System angebunden ist. Diese Evaluation ist für das Sommersemester 2024 geplant. Die Prüfenden hatten die Möglichkeit, die E-Portfolios sowohl im (Entwicklungs-)Mahara als auch in der angebundenen Web-Application zu betrachten. Mit letzterer wurden die E-Portfolios analysiert und deren interaktive Dashboards sowie Portfolionavigation zur Bewertung herangezogen. Auf diese Weise konnten zehn E-Portfolios importiert und vollständig in der Mahara-Testumgebung sowie Web-Application betrachtet werden. Die E-Portfolios wurden von insgesamt vier Gutachterinnen und Gutachtern mit Hilfe des Bewertungsrasters überprüft. Zwei der Teilnehmenden leiteten die zu prüfenden Seminare im laufenden Semester, während die anderen beiden in den vorhergehenden Semestern zuständig waren. Eine der beiden Gutachterinnen nahm zudem in der Rolle als UX-Expertin an der Evaluation teil. Alle Teilnehmenden teilen dasselbe informationstechnische Wissen.

Folgende Fragestellungen sollten mit der Evaluation beantwortet werden:

- Ist der mit der webbasierten Infrastruktur umgesetzte Workflow funktionsfähig? Ist insbesondere eine reibungslose Integration des E-Portfoliosystems, der NLP-Pipeline und des Lehrenden-Dashboards über die verfügbaren Schnittstellen in der Praxis umsetzbar und das System für das vorgesehene Bewertungsszenario nutzbar?
- Ist die Concept-Map-basierte Navigation und die um die Ergebnisse der Textanalyse angereicherte Darstellung der E-Portfolios intuitiv und einfach zu verwenden?
- Wie gut nehmen die Lehrenden das interaktive Dashboard an?
- Welchen Mehrwert bieten die im Lehrenden-Dashboard dargestellten Ergebnisse der Textanalyse bei der Bewertung von E-Portfolios?

4.1 Methodik und Durchführung

Um die obenstehenden Fragen zu beantworten, wurde ein halbstrukturiertes Interview mit den an der Evaluation teilnehmenden Lehrenden und einer Expertin im Bereich User-

Experience-Design (UX-Design) durchgeführt. Das Interview wurde auf der Grundlage eines Interviewleitfadens durchgeführt, der basierend auf Konstrukten der Usability [De18], User Experience [De19a] und des Technologieakzeptanzmodells (TAM) [Da85] konzipiert wurde. Somit konnten die Gebrauchstauglichkeit und das Benutzererlebnis der Befragten Aufschluss darüber geben, inwieweit sie ihre Ziele bei der Verwendung der Web-Application effektiv, effizient und zufriedenstellend erreichen konnten und welche Vorstellungen, Wahrnehmung und Vorgehensweisen sie hatten. Um Aufschluss über die Akzeptanz und den potenziellen Mehrwert des Lehrenden-Dashboards zu erhalten, wurden außerdem Fragen zur wahrgenommenen Nützlichkeit (Perceived Usefulness, PU) und zu möglichen Erweiterungen des Dashboards in den Leitfaden mit aufgenommen. Das halbstrukturierte Interview wurde als Evaluationsmethodik gewählt, um möglichst tiefgreifende Einblicke in die Erfahrungen der Lehrenden sowie die Beurteilung der UX-Expertin zu erhalten und bei Bedarf Rückfragen stellen zu können. Die Interviews wurden in einem zeitlichen Abstand von einer Woche nach der Nutzung der Web-Application in einer Online-Session geführt. Die qualitativ erhobenen Daten wurden durch eine Inhaltsanalyse, in welcher Kategorien gebildet wurden, ausgewertet.

Die Durchführung der Bewertung wurde von jedem Prüfenden unabhängig vorgenommen, d.h. es fanden keine Absprachen zwischen den Prüfenden statt. Zur Einführung in die Web-Application erhielten die Lehrenden vorab eine videobasierte Anleitung (Anmeldung, grundlegenden Nutzung). Informationen zur Concept-Map-basierten Navigation und der interaktiven Visualisierung der Analyseergebnisse wurden nicht gegeben (Gegenstand der Evaluation). Um den datenschutzrechtlichen Anforderungen gerecht zu werden, wurden nur solche E-Portfolios gesichtet, für die eine schriftliche Einwilligung der Studierenden vorlag.

4.2 Ergebnisse

Für die Nutzung der Web-Application mussten sich die Teilnehmenden zunächst bei Mahara einloggen. Dieser Weg war manchen nicht bewusst und hat den Einstieg in die Evaluation erschwert. In der Web-Application angekommen, konnten alle Teilnehmenden zügig zur Prüfungsgruppe wechseln und darüber die Dashboards der Studierenden aufrufen. Alle konnten die Analyse eines E-Portfolios anstoßen, wurden durch eine durchschnittliche Wartezeit von zwei Minuten allerdings verunsichert und hätten sich Feedback über den Analysefortschritt gewünscht. Alle Teilnehmenden konnten die zur Analyse erzeugte Übersicht öffnen. Die enthaltenen Visualisierungen wurden positiv bewertet, deren Datenvielfalt allerdings bemängelt. So hat ein Teilnehmer beispielsweise kritisiert, dass lediglich die lexikalische Dichte berechnet wurde und weitere Textmetriken nicht angezeigt wurden.

Die Benutzeroberfläche der Portfolioansicht wurde positiv aufgenommen. In Bezug auf die Produktivität gaben die Teilnehmenden an, dass sie schnell zwischen den bereitgestellten Auswertungen (z.B. Inhaltsverzeichnis, Konzepte, Hyperlinks, Bilder) wechseln und mit Hilfe der internen Verlinkungen zu den entsprechenden Trefferstellen im E-Portfolio springen konnten. Die Teilnehmenden haben die Möglichkeit der Navigation über die interaktive

Concept Map rege in Anspruch genommen, bemängelten aber eine fehlende Systematik z.B. durch eine Abhebung von bereits inspizierten Konzepten von den verbleibenden Konzepten, um den Assessmentprozess effektiver zu gestalten. Die Effektivität wurde laut den Teilnehmenden ebenfalls durch eine unpassend gewählte Farbgebung eingeschränkt, sodass die Teilnehmenden nicht erkennen konnten, welche Konzepte im E-Portfolio nicht abgedeckt waren. Zudem fühlten sich die Teilnehmenden unsicher, welchen Mehrwert die für jedes Konzept berechneten Kennzahlen für ihr Assessment bieten. Um die Nützlichkeit dieser Auswertungen zu steigern, wurde eine strengere Strukturierung der Analyseergebnisse entlang des Kompetenzrasters gefordert. In Bezug auf die wahrgenommene Gebrauchstauglichkeit der Web-Application wurde die Concept Map als zu klein bewertet und die Möglichkeit gefordert, die statische durch eine benutzerdefinierte Aufteilung des Viewports abzulösen. Fachfremde Teilnehmende haben die Concept Map zudem als sehr komplex beschrieben und gaben an, mehr Zeit zur Orientierung zu benötigen. Außerdem bedürfen die für jedes Konzept berechneten Kennzahlen einer Erklärung - so einer der Teilnehmenden. Beispielsweise erlaubt die Portfolionavigation das Springen zu dem Paragraphen, der von der Textanalyse mit dem höchsten Wahrscheinlichkeitswert bewertet wurde - eine Funktionalität, die nicht alle Teilnehmenden verstanden haben. Durch eine abweichende Sortierung der Analyseergebnisse von den Indikatoren im Bewertungsraster konnten die Teilnehmenden nur mit Verzögerung die Kriterien des Rasters mit den dargestellten Werten in der Benutzeroberfläche der Web-Application in Verbindung bringen.

Darauf angesprochen, stellten die Teilnehmenden folgende Vorteile der Web-Application heraus: In der Möglichkeit, die Ergebnisse unterschiedlicher Analysen direkt im E-Portfolio zu verorten, sehen sie den Vorteil, sich schnell einen Überblick über das Dokument zu verschaffen. Durch das Hervorheben der entsprechenden Fundstellen wirft die Portfolionavigation den Fokus auf bestimmte Aspekte des E-Portfolios. Laut den Prüfern fällt es ihnen so leichter, dem E-Portfolio zusätzliche Informationen zu entnehmen, die ihrem Assessment zugutekommen. Sie bewerten die Web-Application als hilfreich, sehen in ihr ein sinnvolles Tool für das Assessment von E-Portfolios und trauen ihr bei zukünftiger Verwendung eine Zeitersparnis zu. Alle Teilnehmenden können sich eine Nutzung der Web-Application in kommenden Portfolioprüfungen vorstellen.

Die durchgeführte Evaluation liefert bereits einen Einblick, welchen Mehrwert die Web-Application in der Praxis bietet. Auch wenn der Autorisierungsprozess noch Verbesserungspotenzial aufweist, konnten alle Teilnehmenden den gesamten Workflow aus Login, Portfolioanalyse und Betrachtung der Ergebnisse durchlaufen. Die Teilnehmenden haben die Web-Application als Werkzeug verstanden, das sie während des Bewertungsszenarios unterstützend begleitet hat. Obwohl die Concept Map noch nicht alle gewünschten Informationen abbildet, zeigt sie bereits jetzt ihr Potenzial der vereinfachenden Navigation innerhalb eines E-Portfolios. Die derzeit im Lehrenden-Dashboard integrierten Visualisierungen können hingegen in verschiedenen Use Cases (z.B. vergleichende Benotung) das Potential noch nicht vollständig entfalten.

5 Zusammenfassung und Ausblick

Dieser Beitrag stellt ein KI-gestütztes Portfolio-Analysesystem vor, das Lehrende im Assessment von E-Portfolios unterstützen kann. Mit der Evaluation legt das Projekt sowohl erste Erkenntnisse zum Mehrwert des Systems als auch eine Machbarkeitsstudie vor: Mit der Kopplung an eine Mahara-Instanz kann das vorgestellte Softwaresystem auf deren Portfolioinhalte autorisiert zurückgreifen. Die empfangenen E-Portfolios lassen sich mit Hilfe spezieller Analyse-Skripte auswerten und die Ergebnisse werden in Dashboards und der Portfolionavigation zugänglich gemacht. Mit der durchgeführten Evaluation konnte der ganzheitliche Workflow aus Login, Portfolioexport, Analyse, Aggregierung und Visualisierung erprobt werden. Dennoch hat die Evaluation weiteres Verbesserungspotenzial aufgedeckt, das im Folgenden überblicksartig dargelegt word.

Aufgrund seines begrenzten Label-Sets entdeckt das vom Projekt trainierte ML-Modell nur die Konzepte des Expertenmodells. Mit dieser Vorgehensweise lassen sich zwar die vorhandenen Konzepte in einem E-Portfolio, nicht aber über das Themenspektrum hinausragende Inhalte identifizieren. Als Ergänzung zur geschlossenen kommt deshalb auch eine offene Themenerkennung in Betracht. Außerdem ist der Einsatz künstlicher Intelligenz in AISOP noch auf die Analyse der Portfoliotexte begrenzt. In Zukunft könnten auch Abbildungen und Videos sowie Code- und PDF-Einreichungen analysiert werden. Darüber hinaus soll überprüft werden, ob sich weitere Textmetriken zur Erhebung qualitativer Indikatoren in die Analysen integrieren lassen (z.B. [Mo23], [HOE23]). Eine über das Dashboard nachvollziehbare Zuordnung dieser Indikatoren zu den im Bewertungsraster hinterlegten Kriterien soll ein effizientes Assessment der Dozierenden ermöglichen. Für eine Verbesserung des trainierten ML-Modells sollen die E-Portfolios der kommenden Kohorten ebenfalls anonymisiert, annotiert und in das Korpus übernommen werden. Sowohl das Korpus als auch die trainierten Modelle werden mit Projektende veröffentlicht.

Da die Testumgebung auf derselben Version gehalten wird, wie das Mahara der Hochschule, kann mit deren finaler Zustimmung die AISOP Web-Application für die gesamte Institution angeboten werden. Mit der Kopplung beider Softwaresysteme kann eine umfassendere Evaluation geplant werden, die auch Nutzertests mit Studierenden umfasst.

Die Web-Application, Analyse-Skripte sowie die trainierten ML-Modelle sind auf GitLab[5] öffentlich zugänglich und werden bis zum Projektende laufend aktualisiert. Mit Hilfe der ausführlichen Dokumentation kann die Web-Application an Mahara-Systeme anderer Hochschulen angeschlossen und für den dort gegebenen Anwendungsfall erweitert werden. Lediglich ein ML-Modell muss für den eigenen Anwendungsfall selbst trainiert werden, alle anderen Komponenten der entwickelten Infrastruktur können direkt eingesetzt werden. Das im Experiment verwendete Modell wurde mit Hilfe annotierter E-Portfolios und Foliensätze trainiert; das Korpus und das Modell werden in einem RADAR-Repository[6] veröffentlicht, sobald Klarheit über das anzuwendende Anonymisierungsverfahren herrscht.

5 https://gitlab.com/aisop

6 https://www.radar-service.eu/radar/de/workspace/OcssWywfBlOotMSi

Literaturverzeichnis

[BC15] Brookhart, Susan M.; Chen, Fei: The quality and effectiveness of descriptive rubrics. Educational Review, 67(3):343–368, Juli 2015. https://doi.org/10.1080/00131911.2014.929565.

[Ca22] Catalyst.Net Limited: , Mahara ePortfolio System, Oktober 2022.

[Ch23] Chan, Branden; Möller, Timo; Pietsch, Malte; Soni, Tanay: , German BERT, April 2023. https://huggingface.co/google-bert/bert-base-german-cased.

[Co19] Conneau, Alexis; Khandelwal, Kartikay; Goyal, Naman; Chaudhary, Vishrav; Wenzek, Guillaume; Guzmán, Francisco; Grave, Edouard; Ott, Myle; Zettlemoyer, Luke; Stoyanov, Veselin: Unsupervised Cross-lingual Representation Learning at Scale. arXiv preprint arXiv:1911.02116, 2019.

[Da85] Davis, Fred D.: , A technology acceptance model for empirically testing new enduser information systems: Theory and results, 1985.

[De18] Deutsches Institut für Normung e. V.: , Ergonomie der Mensch-System-Interaktion - Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte, 2018.

[De19a] Deutsches Institut für Normung e. V.: , Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme, 2019.

[De19b] Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In (Burstein, Jill; Doran, Christy; Solorio, Thamar, Hrsg.): Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, Minneapolis, Minnesota, S. 4171–4186, Juni 2019.

[Eh20] Ehlers, Ulf-Daniel: Future Skills: Lernen der Zukunft - Hochschule der Zukunft. Springer VS, Wiesbaden, Januar 2020.

[Ex23a] Explosion AI GmbH: , Prodigy. An annotation tool for AI, Machine Learning & NLP, November 2023.

[Ex23b] Explosion AI GmbH: , SpaCy Model de_core_news_md. German pipeline optimized for CPU., Oktober 2023. https://spacy.io/models/de#de_core_news_md.

[HOE23] Hansen, Lasse; Olsen, Ludvig Renbo; Enevoldsen, Kenneth: TextDescriptives: A Python package for calculating a large variety of metrics from text. Journal of Open Source Software, 8(84):5153, April 2023.

[KB20] Kumar, Vivekanandan; Boulanger, David: Explainable Automated Essay Scoring: Deep Learning Really Has Pedagogical Value. Frontiers in Education, 5:572367, Oktober 2020.

[Li19] Liu, Yinhan; Ott, Myle; Goyal, Naman; Du, Jingfei; Joshi, Mandar; Chen, Danqi; Levy, Omer; Lewis, Mike; Zettlemoyer, Luke; Stoyanov, Veselin: , RoBERTa: A Robustly Optimized BERT Pretraining Approach, Juli 2019. https://doi.org/10.48550/arXiv.1907.11692.

[Mo23] Morse, Colin: , ReadabilityAnalyserDE, Juni 2023.

[MPA07] Miles, Alistair; Pérez-Agüera, José R.: SKOS: Simple Knowledge Organisation for the Web. Cataloging & Classification Quarterly, 43(3-4):69–83, April 2007. https://doi.org/10.1300/J104v43n03_04.

[MRL17] Müller, Wolfgang; Rebholz, Sandra; Libbrecht, Paul: Automatic Inspection of E-Portfolios for Improving Formative and Summative Assessment. In (Wu, Ting-Ting; Gennari, Rosella; Huang, Yueh-Min; Xie, Haoran; Cao, Yiwei, Hrsg.): Emerging Technologies for Education, Jgg. 10108, S. 480–489. Springer International Publishing, Cham, 2017. Series Title: Lecture Notes in Computer Science.

[Mu23] Muecke, Justin; Waldow, Daria; Metzger, Luise; Schauz, Philipp; Hoffman, Marcel; Lell, Nicolas; Scherp, Ansgar: , Fine-Tuning Language Models for Scientific Writing Support, Juni 2023. https://doi.org/10.48550/arXiv.2306.10974.

[Ra05] Ravet, Serge: ePortfolio for a learning society. E Learning Conference, Brussels. Retrieved February, Januar 2005.

[SAH07] Steiner, C.; Albert, Dietrich; Heller, Juergen: Concept mapping as a means to build e-learning. In (Buzzetto-More, Nicole A., Hrsg.): Advanced Principles of Effective e-Learning, S. 59–111. Informing Science Press, Santa Rosa, Januar 2007.

[Sc23] Schlippe, Tim; Stierstorfer, Quintus; Koppel, Maurice ten; Libbrecht, Paul: Explainability in Automatic Short Answer Grading. In (Cheng, Eric C. K.; Wang, Tianchong; Schlippe, Tim; Beligiannis, Grigorios N., Hrsg.): Artificial Intelligence in Education Technologies: New Development and Innovative Practices. Springer Nature Singapore, Singapore, S. 69–87, 2023.

[So23] Solopova, Veronika; Gruszczynski, Adrian; Rostom, Eiad; Cremer, Fritz; Witte, Sascha; Zhang, Chengming; Plößl, Fernando Ramos López Lea; Hofmann, Florian; Romeike, Ralf; Gläser-Zikuda, Michaela; Benzmüller, Christoph; Landgraf, Tim: , PapagAI:Automated Feedback for Reflective Essays, Juli 2023. https://doi.org/10.48550/arXiv.2307.07523.

[SR19] Seidel, Niels; Rieger, Christian Moritz: Semantic Textual Similarity von textuellen Lernmaterialien und ihre Anwendungen. In: DELFI 2019 - Die 17. Fachtagung Bildungstechnologien. Gesellschaft für Informatik e.V., Berlin, 2019. https://doi.org/10.18420/delfi2019_360.

[TKSDM03] Tjong Kim Sang, Erik F.; De Meulder, Fien: Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. In: Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003. S. 142–147, 2003.

[Wu23] Wulff, Peter; Westphal, Andrea; Mientus, Lukas; Nowak, Anna; Borowski, Andreas: Enhancing writing analytics in science education research with machine learning and natural language processing—Formative assessment of science and non-science preservice teachers' written reflections. Frontiers in Education, 7, Januar 2023. Publisher: Frontiers.

Learning Experience durch Content Curation und KI-basierte Lernempfehlungen

Teresa Sedlmeier [1], Volker Sänger [1], Claudia Schmidt [1], Michael Canz [1], Gisela Hillenbrand [1], Katrin Bauer [1], Prabin Dahal [1] und Saptadi Nugroho [1]

Abstract: Open Educational Resources (OER) fördern Lehren und Lernen. Content Curation unterstützt die Integration von OER in Lernumgebungen. Allerdings wächst durch die Einbindung von OER die Zahl an Lerninhalten an, so dass Lernende beim Auffinden passender Lerninhalte unterstützt werden sollten. Zu diesem Zweck bieten sich KI-basierte Lernempfehlungen an, die Lernende in ihrem Lernprozess auf relevante Inhalte hinweisen. In diesem Artikel wird die Erweiterung der Lernplattform Moodle hin zu einer Learning Experience Plattform (LXP) beschrieben, die individuelle Lernerfahrungen für Studierende schaffen soll, indem Content Curation gemeinsam mit KI-basierten Lernempfehlungen integriert werden.

Keywords: Content Curation, KI-basierte Lernempfehlungen, Learning Experience, Open Educational Resources

1 Einführung

OER werden als ein elementarer Baustein einer zukünftigen Bildungsoffensive angesehen. Dies reflektiert auch die Stellungnahme der UNESCO, dass offene Lernressourcen das Lernen und den Wissensaustausch in besonderer Weise unterstützen und damit eine sehr große Bedeutung für Wissensgesellschaften besitzen [UN22]. Mit der Verbreitung von OER entsteht jedoch zugleich die Herausforderung, aus der Vielzahl an öffentlich zugänglichen Lernmaterialien den einzelnen Lernenden die für sie relevante Information zur Verfügung zu stellen. Eine Lösung bietet Content Curation, das Finden, Auswählen, Gruppieren, Anreichern und Teilen von Lernressourcen zu speziellen Themen [KG20]. Das Kuratieren der Lernressourcen reduziert nicht nur die Informationsüberlastung für Lernende, sondern kann gezielt ihre Präferenzen und den aktuellen Lernstand unterstützen. Das Kuratieren von OER ermöglicht die Erstellung eines Materialpools, der als Basis für Lehrveranstaltungen genutzt werden kann, denn durch die Integration kuratierter Inhalte kann das Engagement und Verständnis der Studierenden verbessert werden, insbesondere wenn bei der Gestaltung

1 Hochschule Offenburg, Fakultät Medien, Badstraße 24, 77652 Offenburg, Deutschland,
teresa.sedlmeier@hs-offenburg.de, https://orcid.org/0009-0004-7477-7979;
volker.saenger@hs-offenburg.de, https://orcid.org/0009-0008-2699-0694;
c.schmidt@hs-offenburg.de, https://orcid.org/0009-0005-5849-4992;
michael.canz@hs-offenburg.de, https://orcid.org/0009-0008-5498-4814;
gisela.hillenbrand@hs-offenburg.de, https://orcid.org/0009-0006-8581-7328;
katrin.bauer@hs-offenburg.de, https://orcid.org/0009-0004-0615-0146;
prabin.dahal@hs-offenburg.de, https://orcid.org/0009-0009-9090-2764;
saptadi.nugroho@hs-offenburg.de, https://orcid.org/0000-0001-6364-7133

doi:10.18420/delfi2024_32

von Lernaktivitäten diese dazu angeregt werden, sich mit den kuratierten Inhalten aktiv auseinanderzusetzen und diese kritisch zu analysieren.

Content Curation ist demnach eine wichtige Komponente von LXPs, bei denen neben der Unterstützung von Lernprozessen, das Lernerlebnis im Fokus stehen soll. Um Lernenden in der Vielzahl vorhandener Lernressourcen, die für sie relevanten bereitzustellen, finden Lernempfehlungen Anwendung, vergleichbar mit Empfehlungen in Online-Shops, die dem Kunden zu ihm passende Produkte anbieten. Weitere Komponenten von LXPs sind Microlearning, Mobile Learning, Gamification und ein ansprechendes User Interface [St].

2 Content Curation und KI-basierte Empfehlungen in der Lehre

Content Curation umfasst das geprüfte Auswählen von relevanten Ressourcen zu spezifischen Themen aus verschiedenen Quellen, das qualitative Bewerten, das Kommentieren, um Bedeutung und Kontext bereitzustellen, das Organisieren zur Erleichterung der Systematik, das Speichern für den späteren Zugriff sowie das Löschen von irrelevanten Inhalten [Va19]. Diverse Modelle beschreiben den Content- Curation-Prozess aus verschiedenen Perspektiven, jedoch unterscheiden alle mehrere (ähnliche) Phasen [BA15; Ur22], siehe auch Abb. 2. Search, die erste Phase umfasst die Suche nach Inhalten [GC23]. Dabei werden Inhaltsquellen (z.B. OER-Repositories) lokalisiert und passende Lernressourcen gesammelt [Ag22]. In der zweiten Phase, Select, agieren Kuratierende als eine Art Filter, indem sie die Inhalte hinsichtlich ihrer Relevanz, der Aktualität, der Zuverlässigkeit der Quellen und Passung in den Kontext bewerten und auswählen [BA15]. In der dritten, entscheidenden Phase der Content Curation, dem Sense-Making, verleihen die Kuratierenden den selektierten Inhalten einen Mehrwert. Den ausgewählten Inhalten wird eine konkrete Bedeutung in dem spezifischen Kontext zugeschrieben [GC23]. Das Kuratieren umfasst somit spezielle Tätigkeiten, insbesondere solche, die Inhalte durch Hinzufügen von Kommentaren, Titel, Erläuterungen, Klassifikationen und Tags kontextualisieren. Die letzte Phase, Share, bezieht sich auf die Veröffentlichung der kuratierten Inhalte für die Zielgruppe [GC23], wobei Lernmanagementsysteme eine passende Umgebung darstellen können. Einige Modelle fügen der Content Curation noch einen abschließenden, aber wesentlichen Schritt hinzu: die Bewertung [GC23]. Feedback wie Kommentare und Bewertungen von Nutzenden können im Besonderen die Suche nach passenden Lerninhalten erleichtern [Un16].

Ein erfolgreicher Content Curation-Prozess führt zu einer stetig wachsenden Zahl an Lernressourcen, wodurch Lernende oft nicht wissen können, welche Ressourcen für sie relevant sind. Sie benötigen Hilfe, um die für ihren Lernstatus passenden Lernressourcen zu finden, wobei KI-basierte Lernempfehlungen [ZLZ21] eine adäquate technologische Unterstützung bieten können. Unterschieden werden inhaltsbasierte, kollaborative und hybride Empfehlungen [AT05]. Inhaltsbasierte Empfehlungen verwenden ausschließlich Informationen, die den Inhalt von Lernressourcen beschreiben, z.B. Tags, Schlüsselworte oder Kurzbeschreibungen der Inhalte, also genau die Informationen, die beim Prozess des Kuratierens hinzugefügt werden (siehe Abb. 2). Kollaborative Empfehlungen dagegen

basieren auf dem Verhalten der Lernenden, indem sie Bewertungen der Lernenden oder Zugriffszahlen auf Lernressourcen nutzen. Für andere kollaborative Empfehlungen wird das Verhalten der Lernenden untereinander verglichen. Hybride Empfehlungen kombinieren kollaborative und inhaltsbasierte Empfehlungen und integrieren ggf. zusätzliches Wissen über persönliche Interessen und Präferenzen der Lernenden.

3 Konzeption und Umsetzung von KI-basierten Lernempfehlungen und Content Curation

3.1 KI-basierte Lernempfehlungen

Sowohl die Content Curation als auch KI-basierte kollaborative Lernempfehlungen nutzen das Feedback der Lernenden. Daher wurde im Projekt als Basis ein Bewertungssystem entwickelt, das Lernenden erlaubt, interne und externe Lernressourcen zu bewerten und zusätzlich über einen Kommentar ein persönliches Feedback zu geben [Nu23]. Das Rating nutzt analog zu existierenden Systemen eine Skala von einem bis fünf Sternen. Alle Bewertungen und Kommentare sind pseudonymisiert, um die Vertraulichkeit zu unterstützen und eine offene Kommunikation zu ermöglichen. Basierend auf den existierenden Bewertungen wird bei jeder Lernressource der aktuelle Mittelwert und die Anzahl der Bewertungen angezeigt (siehe Abb. 1a).

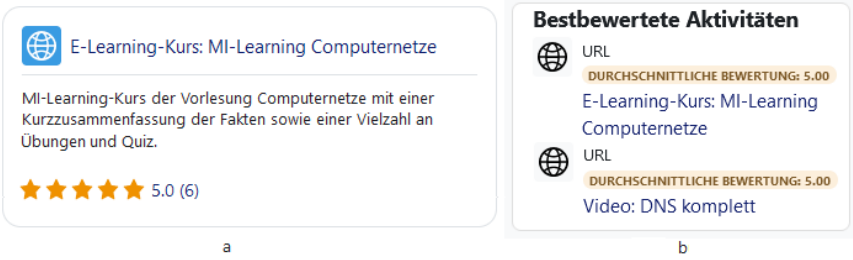


Abb. 1: a) Beispielhafte Lernressource mit Bewertung b) Beste Bewertungen

Neben dem Bewertungssystem unterstützen inhaltsbasierte und kollaborative Lernempfehlungen die Studierenden beim Lernen zielgerichtet. Die Präsentation der Lernempfehlungen erfolgt über Moodle-Blöcke, die Kursen auf der rechten oder linken Seite hinzugefügt werden können. Jede Empfehlungsart wird über ein separates Plugin implementiert und in einem eigenen Block präsentiert, um die Modularität zu unterstützen und gleichzeitig die Transparenz und Erklärbarkeit der KI-Algorithmen und Datennutzung zu gewährleisten [Li18]. Vertrauenswürdigkeit und Akzeptanz werden durch kurze Beschreibungen gefördert, die angeben, welche persönlichen Daten genutzt und welche Algorithmen bzw. KI-Methoden eingesetzt werden.

Zum aktuellen Zeitpunkt sind drei unterschiedliche Empfehlungsarten konzipiert und in Moodle umgesetzt: bestbewertete Lernressource, ähnliche Lernressourcen und "Trending

Items". Die Empfehlung “Beste Bewertung” greift auf die Mittelwerte der aktuellen Bewertungen und das Feedback zu. In Abbildung 1b ist ein Block-Plugin in Moodle zu sehen, das Studierenden anzeigt, welche Ressourcen gerade von ihrer Lerngruppe am besten bewertet werden. Inhaltsbasierte Lernempfehlungen berechnen auf Basis des TF-IDF Maßes die Cosinus-Ähnlichkeit zwischen verschiedenen Lernressourcen im Kursmodul. Der Algorithmus verarbeitet dazu Metadaten wie Ressourcenname, Ressourcentyp und Inhaltsbeschreibungen (siehe Abb. 2). "Trending Items"weist Studierende auf aktuell relevante Lernmaterialien hin, indem die Anzahl der Klicks, die Klickrate (Anzahl der Klicks pro Zeitspanne) und das Basislevel der Klicks ausgewertet werden [Ma20]. Visuell sind die beiden letzten Empfehlungsarten sehr ähnlich wie die bestbewerteten Aktivitäten aufbereitet, und sie verwenden ausschließlich anonyme Daten [Nu23].

3.2 Content Curation

Viele OER-Repositorien folgen einem referenzbasierten Ansatz, in dem nicht die Ressourcen an sich gehostet werden, sondern lediglich deren Metadaten sowie die Hyperlinks zu den jeweiligen Original-Ressourcen [PE23]. Für das vorliegende Projekt wurde ebenfalls ein derartiger referenzbasierter Ansatz gewählt, um auf einfache Weise auf OER zu verweisen.

Da Metadaten hinsichtlich offener Lernressourcen zentral sind, wurde ein Metadatenschema entwickelt [Me23], bei dem Titel, inhaltliche Beschreibung, Themenbereich, Lizenz, unter der die Ressource veröffentlicht ist, Urheber, Format (Video, Podcast, Paper, Webseite, etc.) und der Link zur Ressource verpflichtend angegeben werden müssen. Optional ergänzt werden können die Sprache, frei wählbare Tags, ein Bild als Eintrags-Cover, Veröffentlichungsdatum und spezifische Nutzungsvoraussetzungen. Entscheidend für die Auswahl der Kriterien und die Einteilung in verpflichtende und optionale Kriterien waren die geplanten Einsatzszenarien, bei denen KI-basierte Lernempfehlungen vorgesehen sind, die explizit Titel, Zusammenfassung, Format und Themenbereich benötigen.

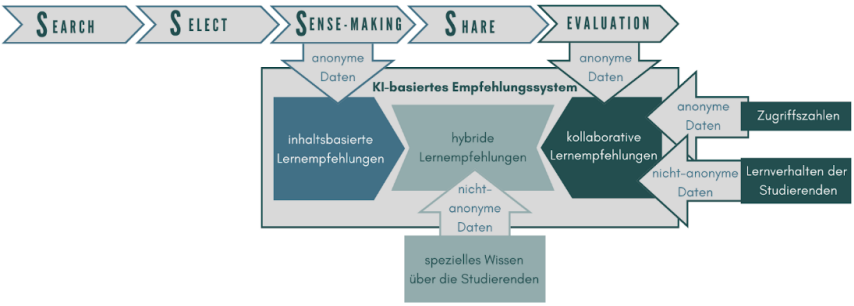


Abb. 2: Zusammenspiel Content Curation Prozess und KI-basierte Lernempfehlungen

Umgesetzt wird das Repository durch die Moodle-Aktivität Datenbank. Lehrende und Lernende können Lernressourcen dort integrieren und in der Sense-Making-Phase basierend auf dem beschriebenen Metadatenschema die Ressourcen mit Metadaten ergänzen. Die

annotierten Lernressourcen können nach Themen oder Medientypen gefiltert werden. Studierende können Kommentare zu den Lernressourcen und wie bereits beschrieben eine Sterne-Bewertung abgeben. Zusätzlich werden die in der Datenbank aufgeführten Ressourcen auch von dem beschriebenen KI-basierten Empfehlungssystem genutzt und in den Empfehlungen angezeigt.

Bisher wurden die Lernressourcen manuell von den Dozierenden gesucht, bewertet und annotiert in die Datenbank integriert. Ab dem nächsten Semester ist geplant, in geeigneten Kursen Content Curation durch Studierende in das didaktische Konzept einzubinden. Ein erster Versuch mit Studierenden zeigt, dass besonders das Sense-Making durch didaktische Methoden unterstützt werden sollte, damit durch die kuratorischen Tätigkeiten ein tieferes Verständnis der Thematik gefördert wird. Technisch bietet die gewählte Realisierung zusätzlich einen Workflow an, bei dem Studierende OER in die Datenbank annotiert integrieren und anschließend Lehrende als Inhalts-Experten diese prüfen, ergänzen sowie für die Lerngemeinschaft freigeben, so dass sie für das KI-basierte Empfehlungssystem nutzbar sind.

3.3 Erste Erfahrungen

Seit zwei Semestern wurden insgesamt acht Moodlekurse manuell von den Lehrenden mit offenen Lernressourcen aus dem Internet und selbst entwickelten interaktiven E-Learning-Anwendungen, z.B. MI-Learning [SS11], angereichert. Anzumerken ist, dass von den Lehrenden mit MI-Learning drei komplette Kurse zu Computernetzen, Datenbanken und Software-Engineering mit Übungen und Quiz als OER angeboten werden.

Hinsichtlich der Content Curation war es in der Search-Phase wichtig, OER zu identifizieren, die inhaltlich zur Thematik und dem Niveau der Lehrveranstaltung passen und von zuverlässigen Herausgebern stammen. Obwohl immer mehr OER-Repositorien existieren, ist speziell im Hochschulkontext die manuelle Suche ein sehr aufwendiger Prozess, der nicht immer das gewünschte Ergebnis erzielt. Denn die Erfahrung lehrt, dass die externen Ressourcen hochwertig sein müssen und sie müssen exakt zu den Inhalten der Lehrveranstaltung passen. Zudem müssen die Syntax und die Begriffswelt ebenfalls der Lehrveranstaltung entsprechen. Ansonsten sind die Studierenden irritiert und die integrierten OER erbringen keinen Mehrwert. Dies unterstreicht die Bedeutung der Select-Phase der Content Curation, bei der Ressourcen anhand von Qualität und Passung im Kontext bewertet werden. Als Resultat mussten viele gefundene Ressourcen als ungeeignet klassifiziert werden. Dennoch konnten alle Kurse mit hochwertigen OER angereichert werden.

Positives Feedback der Studierenden erhielten die selbst entwickelten interaktiven E-Learning-Anwendungen, aber auch Lerneinheiten aus Videokursen (z.B. LinkedIn Learning), die durch die Bibliothek der Hochschule angeboten werden. Generell bestätigt das Feedback der Studierenden die Erfahrungen aus früheren Evaluationen, dass Lernvideos, interaktive Übungen und Quiz zum Lernen bevorzugt werden [SUS21].

Darüber hinaus zeigten erste Analysen der Klickzahlen in ausgewählten Kursen, dass externe Ressourcen durchaus zum Lernen genutzt werden, aber nur von einem Teil der Lernenden. Ein größerer Anteil nutzt die bereits erwähnten E-Learning Anwendungen, weil die Inhalte von Dozenten der Module erstellt worden und eng am Stoff der Vorlesungen ausgerichtet sind, weshalb sich Studierende hiervon einen höheren Nutzen ausrechnen als von externen Ressourcen. Jedoch verlangen viele Befürworter mehr als die reine Konzentration auf OER [Ot21]. Sie befürworten die Einbindung von OER in offene Bildungspraktiken, die auf Offenheit, Selbstorganisation und Partizipation basieren. OER sind oft nicht für einen speziellen didaktischen Zweck konzipiert, weshalb sie didaktisch adäquat in das Lernszenario eingebunden werden müssen.

Sehr positiv sind die bisherigen Kommentare bei den Ratings zu bewerten, denn sie waren immer konstruktiv. Entweder geben sie anderen Lernenden inhaltliche Hinweise oder sie beschreiben den Lehrenden, wo Probleme auftreten - die damit auch leicht behoben werden können.

3.4 Ausblick – weitere Schritte

Im Rahmen des Projektes ist geplant, das bisher manuelle Kuratieren durch KI-gestützte Suche nach OER zu ersetzen, um den Suchprozess zu optimieren. In einem weiteren Schritt sollen KI-basierte persönliche Empfehlungen für Lernende ermöglicht werden. Da diese jedoch auf persönlichen Daten der Lernenden basieren, muss die Datenschutzkomponente ausgebaut werden.

Das Projekt KompiLe wird vom BMBF und MWK Baden-Württemberg im Rahmen der Bund-Länder-Initiative Förderung der Künstlichen Intelligenz in der Hochschulbildung unter der Reg. Nr.: 16DHBKI074 gefördert.

Literaturverzeichnis

[Ag22] Aguilar-Peña, J. D.; Rus-Casas, C.; Eliche-Quesada, D.; Muñoz-Rodríguez, F. J.; La Rubia, M. D.: Content Curation in E-Learning: A Case of Study with Spanish Engineering Students. Applied Sciences 12 (6), 2022, ISSN: 2076-3417, DOI: 10.3390/app12063188, URL: https://www.mdpi.com/2076-3417/12/6/3188.

[AT05] Adomavicius, G.; Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17 (6), S. 734–749, 2005, DOI: 10.1109/TKDE.2005.99.

[BA15] Betts, B.; Anderson, A.: Ready, Set, Curate: 8 Learning Experts Tell You How. American Society for Training & Development, 2015, ISBN: 9781562869403.

[GC23] de Godoi e Silva, K. A.; Costa, A. P.: Study on Contexts and Stages of Digital Content Curation Models: Guidelines for Use in Qualitative Analysis Software. The Qualitative Report 28 (10), S. 2980–2994, 2023, URL: https://doi.org/10.46743/2160-3715/2023.6684.

[KG20] Kröpfl, C.; Geyer, B.: Content Curation im E-Learning. In. S. 1–17, 2020.

[Li18] Lipton, Z. C.: The mythos of model interpretability. Commun. ACM 61 (10), S. 36–43, 2018, ISSN: 0001-0782, DOI: 10.1145/3233231, URL: https://doi.org/10.1145/3233231.

[Ma20] Marrone, M.: Application of entity linking to identify research fronts and trends. Scientometrics 122, S. 357–379, 2020, URL: https://doi.org/10.1007/s11192-019-03274-x.

[Me23] Menzel, M.: Developing a Metadata Profile for Higher Education OER Repositories. In (Otto, D.; Scharnberg, G.; Kerres, M.; Zawacki-Richter, O., Hrsg.): Distributed Learning Ecosystems : Concepts, Resources, and Repositories. Springer Fachmedien Wiesbaden, Wiesbaden, S. 263–278, 2023, ISBN: 978-3-658-38703-7, DOI: 10.1007/978-3-658-38703-7_14, URL: https://doi.org/10.1007/978-3-658-38703-7_14.

[Nu23] Nugroho, S.; Dahal, P.; Hillenbrand, G.; Bauer, K.; Sedlmeier, T.; Schlemmer, D.; Schmidt, C.; Sänger, V.: From LMS to LXP: Extending Moodle with AI-based Recommendations for Learning. In (Lauffenburger, J.-P.; Weber, J., Hrsg.). The Upper-Rhine Artificial Intelligence Symposium (URAI 2023): AI for Time Seriers, Robotics and Beyond, Presented at URAI 2023 Symposium on November 16-17, 2023 at ENSISA; IRIMAS of the Université de Haute-Alsace, The proceedings document is still under process of publication, 2023, URL: https://urai2023.sciencesconf.org.

[Ot21] Otto, D.; Schröder, N.; Diekmann, D.; Sander, P.: Offen gemacht: Der Stand der internationalen evidenzbasierten Forschung zu Open Educational Resources (OER). Zeitschrift für Erziehungswissenschaft 24, S. 1061–1085, 2021, DOI: 10.1007/s11618-021-01043-2.

[PE23] Perifanou, M. A.; Economides, A. A.: Analyzing repositories of OER using web analytics and accessibility tools. Univers. Access Inf. Soc. 22 (4), S. 1243–1257, 2023, DOI: 10.1007/S10209-022-00907-6, URL: https://doi.org/10.1007/s10209-022-00907-6.

[SS11] Schmidt, C.; Sänger, V.: Fünf Jahre E-Learning in der Informatiklehre – Einsichten und Erkenntnisse. Ludwigsburger Beiträge zur Medienpädagogik 14, S. 1–8, 2011, DOI: 10.21240/lbzm/14/06, URL: https://www.medienpaed-ludwigsburg.de/article/view/249.

[St] Stoller-Schai, D.: Was machen wir mit "Learning Experience Platforms"?, https://www.elearning-journal.com/2020/08/12/was-machen-wir-mit-lxp/, [Accessed 10-01-2024].

[SUS21] Schmidt, C.; Utz, M.; Sänger, V.: Digitale Lernszenarien in der Hochschullehre. Bedeutung und Funktion aus Sicht von Studierenden. Ludwigsburger Beiträge zur Medienpädagogik 21, S. 1–14, 2021, DOI: 10.21240/lbzm/21/12, URL: https://www.medienpaed-ludwigsburg.de/article/view/406.

[Un16] Ungerer, L. M.: Digital Curation as a Core Competency in Current Learning and Literacy: A Higher Education Perspective. The International Review of Research in Open and Distributed Learning 17 (5), 2016, DOI: 10.19173/irrodl.v17i5.2566, URL: https://www.irrodl.org/index.php/irrodl/article/view/2566.

[UN22] UNESCO: The 2019 UNESCO Recommendation on Open Educational Resources (OER): supporting universal access to information through quality open learning materials, [Accessed 10-01-2024], 2022, URL: https://unesdoc.unesco.org/ark:/48223/pf0000383205.

[Ur22] Urenda, S. D. R.; Yparrea, N. G.; Moreira, E. L.; Uicab, M. C.; Grez, A. G. In: 2022 XII International Conference on Virtual Campus (JICV). S. 1–4, 2022, DOI: 10.1109/JICV56113.2022.9934776.

[Va19] Vampugani, V. S.: Approaches for curation of open educational resources. International Journal of Research and Analytical Reviews 6, S. 44–47, 2019, URL: https://www.ijrar.org/papers/IJRAR19L1008.pdf.

[ZLZ21] Zhang, Q.; Lu, J.; Zhang, G.: Recommender Systems in E-learning. Journal of Smart Environments and Green Computing 1 (2), 2021, ISSN: 2767-6595, DOI: 10.20517/jsegc.2020.06, URL: https://www.oaepublish.com/articles/jsegc.2020.06.

Term Extraction for Domain Modeling

Theresa Kruse [1], Dominic Lohr [2], Marc Berges [2], Michael Kohlhase [1], Halimeh Moghbeli[1], and Marcel Schütz [1]

Abstract: Adaptive learning systems need to use domain and learner models to provide meaningful support for learners. Building fine-grained domain models by hand is very time-consuming, so the demand for partial automation is high. This paper investigates how term extraction tools can support constructing a domain model. Therefore, we study if different automatic term extraction tools give comparable results to a human annotator. Our results show that the current extraction tools support the process, but their results are not directly usable and still need human adjustments.

Keywords: domain modeling, term extraction, adaptive learning system, programming

1 Introduction

Addressing learners' individual needs in large lectures is challenging, especially if we want to tailor teaching materials, learning interventions, and feedback to individual students or sub-cohorts with special needs and educational biographies [DZ11]. AI-based adaptive learning systems (ALS) can help. Given a fine-grained domain model, we work on a system that can maintain a learner model and use it to generate learner-adapted material: targeted explanations, flashcards, or quizzes from collections of learning objects that reference the domain model [Kr23]. In this system, the domain model consists of a knowledge graph of theories, each introducing a set of concepts and their definitions, properties, and relations. The conceptual and semantic relations – i.e., the terminological dependency relation between concepts – in the knowledge graph, together with a model of cognitive processes, shape the learner model. In our experience, given sufficient student activity in the system, the quality and coverage of the domain model are key determinants of the quality of the learner-adaptivity of the system.

The domain model is essential to the system and requires a large investment. Concept definitions and statements of properties and relations can be taken from course material or textbooks. In other words, the development of domain models should be computer-assisted, not only for quality assurance but also for efficiency reasons. In this paper, we present and evaluate experiments on using natural language processing (NLP) techniques, in particular

1 Friedrich-Alexander-Universität Erlangen-Nürnberg, Wissensrepräsentation und -verarbeitung, Martensstraße 3, 91058 Erlangen, Deutschland, theresa.kruse@fau.de, https://orcid.org/0000-0003-4613-2060; michael.kohlhase@fau.de, https://orcid.org/0000-0002-9859-6337; halimeh.moghbeli@fau.de; marcel.schuetz@fau.de, https://orcid.org/0000-0002-5386-5134

2 Friedrich-Alexander-Universität Erlangen-Nürnberg, Didaktik der Informatik, Martensstraße 3, 91058 Erlangen, Deutschland, dominic.lohr@fau.de, https://orcid.org/0000-0002-6330-2327; marc.berges@fau.de, https://orcid.org/0000-0002-9982-547X

doi:10.18420/delfi2024_33

term extraction, to generate term lists for domain models. We focus on the following research question: *To what extent can automatic keyword extraction support domain model creation?* To answer this question, we extract terms from an introductory computer science course to compare the results of different automatic term extraction tools with those of manual term extraction.

2 Background

2.1 Domain Modeling for Adaptive Learning Systems

Representing the content of a domain model as a theory graph [Fa18, RK13] allows to keep track of all dependencies automatically and instances of the concepts introduced and presumed in a course offered by an ALS. Together with a model of the competencies of each user, this approach automatically identifies their ability to handle specific concepts. This enables us to provide learners with personalized learning materials that meet their individual needs [AK09, Be23].

2.2 Automatic Term Extraction

For a long time, term extraction has been done based on a combination of linguistic, statistical, and pattern-based methods – combining different metrics based on word frequencies. Thus, the most common words in a terminological text are compared with the most common words in general language. Another metric is term-frequency-inverse document frequency: based on this comparison, values for the termhood of words can be calculated [CP14]. Recently, methods based on large language models (LLM) have become increasingly popular, where word embeddings and cosine similarity are used to estimate termhood values [DNMS23].

Regardless of the extraction method, the evaluation measures remain the same, and the quality of a term extraction method in information retrieval is usually evaluated by the precision, recall, and f-measure [Ch92]. To that end, a gold standard is created, usually manually, to serve as a baseline for the evaluation. *Precision* is the percentage of extracted term candidates that are actual terms, while *recall* is the percentage of how many of the terms to be found were extracted. Usually, precision can be improved at the expense of recall and vice versa. The *f-measure* is the harmonic mean between precision and recall and is used to compare systems.

2.3 Term Extraction for Adaptive Learning Systems

Most of the work on term extraction for ALS and domain modeling is based on unsupervised methods because of the large amount of annotated training data required. [Sa19] did a

systematic review of NLP techniques that are used to create concept maps. They show that there is no objective boundary between terms and non-terms when evaluating term extraction tools, and that human experts can disagree on which candidates should be classified as terms. In the following, we present selected work on term extraction for ALS and domain modeling.

[Xu02] follow a frequency-based statistical approach based on GermaNet. Their best precision values vary between 0.52 and 0.63, depending on the corpus size. [BBP13] also use a frequency-based method, which yields a precision of about 0.6 with the best settings. [GN09] take an unsupervised approach based on n-grams. Their results are evaluated against a specialized dictionary. They get an average recall of 0.235 but a precision of 1 with an average f-measure of 0.84. [GA10] extract concepts in mathematics with combined frequency values, also considering how many documents a term candidate appears in. They achieve $f_{0.2}$-measures between 0.67 and 0.94 depending on the document. [ZZY17] use a semantic linked network and get 0.65 for both precision and recall. [BOR23] use different string-matching approaches and achieve the best f-measure of 0.38 with a precision of 0.26 and a recall of 0.36. [So23] extracts terms from PDF textbooks. The texts are split into sentences from which definitions are extracted pattern-based. Similarly, the term candidates are extracted pattern-based from these sentences. The list of terms in the index of the investigated books is used as a gold standard. They get an average recall of 0.647 and a precision of 0.16. To ease the following step of creating a domain model based on the extracted terms, [Ol10] use an approach focused on relations between the concepts that they apply to the domain of biology. Similar work has been done by [Ch21] to create adaptive textbooks and learner models.

3 Methodology

3.1 Extraction Tools

We compare three different term extraction tools that have not been used for the purpose of domain modeling yet: YAKE, D-TERMINER, and KEYBERT. All three are directly usable without further training, which is essential for our use case. In building domain models for an ALS, creating training data for each subdomain is not practical.

YAKE uses an unsupervised approach combining different frequency values. Language adaptivity is provided by different stop word lists [Ro10, Ca20]. The scores considered are casing, position, frequency, and distribution. Casing is because uppercase words are more likely to be a term, which is true for English but cannot be directly transferred to German. Regarding position, the authors assume that term candidates from the beginning of a document tend to be more important than those from the end. We need to investigate if that also holds for our text type, namely lecture slides, where concepts introduced at the end of the semester are considered equally important. It is also assumed that more frequent

terms are more important. Regarding the distribution, two aspects are considered: First, the contextuality, i.e., how many different terms a candidate appears within its neighborhood, and second, the distribution over sentences.

D-TERMINER [Ri21, RTHL22] is also based on statistical measures, but they are combined with neural network approaches. The data was trained on manually annotated corpora from the domains corruption, equitation, heart failure, and wind energy. Based on traditional frequency-related methods, the extraction is done using conditional random fields and a recurrent neural network with word embeddings. Due to the combination of different domains in the training process, the method should be transferable to new domains.

As a third approach, we apply KEYBERT, which uses the embeddings of the language model Bidirectional Encoder Representation (BERT) [Gr20, De19]. The embeddings are combined with cosine similarities.

3.2 Manual extraction

For our manual extraction, an expert in the domain and its education reads the slides and marks all terms they consider necessary for a domain model. The annotation process is accompanied by several discussions with experts from the domain and from education, as well as knowledge representation to address coding quality issues (i.e., reliability or validity). In these discussions, the coder presents their work and adjusts the process according to the comments, resulting in a final term list.

4 Experiment: Dataset and Results

Our experiment is a case study due to its small scope, focusing on the *Introduction to Programming* as it is part of common undergraduate studies in computer science. We work with a collection of slides from a lecture in German. It contains 892 slides spread over 13 sessions and comprises about 64,300 tokens. The manual annotation yields 1195 terms, out of which 853 are single-word terms (SWT) and 342 multi-word terms (MWT). We further split the annotated MWT into SWT to avoid different criteria being used by the tools to combine single-word terms into multi-word terms. This leaves 1232 terms as a baseline for the automatic extraction.

We apply the three extraction methods from section 2.3 and calculate precision, recall, and f-measure. We extract 2464 SWT with YAKE, i.e., twice as many terms as the expert chose since our goal is a higher recall. Recall is more important than precision in our case because it is less effort for instructors to eliminate false positives (due to a lower precision) than having to think of more concepts for which denominations are missing. YAKE does not lemmatize the words, i.e., it is possible that, for example, one word is extracted twice, in singular and in plural. Still, for the use case of a domain model, we are only interested in

one of these forms, so we regard different forms as a kind of doublet and do a semi-manual lemmatization of the results. Afterward, 1994 term candidates remain. For the extraction with D-Terminer, we use the monolingual extraction with the recommended default settings, i.e., we extract specific terms, common terms, out-of-domain terms, and named entities. We also use the training data from all provided domains (corruption, equitation, heart failure, and wind energy) as none of them is our domain. In D-Terminer, it is impossible to extract only SWT and MWT, nor is it possible to specify the number of term candidates to be extracted. We reduce the extracted list of term candidates by the measures already described for the manual extraction. The extraction yields 1681 term candidates (360 MWT and 1321 SWT); after the reduction, 1425 candidates remain. In KeyBERT, we use the standard configuration to extract 2500 SWT, of which 1921 remain after lemmatization.

	Precision	**Recall**	**F-measure**
YAKE	0.3225	0.5297	0.4009
D-Terminer	0.3818	0.4481	0.4123
KeyBERT	0.2806	0.4523	0.3457

Tab. 1: Results of the automatic extraction compared to the manual work

The precision, recall, and f-measure results are given in Table 1 and are somewhat lower than most of those presented in the related work. We achieved our goal of having recall higher than precision as we extracted way more term candidates than had been manually annotated. We also see that the extraction with KeyBERT gives the lowest f-measure but a recall similar to that of D-Terminer.

5 Discussion

By qualitatively evaluating the results, we find some systems in the false negatives, which are visible throughout the whole dataset, and we do not see systematic mistakes made by one tool or the other. Some errors are due to peculiarities of the German language. For example, the concept *Präzedenz-/Vorrangregel* is introduced by giving two synonymous terms. A human can easily deduce that both, *Präzedenzregel* and *Vorrangregel*, can be used, whereas the automatic extraction can only find *Vorrangregel*. Something similar holds for syntactic variants such as *Name einer Variable* and *Variablenname*, where MWTs are used differently. Another aspect is terms that have a meaning in general language, too, and are thus not recognized as terms, e.g., *faule Auswertung* (Engl. *lazy evaluation*).

We also find a system for classifying false positives. One issue is extra-domain terms in the sense of [Ro20], i.e., words that are terms but not of the domain we are looking at. In our case, these are terms from the domain of academic examinations, e.g., *klausurrelevant* (Engl. *relevant for the exam*), *Tutorin* (Engl. *tutoress*) or *Prüfungsleistung* (Engl. *examination requirement*). This could be avoided by relying on text types other than lecture slides and using only textbooks, although it might take more effort to prepare such data for extraction. Another solution could be domain-specific stop word lists, e.g., one for academic

examination, which can also be reused in other domains. A third aspect is words used in examples without meaning in the domain itself. In our data, a library was used for designing example programs, and thus, words like *Ausleihen* (Engl. *borrow*) or *Bibliotheksmitglied* (Engl. *library member*) were extracted by the tools. It might help to annotate example environments in the data and exclude those parts in the automatic extraction.

The results show a trade-off between the completeness and accuracy of the domain model and the effort required to eliminate unneeded terms. If a high recall is aimed for, as discussed above, the automatically extracted list will contain many terms that need to be removed afterward. On the other hand, if a higher precision is aimed at, concepts will be missing from the list. In our view, adding missing concepts to an extracted list is more time-consuming and error-prone than deleting terms. We believe the precision can be further improved by a stop list adapted to the use case. However, our scores are pretty low, and further research is needed to determine if the tools are a valuable aid for building domain models, i.e., if the time saved by the automatic extraction is worth the errors.

6 Conclusion and Future Work

We presented an experiment to support the creation of domain models with term extraction tools from NLP. It is important to note that the automatic analysis can only consider the terminological level, not the conceptual level. For a full comparison between the automatic and the manual extraction, it would be necessary to create the complete domain model based on both lists and to investigate if the domain model created from the automatically extracted terminology fully covers the concepts from the model based on the manual extraction. Building a complete domain model is tedious, so this comparison is beyond this paper's scope. For these reasons, we cannot finally answer our research question, but we state that automatic extraction could be helpful to ease the creation of domain models. As the manual collection of concepts through extensive reading of the material takes a lot of time, it could be faster if the results of an automatic extraction are used as a baseline to find more terms and build the domain model.

Another area where NLP techniques can help with semantic modeling in ALS is annotating learning objects with domain model URI references for semantic processing. For our system, we have independently developed an LLM-based named entity recognizer for terms and integrated it into the IDE to edit learning objects. This can be seen as another source of coverage requirements, and it would be interesting to combine the two approaches. A practically significant difference between the two approaches is that the requirements from learning object annotation are only generated when the learning model is "used", so it is more of a post-hoc quality evaluation or an incremental way of naturally growing a domain model rather than a "table of contents" from which a domain model can be assembled in parallel as the term lists presented in this paper. We expect the practical change management challenges posed by the two approaches will be quite different.

Bibliography

[AK09] Anderson, Lorin W.; Krathwohl, David R.: A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman, New York, 2009.

[BBP13] Bordea, Georgeta; Buitelaar, Paul; Polajnar, Tamara: Domain-independent term extraction through domain modelling. In: TIA 2013 - 10th International Conference on Terminology and Artificial Intelligence. Paris, France, 2013.

[Be23] Berges, Marc; Betzendahl, Jonas; Chugh, Abhishek; Kohlhase, Michael; Lohr, Dominic; Müller, Dennis: Learning Support Systems Based on Mathematical Knowledge Management. In (Dubois, Catherine; Kerber, Manfred, eds): Intelligent Computer Mathematics. Springer Nature Switzerland, Cham, pp. 84–97, 2023.

[BOR23] Banjade, Rabin; Oli, Priti; Rus, Vasile: Automated Extraction of Domain Models from Textbook Indexes for Developing Intelligent Tutoring Systems. In (Frasson, Claude; Mylonas, Phivos; Troussas, Christos, eds): Augmented Intelligence and Intelligent Tutoring Systems. Springer Nature Switzerland, Cham, pp. 124–136, 2023.

[Ca20] Campos, Ricardo; Mangaravite, Vítor; Pasquali, Arian; Jorge, Alípio; Nunes, Célia; Jatowt, Adam: YAKE! Keyword extraction from single documents using multiple local features. Information Sciences, 509:257–289, 2020.

[Ch92] Chinchor, Nancy: MUC-4 Evaluation Metrics. In: Fourth Message Understanding Conference (MUC-4): Proceedings of a Conference Held in McLean, Virginia, June 16-18, 1992. 1992.

[Ch21] Chau, Hung; Labutov, Igor; Thaker, Khushboo; He, Daqing; Brusilovsky, Peter: Automatic Concept Extraction for Domain and Student Modeling in Adaptive Textbooks. International Journal of Artificial Intelligence in Education, 31(4):820–846, December 2021.

[CP14] Cabre, Maria Teresa; Palatresi, Jorge Vivaldi: 110. Acquisition of terminological data from text: Approaches. In (Gouws, Rufus Hjalmar; Heid, Ulrich; Schweickard, Wolfgang; Wiegand, Herbert Ernst, eds): Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography. De Gruyter Mouton, Berlin, Boston, pp. 1486–1497, 2014.

[De19] Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In (Burstein, Jill; Doran, Christy; Solorio, Thamar, eds): Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, Minneapolis, Minnesota, pp. 4171–4186, June 2019.

[DNMS23] Di Nunzio, Giorgio Maria; Marchesin, Stefano; Silvello, Gianmaria: A systematic review of Automatic Term Extraction: What happened in 2022? Digital Scholarship in the Humanities, 38(Supplement 1):i41–i47, 06 2023.

[DZ11] Dosch, Margaret Fuller; Zidon, Margaret: The Course Fit Us": Differentiated Instruction in the College Classroom. The International Journal of Teaching and Learning in Higher Education, 26:343–357, 2011.

[Fa18] Farmer, William M.: A New Style of Proof for Mathematics Organized as a Network of Axiomatic Theories. 2018.

[GA10] Günel, Korhan; Aşlıyan, Rıfat: Extracting learning concepts from educational texts in intelligent tutoring systems automatically. Expert Systems with Applications, 37(7):5017–5022, 2010.

[GN09] Gunel, Korhan; Nuriyev, Urfat: Relation extraction among learning concepts in intelligent tutoring systems. In: 2009 International Conference on Application of Information and Communication Technologies. pp. 1–5, 2009.

[Gr20] Grootendorst, Maarten: KeyBERT: Minimal keyword extraction with BERT., 2020.

[Kr23] Kruse, Theresa; Berges, Marc-Pascal; Betzendahl, Jonas; Kohlhase, Michael; Lohr, Dominic; Müller, Dennis: Learning with ALeA: Tailored experiences through annotated course material. In (Klein, Maike; Krupka, Daniel; Winter, Cornelia; Wohlgemuth, Volker, eds): INFORMATIK 2023. Designing Futures: Zukünfte gestalten. volume 337 of Lecture Notes in Informatics (LNI) - Proceedings, Köllen Druck+Verlag, Bonn, pp. 395–398, 2023.

[Ol10] Olney, Andrew M.: Extraction of Concept Maps from Textbooks for Domain Modeling. In (Aleven, Vincent; Kay, Judy; Mostow, Jack, eds): Intelligent Tutoring Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 390–392, 2010.

[Ri21] Rigouts Terryn, Ayla: D-TERMINE : data-driven term extraction methodologies investigated. PhD thesis, Ghent University, 2021.

[RK13] Rabe, Florian; Kohlhase, Michael: A Scalable Module System. Information & Computation, 0(230):1–54, 2013.

[Ro10] Rose, Stuart; Engel, Dave; Cramer, Nick; Cowley, Wendy: Automatic Keyword Extraction from Individual Documents. In: Text Mining. John Wiley & Sons, pp. 1–20, 2010.

[Ro20] Roelcke, Thorsten: Fachsprachen. Grundlagen der Germanistik 37. Erich Schmidt Verlag, Berlin, fourth edition, 2020.

[RTHL22] Rigouts Terryn, Ayla; Hoste, Veronique; Lefever, Els: D-Terminer: Online Demo for Monolingual and Bilingual Automatic Term Extraction. In (Costa, Rute; Carvalho, Sara; Anić, Ana Ostroški; Khan, Anas Fahad, eds): Proceedings of the Workshop on Terminology in the 21st century: many faces, many places. European Language Resources Association, Marseille, France, pp. 33–40, June 2022.

[Sa19] Santos, Vinicius; Érica de Souza; Felizardo, Kátia; Watanabe, Willian; Vijaykumar, Nandamudi; Aluizio, Sandra; Júnior, Arnaldo: Conceptual Map Creation from Natural Language Processing: a Systematic Mapping Study. Revista Brasileira de Informática na Educação, 27(03):150–176, 2019.

[So23] Soliman, Ashraf: An unsupervised linguistic-based model for automatic glossary term extraction from a single PDF textbook. Education and Information Technologies, May 2023.

[Xu02] Xu, Feiyu; Kurz, Daniela; Piskorski, Jakub; Schmeier, Sven: A Domain Adaptive Approach to Automatic Acquisition of Domain Relevant Terms and their Relations with Bootstrapping. In: Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02). European Language Resources Association (ELRA), Las Palmas, Canary Islands - Spain, May 2002.

[ZZY17] Zhu, Mengmeng; Zhao, Defang; Yang, Juan: A Cognitive-Related Entity and Relation Extraction Model for Online Tutoring Systems. In: 2017 13th International Conference on Semantics, Knowledge and Grids (SKG). pp. 113–119, 2017.

Lernunterstützung durch Bildungstechnologien

Supporting Constructive Alignment: Visualization of the Alignment of Teaching Goals and Developed Assignments

Imen Azaiz [1] and Sven Strickroth [1]

Abstract:

In higher education, aligning educators' goals with students' diverse needs is a major challenge, especially in large classes that make it difficult for educators to monitor students' progress. This paper introduces a novel approach grounded in constructive alignment principles, enhancing the integration of learning objectives and competency models. By aligning teaching goals and developed assignments, our approach fosters deep insights, enabling adaptive teaching and promoting differentiated learning activities. It demonstrates whether assignments comprehensively cover the course's learning objectives based on assignment annotations. This comprehensive alignment analysis provides a comprehensive overview, aiding educators in developing assignments that match the learning objectives and in planning targeted interventions to support students. The feasibility of the general approach was tested using data from an introductory computer science course.

Keywords: Teaching Analytics, Instructional Design Models, E-Assessment, Competence Assessment, Constructive Alignment, Assessment Analytics

1 Motivation

Scalability and interactivity pose major challenges for educators in higher education, as student enrollments increase and the expectations for personalized feedback persist. This is particularly significant when several hundred students attend a (Bachelor's) course which is not uncommon in computer science education [SB22]. Additionally, the challenges of designing and revising course curricula to align with evolving trends [Ba20], accurately identifying students' educational needs, and developing assignments for systematic learning assessment are substantial. Disregarding these aspects can have a negative impact on students' learning processes and educators' ability to maintain students' attention and motivation, ultimately leading to a decline in the educational quality, student neglect, and early drop-outs [SB22]. Many e-assessment systems have been developed to support educators with automated grading and feedback provisioning in programming education [SS22]. While these systems enhance online learning efficiency and scalability, the generated data and user traces are often unused [St22]. These systems often lack learning analytics (LA) capabilities, do not provide a comprehensive overview of students' learning progress, and do not support data-driven assessments [HI23]. Thus, harnessing the available data and applying LA can be beneficial for both teaching and learning [HI23; Om23].

1 LMU Munich, Institute for Informatics, Oettingenstraße 67, 80538 Munich, Germany,
imen.azaiz@ifi.lmu.de, https://orcid.org/0009-0005-6458-4169;
sven.strickroth@ifi.lmu.de, https://orcid.org/0000-0002-9647-300X

doi:10.18420/delfi2024_34

The motivation for this work stems from an often observed discrepancy between learning objectives and assignments developed for teaching. Data generated by e-assessment systems is not automatically matched to learning objectives, as the tasks are not linked to these objectives and assignments are often not systematically developed with learning outcomes and goals, and the progress of the learners in mind. Therefore, our vision is to provide a targeted and systematic analysis of the learning objectives, the developed assignments, and e-assessment (performance) data of students to bridge this gap. We aim to use instructional metadata and e-assessment data effectively to support educators in their course design, and to provide insights into students' learning progress. Towards our vision, this paper answers the following research question: *How can the alignment between learning objectives and assessment assignments be analyzed and visualized to support educators?* A well-founded modeling of learning objectives is a prerequisite to ensure systematic analysis. The contributions are the visualization and the foundational modeling of learning objectives.

2 Related Work

Defining clear learning objectives (LO) is considered a crucial step when designing and planning courses [Ki24]. Kiesler identified this to be a significant challenge for educators, especially for computer science courses [Ki24]. To support computer science educators in the design of programming classes and assessments, Kiesler [Ki20a] proposed a competency model based on Bloom's Revised Taxonomy. Despite this, the systematic and explicit inclusion of learning goals in e-learning systems and the elucidation of these goals through technological means still presents a gap in the Technology-Enhanced-Learning research field [Fe21]. Addressing this gap, Fessl et al. [Fe21] propose a visualization in form of a Learning Goal Widget that aids in the systematic alignment and clarification of learning goals, enhancing both teaching and learning environments. However, using LO in LA remains underexplored. Constructive Alignment (CA) [Bi96] is a crucial pedagogical approach in higher education, and aims to integrate learning objectives with teaching methods and assessments to effectively support the intended learning outcomes. For example, ensuring constructive alignment [Ch23] classified learning outcomes to align with program outcomes and used this alignment to enhance teaching strategies and assessment methods. The classification of LOs according to Anderson Krathwohl's taxonomy to enhance course design has also been proposed in [Ki20b].

The transition to data-driven analytics represents a significant shift towards leveraging e-assessment data and employing LA to enhance student performance predictions using e-assessment data and concept maps, as seen in the work of Omer [Om20]. Using e-assessment data to gain insight into student needs has opened new opportunities for educators. For example, common programming errors can be automatically derived from student submissions in computer science courses [St24]; however, there is no linkage to LOs. As a result, this data is often not systematically used to support CA, particularly the data generated by e-assessment systems [St22]. Despite the recognized importance of CA

in educational theory, its systematic analysis and visualization, particularly in practical course design to support educators, remains underexplored in teaching analytics. Providing multiple representations has been shown to be beneficial for understanding [Ai06].

Our research aims to fill this gap by developing a framework based on competency modeling that follows the principles of CA. This framework is designed to enable advanced data analytics and the visualization of analyses for systematically and automatically aligning and assessing LO and teaching methods. Our overall vision is to dynamically visualize the alignment between educators' goals and students' needs. By systematically aligning and matching LOs with assignment assessments, our research facilitates dynamic teaching strategies that adapt to learning outcomes. This paper addresses the research gap of a missing support of educators in applying CA in their courses by visualizing the alignment of defined LO and developed assignments for the students as a first step towards our vision.

3 Approach

Our study aims to align educators' goals and developed assignments by evaluating how well instructional objectives align with assignments. This requires a systematic evaluation of LOs that adhere to the principles of CA, and the subsequent matching of these LOs to determine how course objectives are covered and assessed in assignments. LOs are typically found in lecture notes or, at most, in the documentation of courses (aka. the module handbook), but not in a machine-readable format. Therefore, we started by preparing the extraction of LO to develop a corresponding data model. We analyzed these objectives to model them into a hierarchical framework, distinguishing between coarse-grained LO for the entire course, fine-grained LO for each topic, and task-specific LO derived from the fine goals, but tailored specifically for individual tasks. A tree data structure was utilized to organize these objectives. Fig. 1 shows an excerpt from the developed tree of LOs.

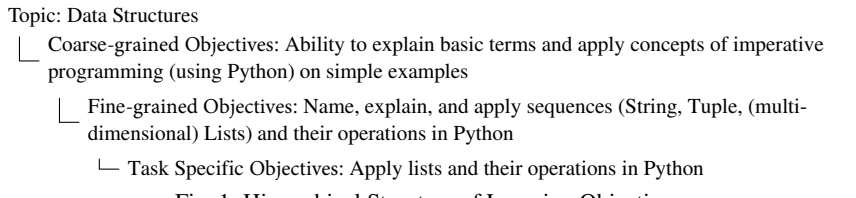


Fig. 1: Hierarchical Structure of Learning Objectives

The assignments were manually annotated with LOs, including substasks, in order to link each individual assessment to the LOs from the course lateron. Where each LO was accessible through its ID, we modeled an indexing system to facilitate the matching process. Now that our data model has been assembled, we can apply our matching algorithm: The first step involves aligning the LOs of the course content with the LOs of the assignments. This was achieved by checking the similarity between each task's goal and all course objectives, using a similarity matrix (cf. Fig. 2a) to match the LOs from the course topics with those of

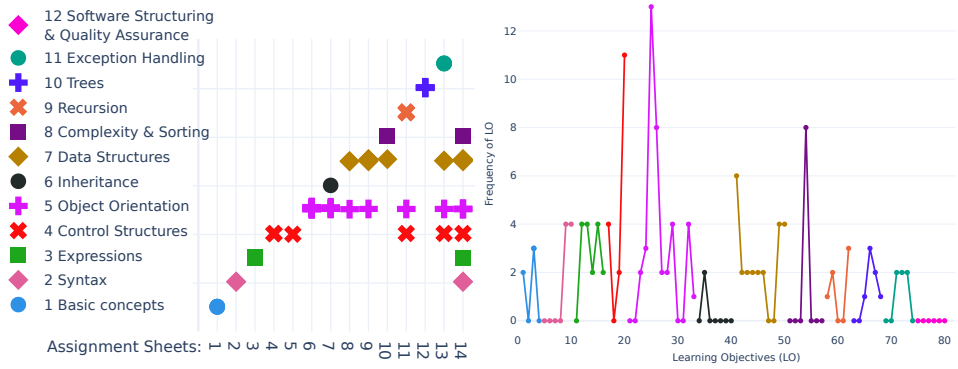
the assignment sheets. This process helps to identify which goals were not addressed in the assignments and to calculate the occurrence of these goals as well as their distribution across assignments (see Fig. 2b). The second step is to determine the coverage of the LOs across all assignment sheets and to visualize the distribution of these goals across all assignments (see Fig. 2a). This analysis helps to determine which LOs require further reinforcement, potentially through the provision of new exercises or by specifically incorporating these objectives into upcoming tasks. By identifying these needs early, we can facilitate timely interventions and corrections. This ensures the goal-oriented development of assignment sheets and assessments that aligns with the COs.

To assess the feasibility of this concept, we gathered instructional goals from the educator of the course "Introduction to Computer Science: Programming and Software Development" for CS minors in winter term 2023/24 at LMU Munich, Germany. For our case study, the assignments were manually annotated with the LOs by the educators and incorporated into the used e-assessment and feedback system GATE [SOP11] to systematically match specific LOs with assignments, making these LOs visible to students within the system. The student submissions were assessed semi-automatically by combining personalized tutor feedback with automated (functional) tests performed by GATE. The course structure includes 12 topics with 6 coarse-grained course objectives (CO) (based on the module handbook [LM23]), 82 fine-grained LOs, and 17 task-specific LOs. There are 13 weekly assignments sheets comprising of 66 exercises overall, including subtasks, and an additional midterm assessment (sheet 14) with personalized feedback.

4 Results

In our ongoing work, we are developing an interactive dashboard to visualize the alignment of course LOs with the assessments tasks, highlighting their distribution across topics and how they are covered by assessment tasks. Fig. 2a illustrates a scatter plot that maps the distribution of LOs across the assignment sheets, categorized by topics. Each point on the scatter plot represents the intersection of a LO addressed in an assignment with a LO defined in a topic and is coded by different color and shapes to correspond to the lecture topics such as "Basic Concepts", "Syntax", and "Exceptions". The plot reveals the diversity and frequency of these LOs in the assignments, highlighting the emphasis placed on each topic of the course. The horizontal axis labels the assignment sheet, while the vertical axis labels the LOs categorized by topic. The diagram is designed to be interactive, enabling educators to directly access content relevant to each topic by its LOs. This granular breakdown helps to identify which topics are more or less heavily emphasized within each assignment sheet and facilitates the identification of potential gaps in the course and assignment design. LOs with a frequency of zero are those that were not addressed in any of the assignment sheets. This helps identify areas that may need greater coverage. This is the case, for instance, for topic 12, "Software Structuring & Quality Assurance " which is not addressed on any assignment sheet. The most frequently addressed topic is "Object Orientation", which is

marked with purple color and clearly visible, followed by “Data Structures” and “Control Structures”. This indicates a higher emphasis on these areas within the assignments. The diversity of topics and their respective coverage are visible at a glance, enabling educators and course designers to assess the balance of educational content across the course.



(a) Distribution of LO across the Assignments by Topics (b) Coverage of LO in Assignment Sheets

Fig. 2: Analysis of Learning Objectives Distribution and Coverage in Assignments

Fig. 2b presents a line graph that illustrates the coverage of LO in assignment sheets by showing their frequencies. The x-axis represents the different fine-grained LOs, while the y-axis indicates how often each LO is actually addressed in the assignment sheets, providing a measure how thoroughly these LOs are covered through practical assignments. LOs with a frequency of zero indicate objectives that are not (yet) addressed in the assignment sheets. Fig. 2b illustrates that many of the defined LOs lack corresponding assignments, highlighting significant gaps and potential areas for improvement in the assignment development. This visual distinction serves as an indicator of areas within the course that may require additional focus to ensure a comprehensive coverage of the intended learning outcomes. These visualizations allow for a direct comparison between the intended course coverage and the actual emphasis placed on each LO in practice. They highlight discrepancies and areas where the course implementation may need optimization to better align with the set educational goals. Again, the visualization is designed to be interactive: Hovering on any data point will reveal detailed information about that objective, including its description, frequency, and associated topic.

The alignment of the topic’s LOs and COs based on the relations in the developed tree is depicted as a heatmap in Fig. 3. This shows that each CO is addressed in the course and the course focuses on key objectives like imperative programming concepts (with 25 LOs) and developing simple algorithmic solutions (with 21 LOs). Less emphasis is placed on assessing algorithm complexity (4 LOs), highlighting areas for potential improvement. This visualization helps educators identify well-covered COs and areas needing more focus, ensuring comprehensive course coverage.

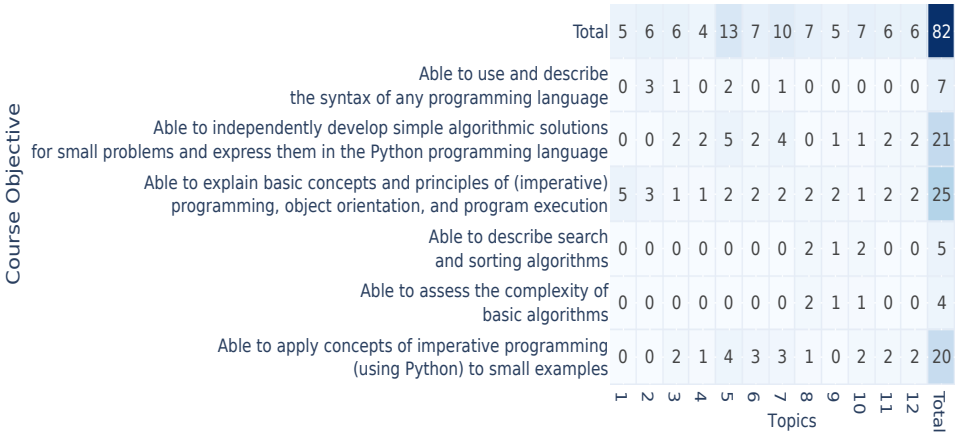


Fig. 3: Heatmap of Learning Objectives Classified by Course Objectives Aligned Across Topics

5 Discussion, Conclusions, and Future Work

It is essential to note that our investigation does not delve into the specifics of course content, which can vary widely from one instructor to another. Instead, a generic approach for the systematic application of CA and the matching of LOs to assignments is presented in this paper. This approach is not limited to computer science courses. By matching fine-grained LOs to assignments, our analysis and visualizations allow educators to unveil unmet LOs in assignments and alignment gaps in their course design. Addressing these gaps is essential for ensuring that all LOs are adequately covered, which can lead to improved educational outcomes.

In a case study, our proposed visualization highlighted disparities in the application of LOs, particularly in Object Orientation and Data Structures in an introductory programming course. While our method can make alignment issues visible, it does not outline other issues of course quality that depend on teaching methods or student engagement. The approach depends on having course learning objectives as well as fine-grained LOs and annotated assignments. Here, tool support for these tasks is needed to support educators.

Future research will extend our data model and matching algorithm to include assessment results. By reflecting on how LOs are covered and aligned with learning outcomes, we aim to identify where intervention is needed to achieve alignment between educators' goals and students' needs. Knowledge on disparities may also be used to enhance automatically generated feedback.

Acknowledgements: We thank Prof. Marie-Christine Jakobs and Dr. Philipp Wendler for their valuable support. This research is part of the project AIM@LMU funded by the German Federal Ministry of Education and Research (BMBF) under the grant no. 16DHBKI013.

References

[Ai06] Ainsworth, S.: DeFT: A conceptual framework for considering learning with multiple representations. Learning and Instruction 16 (3), pp. 183–198, 2006, DOI: 10.1016/j.learninstruc.2006.03.001.

[Ba20] Bain, A.: Addressing the Challenges of Program and Course Design in Higher Education with Design Technologies. Journal of Applied Instructional Design 9, 2020, DOI: 10.51869/92ab.

[Bi96] Biggs, J.: Enhancing Teaching Through Constructive Alignment. Higher Education 32, pp. 347–364, 1996, DOI: 10.1007/BF00138871.

[Ch23] Chor, W. T.; Goh, K. M.; Lim, L. L.; Lum, K. Y.; Chiew, T. H.: Towards a machine learning-based constructive alignment approach for improving outcomes composure of engineering curriculum. en, Educ. Inf. Technol. 2023.

[Fe21] Fessl, A.; Maitz, K.; Dennerlein, S.; Pammer-Schindler, V.: The Impact of Explicating Learning Goals on Teaching and Learning in Higher Education: Evaluating a Learning Goal Visualization. In: Technology-Enhanced Learning for a Free, Safe, and Sustainable World: 16th European Conference on Technology Enhanced Learning, EC-TEL 2021, Bolzano, Italy, September 20-24, 2021, Proceedings. Springer-Verlag, Berlin, Heidelberg, pp. 1–15, 2021, ISBN: 978-3-030-86435-4, DOI: 10.1007/978-3-030-86436-1_1.

[HI23] Heil, J.; Ifenthaler, D.: Online Assessment in Higher Education: A Systematic Review. Online Learning 27 (1), 2023, DOI: 10.24059/olj.v27i1.3398.

[Ki20a] Kiesler, N.: Towards a Competence Model for the Novice Programmer Using Bloom's Revised Taxonomy - An Empirical Approach. In: Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education. ITiCSE '20, Association for Computing Machinery, Trondheim, Norway, pp. 459–465, 2020, ISBN: 9781450368742, DOI: 10.1145/3341525.3387419, URL: https://doi.org/10.1145/3341525.3387419.

[Ki20b] Kiesler, N.: Zur Modellierung und Klassifizierung von Kompetenzen in der grundlegenden Programmierausbildung anhand der Anderson Krathwohl Taxonomie. CoRR abs/2006.16922, 2020, arXiv: 2006.16922, URL: https://arxiv.org/abs/2006.16922.

[Ki24] Kiesler, N.: Modeling programming competency: A qualitative analysis. Springer Nature, Cham, Switzerland, 2024.

[LM23] LMU München: Modulhandbuch Nebenfach: Informatik im Umfang von 60 ECTS-Punkten für Bachelorstudiengänge, Version 2023-07-15, 2023, URL: https://www.ifi.uni-muenchen.de/studium/studiengaenge/bachelor/informatik_nf_bachelor/2023-07-15-ba-nf-informatik-60-ects_modulhandbuch-z-psto-2022-bei-beginn-im-wise.pdf.

[Om20] Omer, U.: Cognitive Learning Analytics Using Assessment Data and Concept Map: A Framework-Based Approach for Sustainability of Programming Courses. Sustainability 12, 2020, DOI: 10.3390/su12176990.

[Om23] Omer, U.; Tehseen, R.; Farooq, S.; Abid, A.: Learning analytics in programming courses: Review and implications. Education and Information Technologies 28, pp. 1–47, 2023, DOI: 10.1007/s10639-023-11611-0.

[SB22] Strickroth., S.; Bry., F.: The Future of Higher Education Is Social and Personalized! Experience Report and Perspectives. In: Proceedings of the 14th International Conference on Computer Supported Education - Volume 1: CSEDU. INSTICC, SciTePress, pp. 389–396, 2022, DOI: 10.5220/0011087700003182.

[SOP11] Strickroth, S.; Olivier, H.; Pinkwart, N.: Das GATE-System: Qualitätssteigerung durch Selbsttests für Studenten bei der Onlineabgabe von Übungsaufgaben? In: Proc. DeLFI. Gesellschaft für Informatik e.V., Bonn, pp. 115–126, 2011.

[SS22] Strickroth, S.; Striewe, M.: Building a Corpus of Task-Based Grading and Feedback Systems for Learning and Teaching Programming. International Journal of Engineering Pedagogy (iJEP) 12 (5), pp. 26–41, 2022, DOI: 10.3991/ijep.v12i5.31283.

[St22] Striewe, M.: Where Does All the Data Go? A Review of Research on E-Assessment Data. In: Proceedings of the 14th International Conference on Computer Supported Education. SCITEPRESS - Science and Technology Publications, pp. 157–164, 2022, DOI: 10.5220/0010879700003182.

[St24] Strickroth, S.: Scalable Feedback for Student Live Coding in Large Courses Using Automatic Error Grouping. In: Proceedings of the 2024 Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2024), July 8–10, 2024, Milan, Italy. Association for Computing Machinery, 2024, DOI: 10.1145/3649217.3653620.

Design, Realisierung und Evaluation von Scaffolds zur Unterstützung von Lernprozessen im Hochschulkontext

Lea Reinartz [1], Sarah Schneeweiß [1] und Andreas Harrer [1]

Abstract:

Die Beherrschung systematischer Vorgehensmodelle gehört in zahlreichen wissenschaftlichen Disziplinen zu deren wesentlichen Methodenkompetenzen. Unerfahrene Studierende haben bei der eigenständigen Anwendung häufig Probleme. Zu diesem Zweck entwickeln wir im Forschungsprojekt IPPOLIS einen Unterstützungsansatz, der die Ideen des Scaffoldings, im Sinne eines Stützgerüsts, mit adaptivem Feedback kombiniert. Dieser Ansatz wurde in einer ersten Feldstudie im Rahmen einer Softwaretechnik-Veranstaltung für die systematische Modellierung von Diagrammen in der Unified Modeling Language (UML) praktisch erprobt und evaluiert. Dabei haben wir wertvolle Erkenntnisse erhalten, wie hilfreich das Scaffolding-Feedback wahrgenommen wird. Erweiterungen in den Bereichen der Feedbackgestaltung und Aktionsanalyse mit Methoden künstlicher Intelligenz (KI) werden für weitere Studien skizziert.

Keywords: Scaffolding, Lernprozessunterstützung, Feldstudie, adaptives Feedback

1 Einleitung

In zahlreichen wissenschaftlichen Disziplinen gibt es systematische Vorgehensweisen, die entweder als theoretische Modelle begründet werden können oder die Praxiserfahrung und etablierte Methodik darstellen. Beispielsweise gibt es im Software-Entwurf sogenannte Vorgehensmodelle, wie das Wasserfallmodell, Spiralmodell, V-Modell und die Familie der agilen Vorgehensmodelle, die jeweils unterschiedliche Phasen und Praktiken samt organisatorischer Aspekte in Bezug setzen. Die Beherrschung solcher Vorgehensweisen gehört wesentlich zu den Methodenkompetenzen der jeweiligen Disziplin. Für unerfahrene Studierende ist die praktische und eigenständige Beherrschung dieser Vorgehensweisen oft eine große Herausforderung, auch wenn projektorientierte Lehre dabei hilft, die Lücke zwischen Theorie und Praxis zu verringern. Mit zunehmender Kompetenz wird der Bedarf an Unterstützung geringer und diese möglicherweise sogar als aufdringlich empfunden.

An dieser Stelle kann die didaktische Theorie und Praxis mit dem sogenannten „Scaffolding“, im Sinne eines Stützgerüsts für Lernaktivitäten, einen Ansatz bereitstellen, der bereits seit einigen Jahrzehnten etabliert und gründlich untersucht ist [WBR76] [Qu04] [Qu21]. Bei zunehmender Kompetenz wird das Stützgerüst schrittweise entfernt, was als

1 Fachhochschule Dortmund, Fachbereich Informatik, Emil-Figge-Strasse 42, 44227 Dortmund, Deutschland,
lea.reinartz@fh-dortmund.de, https://orcid.org/0009-0000-8014-7624;
sarah.schneeweiss@fh-dortmund.de, https://orcid.org/0009-0002-4281-8356;
andreas.harrer@fh-dortmund.de, https://orcid.org/0009-0001-1076-7964

doi:10.18420/delfi2024_35

„Fading“ bezeichnet wird. Ähnliche Ansätze finden sich im Bereich der lernunterstützenden Computersysteme ursprünglich unter den Begriffen „adaptive Lernunterstützung“ und „Intelligent Tutoring Systems“ (ITS), heute zumeist als spezieller Fall eines „Learning Analytics“ mit unmittelbar geschlossener Feedbackschleife. Schon vor Jahrzehnten wurden dazu Methoden der künstlichen Intelligenz eingesetzt, die eher dem Bereich des Knowledge Engineering als den heute weitgehend dominierenden Machine Learning, Deep Learning und neuen neuronalen Netzwerk-Architekturen zuzuordnen sind. Es bietet sich daher an, die Eignung traditioneller und neuer KI-Methoden für die adaptive Lernunterstützung der eingangs beschriebenen Vorgehensweisen zu prüfen.

2 Projektvorstellung „Scaffolds“

Im BMBF-geförderten Projekt IPPOLIS steht die praktische Unterstützung von Lernaktivitäten durch KI-Methoden im Hochschulbereich im Zentrum. Dazu gehören beispielsweise Analyse und Feedback zu schriftlichen Arbeiten, Erstellen und Lösen von fallbasierten Statistikaufgaben und Bereitstellung eines Chatbots für Studienfragen. In unserem Teilprojekt „Scaffolds“ befassen wir uns mit der Thematik der adaptiven Unterstützung von Lernenden bei der Nutzung von Prozessmodellen, die in den jeweiligen Fachdomänen zur guten Praxis gehören, von Anfängern aber oft nur unzureichend selbstkompetent verwendet werden. Zu diesem Zweck verknüpfen wir in unserem Teilprojekt Aspekte der Prozessmodellierung mit der Analyse von Aktionen der Nutzenden und adaptiver Generierung von Feedback nach Prinzipien von Scaffolds und Fading. Um den Ansatz praktisch zu erproben und auch über enge Domänengrenzen hinaus zu transferieren, sind Anwendungsfälle in den Disziplinen (Software) Engineering, Data Science und wissenschaftlichem Arbeiten geplant.

In diesem Beitrag beschränken wir uns auf Aktivitäten im Fall „Software Engineering“ für den ersten Praxiseinsatz im realen Feld einer Lehrveranstaltung und den daraus gewonnenen Erkenntnissen für den zweiten Praxiseinsatz im nächsten Jahresturnus. Im Fokus der ersten Iteration steht dabei die Untersuchung des Feedbacks, das zunächst objektivierbare Metriken und Richtlinien der Lehrveranstaltung zu UML-Diagrammen umfasst, und dessen Wirkung bezüglich Usability und empfundener Nützlichkeit. Die Generierung adaptiven und korrektiven Feedbacks sowie Auswirkungen auf den Lernerfolg werden hier noch nicht verfolgt und Bestandteil der weiteren Iterationen sein (siehe Abschnitt 5).

3 Technische Umsetzung eines Scaffolding-Systems

Unser Scaffolding-System wurde als Webanwendung umgesetzt. Dafür wurde aus Gründen der Datensouveränität eine Fat-Client Architektur gewählt, damit sämtliche Daten der Nutzenden lokal auf deren Endgerät bleiben. Daten werden zunächst ausschließlich clientseitig gespeichert und auch Funktionen mithilfe von JavaScript im Client ausgeführt. Jegliche Eingaben und Prozessfortschritte der Nutzenden werden im Local Storage des Browsers

über Sessions hinweg gespeichert. Außerdem wird optional das Verhalten der Nutzenden geloggt und ebenfalls im Local Storage des Browsers gespeichert. Erst nach explizitem Export der Nutzenden werden diese Daten verschickt.

Die Prozessmodelle werden explizit in eigenen Dateien spezifiziert und sind somit ohne Code-Änderungen austauschbar. Dadurch ist die Anpassung eines Prozesses und die Erweiterung des Systems um beliebige weitere Prozessmodelle jederzeit möglich. Die Modelle können über Verzweigungen verfügen und auch in mehreren Iterationen durchlaufen werden. Die grafische Bedienoberfläche wird aus der Spezifikation dynamisch generiert.

4 Pilotstudie „Software Engineering“

In der ersten Pilotstudie im Wintersemester 2023/2024 wurde die Umsetzung eines Engineering Prozesses evaluiert. Dazu wurde das Scaffolding-System im Rahmen eines Feldversuchs in einer Lehrveranstaltung des dritten Bachelor-Semesters mit ca. 120 Informatik-Studierenden eingesetzt. Ziel dieser Studie war es herauszufinden, ob die Unterstützung genutzt wird und welche Aspekte davon als hilfreich erachtet werden.

4.1 Einbindung in den „Softwaretechnik 1“ Kurs

Der erstellte Prozess ist ein Software-Engineering-Prozess, der den Abschnitt „Modellierung“ in der Veranstaltung „Softwaretechnik 1“ unterstützt. Die Anwendung wurde den Studierenden zu Beginn des Veranstaltungsabschnitts vorgestellt und zusätzlich in den Praktika als Werkzeug zur Unterstützung beworben. Der spezifisch an die Lehrveranstaltung angepasste Prozess besteht aus zwei Prozessschritten mit mehreren Unteraktivitäten. Dadurch bedingt liegt der Fokus in diesem Scaffold zunächst mehr auf der inhaltlichen Seite als bei der Prozessunterstützung. Die Studierenden wurden bei der Erstellung von UML-Diagrammen und im zweiten Schritt bei der Konsistenzprüfung zwischen Diagrammen unterstützt.

4.2 Aufbau der Oberfläche des Scaffolding-Systems

Auf der Hauptseite der Anwendung wird der Prozess dargestellt. Jeder Prozessschritt ist ausklappbar und enthält mehrere Aktivitäten, die benötigt werden, um den Prozessschritt abzuschließen (siehe Abbildung 1). Die Aktivitäten können zweimal abgehakt werden, zunächst links, wenn die Studierenden die Aktivität abgeschlossen haben und zusätzlich rechts, wenn die Ergebnisse der Aktivität von einer Lehrperson überprüft wurden. Bei den jeweiligen Prozessschritten wird mithilfe einer zirkulären Fortschrittsanzeige dargestellt, wie viel dieses Schrittes jeweils bereits erledigt wurde.

Jede Aktivität verfügt über eine Unterseite, auf der Hilfestellungen zur Bewältigung dieser Aktivität sind. Auf diesen Unterseiten wird als erste Stütze die Aktivität mit den

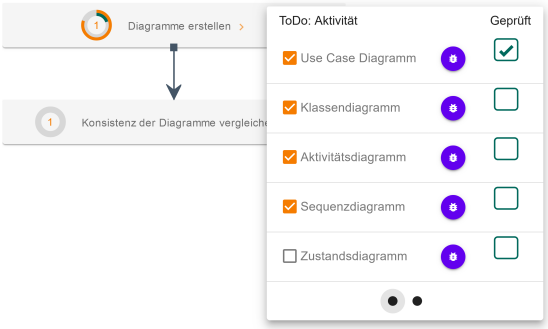


Abb. 1: Prozessschritte und Aktivitätsfortschritte im Softwaretechnik-Scaffolding-System

wichtigsten Punkten erklärt. Eine weitere Stütze ist ein Algorithmus, der die Diagramme prüft und dazu Feedback gibt. Dazu werden die erstellten UML-Diagramme als XML-Datei hochgeladen. Der JavaScript-Algorithmus durchläuft das dazugehörige Document Object Model und prüft im ersten Prozessschritt auf Richtigkeit im Sinne der Vorlesung und der Rahmenbedingungen der Praktikumsaufgaben. Im zweiten Prozessschritt werden zwei Diagramme auf ihre Konsistenz überprüft. Beispielsweise wird bei einem Klassen- und einem Sequenzdiagramm kontrolliert, ob alle benutzten Methoden des Sequenzdiagramms aus dem Klassendiagramm stammen. Das Feedback ist bisher bewusst simpel gehalten, um im ersten Durchlauf die Prinzipien von Prozess-Scaffolding grundsätzlich zu testen. In Abbildung 3 ist auf der linken Seite das simpel gehaltene Feedback zu sehen.

4.3 Datenerhebung und -auswertung

Zum Ende der Pilotstudie wurde das Scaffolding-System von den Studierenden mithilfe eines schriftlich auszufüllenden Fragebogen bewertet. Dieser überprüft, inwiefern die Anwendung als Gesamtes und ihre einzelnen Funktionalitäten verwendet, verstanden und als hilfreich wahrgenommen werden. Von 119 ausgefüllten Fragebogen resultieren im Folgenden N = 65 Studierende, die die Anwendung verwendet haben. Studierende, die die Anwendung nicht verwendet haben, geben an, dies zukünftig nachholen zu wollen, dass lediglich ein anderer Teil der Arbeitsgruppe diese verwendet hat oder, dass ihnen die Anwendung zuvor nicht bekannt war. Unter den 65 Nutzenden wird das Feedback innerhalb des Scaffolding-Systems insgesamt als verständlich und sehr gut anwendbar bewertet. Gleichzeitig wird eine strukturiertere Anordnung und mit Farben gestaltete Rückmeldung gewünscht.

In Abbildung 2 ist die Bewertung der gesamten Anwendung als Hilfestellung anhand dreier - durch den User Experience Questionnaire (UEQ) [LHS08] inspirierten - Kategorien in Form eines „Diverging Stacked Bar Chart“ [HR14] dargestellt. Beobachtungen ohne Angabe wurden jeweils herausgenommen. In der Darstellung sticht deutlich heraus, dass die Anwendung ihr Ziel als Unterstützungshilfe bereits erreicht. Die davon abweichende

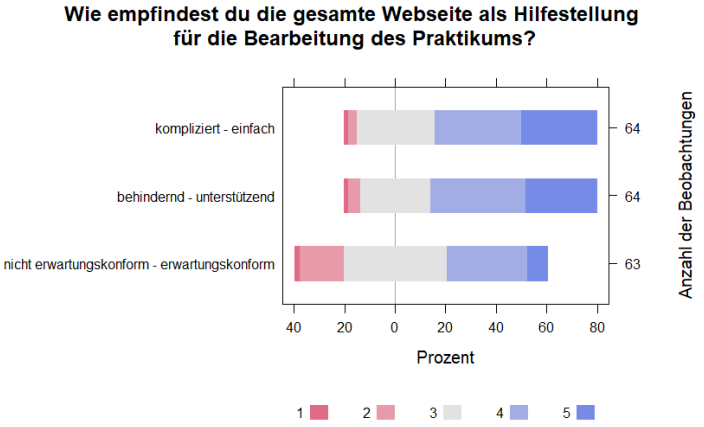


Abb. 2: Diverging Stacked Bar Chart für Bewertung der Hilfestellung (5-Punkt-Skala)

Bewertung im Bereich Erwartungskonformität führen wir anhand ergänzender Freitextantworten darauf zurück, dass sich die Unterseiten mit inhaltlichen Erklärungen und der Diagrammüberprüfung hinter dem Begriff „Hilfestellung“ verbergen, welcher bei den Nutzenden teils andere Assoziationen geweckt hat.

4.4 Ausblick auf den zweiten Durchlauf

Das Scaffolding-System wird im Wintersemester 2024/2025 in der gleichen Lehrveranstaltung erneut eingesetzt werden. Um das System innerhalb der Veranstaltung bekannter zu machen, wird es bereits in die Aufgabenstellungen des Praktikums eingebunden werden. Darüber hinaus werden wir eine Anleitung bereitstellen, um verstärkt auf die verschiedenen Funktionalitäten der Anwendung aufmerksam zu machen. Wesentliche Veränderungen im System nehmen wir beim - zuvor bewusst sehr simpel gehaltenen - Feedback vor, wie im folgenden Abschnitt erläutert wird.

5 Nächste Schritte zur Feedback-Gestaltung

Die Gestaltung von Analyse und Feedback fand ganz bewusst im ersten Praxiseinsatz auf recht allgemeiner und durch objektivierbare Metriken, wie Anzahl der Use Cases und Klassen in den jeweiligen Diagrammen sowie deren Konsistenz statt. Zum einen konnten wir dadurch eine Baseline für Feedbackqualität und -nützlichkeit festlegen, zum anderen wollten wir jegliche Intervention in die Autorität der Lehrenden vermeiden. Allerdings wurden bereits in der Evaluation Anregungen wie farbliche Kategorisierung des Feedback geäußert bzw gewünscht. Nach der Konzeption Patrick Jermans der unterschiedlichen

Ebenen von regulatorischem Feedback [Je04] betrachten wir unser bisheriges Feedback als „Mirroring“ und planen als Schritte für den nächsten Praxiseinsatz die Gestaltung und Ausarbeitung des Feedbacks in Richtung meta-kognitiver Unterstützung (von uns im Weiteren als „normatives Feedback“ bezeichnet [HMD09]) bis hin zum „Guiding“ in Form von konkreten Verbesserungsschritten.

Die Abbildung 3 stellt für ein UML-Diagramm das neutrale bisherige Mirroring-Feedback (links) und einen Entwurf für ein Dashboard mit normativem Feedback (rechts) gegenüber: die visuelle Gestaltung von wünschenswerten und nicht wünschenswerten Metrikbereichen zusätzlich zu den analysierten Werten ist ein erster Schritt, das Feedback reicher zu gestalten. Das Dashboard mit den verschiedenen Symbolen wird als Grundlage für die neue Feedbackgestaltung gewählt, um den Studierenden einen schnellen Überblick zu ermöglichen. Ein weiterer Schritt wäre die Generierung konkreter Handlungsempfehlungen an die Lernenden auf der Ebene des Guiding, wie z.B. „Das Hinzufügen von Akteuren würde klarer machen, wer in diesen Use Cases als Nutzende handeln.“

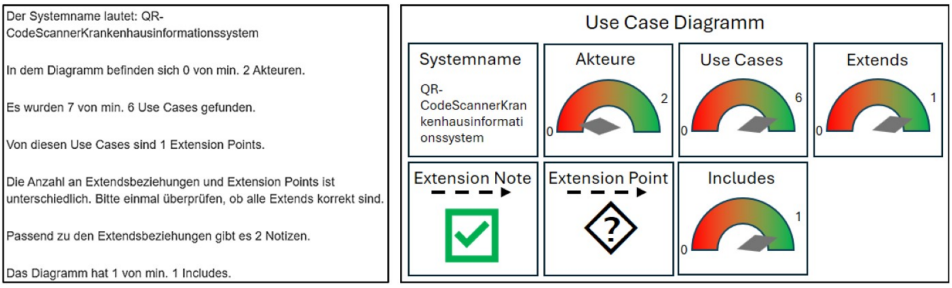


Abb. 3: Feedback für UML-Analysen: Mirroring (links) und normatives Feedback (rechts) - jede auswertbare Information ist als Widget mit normierter Visualisierung repräsentiert

6 Fazit und Ausblick

In diesem work-in-progress Bericht haben wir unseren Ansatz eines Scaffoldings von systematischen Vorgehensweisen bzw. Lernprozessen vorgestellt und erste Ergebnisse und Erkenntnisse aus unserem ersten Feldversuch im Rahmen einer Softwaretechnik-Lehrveranstaltung präsentiert. Wesentliche Resultate bisher und Ausgangspunkt unserer weiteren Schritte sind die praktische Realisierung als Webanwendung, die flexible Austauschbarkeit der Prozessdefinitionen und die Erkenntnisse zur Feedbackgestaltung. Für unser Projektziel der adaptiven Unterstützung und insbesondere des Fadings der Scaffolds sind verallgemeinerbare Modelle für die Festlegung des aktuellen Feedbacklevels geplant, die Aktionen der Nutzenden auf Kompetenzlevel abbilden (je höher die Kompetenz, desto geringer das Level an Feedbackunterstützung bzw. stärkeres Fading). Für diese Aktionsanalyse und Kompetenzmodellierung werden wir stärkere KI-Methoden einsetzen und explorieren, ob klassische Methoden aus dem Bereich ITS (wie „model tracing“ [Bl07]) oder aktuelle Forschungslinien der KI für unsere Zwecke besser geeignet sind.

Literatur

[Bl07] Blessing, S.; Gilbert, S.; Ourada, S.; Ritter, S.: Lowering the bar for creating model-tracing intelligent tutoring systems. In. Bd. 158. Frontiers in Artificial Intelligence and Applications, IOS Press, S. 443–450, 2007.

[HMD09] Harrer, A.; Martinez-Mones, A.; Dimitracopoulou, A.: Users' data: Collaborative and social analysis. In (Balacheff, N.; Ludvigsen, S.; de Jong, T.; Lazonder, A.; Barnes, S., Hrsg.): Technology-Enhanced Learning – Principles and Products. Springer, Berlin, Kap. 11, S. 175–193, 2009.

[HR14] Heiberger, R. M.; Robbins, N. B.: Design of Diverging Stacked Bar Charts for Likert Scales and Other Applications. en, Journal of Statistical Software 57/5, 2014, ISSN: 1548-7660.

[Je04] Jermann, P.: Computer Support for Interaction Regulation in Collaborative Problem-Solving, Diss., Universite de Geneve, 2004.

[LHS08] Laugwitz, B.; Held, T.; Schrepp, M.: Construction and Evaluation of a User Experience Questionnaire. In (Holzinger, A., Hrsg.): HCI and Usability for Education and Work. Bd. 5298, Springer, Berlin, Heidelberg, S. 63–76, 2008, ISBN: 978-3-540-89349-3.

[Qu04] Quintana, C.; Reiser, B.; Davis, E.; Krajcik, J.; Fretz, E.; Duncan, R.; Kyza, E.; Edelson, D.; Solloway, E.: A Scaffolding Design Framework for Software to Support Science Inquiry. The Journal of the Learning Sciences 13/3, S. 337–386, 2004.

[Qu21] Quintana, C.: Scaffolding Inquiry. In (Duncan, R. G.; Chinn, C. A., Hrsg.): International Handbook of Inquiry and Learning. 1. Aufl., Routledge, New York, NY: Routledge, 2021., S. 174–188, 2021, ISBN: 978-1-315-68577-9.

[WBR76] Wood, D.; Bruner, J. S.; Ross, G.: The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry 17/2, S. 89–100, 1976.

Challenges of Computer Science Tutors in Their Teaching: How can Educational Technology help?

Armin Egetenmeier[1] and Sven Strickroth[1]

Abstract: Tutorial sessions led by student tutors are an essential part of teaching in large classes, e. g. in Computer Science (CS), to provide personal support and to guide learning. Although universities offer training and educators conduct regular meetings to assist tutors, the use of technology in their teaching practice is mostly unknown. This paper presents exploratory insights from focus group interviews with CS tutors to reveal tutors' challenges, needs and current usage of educational technology in their teaching. The results show a lack of use of technology by tutors to improve their teaching or to gather useful information about their group in a structured way. The potential of educational technology to enhance tutorial sessions remains largely untapped. Possible technology-driven approaches and ideas proposed by tutors are discussed. Based on these initial results, the goal is to develop tailored technology-enhanced support and provide relevant information to tutors in the future.

Keywords: educational technology, technology-enhanced learning, tutor, computer science

1 Introduction

Educational technologies play an important role in higher education institutions (HEI), serving both administrative and pedagogical functions like learning management systems (LMS). For instance, these technologies are used by lecturers to enable (digital) teaching formats, or learning at scale [SB22]. Technology support is necessary, since large classes with hundreds of students are common in HEI at the beginning of degree courses such as Computer Science (CS) [Ri21]. Therefore, many Technology-Enhanced Learning (TEL) approaches such as Audience Response Systems (ARS) [FM06; KS17; MB19] or e-assessment systems [SS22] have been developed to gain insight into the learning process, stimulate interaction, or facilitate immediate feedback for the primary stakeholders (educators and learners). In addition, lectures are often accompanied by smaller, more interactive tutorial sessions, which can be attended by usually up to 35 students. In Germany, tutors are often responsible for planning and conducting weekly sessions throughout the semester (cf. [KO23; RG18]). These casual, short-term "(undergraduate) teaching assistants" [Da17; Mi19] or "learning assistants" [Lo20; Se17] are usually senior students, who already passed the course exam and are often hired based on the achieved grades or on personal recommendation [PTR16].

Tutoring has a positive effect on learning outcomes and has proven to be helpful in large classes [Le11]. Despite their crucial role in supporting large lectures, the perspective of tutors regarding helpful technology for their teaching task is often untapped. Moreover, tutors are

1 LMU Munich, Institute for Informatics, Oettingenstr. 67, 80538 Munich, Germany,
armin.egetenmeier@ifi.lmu.de; sven.strickroth@ifi.lmu.de, https://orcid.org/0000-0002-9647-300X

doi:10.18420/delfi2024_36

rarely involved in the decision-making process regarding the selection and implementation of technology. Instead, they are typically provided with tools and systems readily available (like LMS) or that the lecturer considers necessary (e. g., a specific IDE) or useful, such as e-assessment systems that offer opportunities to review, grade and give feedback (at scale). Many of these systems focus on the distribution of work among the tutors and to simplify tasks. The results derived from e-assessment systems are mainly for students and educators, while not being tailored or even inaccessible for tutors. Although the tutors often build the backbone of teaching (such that some students see them as the main teachers [RS20]), the benefits of utilizing these systems in their teaching practice is unknown or overlooked. There is not much knowledge about what technology the tutors actually use in tutorial sessions and what tools (or features) are missing to overcome problems in their teaching. Ley et al. [Le23] stated that research should focus on the interplay of technology and the teacher to guide student learning. Also, it is unclear what information are provided to tutors and what would be helpful (or desired). This knowledge gap about the tutors' view on technology motivated this research. The goal is to provide tutors with useful tools to enhance their teaching. This led to the following questions: (RQ1) What tools are used by tutors? (RQ2) What challenges do tutors face? (RQ3) How and from where do tutors get information about the group? (RQ4) What features or tools could be helpful in tutors' teaching practice?

2 Related research

Various technologies have been researched and developed to support educators and students. LMS are widely used in practice. Current research aims to optimize administrative tasks, to incorporate learning analytics (LA) [ZGS20] or AI [AA20], and to support (interactive) teaching designs [HMB19]. Additionally, various e-assessment systems have been developed to support educators and students, for example, to assess programming assignments [KJH18; SS22]. Current research investigates how to relieve teaching assistants from laborious correction (e. g., by providing automatic pre-corrections [SH22]), or to automatically generate assignments and feedback (e. g., by using AI [AKS24]). Tools usually provide overviews of awarded points to students. Furthermore, some LMS and e-assessment systems are incorporating LA and provide dashboards [Ji18]. Most dashboards seem to target teachers and students. However, it often remains unclear whether student tutors are also considered teachers, what information is provided to them about "their" students (if any [Sc17]), and how this impacts the teaching practice (if given to tutors [Ol21; Wa19]). Finally, audience response systems are a proven approach to activate students and to gain insight into their learning during lectures [Ló22; MB19]. Current research mainly focuses on investigating [WS20] and improving their impact in large classes (e. g., using gamification [Ló22]) and specialized approaches such as showing common errors of in-class programming tasks [St24]. There also is research on smaller lectures comprising 20 to 40 students (cf. [WS20]), however, no research on tutorial sessions in recent years is known to the authors.

Riese et al. [Ri21] investigated challenges of CS tutors in three different European universities. They identified five main challenges: *becoming a professional teaching assistant (TA)*,

student focused challenges, *assessment*, *defining and using best practice*, and *threats to best practice*. Danielsiek et al. developed a competence model for tutors to implement a training program [Da17]. In some cases, the use of technology is also part of the program, but often limited to the use of an LMS [RG18] or communication tools [RM09]. There is also research on tutorials, focusing on aspects such as the effectiveness of teaching strategies [Mc16; MRG18] or the implementation, distribution and improving of training programs [Mi19; RG18], but again technology such as TEL and LA only play a minor role (cf. [Ol21]).

To sum up, research on TEL is mostly driven by teaching staff and focuses on optimizing their teaching. Training programs to support novice tutors are widely discussed in literature. However, TEL research with tutors mostly focuses on relieving tutors from tedious (grading) work, but rarely on technology use for (preparing) their tutorials. This leads to the research gap: What tools/information do tutors actually use and wish for (preparing) their tutorials?

3 Study Design and Methodology

At the beginning of winter term 2023/24, three focus groups (with overall 19 tutors from three different first-year CS lectures: group G1: practical session, G2 and G3: introductory courses) were held at LMU Munich by one researcher. All participants were recruited through personal contact. Participation in the focus groups was voluntary and took about one hour. The focus groups were held in German, recorded after consent was given, and transcribed based on personal notes from the researcher and the recordings in a simplified form (without distinguishing individuals). The answers were analyzed and incrementally summarized using Mayring's thematic analysis method [Ma14]. Before the interview, all participants answered a small paper-based questionnaire asking for general information (e. g., age) and contextual aspects of the work as a tutor (including teaching experience, tasks, given support, media use). The interview began with asking the tutors to describe their teaching practices along a typical course of a tutorial session. Next, "pain points" were explored. Then, the focus was on tutors' knowledge on their student group and what would be helpful to know (particularly regarding their preparation). Lastly, the digital tools used and what would be useful were discussed. The interview concluded with an open discussion.

4 Results

Most of the tutors are novice tutors, since they have no or limited (one to two semesters) experience in teaching. Only three persons participated in a training program. Six persons are more experienced due to recurring tutoring in the same lecture. Remarkable is that all but one of the tutors are attending tutorial sessions themselves as students in other courses. Their duties include (digital) communication with the students, helping with the exams (proctoring, grading), and, in some cases, reviewing homework. Specifications regarding teaching are at a minimum by the lecturers like discussing the solutions instead of just handing them out or

presenting these. Typically, the students enroll themselves to one tutorial session (with up to 18 students in G1, about 30 to 40 in group G2 and up to 60 students in G3). The majority of the tutorials are in-person, with only one being fully online. Usually, tutors receive learning material from the lecturer (e. g., slides) and exercise sheets (with solutions). There are also weekly meetings with the educators to give feedback on the previous week's tutorials and receive helpful information on the next week's session. Regarding tools (RQ1), the use of laptops or tablets is common to present (partly self-developed) material (e. g., with PowerPoint or OneNote). In addition, different tools were mentioned as relevant for tutors to exchange data (e. g., Moodle), to communicate (e. g., chat or conferencing tools), specific tools for programming (e. g., IDEs), and to give feedback (e. g., via Moodle or GitLab). Groups G2 and G3 named specific e-assessment systems for reviewing programming tasks.

Tutors face different challenges in relation to their tutorial sessions (RQ2): During the sessions, the students hardly ask questions or provide any feedback. One tutor (G2) stated, *"you can't tell anything from facial expressions – everyone is just 'there', no feedback as to whether the content was understood"*. Each focus group assumed that the students are timid or anxious to ask potentially "stupid or too simple" (G3) questions, maybe due to a lack of familiarity with peer students (G2) or since the concept of a tutorial session is not fully understood yet (G3) – particularly at the start of the term. One tutor (G2) added that phrasing a concrete question may be overwhelming ("Sometimes it is unclear what the students want to express with the question"). Due to the missing feedback from the students, tutors need to constantly nudge participation ("after each part, ask whether there are any comments or questions", G3). Different and unknown levels of prior knowledge, programming experience, and preparation of the current exercises make the teaching quite challenging for the tutors. Therefore, most tutors begin their tutorial sessions by asking if the audience has any questions ("which is rarely the case", G2), followed by working with active individuals. They circulate among the students, pose and answer individual questions on a small scale, collect and resolve recurring issues for the whole group and keep the group on task. This approach is seen to encourage interaction among the students, but keeps the tutor constantly busy. The sessions of the practical provide some specific challenges such as reviewing code from each student upfront with diverse coding approaches. The complexity of the coding task itself makes the preparation difficult.

Information about the group is considered essential by tutors (RQ3). At the beginning of the semester, all tutors have only little knowledge about the student group (e. g., study course, semester), mostly through a brief introduction in the first session (G1, G2), or over time by e. g., informal communication after the tutorials (G3). Tutors become familiar with individual students, but rarely in a structured way. Almost all tutors assumed that this knowledge can be beneficial, but is not sufficient for their teaching practice (G1, G2). Information about the groups such as the major/minor (e. g., to group people or to show connections to upcoming lectures, G2) or the current semester (e. g., to estimate prior knowledge or to tailor material, G2) is assumed to be helpful. Insights into the students' learning progress and understanding or common issues on the current exercise sheets would

be helpful to structure the session, but are not available (G2). Important, recurring issues from the lecture are provided to some tutors (G1). Additional, unresolved questions collected in the lecture may be useful to initiate discussion during the tutorial session, but are not shared (G2). Thus, tutors rely on their own experience or the information from the weekly meeting with the educators for their preparation.

Finally, ideas were discussed to overcome challenges in the focus groups (RQ4). These can be divided into three categories: improvements to existing systems, incorporation of "new" tools, and ideas or new features on unspecified tools: First, there were improvements for the used e-assessment system and communication tools. Group G2 suggested for the e-assessment system to make relevant information easier accessible (e. g., detailed evaluation on MC tasks, filtering/highlighting of their group) or to provide tailored ARS features for tutorials (e. g., collecting and discussing existing MC questions). G1 mentioned that the main communication system (Zulip chat) was hardly used and the students may miss relevant information or news. Thus, this system could be expanded by default to provide unseen messages via mail (e. g., as a weekly summary) or through an integrated messenger platform like Telegram (G3). Second, "new" (commercial) tools were proposed by the tutors to make communication easier and more accessible, such as using instant messaging apps like Discord or WhatsApp, but provided by the HEI (G1), or the collection of (anonymous) feedback using GoogleForms or Etherpad (G3). ARS tools such as Kahoot were also mentioned here (G3), however, the additional effort for using them in the tutorials was seen critically by the group. Third, general ideas of the tutors are presented: A recurring idea in all groups was to have a tool to easily collect common questions (e. g., asked in mails to build a FAQ) or anonymous feedback, since only constant asking and nudging leads to increased activity and discussion. One tutor (G2) expressed concerns about such a tool *"because it is difficult to write down questions, especially if the basics are not understood"*.

5 Discussion, Conclusions, and Outlook

In this paper, focus groups were conducted with CS tutors (senior student teaching assistants) in which their technology use, challenges, information need as well as ideas on potential helpful tools for their teaching practice were explored. In terms of the use of digital tools (RQ1), most tutorial sessions are conducted face-to-face and supplemented selectively with digital systems to communicate or present material. Apart from that, tools selected by the lecturer such as the LMS and e-assessment systems are used to give feedback and to review assignments. Specialized TEL approaches such as ARS, which are quite common in lectures, are not used in CS tutorial sessions. A reason may be the smaller number of students in the tutorials (relative to the lecture), time constraints, and a lack of support or provision of these tools by the HEI. The question on the challenges (RQ2) revealed that tutors face similar problems regarding the heterogeneity and passivity of students compared to the teaching staff in larger lectures as well as the lack of feedback (especially from freshmen). Similar results about learner related challenges and effective teaching practice

were reported in [Ri21]. Although information about the group is considered helpful (RQ3), tutors only have little and unstructured knowledge about their students primarily gathered by themselves during the tutorials. Due to various duties, tutors barely have the time to collect more. However, more and aggregated information about their students and their performance (apart from individual results collected within e-assessment systems) is considered helpful for preparing or structuring the tutorial sessions. The tutors had specific ideas on integrating existing tools into their teaching and to enhance already used tools to support them (RQ4). Interestingly, the tutors in the focus groups already knew tools such as Etherpad and ARS (Kahoot), e. g., from other courses they take as a student, that are commonly recommended in higher education pedagogy, however, they do not use them and expect the effort as too high. Also, tutors named specific tools that may be more convenient for younger people to communicate and wished comparable privacy conform offers from the HEI.

The tutors expressed the need for more detailed insights into the results of the exercise sheets, such as how the MC tasks were solved and what were the most common errors on Java programming tasks (cf. [St24]) in their group. Discussed and open issues from the lecture are also beneficial for tutors. The focus groups revealed that (experienced) tutors have a clear opinion on suitable technology and how it can be used in a learning design. This opens up the perspective on Learning Analytics in courses. Here, tutors (as users) need to be acknowledged and honored as central stakeholders (cf. [Ol21; SKB20]). This would also allow to link the tutorials more closely to the lectures than just by having “the same topics”.

A review of Mirza et al. shows that most tutor training programs focus (only) on pedagogy [Mi19]. However, a combination of pedagogy and technology is useful and necessary [Le23] to address some of the tutors’ challenges. For example, ARS may be more easily integrated into tutorials if not every tutors needs to invent questions. This task can be distributed among (all) tutors and/or questions are provided and discussed with teaching staff. Furthermore, this also outlines the need of HEIs to improve the visibility, accessibility and usability of offered/available tools to tutors and also to provide appropriate training (material).

This is a first explorative study based on three different focus groups from the same faculty. Overall 19 tutors provided an authentic and reflected overview on their practice, challenges, and opinions. Tutors may only have limited knowledge about existing teaching tools and potentially available information about learners. Yet, the results show that tutors have a good sense of what they may need and how it can help them in their practice. A larger, follow-up questionnaire based on these results is planned with tutors from a wider set of disciplines and universities. Knowing the challenges that tutors face and the support they currently receive makes it possible to identify and address their needs. In particular, tutors’ activities are of interest and need to be further investigated in order to provide them with appropriate and optimized tools and features. Further research is needed to investigate not only what tools are supportive, but also what information, and how it can be provided with ease.

Acknowledgement: This research is part of the project AIM@LMU funded by the German Federal Ministry of Education and Research (BMBF) under the grant number 16DHBKI013.

References

[AA20] Aldahwan, N. S.; Alsaeed, N. I.: Use of Artificial Intelligent in Learning Management System (LMS): A Systematic Literature Review. International Journal of Computer Applications 175 (13), pp. 16–26, 2020, DOI: 10.5120/ijca2020920611.

[AKS24] Azaiz, I.; Kiesler, N.; Strickroth, S.: Feedback-Generation for Programming Exercises With GPT-4. In: Proceedings of the 2024 Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2024), July 8–10, 2024, Milan, Italy. Association for Computing Machinery, 2024, DOI: 10.1145/3649217.3653594.

[Da17] Danielsiek, H.; Vahrenhold, J.; Hubwieser, P.; Krugel, J.; Magenheim, J.; Ohrndorf, L.; Ossenschmidt, D.; Schaper, N.: Undergraduate teaching assistants in computer science: Teaching-related beliefs, tasks, and competences. In: Proceedings of 2017 IEEE Global Engineering Education Conference (EDUCON). IEEE, Piscataway, NJ, pp. 718–725, 2017, DOI: 10.1109/EDUCON.2017.7942927.

[FM06] Fies, C.; Marshall, J.: Classroom Response Systems: A Review of the Literature. J Sci Educ Technol 15 (1), pp. 101–109, 2006, DOI: 10.1007/s10956-006-0360-1.

[HMB19] Heller, N.; Mader, S.; Bry, F.: More than the sum of its parts. In: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing. Pp. 2473–2476, 2019, DOI: 10.1145/3297280.3297638.

[Ji18] Jivet, I.; Scheffel, M.; Specht, M.; Drachsler, H.: License to Evaluate: Preparing Learning Analytics Dashboards for Educational Practice. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge. ACM, pp. 31–40, 2018, DOI: 10.1145/3170358.3170421.

[KJH18] Keuning, H.; Jeuring, J.; Heeren, B.: A Systematic Literature Review of Automated Feedback Generation for Programming Exercises. ACM Transactions on Computing Education 19 (1), 2018, DOI: 10.1145/3231711.

[KO23] Kontorovich, I.; Ovadiya, T.: How narratives about the secondary-tertiary transition shape undergraduate tutors' sense-making of their teaching. Educational Studies in Mathematics 113 (1), pp. 125–146, 2023, DOI: 10.1007/s10649-023-10211-6.

[KS17] Kiy, A.; Strickroth, S.: Potentiale aufzeigen und Synergien nutzen: Audience Response Systeme als ein Anwendungsgebiet hochschulübergreifender Kooperationen. In: Joint Proceedings of the Pre-Conference Workshops of DeLFI and GMW 2017. Vol. 2092. CEUR Workshop Proceedings, CEUR-WS.org, Bonn, Germany, 2017.

[Le11] van Lehn, K.: The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist 46 (4), pp. 197–221, 2011, DOI: 10.1080/00461520.2011.611369.

[Le23] Ley, T.; Tammets, K.; Pishtari, G.; Chejara, P.; Kasepalu, R.; Khalil, M.; Saar, M.; Tuvi, I.; Väljataga, T.; Wasson, B.: Towards a partnership of teachers and intelligent learning technology: A systematic literature review of model–based learning analytics. Journal of Computer Assisted Learning 39 (5), pp. 1397–1417, 2023, DOI: 10.1111/jcal.12844.

[Lo20] Lorås, M.: From teaching assistants to learning assistants – lessons learned from learning assistant training at Excited. Læring om læring 4 (1), 2020.

[Ló22] López-Jiménez, J. J.; Fernández-Alemán, J. L.; González, L. L.; Sequeros, O. G.; Valle, B. M.; García-Berná, J. A.; Idri, A.; Toval, A.: Taking the pulse of a classroom with a gamified audience response system. Computer methods and programs in biomedicine 213, p. 106459, 2022, DOI: 10.1016/j.cmpb.2021.106459.

[Ma14] Mayring, P.: Qualitative content analysis: theoretical foundation, basic procedures and software solution. Klagenfurt, 2014.

[MB19] Mader, S.; Bry, F.: Audience Response Systems Reimagined. In: Advances in Web-Based Learning – ICWL 2019. Cham, pp. 203–216, 2019, DOI: 10.1007/978-3-030-35758-0_19.

[Mc16] McKay, T. M.: Do tutors matter? Assessing the impact of tutors on firstyear academic performance at a South African university. Journal of Student Affairs in Africa 4 (1), 2016, DOI: 10.14426/jsaa.v4i1.144.

[Mi19] Mirza, D.; Conrad, P. T.; Lloyd, C.; Matni, Z.; Gatin, A.: Undergraduate Teaching Assistants in Computer Science. In: Proceedings of the 2019 ACM Conference on International Computing Education Research. ACM, pp. 31–40, 2019, DOI: 10.1145/3291279.3339422.

[MRG18] Maharjan, N.; Rus, V.; Gautam, D.: Discovering Effective Tutorial Strategies in Human Tutorial Sessions. In: The Thirty-First International Flairs Conference. 2018.

[Ol21] Olney, T.; Walker, S.; Wood, C.; Clarke, A.: Are We Living In LA (P)LA Land? Reporting on the Practice of 30 STEM Tutors in their Use of a Learning Analytics Implementation at the Open University. Journal of Learning Analytics 8 (3), pp. 45–59, 2021, DOI: 10.18608/jla.2021.7261.

[PTR16] Philipp, S.; Tretter, T.; Rich, C.: Undergraduate teaching assistant impact on student academic achievement. Electronic Journal of Science Education 20 (2), pp. 1–13, 2016.

[RG18] Rößling, G.; Gölz, J.: Preparing first-time CS student teaching assistants. In: Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education. ACM, p. 376, 2018, DOI: 10.1145/3197091.3205829.

[Ri21] Riese, E.; Lorås, M.; Ukrop, M.; Effenberger, T.: Challenges Faced by Teaching Assistants in Computer Science Education Across Europe. In: Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1. ACM, pp. 547–553, 2021, DOI: 10.1145/3430665.3456304.

[RM09] Reinhardt, W.; Magenheim, J.: Modulares Konzept für die Tutorenschulung in der universitären Informatikausbildung. Commentarii informaticae didacticae: (CID) (1), pp. 105–118, 2009.

[RS20] Riese, E.; Stenbom, S.: Experiences of Assessment in Introductory Programming From the Perspective of NonComputer Science Majors. In: 2020 IEEE Frontiers in Education Conference (FIE). IEEE, pp. 1–9, 2020, DOI: 10.1109/FIE44824.2020.9274060.

[SB22] Strickroth, S.; Bry, F.: The Future of Higher Education is Social and Personalized! Experience Report and Perspectives. In: Proceedings of the 14th International Conference on Computer Supported Education - Volume 1: CSEDU. Vol. 1, INSTICC, SciTePress, pp. 389–396, 2022, DOI: 10.5220/0011087700003182.

[Sc17] Schwendimann, B. A.; Rodriguez-Triana, M. J.; Vozniuk, A.; Prieto, L. P.; Boroujeni, M. S.; Holzer, A.; Gillet, D.; Dillenbourg, P.: Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. IEEE Transactions on Learning Technologies 10 (1), pp. 30–41, 2017, DOI: 10.1109/tlt.2016.2599522.

[Se17] Sellami, N.; Shaked, S.; Laski, F. A.; Eagan, K. M.; Sanders, E. R.: Implementation of a Learning Assistant Program Improves Student Performance on Higher-Order Assessments. CBE life sciences education 16 (4), 2017, DOI: 10.1187/cbe.16-12-0341.

[SH22] Strickroth, S.; Holzinger, F.: Supporting the Semi-Automatic Feedback Provisioning on Programming Assignments. In: Methodologies and Intelligent Systems for Technology Enhanced Learning, 12th International Conference. MIS4TEL, Springer International Publishing, Cham, pp. 13–19, 2022, DOI: 10.1007/978-3-031-20617-7_3.

[SKB20] Shibani, A.; Knight, S.; Buckingham Shum, S.: Educator perspectives on learning analytics in classroom practice. The Internet and Higher Education 46, p. 100730, 2020, DOI: 10.1016/j.iheduc.2020.100730.

[SS22] Strickroth, S.; Striewe, M.: Building a Corpus of Task-based Grading and Feedback Systems for Learning and Teaching Programming. International Journal of Engineering Pedagogy (iJEP) 12 (5), pp. 26–41, 2022, DOI: 10.3991/ijep.v12i5.31283.

[St24] Strickroth, S.: Scalable Feedback for Student Live Coding in Large Courses Using Automatic Error Grouping. In: Proceedings of the 2024 Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2024), July 8–10, 2024, Milan, Italy. ACM, 2024, DOI: 10.1145/3649217.3653620.

[Wa19] Walker, S.; Olney, T.; Wood, C.; Clarke, A.; Dunworth, M.: How do tutors use data to support their students? Open Learning: The Journal of Open, Distance and e-Learning 34 (1), pp. 118–133, 2019, DOI: 10.1080/02680513.2018.1554476.

[WS20] Wood, R.; Shirazi, S.: A systematic review of audience response systems for teaching and learning in higher education: The student experience. Computers & Education 153, p. 103896, 2020, DOI: 10.1016/j.compedu.2020.103896.

[ZGS20] Zhang, Y.; Ghandour, A.; Shestak, V.: Using Learning Analytics to Predict Students Performance in Moodle LMS. International Journal of Emerging Technologies in Learning (iJET) 15 (20), p. 102, 2020, DOI: 10.3991/ijet.v15i20.15915.

Entwicklung, Evaluation und Distribution einer Selbstlernakademie zum Thema IT-Sicherheit

Miriam Naß[1], Tabea Dobbrunz[2], Sina Warmer[2], Andreas Harrer[2] und Norbert Pohlmann[1]

Abstract: Im Rahmen des vom Ministerium für Kultur und Wissenschaft des Landes NRW und von der Digitalen Hochschule NRW (DH.NRW) geförderten Projekts SecAware.nrw wurde eine Selbstlernakademie zum Thema IT-Sicherheit entwickelt. Mit dem Ziel Bewusstsein zu schaffen, stellt dieses Lernangebot eine Präventionsmaßnahme für die vermehrt stattfindenden Cyber-Angriffe auf Hochschulen dar. Das Angebot setzt dabei auf ein multimediales didaktisches Konzept, bestehend aus informativ aufbereiteten und zur Exploration anregenden Elementen. Ergebnisse aus Evaluationen deuten darauf hin, dass die Zielgruppen eine Mischung aus Videos, Fachtexten und interaktiven Elementen als effektiv und motivierend wahrnehmen. Die Verbreitung der Online-Selbstlernakademie unter Hochschulangehörigen erfolgt aktuell durch eine zusätzliche Kommunikationsstrategie, welche auf eine maximale Reichweite und eine starke Nutzendenbindung abzielt.

Keywords: Selbstlernkurs, IT-Sicherheit, Multimediales Lernangebot, Laborstudie

1 Einleitung

Cyberkriminelle bevorzugen Universitäten und Hochschulen als Angriffsziele aufgrund der umfangreichen, sensiblen Daten, wie die „Jahresberichte 2022 und 2023 des Bundesamtes für Sicherheit in der Informationstechnik (BSI)“ belegen [Bu]. Besonders im Jahr 2023 zeigt sich der Fokus auf wissenschaftliche Einrichtungen: Von 23 registrierten Ransomware-Attacken im Bildungssektor zielten 13 auf Hochschulen und Forschungseinrichtungen ab. Die fortschreitende Digitalisierung erhöht die Herausforderungen für diese Institutionen und die ansässige Hochschulgemeinschaft, da stetige Anpassungen an neue Technologien und das Internet notwendig sind.

Der „Faktor Mensch“ wird als entscheidender Risikofaktor innerhalb der IT-Sicherheit identifiziert. Faktoren wie Zeitmangel, fehlende Lernmöglichkeiten und unzureichendes Fachwissen erhöhen nicht nur das Risiko unbefugter Zugriffe durch externe Angreifende, sondern fördern auch unbeabsichtigt riskantes Verhalten seitens der Nutzenden. Cyberkriminelle nutzen diese menschlichen Schwachstellen zunehmend aus. Daher besteht die Notwendigkeit einer stetigen Überprüfung und Anpassung der Sicherheitsstrategien, um den dynamischen Cyberbedrohungen effektiv entgegentreten zu können.

1 Westfälische Hochschule Gelsenkirchen, Institut für Internet-Sicherheit, Neidenburger Straße 43, 45897 Gelsenkirchen, Deutschland, miriam.nass@w-hs.de; pohlmann@internet-sicherheit.de, https://orcid.org/0009-0007-6221-7327

2 Fachhochschule Dortmund, Institut für Digitalisierung von Arbeits- und Lebenswelten, Otto-Hahn-Straße 23 & 27, 44227 Dortmund, Deutschland, tabea.dobbrunz@fh-dortmund.de; sina.warmer@fh-dortmund.de; andreas.harrer@fh-dortmund.de, https://orcid.org/0009-0001-1076-7964

doi:10.18420/delfi2024_37

Diesen Herausforderungen begegnet das Projekt SecAware.nrw mit einem speziell auf Studierende und akademisches Personal zugeschnittenen Selbstlernangebot in deutscher und englischer Sprache. Das Angebot zielt darauf ab, IT-Kompetenzen und Cyber-Sicherheitsbewusstsein im akademischen Umfeld zu stärken. Die partizipative Entwicklungsmethode („User Centered Design“) und kontinuierliche qualitative Evaluationen gewährleisten die Relevanz und Effektivität des Angebots.

2 Related Work

Die digitale Bildungslandschaft entwickelt sich durch Initiativen, die IT-Kompetenzen und Cybersecurity-Bewusstsein fördern, stetig weiter. Dieses Kapitel stellt drei aktuelle Bildungsprogramme vor:

Erstens, **DIGI-V.nrw**, initiiert von der DH-NRW, zielt darauf ab, die digitalen Kompetenzen von Personal in Hochschulverwaltungen zu fördern. Es bietet Fortbildung und Vernetzung mit einem Fokus auf praktische und theoretische Kenntnisse zur Integration neuer Technologien in den Arbeitsalltag, unterstützt durch ein Konzept bestehend aus Fachtexten und Quizzen [Di]. Zweitens, der **Digitale Führerschein (DiFü)**, unterstützt vom Bundesministerium des Innern und für Heimat (BMI), ist ein Projekt von Deutschland sicher im Netz e.V., das digitales Grundlagenwissen und Sicherheit im Umgang mit Technologien vermittelt. Durch ein interaktives System profitieren Menschen aller Altersgruppen [De]. Drittens, die **IT-Grundschutz-Schulungen des BSI** zielen darauf ab, ein angemessenes Sicherheitsniveau für Organisationen und Einzelpersonen herzustellen [Bu18]. Alle drei Angebote besitzen Möglichkeiten zur Selbstüberprüfung und weisen modulare Konzepte auf.

Obwohl diese Programme bereits Beiträge zur digitalen Bildung leisten, adressieren sie allgemeine Zielgruppen oder spezifische Berufsgruppen. Es mangelt jedoch an praxisbezogenen Angeboten, die auf die Förderung von Cyber- und IT-Sicherheit besonders gefährdeter Personen im Hochschulkontext fokussiert sind. Damit sind als Zielgruppen die Professorenschaft, Studierende und wissenschaftliche Mitarbeitende bisher nicht angesprochen.

Das Projekt SecAware.nrw hat die maßgeschneiderte Ausrichtung auf diese Zielgruppen durch ein kostenfreies Angebot zum Ziel. Das umfangreiche Lernangebot zeichnet sich durch eine stringente Logik aus und kann nach Bedarf flexibel modular eingesetzt werden.

3 Didaktisches Design

Bei der Konzeption wurden besonders frühere Arbeiten nach Bergmann et al. [Be21] einbezogen. Bergmann et al. entwickelten die Lernplattform *FALEDIA*, die speziell auf die Bedürfnisse von Lehramtsstudierenden zugeschnitten ist und diesen dabei hilft, ihre Fähigkeiten in der Diagnose und Förderung für die Primarstufe zu entwickeln. Diese Arbeit diente als Grundlage für die Entwicklung des didaktischen Konzepts.

Um höhere Lernzielebenen [AK01] und damit eine kritische Reflektion und nicht nur die Kenntnis von Fakten erreichen zu können, wurde Anwendung und Analyse der Inhalte durch Aktivitäten der Lernenden unterstützt. Dazu dienen interaktive Elemente, die auch die Selbstwirksamkeit und Motivation fördern [Ke18]. Impulse zur Gestaltung und Implementierung der interaktiven Elemente wurden aus Bayrak et al. [BRS21] aufgenommen, die in ihrem Beitrag ein interaktives E-Mail-Interface für Lernspiele zur Sensibilisierung gegen Phishing-Angriffe vorstellten.

Bei der Konzeption der Lehrinhalte stellt besonders die Heterogenität innerhalb der und zwischen den Zielgruppen eine Herausforderung dar. Die multimedialen Lehrinhalte und redundant gestalteten *Videos* sowie *Fachtexte* sprechen daher ein möglichst breites Publikum an, unterstützen unterschiedliche Vorgehensweisen und sorgen neben Untertiteln für Barrierefreiheit. Da es sich um ein freiwilliges Lernangebot handelt, wurden Fachtexte in einfacher Sprache verfasst und Videos mit maximaler Länge von fünf Minuten kurz und prägnant gehalten [Da20; GKR14], um die Motivation zu erhalten.

Das Lernangebot unterteilt sich in insgesamt acht *Module* (Stand: März 2024): Ein Einleitungsmodul, welches durch ein fiktives Beispiel Interesse weckt und sieben inhaltliche Module, z.B. Social Engineering, Urheberrecht und Lizenzen. Diese umfassen ein bis vier Themenseiten und je ein abschließendes *Modulquiz*.



Abb. 1: Interaktives Element, in dem Lernende personenbezogene Informationen aus einem Social Media Feed sammeln und beispielhaft für die Formulierung eines Phishing Angriffs nutzen

Animierte Videos stellen die Themen vor und fassen die wesentlichen Aspekte unter den Rubriken „`BE AWARE`“ und „`BE SMART`“ zusammen. Im Anschluss werden die Inhalte nochmals als Fachtext dargeboten. Bei ausgewählten Themen folgt mit „`BE ACTIVE`“ ein zur Exploration anregendes Element in Form einer interaktiven Animation oder einer kurzen Aufgabe zur Selbstüberprüfung (siehe Abb. 1). Am Ende einer Seite werden die wichtigsten Punkte sowie weiterführende Links zur Vertiefung des Themas aufgelistet. Der didaktische Aufbau aus informierenden Inhalten und anschließenden interaktiven Elementen folgt den Erkenntnissen aus Bergmann et al. [Be21].

Im abschließenden Modulquiz können die Lernenden ihr Wissen mittels einfachen Single- und Multiple-Choice-Fragen sowie Zuordnungs- oder Freitextaufgaben überprüfen. Nach Abschluss wird den Lernenden visualisiert, wie viel Prozent der Fragen richtig beantwortet worden sind. Das Quiz kann dabei beliebig oft wiederholt werden.

4 Produktionsprozess und Umsetzung

Der auf Kommunikation und Journalismus spezialisierte Teil des Projektteams erstellt *Skripte* für Lernvideos und Fachtexte basierend auf der Grundlage geprüfter Quellen und sorgfältiger Recherche. Die Redaktion verfasst die Skripte, welche iterativ für die Videoproduktion angepasst und von externen Fachexperten überprüft werden. Die finalen Versionen werden von professionellen Sprechenden aufgenommen. Komplexe Sicherheitskonzepte werden über *Videos* mit fiktiven Szenarien erklärt, wobei Begriffe wie „Socialgram“ (siehe Abbildung 1) verwendet werden, um konkrete Plattformen nicht zu exponieren und mit ihren etwaigen Sicherheits- und Datenschutzproblemen in Bezug zu stehen. Zudem werden unterstützende *Grafiken* zur Veranschaulichung des Lernmaterials erstellt.

Die Entwicklung des E-Learning-Contents erforderte insgesamt 69 Personenmonate, also 9.889,77 Arbeitsstunden. Diese wurden auf 21 Themenseiten aufgeteilt, was einen Durchschnitt von etwa 470,94 Stunden pro Thema (ca. 20 Arbeitstagen) entspricht. Das stimmt mit früheren Schätzungen, wie Chapmans Durchschnitt von 490 Stunden für komplexe E-Learning-Module überein. Angesichts der durchschnittlichen Nutzungsdauer von 20-25 Minuten pro Thema ergibt sich ein Verhältnis von Produktionszeit zu Nutzungszeit von 1129,74 bis 1413,06. Laut Chapman kann für komplexe Projekte einen Faktor ab 716 entstehen [Ch10]. Durch die ermittelten Faktoren in diesem Projekt wird der hohe Aufwand für den Produktionsprozess unterstrichen.

Für die iterative Umsetzung der Lernplattform inklusive der Quizze, wurde das Autorentool *Articulate Rise 360*[3] verwendet, welches die kollaborative Erstellung optisch ansprechender E-Learning-Kurse vereinfacht. Darüber hinaus ermöglicht es den Export als SCORM-Modul[4]. Die eingesetzte Software ermöglicht zudem schnell eine Vorschau des Kurses zu erstellen, welche externen Personen für eine Revision zur Verfügung gestellt werden kann.

Mithilfe der Zusatzsoftware *Articulate Storyline 360*[5] wurden die interaktiven Elemente erstellt. Diese werden in Form von Folienpräsentationen designt, welche dabei diverse Storypfade abbilden können. Durch eine Vielzahl an Events können Ebenen auf den Folien angezeigt oder ausgeblendet werden. Es ist möglich den Folien Komponenten zur Wissensüberprüfung hinzuzufügen, die das Ergebnis an ein *Lernmanagementsystem (LMS)* übermitteln.

3 https://www.articulate.com/360/rise/
4 E-Learning-Standard SCORM: https://scorm.com/
5 https://www.articulate.com/360/storyline/

5 Evaluation

Die SecAware.nrw-Selbstlernakademie wurde über einen Zeitraum von zwei Jahren kontinuierlich evaluiert. In regelmäßigen Abständen erfolgte die Bereitstellung von Inhalten für Expertenfeedback an die Chief Information Security Officers (CISOs) der Hochschulen in NRW. Darüber hinaus wurden zwei Laborstudien mit zehn beziehungsweise zwanzig Teilnehmenden durchgeführt. Die Studien umfassten außerdem die Beantwortung von Fragebögen sowie die Durchführung von Interviews. Die erste Studie fokussierte sich auf Videos und das didaktische Konzept. Im Gegensatz dazu zielte die zweite Studie auf die Bewertung des überarbeiteten Konzepts, der resultierenden Selbstlernakademie sowie die interaktiven Elemente ab. Das Feedback der CISOs und der Studienteilnehmenden fließt beständig in die Weiterentwicklung der Selbstlernakademie ein.

Die an der Laborstudie teilnehmenden Personen wurden gebeten, unter Anwendung der *Think-Aloud-Methode* jeweils drei ausgewählte Themenseiten mit interaktiven Elementen sowie ein Quiz zu bearbeiten. Im Anschluss haben die Teilnehmenden die Gebrauchstauglichkeit sowie ihre persönliche Erfahrung mit der Selbstlernakademie evaluiert. Eingesetzt wurden hierfür der reduzierte *User Experience Questionnaire (UEQ-S)* [SHT17] sowie vier Subskalen des *Intrinsic Motivation Inventory (IMI)* [RMK83; Ry82]. Des Weiteren schätzten die Teilnehmenden ihre Erfahrung mit dem Thema IT-Sicherheit und dem Umgang mit Lernplattformen ein. Während viele der Teilnehmenden ihre Erfahrung mit Lernplattformen eher hoch bewerteten, sind die Erfahrungen mit dem Thema IT-Sicherheit gemischt.

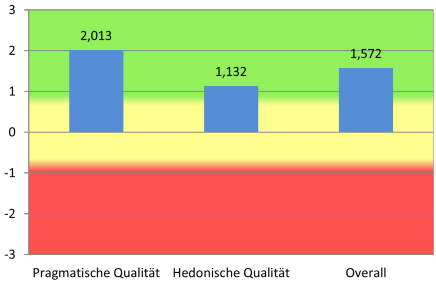


Abb. 2: Bewertung der Usability (UEQ-S)

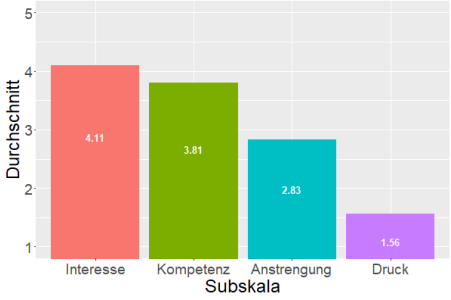


Abb. 3: Bewertung mittels IMI

Die Gebrauchstauglichkeit der Selbstlernakademie wurde durch den UEQ-S mit einem Wert von 1, 572, wie Abbildung 2 dargestellt, insgesamt sehr gut bewertet. Dabei wurde die reine Funktionalität durch die *pragmatische Qualität* (PQ = 2, 013) höher bewertet, als das empfundene Vergnügen durch die *hedonische Qualität* (HQ = 1, 132).

Der IMI ergab eine hohe Bewertung auf der Subskala „Interesse/Vergnügen“, welche mit der intrinsischen Motivation in Verbindung gebracht wird (Siehe Abb. 3). Auch die Skala „wahrgenommene Kompetenz“ wurde als gut bewertet und gilt als positiver Faktor auf die intrinsische Motivation. Der negative Faktor „Druck/Anspannung“ wurde dabei als niedrig evaluiert. Das Ergebnis der Subskala „Anstrengung/Wichtigkeit“ liegt im mittleren Bereich.

6 Distribution

Das Lernangebot steht auf der `Webseite` (Open Access) zur Verfügung und kann für die Integration in LMS wie Moodle oder ILIAS heruntergeladen werden. Für eine effektive Verbreitung unter Hochschulangehörigen sind jedoch verstärkte Anstrengungen in den Bereichen Marketing, Kommunikation und Netzwerkbildung erforderlich, um eine optimale Reichweite und Nutzendenbindung zu erreichen. Als Teil dieser Bemühungen wurden bereits erste Implementierungsschritte unternommen:

- **Öffentlichkeitsarbeit und Social Media:** Um die Aufmerksamkeit innerhalb der Hochschulgemeinschaft zu erhöhen, wurden zwei Pressemitteilungen und ein Instagram-Video veröffentlicht, welche das Angebot breit kommunizieren und zielgruppenspezifisch zugänglich machen.
- **Vorträge, Seminare und Konferenzen:** Durch Teilnahme an Veranstaltungen und Workshops, die die Anwendung der Plattform demonstrieren, wurde ein erster Eindruck vermittelt und ein direkter Austausch mit Interessierten gefördert.
- **Kooperationen:** Die Einbindung von Stakeholdern, wie beispielsweise CISOs als Multiplikatoren, verstärken die Reichweite der Inhalte.

Die Kombination verschiedener Formate zielt darauf ab, Reichweite in der Distribution für SecAware.nrw zu realisieren. Der Erfolg dieser Maßnahmen wird mittels technischer Reporting-Ansätze sowie weiterführender Evaluationsaktivitäten überwacht und bewertet.

7 Fazit und Ausblick

Der hohe Arbeitsaufwand bei der Entwicklung und Umsetzung der Lernmaterialen resultieren in einer Selbstlernakademie, die eine insgesamt positiv bewertete Gebrauchstauglichkeit und wahrgenommene intrinsische Motivation aufweist. Der gemessene zeitliche Aufwand unterstreicht die Notwendigkeit einer guten Projektplanung bei der Erstellung von E-Learning-Kursen. Besonders bei der Videoproduktion mit externen professionellen Sprechenden ist ein komplexer Prozess zu durchlaufen. Der Einsatz von Autorentools für die Generierung von Modulen kann sich dagegen positiv auf die benötigte Zeit auswirken, da die Einhaltung des SCORM-Standards sowie eine einheitliche Gestaltung größtenteils durch das Tool gewährleistet wird. Jedoch sind hier Limitationen in der Umsetzungsfreiheit sowohl im Design (Farb-, Symbol- und Dialoggestaltung) als auch in der Erhebung von Nutzendendaten zur Analyse von Lernaktivitäten zu beachten. Darüber hinaus wird die Anpassbarkeit durch Dritte im Sinne von Open Educational Ressources beeinträchtigt.

Da das Thema IT-Sicherheit ein sich stetig änderndes breites Feld darstellt, wird die Selbstlernakademie innerhalb der nächsten drei Jahre kontinuierlich aktualisiert und um neue Module erweitert. Gleichzeitig wird die Wirksamkeit des Angebots quantitativ evaluiert, um die Nutzungserfahrung zu verbessern.

Literaturverzeichnis

[AK01] Anderson, L. W.; Krathwohl, D. R., Hrsg.: A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives. Allyn & Bacon, New York, 2001, ISBN: 978-0801319037.

[Be21] Bergmann, A.; Dobbrunz, T.; Harrer, A.; Huethorst, L.; Walter, D.; Gutscher, A.; Selter, C.: FALEDIA: Eine Lernplattform für Lehramtsstudierende zum Erwerb von Diagnose- und Förderfähigkeit. 2021.

[BRS21] Bayrak, D.; Röpke, R.; Schroeder, U.: Konzeption und Entwicklung eines interaktiven E-Mail- Interface für Anti-Phishing Lernspiele. 2021.

[Bu] Bundesamt für Sicherheit in der Informationstechnik (BSI): Die Lage der IT-Sicherheit in Deutschland 2023, Bundesamt für Sicherheit in der Informationstechnik, URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2023.html?nn=132646, Stand: 07. 03. 2024.

[Bu18] Bundesamt für Sicherheit in der Informationstechnik (BSI): Online-Kurs IT-Grundschutz - Druckversion. S. 1–103, 2018, URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/Webkurs/onlinekurs2018.pdf?__blob=publicationFile&v=1, Stand: 12. 03. 2024.

[Ch10] Chapman, B.: How Long Does it Take to Create Learning? Chapman Alliance, 2010, URL: https://www.cedma-europe.org/newsletter%20articles/misc/How%20long%20does%20it%20take%20to%20develop%20training%20by%20Brian%20Chapman%20(Sep%2010).pdf.

[Da20] Dart, S.: Khan-Style Video Engagement in Undergraduate Engineering: Influence of Video Duration, Content Type and Course. In: Proceedings of the 31st Annual Conference of the Australasian Association for Engineering Education (AAEE 2020). Conference Name: AAEE - Annual Conference of Australasian Association for Engineering Education Meeting Name: AAEE - Annual Conference of Australasian Association for Engineering Education, Engineers Australia, Australia, 2020, URL: https://www.aaee2020.com.au/wp-content/uploads/2020/11/AAEE2020_paper_4.pdf, Stand: 14. 03. 2024.

[De] Deutschland sicher im Netz: Lernzentrale, Digitalführerschein (DiFü), URL: https://difue.de/lernzentrale/, Stand: 12. 03. 2024.

[Di] Digi-V.dh.nrw: digi-v.dh.nrw, URL: https://digi-v.dh.nrw/, Stand: 11. 03. 2024.

[GKR14] Guo, P. J.; Kim, J.; Rubin, R.: How video production affects student engagement: an empirical study of MOOC videos. In: Proceedings of the first ACM conference on Learning @ scale conference. L@S 2014: First (2014) ACM Conference on Learning @ Scale. ACM, Atlanta Georgia USA, S. 41–50, 2014, ISBN: 978-1-4503-2669-8, DOI: 10.1145/2556325.2566239, URL: https://dl.acm.org/doi/10.1145/2556325.2566239, Stand: 14. 03. 2024.

[Ke18] Kerres, M.: Mediendidaktik: Konzeption und Entwicklung digitaler Lernangebote. Publication Title: Mediendidaktik, De Gruyter Oldenbourg, 2018, ISBN: 978-3-11-045683-7.

[RMK83] Ryan, R. M.; Mims, V.; Koestner, R.: Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. Journal of Personality and Social Psychology 45 (4), Place: US Publisher: American Psychological Association, S. 736–750, 1983, ISSN: 1939-1315, DOI: 10.1037/0022-3514.45.4.736.

[Ry82] Ryan, R. M.: Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. Journal of Personality and Social Psychology 43 (3), Place: US Publisher: American Psychological Association, S. 450–461, 1982, ISSN: 1939-1315, DOI: 10.1037/0022-3514.43.3.450.

[SHT17] Schrepp, M.; Hinderks, A.; Thomaschewski, J.: Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). International Journal of Interactive Multimedia and Artificial Intelligence 4, S. 103–108, 2017, DOI: 10.9781/ijimai.2017.09.001.

Design und Implementierung einer interaktiven Lernumgebung im Bereich der IT-Sicherheit

Michael Zohner [1] und Philipp Brettschneider[1]



Abstract: IT-Sicherheit ist ein ständiger Begleiter von sich verbreitenden und innovativen IT-Systemen. Eine große Herausforderung der IT-Sicherheit besteht darin, dass jede Technologie unterschiedliche Schwachstellen birgt, deren Zusammenspiel zum Entwicklungszeitpunkt eines IT-Systems schwierig vorhersagbar ist. Die Kompetenz,unbekannte Schwachstellen in Technologien zu identifizieren nimmt somit einen zentralen Stellenwert ein, kann aber mit einem inhaltszentrierten Frontalunterricht, bei dem der Fokus auf der Funktionsweise von Sicherheitsmechanismen liegt, nicht aufgebaut werden. Das Ziel dieser Arbeit ist es eine Lernumgebung zu entwickeln, die es Studierenden in einem Bachelor-Studium der Informatik im Rahmen einer Lehrveranstaltung IT-Sicherheit ermöglicht, die Kompetenz zur Erkennung von Schwachstellen aufzubauen. Dazu wird die Lernumgebung so konzipiert und umgesetzt, dass die Studierenden auf Basis eines kompetenzbasierten Ansatzes Schwachstellen in Technologien, die sie aus ihrem Studium kennen, identifizieren und ausnutzen sollen. Die Erreichung der so zu erlangenden Erkenntnisse wird im Rahmen der Lehrveranstaltung mittels einer Reflexion der teilnehmenden Studierenden evaluiert.

Keywords: IT-Sicherheit, Interaktive Lernumgebung, Kompetenzbasiertes Lernen

1 Einführung

IT-Sicherheit ist ein Thema, das in vielen Studiengängen der Informatik einen festen Platz im Curriculum einnimmt. Ähnlich wie bei vielen Themen der Informatik ist die IT-Sicherheit sehr schnelllebig und relevante Aspekte wie Schwachstellen oder anzuwendende Mechanismen wandeln sich mit den eingesetzten Technologien, womit das vermittelte Wissen im Studium während des Berufslebens der Studierenden schnell an Relevanz verliert. Eine Möglichkeit, um diesem alterungsbedingten Relevanzverlust zu begegnen, ist der Aufbau der Kompetenz zur Identifikation von Schwachstellen in Technologien, anstatt der inhaltszentrierten Vermittlung aktueller Schwachstellen. Um diese Kompetenz aufzubauen, konzipiert und implementiert diese Arbeit eine interaktive Lernumgebung, die von Studierenden im Rahmen einer Lehrveranstaltung zu IT-Sicherheit ergänzend genutzt werden kann. Dazu erhalten die Studierenden im Rahmen der Lernumgebung Aufgaben, um Schwachstellen in Technologien, mit denen sie im bisherigen Studium bereits in Kontakt gekommen sind, zu identifizieren und auszunutzen. Die Lernumgebung wird im Rahmen einer Lehrveranstaltung evaluiert, indem die gewonnenen Erkenntnisse der Studierenden den zu vermittelnden Erkenntnissen gegenüberstellt werden. Die Evaluation zeigt, dass der Großteil der Studierenden erfolgreich die zu vermittelnden Erkenntnisse sowie weitere, unbeabsichtigte, aber thematisch relevante Erkenntnisse gewinnen konnte.

1 Hochschule Fulda, Fachbereich Angewandte Informatik, Leipziger Straße 123, 36037 Fulda, Deutschland, michael.zohner@hs-fulda.de, https://orcid.org/0009-0002-1188-1774

doi:10.18420/delfi2024_38

2 Grundlagen und Verwandte Arbeiten

Im Folgenden werden Grundlagen der IT-Sicherheit und des kompetenzbasierten Lernens präsentiert sowie verwandte Arbeiten diskutiert.

IT-Sicherheit. Das Ziel der IT-Sicherheit besteht darin, potenzielle Angreifer daran zu hindern, auf Daten und Informationen zuzugreifen, sie zu manipulieren, zu duplizieren oder zu vernichten [ITZ]. Im Kontext der IT-Sicherheit werden Vorgehensweisen eines Angreifers als **T**aktiken (Ziel des Angreifers), **T**echniken (Methoden zur Erlangung des Ziels) und **P**rozeduren (detailliertes, schrittweises Vorgehen zur Umsetzung der Technik inklusive genutzter Programme), kurz: **TTPs**, beschrieben [Bal]. Die IT-Sicherheit beschäftigt sich mit einer Vielzahl an Themen: Von der Netzwerksicherheit über die Rechtekonfiguration in Betriebssystemen bis hin zur Softwaresicherheit (siehe §3).

Kompetenzbasiertes Lernen. Die Kompetenz ist die Interaktion von kognitiven, motivationalen, volitiven, emotionalen und weiteren Komponenten und wird bei kontextbezogenen Handlungen in Anspruch genommen. Um Kompetenzen zu vermitteln, müssen fachübergreifende und multifunktionale Qualifikationen durch die praktische Anwendung, Reflexion und Weiterentwicklung der erworbenen Fähigkeiten ermöglicht werden. Hierfür wird ein Kompetenzprofil definiert, welches beschreibt, zu was die Studierenden anhand der Lehre im Stande sein sollen. Auf Basis des Kompetenzprofiles werden Erkenntnisse bestimmt, welche die Studierenden durch die Lehre generieren sollen. Durch aktivierende Lehr-Lern-Methoden sollen sich die Studierenden mittels forschendem und problemorientiertem Lernen mit den Lerngegenständen auseinandersetzen, um die vorher definierten Erkenntnisse zu generieren [MR14].

Verwandte Arbeiten. Im Bereich der IT-Sicherheits Awareness-Steigerung befasst sich die Arbeit [Pe16] mit der Stärkung des Bewusstseins von Studierenden zu den Themen der Passwortwahl sowie der Beeinflussung von Menschen (Social Engineering). Zudem sind verschiedene IT-Sicherheits Awareness Kurse [Fo23; TÜ23] online verfügbar. Im Bereich des Penetration Testings, bei dem Kenntnisse zur Durchführung realer und komplexer Angriffe vermittelt werden, befasst sich die Arbeit [Li15] mit der Konzeption eines Curriculums für eine solche Lehrveranstaltung. Zudem existieren diverse Online-Plattformen, wie HackTheBox oder TryHackMe, die Wissen im Bereich Penetration Testing vermitteln und gleichzeitig die technische Infrastruktur zum praktischen Anwenden des erlernten Wissens bereitstellen. Die vorliegende Arbeit ordnet sich zwischen den beiden Polen der Arbeiten zur Awareness-Steigerung sowie den Arbeiten zum Erlernen von tiefgehenden Peneration-Testing-Kenntnissen ein. Insbesondere wird ein Fokus auf relevante Technologien des Studiums gelegt und technisches Vorwissen vorausgesetzt, welches über die Awareness-Arbeiten hinausgeht, aber nicht den technischen Tiefgang erreicht oder den Einarbeitungsaufwand voraussetzt, der im Penetration Testing Bereich benötigt wird.

3 Konzeption und Implementierung der Lernumgebung

Im Folgenden wird ein Überblick über die Rahmenbedingungen sowie die Realisierung der Lernumgebung gegeben.

Rahmenbedingungen. Die Zielgruppe der Lernumgebung und Evaluation sind Bachelor-Studierende der Studiengänge Angewandte Informatik und Wirtschaftsinformatik. Die Studierenden haben in ihrem Studium u.a. Module belegt zu den Themen: (Web-)Programmierung, Datenbanksysteme und Netzwerktechnologie, welche den thematischen Umfang der Lernumgebung vorgeben. Die Lernumgebung soll praktische Übungen für fünf Übungseinheiten von je 1,5 Stunden zur Verfügung stellen, sowie eine Übungseinheit zur Einführung von 0,75 Stunden, der eine theoretische Einführung in die Themen Ethik und Recht der IT-Sicherheit vorausgeht, um den Studierenden die Grenzen der Anwendung der erlernten Inhalte zu vermitteln. Die Studierenden sollen die Lernumgebung sowohl in einem Rechnerlabor unter Betreuung als auch am eigenen Rechner in Eigenarbeit nutzen können.

Realisierung. Die Lernumgebung wird als isoliertes, virtuelles Netzwerk, bestehend aus drei virtuellen Maschinen (VMs) bereitgestellt, um versehentliche Angriffe der Studierenden auf reale Ziele zu vermeiden und eine Bearbeitung am eigenen Rechner zu ermöglichen. Insgesamt werden drei VMs zur Verfügung gestellt: Eine Kali-Linux-VM als angreifender Rechner sowie zwei Ubuntu-Linux VMs als angegriffene Rechner, die für bestimmte Aufgaben miteinander kommunizieren. Die drei VMs werden an den Laborrechnern als nicht-persistent konfiguriert, um Veränderungen durch vorhergehende Studierende zu vermeiden. Um den Studierenden die Möglichkeit zur Lösung der Aufgaben in Eigenarbeit zu geben, wird das benötigte Wissen zur Bearbeitung der Übungen in einer Wissensdatenbank in der Kali-Linux VM zur Verfügung gestellt. Zudem erhalten die Studierenden ein Cheat-Sheet mit den nötigen Befehlen und Beschreibungen der Tools. Die thematischen Schwerpunkte der fünf Übungen ergeben sich aus vorhergehenden Modulen der Studierenden in ihren Studiengängen sowie den Inhalten der parallel laufenden Vorlesungen in IT-Sicherheit: *Netzwerk-Enumeration*, *Passwort Brute-Force*, *Ausnutzen fehlender Eingabevalidierung*, *sichere Kommunikation* und *Rechte-Erweiterung*. Die TTPs dieser fünf Übungen werden im Folgenden grob erläutert. Die Lernumgebung kann unter dem folgenden Link heruntergeladen werden: `www.hs-fulda.de/it-sec-sandbox`.

Im Rahmen der ersten Übung zu *Netzwerk-Enumeration* [Ha18] sollen Studierende im bisher unbekannten, virtuellen Netzwerk durch Host-, Port-, und Service-Discovery mithilfe des Programms „NMap“ [NM24] alle VMs sowie deren angebotene Dienste identifizieren. In den Ausgaben der Netzwerk-Enumeration finden die Studierenden eine Webseite mit Zugangsdaten im Quellcode, die sie für den Zugriff auf einen FTP-Server nutzen können, der auf einem nicht-standardisierten Port betrieben wird. Durch diese Übung soll vermittelt werden, dass Hosts und laufende Dienste in einem Netzwerk identifiziert werden können.

In der zweiten Übung zu *Passwort Brute-Force* [Ta11] soll zu einem gegebenen Passwort-Hash das Klartextpasswort ermittelt werden. Um einen thematischen Übergang von der ersten Übung zu schaffen, wird der Passwort-Hash hierbei durch Ausnutzung einer Schwachstelle in WordPress durch das Tool „Metasploit“ [Ra24] aus der Shadow-Datei einer Ubuntu-VM extrahiert. Im Anschluss müssen die Studierenden das Tool „Hashcat“ [Ha24] nutzen, um mittels eines Wörterbuch-Angriffs das richtige Passwort zu erhalten. Durch diesen Schritt erhalten sie Zugriff auf eine verschlüsselte ZIP-Datei, deren geheimes Passwort sie durch einen Brute-Force-Angriff ohne Wörterbuch ermitteln müssen. Durch diese Übung soll vermittelt werden, dass durch Schwachstellen Zugriff auf sensitive Informationen ermöglicht wird und dass schwache Passwörter auch in gehashter Form gebrochen werden können.

In Übung drei zur *Ausnutzung fehlender Eingabevalidierung* sollen die Studierenden einen SQL-Injection-Angriff auf ein Login-System einer Webseite durchführen, welches keine Eingaben validiert, um Admin-Zugriff auf die Webseite zu erhalten. Daraufhin sollen sie eine File-Upload-Vulnerability [Ch23] ausnutzen, um mittels einer Reverse-Shell Zugriff auf den Webserver zu erhalten, indem sie ein Skript über eine Upload-Funktionalität des Webservers, welche Datei-Endungen nicht validiert, hochladen und dieses dann über die URL aufrufen, sodass es vom Webserver ausgeführt wird. Durch diese Übung soll vermittelt werden, dass die Validierung von Eingaben essenziell zum Schutz vor Angriffen ist.

In Übung vier zur *sicheren Kommunikation* [MC07] sollen Studierende mithilfe der Site-Cloner-Funktionalität des Programmes „SEToolkit“ [tr20] eine gegebene Webseite klonen. Die Studierenden sollen hierbei durch Betrachtung der originalen und der geklonten Seite realisieren, wie eine Phishing-Webseite erkannt werden kann. Zur Umleitung der Kommunikation zwischen den beiden Ubuntu-VMs soll ein Man-in-the-Middle-Angriff mittels ARP-Spoofing [Ko18, 427f.] mit dem Programm „Bettercap“ [Be23] durchgeführt werden. Im nächsten Schritt sollen die Studierenden dann mittels des Programms „MitMProxy“ [Mi24] die umgeleitete Kommunikation abfangen und manipulieren. Durch diese Übung soll die Relevanz der Verschlüsselung von Netzwerkkommunikation sowie die Authentifikation der Gegenstelle vermittelt werden.

Im letzten Modul sollen Ansätze zur *Rechte-Erweiterung* [Ko18] aufgezeigt werden, indem die Studierenden durch Ermitteln einer Rechte-Fehlkonfiguration ein Programm mit root-Rechten ausführen können, obwohl der Benutzer keine root-Rechte besitzt. Auf diese Weise sollen sie einen SSH-Schlüssel kopieren, um sich unter einem anderen Benutzerprofil anzumelden. Hier finden die Studierenden sensible Zugangsdaten in der Bash-History, mit denen sie sich unter einem anderen Benutzerprofil anmelden können, in welchem sie dann mittels eines bereitgestellten Exploits eine veraltete Version eines Programms ausnutzen, um sich root-Rechte auf der Ubuntu-VM zu verschaffen. Durch diese Übung soll vermittelt werden, dass eine saubere Rechtekonfiguration sowie ein besonderer Zugriffsschutz für sensible Daten essenziell zur Eindämmung eines Angriffs sind.

4 Evaluation

Zur Evaluation der Lernumgebung werden die gewonnenen Erkenntnisse der Studierenden mittels einer Reflexion erhoben. Dazu wird die Lernumgebung begleitend zu einer Lehrveranstaltung IT-Sicherheit eingesetzt, wobei die Studierenden in 2-wöchigem Rhythmus die fünf thematischen Übungen inklusive der Einführungsübung bearbeiteten.

Aufbau der Evaluation. Die Evaluation der Erkenntnisgewinnung erfolgte anhand einer qualitativen Fragestellung, indem nach Abschluss jeder der fünf thematischen Übungen die Studierenden nach drei Erkenntnissen gefragt wurden, die sie in dieser Übung gewonnen haben, bzw. die für diese Übung relevant waren und die sie bereits vorher gewonnen haben. Die Studierenden hatten mindestens zwei Wochen Zeit, um die Übung durchzuführen und die Freitextfrage zu beantworten. Für die thematisch passende Beantwortung der Frage wurde den Studierenden je Übung ein Punkt für die Portfolioleistung des Moduls anerkannt (insgesamt also bis zu 5 von 100 erreichbaren Punkten der Portfolioleistung). Am Ende des Moduls hatten die Studierenden die Möglichkeit, der Verarbeitung ihrer Antworten im Rahmen dieser Arbeit zu widersprechen, jedoch wurde dies von keiner Person wahrgenommen. Zur Auswertung wurden die Antworten anonymisiert und aggregiert, um Rückschlüsse auf Einzelpersonen zu verhindern. Die Zahl der Antworten pro Übung variierte zwischen 107 und 118. Zur Auswertung der Erkenntnisse werden die Antworten der Studierenden zu jeder Übung einer Aussagengruppe zugeordnet. Nach Auswertung aller Antworten der jeweiligen Übung wurde dann der prozentuale Anteil der Aussagengruppe in allen gegebenen Antworten ermittelt. Die Ergebnisse für Aussagengruppen mit mehr als 30% Nennung sind in Tabelle 1 dargestellt.

Diskussion der Evaluationsergebnisse. Die Evaluation zeigt, dass die intendierten Erkenntnisse der Übungen auch grob vermittelt wurden, wie anhand der Erkenntnisse #1, #2, #3, #4, #6, #7, #8 und #10 zu sehen ist. Kleinere Abweichungen gibt es z.B. bei der Übung zu Passwort-Brute-Force, bei der der Zugriff auf sensible Informationen durch Schwachstellen nicht explizit erwähnt wurde, sondern nur die Ausnutzung von Schwachstellen beschrieben wurde (#3) sowie in der Übung zu Eingabe-Validierung, bei der die fehlende Validierung des hochgeladenen Dateityps nicht von genügend Studierenden als Erkenntnis genannt wurde.

Die Studierenden antworteten mit zwei weiteren Erkenntnissen, die nicht explizit beabsichtigt waren. Zum einen wurde in #5 der Unterschied zwischen der Remote-Shell und der Reverse-Shell genannt. Diese Erkenntnis kann darauf zurückgeführt werden, dass die Studierenden zur Bearbeitung der Aufgabe zur File-Upload-Vulnerability die Rollenverteilung der Reverse-Shell begreifen mussten, also, dass die Kali-VM in diesem Fall die Rolle des SSH-Servers einnimmt, während sie bei vorherigen Aufgaben immer die Rolle des SSH-Clients einnahm. Zum anderen wurde in #10 angegeben, dass neben den Benutzerrechten explizit auch andere Rechtekonfigurationen (wie z.B. von Programmen) relevant sind. Diese Erkenntnis kann darauf zurückgeführt werden, dass die Studierenden in der Aufgabe zur Ausnutzung einer

root-Fehlkonfiguration ein Programm mit root-Rechten ausführten, obwohl ihr Benutzer keine root-Rechte besaß.

Tab. 1: Zusammengefasste Erkenntnisse, die von den Studierenden genannt wurden, mit anteiliger Nennung in Prozent in allen Antworten pro Übung (Erkenntnisse unter 30% wurden ausgelassen).

#	Übung	Beschreibung der Erkenntnis	Nennung
1	Netzwerk-Enumeration	Hosts, offene Ports und laufende Dienste in einem Netzwerkes können effizient ermittelt werden.	80%
2	Passwort-Brute-Force	Hashes von schwachen Passwörtern oder von Passwörtern in Wörterlisten können leicht gebrochen werden.	68%
3	Passwort-Brute-Force	Software-Schwachstellen können einfach durch Exploits ausgenutzt werden und schwerwiegende Folgen haben.	61%
4	Eingabe-Validierung	Datenbanken können aufgrund einer fehlenden Eingabe-Validierung auf einer Webseite manipuliert werden.	88%
5	Eingabe-Validierung	Unterschiede zwischen Remote-Shell und Reverse-Shell.	59%
6	Sichere Kommunikation	Netzwerkverkehr im lokalen Netzwerk kann mittels ARP-Spoofing via eines Man-in-the-Middle-Angriffs automatisiert umgeleitet werden.	72%
7	Sichere Kommunikation	Webseiten können durch Tools automatisiert nachgestellt werden und Login-Daten abfangen.	56%
8	Rechte-Erweiterung	Ein fahrlässig konfiguriertes Rechtesystem kann zur Übernahme des Systems durch einen Angreifer führen.	48%
9	Rechte-Erweiterung	Neben Nutzer-Rechten sind auch andere Rechtevergaben (z.B. für Programme) relevant	42%
10	Rechte-Erweiterung	Sensible, gespeicherte Daten (z.B. Zugangsdaten und SSH-Schlüssel) können gezielt entwendet werden.	38%

5 Fazit

Im Rahmen dieser Arbeit wurde eine Lernumgebung zur Vermittlung von typischen Angriffen der IT-Sicherheit konzipiert, realisiert und im Rahmen einer Lehrveranstaltung evaluiert, mit dem Ziel, die Kompetenz zur Identifikation von Schwachstellen bei Studierenden aufzubauen. Die Ergebnisse der Evaluation zeigen, dass die intendierten Erkenntnisse der Lernumgebung von den Studierenden in großen Teilen gewonnen werden konnten. Insgesamt war die Rückmeldung der Studierenden zur Lernumgebung positiv, mit dem häufigen Argument, dass sie die besprochenen Themen aus einer alternativen, interessanteren Sicht darstelle, die mehr Motivation zur Beschäftigung mit den Themen liefere. Aus Sicht der Lehrperson war es interessant zu sehen, welche Hürden bei der Anwendung von theoretischem Wissen auftreten können (z.B. bei Erkenntnis #5, bei der viele Studierende erst begreifen mussten, dass die Kali-VM, die sonst als SSH-Client genutzt wurde, auch als SSH-Server fungieren kann). Die entwickelte Lernumgebung soll auf Basis der Rückmeldungen besser an die Vorkenntnisse, Lösungsfindungsstrategien und Bedürfnisse der Studierenden angepasst und in den nächsten Semestern weiter eingesetzt werden.

Literaturverzeichnis

[Bal] Balbix: What are tactics, techniques, and procedures (TTPs) in cyber security?, URL: https://www.balbix.com/insights/tactics-techniques-and-procedures-ttps-in-cyber-security/, Stand: 10. 04. 2024.

[Be23] Bettercap: Home Page, 2023, URL: https://www.bettercap.org/, Stand: 10. 04. 2024.

[Ch23] Chen, Y.; Li, Y.; Pan, Z.; Lu, Y.; Chen, J.; Ji, S.: URadar: Discovering Unrestricted File Upload Vulnerabilities via Adaptive Dynamic Testing. IEEE Transactions on Information Forensics and Security 19, S. 1251–1266, 2023, DOI: 10.1109/TIFS.2023.3335885.

[Fo23] Fortinet: Fortinet Security Awareness Training, 2023, URL: https://www.fortinet.com/de/training/security-awareness-training, Stand: 10. 04. 2024.

[Ha18] Harris, S.; Michalak, E.; Schoonover, K.; Gausmann, A.; Reinbolt, H.; Herman, J.; Tauritz, D.; Rawlings, C.; Pope, A. S.: Evolution of network enumeration strategies in emulated computer networks. Proceedings of the Genetic and Evolutionary Computation Conference Companion, S. 1640–1647, 2018, DOI: 10.1145/3205651.3208270.

[Ha24] Hashcat: Hashcat, 2024, URL: https://hashcat.net/hashcat/, Stand: 10. 04. 2024.

[ITZ] ITZBund: IT-Sicherheit, URL: https://www.itzbund.de/DE/digitalemission/itsicherheit/itsicherheit.html, Stand: 10. 04. 2024.

[Ko18] Koffler, M.; Gebeshuber, K.; Kloep, P.; Neugebauer, F.; Zingsheim, A.: Hacking und Security: Das umfassende Handbuch. Rheinwerk Verlag, Bonn, 2018.

[Li15] Li, C.: Penetration testing curriculum development in practice. Journal of Information Technology Education. Innovations in Practice 14, S. 85, 2015.

[MC07] Moore, T.; Clayton, R.: Examining the impact of website take-down on phishing. Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit, S. 1–13, 2007, DOI: 10.1145/1299015.1299016.

[Mi24] Mitmproxy Project: Introduction, 2024, URL: https://docs.mitmproxy.org/stable/, Stand: 10. 04. 2024.

[MR14] Michelsen, G.; Rieckmann, M.: Kompetenzorientiertes Lehren und Lernen an Hochschulen–Veränderte Anforderungen und Bedingungen für Lehrende und Studierende. Campus Transformation. Education, Qualification & Digitalization, S. 45–65, 2014.

[NM24] NMAP: NMAP.ORG, 2024, URL: https://nmap.org/, Stand: 10. 04. 2024.

[Pe16] Peker, Y. K.; Ray, L.; Da Silva, S.; Gibson, N.; Lamberson, C.: Raising cybersecurity awareness among college students. Journal of The Colloquium for Information Systems Security Education 4 (1), S. 17–17, 2016.

[Ra24] Rapid7: Metasploit Framework, 2024, URL: https://docs.rapid7.com/metasploit/msf-overview/, Stand: 10. 04. 2024.

[Ta11] Tasevski, P.: Password attacks and generation strategies, 2011, URL: https://www.researchgate.net/profile/Predrag-Tasevski/publication/284609462_Password_Attacks_and_Generation_Strategies/links/56556e0d08aefe619b1a6216/Password-Attacks-and-Generation-Strategies.pdf, Stand: 10. 04. 2024.

[tr20] trustedsec: Social Engineering Toolkit - Dokumentation, 2020, URL: https://raw.githubusercontent.com/trustedsec/social-engineer-toolkit/master/readme/User_Manual.pdf, Stand: 10. 04. 2024.

[TÜ23] TÜVNord: TÜVNord Security Awareness Training, 2023, URL: https://www.tuev-nord.de/de/weiterbildung/seminare/security-awareness-training-a/, Stand: 10. 04. 2024.

Programmieren Lehren und Lernen in der Hochschule

A novel project assignment fostering self-competence and reflection using robotics for first-year CS students [2]

Melanie Baur [1], Johanna Sedlmair [1], and Sebastian Speiser [1]

Abstract: We developed and evaluated a novel project assignment tailored to computer science Bachelor freshmen, which challenges them to program a self-driving car using robotics, as part of orientation week at our university. The project assignment is a main element of our orientation week, supporting the transition of our new students into the higher education system. Therefore, we facilitate peer-to-peer interaction, encourage students to reflect on their academic role, and focus on skills they need for studying. A requirements analysis is the basis for the redesign, leading to improvements in achieving the program objectives. Evaluation results support the significant effectiveness of the redesign. The requirements developed can serve as a blueprint for creating engaging project assignments for various other degree programs and beyond. The assignment is available on the ZOERR platform licensed under Creative Commons.

Keywords: project assignment, orientation week, programming novices, educational robotics, self-competence, reflection, evaluation

1 Introduction

Transitioning from secondary school or work to higher education is challenging and a crucial step to enhance academic success [GP14; TW15]. At the Stuttgart University of Applied Sciences we support our new students during this important phase. In 2010 we introduced an orientation week to guide our new students and help them make a successful transition into the higher education system. It is a central part of our on-boarding process for freshmen; hence, it is very crucial to maintain the quality and integrate feedback and evaluation results. In order to improve the components of the week, and in particular the project assignment which serves as the cornerstone of the week, feedback from both students and trainers, as well as evaluation results, need to be taken into account and aligned with the program's specifications. Based on the open text comments in the evaluations of the previous years combined with feedback from trainers and didactic staff, we identified a need for the improvement of the project assignments of certain degree programs, among others in computer science (CS).

In this paper, we investigate necessary requirements for project assignments for the various degrees programs at our university within this specific context. We then implement a novel

1 Hochschule für Technik Stuttgart, Schellingstr. 24, 70174 Stuttgart, Germany,
melanie.baur@hft-stuttgart.de, https://orcid.org/0009-0005-9683-2446;
johanna.sedlmair@hft-stuttgart.de, https://orcid.org/0009-0007-4478-3245;
sebastian.speiser@hft-stuttgart.de, https://orcid.org/0009-0007-6587-1527

2 Supported by Stifterverband and MWK Baden-Wuerttemberg

doi:10.18420/delfi2024_39

assignment tailored to CS and validate the stated requirements. Our methodology involves analyzing the effectiveness of the new approach by comparing the evaluation results of two years and gathering feedback from students who participated in the assignment. Since five of the six requirements we identified are content-independent, while the sixth requirement focuses on program-specific content, the assignment can thus serve as a blueprint for a broad number of use-cases. It can be used to systematically redesign the project assignment independently of the subject. Additional application scenarios of the specific assignment include, e.g., CS team-building events or introductory programming classes.

2 Related Work

Students face many obstacles when first entering a higher education institution: new faces, new surroundings, and a new academic culture [BD14]. Especially at universities for applied sciences they start with diverse educational backgrounds and prior knowledge [Be21; Ha15]. To address these issues we introduced an orientation week for all Bachelor freshmen. The orientation week focuses on five primary objectives to help our students settle into the new environment:

- O_1 : Orientation on the new campus
- O_2 : Motivation for the upcoming study program
- O_3 : Getting to know their peers
- O_4 : Reflecting on the new role as student
- O_5 : Developing learning skills and strategies

Bosse and Mergner [BM19] identified nine types of offerings for the study entry phase focusing on content related, personnel, social, and organizational study requirements. Previous work shows how the five primary objectives and all activities of our orientation week align with Bosse's typology [HMS21; TB16]. This approach illustrates the broad range of study requirements we address during the orientation week, and therefore, the importance of this week.

In order to improve the orientation week, we implement a new assignment for our CS students. In this context we are looking at related work that is crucial for implementing an innovative learning experience for CS freshmen. More references and related work are cited when stating the requirements in Section 4.

Prince [Pr04] reviews literature on different learning scenarios and states that active learning is effective. While all active learning approaches have their problems, our main aspect is to support the objectives of the week which is clearly given through a collaborative learning task. Also the aspect of pair programming needs to be considered in a diverse team of programming beginners [Ha09; HBS21]. Although there are mixed results regarding the use of game-based elements, Kalogiannakis et al. [KPZ21] state in their extensive literature

study, that gamification may improve the learning process in CS education and increase student motivation and engagement. While different digital tools can be used for gamification purposes, among others, it is reported that the use of LEGO® MINDSTORMS® has positive influence on student motivation [Lu18; MK10], also in combination with problem-based learning [Ly14]. In addition, independent learning skills are developed [Lu10]. One way of integrating the use of robots as physical models in introductory programming classes is presented by Lawhead et al. [La02]. Only little research exists so far about LEGO® Education SPIKE™ Prime which is the follow-up model of LEGO® MINDSTORMS® robot. However, Körei and Szilágyi [Kö22] already made good experiences in using Scratch with LEGO® Education SPIKE™ Prime.

3 Orientation week

During orientation week, the president of our university welcomes all new students and the study deans present their specific programs. The freshmen take a tour through the university and library to familiarize themselves with the campus. There is an information market where all central facilities like the students advisory board, the service center for learning & teaching (SkiLL), the sports department as well as the student cafeteria present themselves. Other components of the orientation week include maths and English placement tests, a workshop on study strategies and techniques, and, of course, a final party.

A significant part of the week is spent working in teams on a project assignment which is designed to help first-year students get to know each other, while providing an introduction to the academic aspects of the program. The students have a total of 5x3 hours to prepare their solution to the assignment, with an additional 3 hours session allocated for the final team presentations. The project assignment should engage both students with and without prior knowledge.

A total outline of the orientation week is given in Table 1 where the different parts of the week are mapped to the objectives stated in Section 2. In this paper, we focus on improving the project assignment as this is the main part of the week and the evaluation results of the previous year show that this part has the greatest need for improvement.

4 Requirements for a novel project assignment

In this chapter, we set out the requirements for a successful project assignment so that we can devise the assignment in the next chapter. Subsequently, we evaluate the new assignment against the specified criteria and compare the evaluation results with those of the previous year.

We set up six requirements, listed as as $R_1 \ldots R_6$, that can then be extrapolated to project assignments in other disciplines, thus allowing this framework to serve as a blueprint for other degree programs, faculties or institutions if needed.

Day	Description	Objectives (o)
Monday	General part of the orientation week: - Opening ceremony (University Board) - Tour of the campus - Information market for students - Introduction to the degree program - Monday afternoon (optional): Learning to learn	O_1, O_4, O_5
Tuesday Wednesday Thursday	Program-related part of the orientation week: - (Interdisciplinary) project work in teams - Final presentation of the project results - Supervision and feedback from trainers, student mentors and the study deans	O_2, O_3, O_4, O_5

Tab. 1: Outline of orientation week

Requirements:

R_1 *Team engagement:* Research shows that interaction with peers ensures student engagement and a successful learning outcome [Al19]. The diversity in skills, backgrounds, and perspectives needs to be considered to encourage collaboration, innovation and creativity not only in how the tasks are performed but also in the end result [Pr04; Sn20].

R_2 *Active, game-based learning:* The project assignment should facilitate active, on-site learning combined with a game-based aspect to motivate and engage students in their work [KPZ21; NS00; Pr04].

R_3 *Scope of the assignment:* The new assignment must strike a balance between challenging students to think critically and solve problems, while ensuring that the task is manageable and not overly daunting [NC+02; VK21].

R_4 *Assessment criteria:* The project assignment needs to provide clear and transparent assessment criteria for the intended learning outcome in a problem-based learning scenario [BTK22].

R_5 *Study and examination regulations:* The content of the study and examination regulations (SPO) should be introduced, as it is essential for students to be familiar with them. Experience from the student advisory service at our university has repeatedly shown that unawareness of the SPO-relevant deadlines and rules (e.g. 30 CP rule after two semesters, maximum duration of study) leads to disadvantages such as delays in the course of study and sometimes even to exmatriculation.

R_6 *Subject-specific requirements:* The assignment should be relevant to the students' desired degree, as well as align with the overall curriculum and provide practical application of the concepts that will be learned to strengthen the students' degree satisfaction [RJG18].

5 Proposed solution[3]

In this chapter, we investigate the project assignment tailored to CS. Since CS includes several programming classes that students usually enjoy and look forward, we set a programming task for the subject-specific requirements R_6 using educational robotics. The assignment is also available on the ZOERR[4] platform which is the OER platform of the authors' federal state in German language and it is licensed under Creative Commons such that it can be used by other universities and institutions as well.

Scenario:

You are a young start-up that programs self-driving cars. There is a large market for your technology and expertise. However, you need financial support to get started, as the material and personnel costs exceed your equity capital. In your search for funding, you discover a fundraising competition. The prize for the winning team is not only a financial sum in the 6-digit range but also membership in an international network that offers idealistic support for young entrepreneurs. You decide to participate. The evaluation criteria are as follows:

1. Time limit in the final: 15 minutes per team
2. Which car fulfils the criteria?
3. Which start-up shows the greatest expertise in its own field?
4. Which presentation is the most convincing?

Your tasks:

1. Familiarize yourself with the study and examination regulations of your degree program so that you know what modules you will face and to signal the necessary expertise in your area.
2. Build a car in your group according to the given construction manual using the provided LEGO®[5] Education SPIKE™ Prime sets.[6]
3. Program the car so that:
   A. It can follow a line.
   B. It can pick up a brick at the end with the gripper arm.
   C. It can complete an additional task to match the name and slogan of your start-up.
   Tasks A and B are obligatory and the same for everyone. C is an individual task and can be creative.
4. Think of a name and a slogan for your start-up.

3 Thanks to Kay Lendermann for supporting us in creating this assignment.
4 Zentrales OER-Repositorium der Hochschulen in Baden-Württemberg: `https://www.zoerr.de/`
5 LEGO® is a trademark of the LEGO Group.
6 Detailed materials that we used and links to the construction manuals are provided in the appendix.

5. Prepare a presentation that will convince the jury of your professional and personal suitability. The presentation should include a demonstration of your vehicle, the content of your degree program, a brief introduction to your start-up and your personal computer science experience. Each team has a maximum of 15 minutes for the overall presentation.

General conditions:

All teams consist of four students. Further, there are five working sessions, each lasting three hours, dedicated to solving the assignment. The trainer and a peer mentor are present during all project phases to provide help and support to the students. Finally, there is an additional three hours session designated for the teams' presentations (depending on the number of teams) where the study dean evaluates the presentations according to the criteria stated in the assignment.

6 Evaluation and Results

In this chapter, we evaluate the new project assignment on several levels:

1. The project assignment supports all requirements stated in Section 4, which is shown in Subsection 6.1.
2. The project assignment shows a significant improvement in the evaluation results, which is shown in Subsection 6.2.
3. The project assignment is positively received by first-year students, which is shown in Subsection 6.3.

6.1 Evaluation of requirements

R_1 *Team engagement:* The teams consist of four students each. Numerous studies show that the use of robots, such as LEGO® MINDSTORMS®, engages and motivates students in introductory programming classes and fosters peer interaction, e.g. [MK10; MO09]. Since we have a very small number of female students, we do not take into account splitting teams according to gender or other characteristics.

R_2 *Active, game-based learning:* When building and programming the robot, the students can assemble the bricks either following the construction manual or being creative on their own. In both cases, they need to actively participate in the given assignment. Jormanainen [Jo18] shows in his work that educational robotics support engagement and motivation and Summet et al. [Su09] underpin this requirement that students can demonstrate their personal creative side with the robot.

R_3 *Scope of the task:* The assignment needs to be designed so that the subtasks are adaptive and can be completed in the time allotted. In addition, we address the written comments in the evaluation that the previous project task was too simple, boring and not challenging in terms of the content. Lykke et al. [Ly14] show in their study that a problem-based approach combined with LEGO® MINDSTORMS® helps students to feel neither sad nor to feel bored. The differentiation of our assignment in an obligatory and a voluntary part helps to control the scope such that nobody feels overwhelmed nor bored and all teams members have to play an active role in their team.

R_4 *Assessment criteria:* The assessment criteria are stated within the assignment and include both content-related criteria as well as process-related criteria. The results are assessed in a challenge among the teams at the last day of the orientation week by the study dean of the program according to Table 2 each on a scale from 0 (not fulfilled) to 2 (very good implemented), followed by a celebration of the winning teams to foster a sense of pride and reinforce the value of collaborative effort.

Start-up	**Car**	**Presentation**
Teamname	Follow line	Creativity
Slogan	Pick-up brick	Performance
Expertise	Extra feature	Time

Tab. 2: Assessed criteria

R_5 *Study and examination regulations:* The students are required to engage with the study and examination regulations in order to familiarize themselves with the academic framework of their degree program. This need is underpinned by Trautwein and Bosse who represent the organizational dimension as a critical requirement for first-year students [TB16]. Since dealing with the paragraphs of the study and examination regulations is known as a tiresome but important task, we integrate this need in the game-based assignment that requires detailed exploration of the regulations in a playful way.

R_6 *Subject-specific content:* Noor et al. [NMM20] show in a small study that using LEGO® MINDSTORMS® can help students with no programming skills improve problem-solving and computational thinking skills. The use of the programming language Scratch makes the introduction to programming simple and self-explanatory, while also preparing students for the start of programming lectures the following week [Su09]. Students with programming skills can use more complex programming concepts and be creative in the voluntary task.

6.2 Statistical evaluation

6.2.1 Methodology

To compare the new project assignment with the one of the previous years in different dimensions, we explored two groups of CS students in the orientation week (denoted as

week w_0), one in the fall semester 2022/2023 (denoted as year y_0), the other one in the fall semester 2023/2024 (denoted as year y_1).

A voluntary online questionnaire with the same questions was carried out both times on the last day of the orientation week. In fall semester 2022/2023 we received 29 questionnaires and in fall semester 2023/2024 we ended up with 24 collected questionnaires. The number of surveyed students is denoted by n_0 for y_0 and n_1 for y_1, respectively. Note that not all students answered all questions which results in smaller sample sizes for some of the questions.

The online questionnaire consists of 35 questions, where 8 of these 35 questions are open-text questions, the other 27 questions are recorded on a Likert scale (LS) from 1 ("applies") to 5 ("does not apply") and span the categories

- orientation and overview (9 LS questions),
- motivation and activity (6 LS questions),
- interaction with trainer (4 LS questions),
- own goals (2 LS questions and 5 open-text questions) and
- self-competence and reflection (6 LS questions and 3 open-text questions).

Furthermore, there are three questions in the category *personal information* about gender, age and educational qualification which are not considered in this paper as we have integrated all first-year CS students of the respective semester. We do not have further information about the questionnaire, in particular about the testing of the items.

We are at most interested in changes in the category *self-competence and reflection* as the team assignment contributes mostly to this category. However, we also checked for significant positive as well as negative changes in the other categories.

We first computed descriptive statistics on the data sets such as mean and standard deviation. For all questions of the category *self-competence and reflection* and for further questions with noticeable differences in means, we then tested for statistically significant changes using hypothesis tests. Therefore, we state the two hypotheses:

H_0: There is no difference in the answers between the groups of y_0 and y_1.
H_1: There is a difference in the answers of the groups of y_0 and y_1.

To select a proper hypothesis test, we checked the variables for normal distribution using the Shapiro-Wilk normality test. None of the variables which was selected for further analysis was normally distributed for both years. Since we have a rather small sample size with many ties which is due to the choice of a Likert-Scale, we are using the exact Wilcox-Mann-Whitney U test implemented in the R package *coin*, instead of the standard Mann-Whitney U test for two independent variables. Note that all statistical analyses are conducted using R (version 4.3.3).

6.2.2 Evaluation results

Descriptive statistics: We found better means in 20 of the 35 questions, one mean was the same and five means were (slightly) worse. The comparison of means and standard deviations (SD) of w_0 in y_0 to the answers of w_0 in y_1 is illustrated in Table 3. Note that we analyzed all questions of the category *self-competence and reflection*. The additionally analyzed questions have a change in means of at least 0.5 points on the LS and can be influenced by the content of the project assignment.

The best improvements in means were found in questions number 8 (improvement of 1.4 points), number 2 and 5 (improvement of 1.1 points). Having a closer look at the category *self-competence and reflection*, which is the main category targeted by the project assignment, we observe better means in five of the six questions of this category.

The five deteriorations that we found in all of the 35 questions range from 0.1 to 0.4 points on the LS and were therefore only marginal.

	Year 0			**Year 1**		
	Mean	SD	n_0	Mean	SD	n_1
Self-competence and reflection						
1. I felt up to the tasks during the week.	1.56	0.8	27	1.78	1.17	23
2. I had experiences of success that confirmed my choice of study	3.57	1.29	28	2.43	1.27	23
3. I got to know my own strengths and weaknesses	3.55	1.38	29	2.83	1.23	23
4. I have recognized how I have organized my learning so far and how I can improve it	3.38	1.47	29	3.95	1.05	22
5. I am completely satisfied with the result of the final presentation	2.94	0.83	17	1.82	1.01	22
6. I have learned that interdisciplinary skills are important for the achievement of team assignment	2.46	1.17	28	2.33	1.02	21
Organization and overview						
7. I got enough information about the team work	2.52	0.77	29	1.65	0.76	23
8. I experienced the activities of the entire orientation week as a useful combination	3.6	1.1	29	2.2	0.9	23
9. I realised what was expected of me during the orientation week	2.2	1.2	29	1.7	1.0	23
Motivation and activity						
10. I was able to contribute my own interests and goals into the project assignment and compare them with those formulated by the trainers	3.25	1.15	28	2.00	0.95	22

Tab. 3: Descriptive statistics of evaluated questions

Hypothesis tests: The results of the hypothesis tests are presented in Table 4 which shows the Z and the p-values of the exact Wilcox-Mann-Whitney U test computed by R.

We observe that three of the six questions of the category *self-competence and reflection* have significantly better results. We also observe that the worse result of question 1 is not significantly worse. Further, we found three significant improvements in the category *organization and overview* and *motivation and activity*.

	Results		**Hypotheses**	
	z	p-value	H_0	H_1
Self-competence and reflection				
1. I felt up to the tasks during the week	-0.40701	0.6826	x	
2. I had experiences of success that confirmed my choice of study	2.96	0.002682		x
3. I got to know my own strengths and weaknesses	1.9818	0.04816		x
4. I have recognized how I have organized my learning so far and how I can improve it	1.1787	0.2416	x	
5. I am completely satisfied with the result of the final presentation	3.6485	0.0001112		x
6. I have learned that interdisciplinary skills are important for the achievement of team assignment	0.24134	0.824	x	
Organization and overview				
7. I got enough information about the team work	3.6711	0.0001579		x
8. I experienced the activities of the entire orientation week as a useful combination	4.2215	8.136e-06		x
9. I realised what was expected of me during the orientation week	1.4994	0.1367	x	
Motivation and activity				
10. I was able to contribute my own interests and goals into the project assignment and compare them with those formulated by the trainers	3.5333	0.0002573		x

Tab. 4: Results of hypotheses tests of evaluated questions

6.3 Additional survey results in w_0 of y_1

Since the text comments about the project assignment in y_0 were rather negative, we posed in w_0 of y_1 an additional question to the students to investigate how the assignment is received by them. This additional question, however, was not included in the online questionnaire but conducted separately. The question "how do you rate the assignment" was rated by the students on a Likert scale from 1 to 5 stars, where more stars denote a better result. The results are shown in Table 5.

Number of stars	1	2	3	4	5
Quantity	0	1	1	12	11

Tab. 5: Rating of the new project assignment in stars

Furthermore, the students were asked to give feedback on their preferences and concerns regarding the project assignment in an open text field. They were allowed to enter up to three terms. The wordcloud illustrated in Figure 1 and generated out of 74 answers by Mentimeter is showing the atmosphere of the week.



Fig. 1: Wordcloud generated out of survey results

7 Discussion

In this section we discuss the evaluation results comparing the previous and the newly developed project assignment. The old one consisted of six independent questions requiring only little team interaction and collaboration and thus not fully meet the objectives of the orientation week. When designing the new assignment, we therefore focus on the derived requirements to ensure that the objectives are met.

The changes in the evaluations shown by descriptive statistics illustrate that we have found 20 improvements and only five deteriorations. It can also be seen that some of the improvements are considerable better (up to 1.4 points), while the deteriorations are only marginal (up to 0.4 points). This shows at first glance that we have made a positive change due to the new assignment.

Analyzing the results of the category *self-competence and reflection* with hypothesis tests, we figure out that three improvements are significantly better. In particular, the students have gained more confidence in what they study, get to know their strengths and weaknesses better and are more satisfied with the results of their final presentation. This is a meaningful improvement since the degree of student satisfaction has a positive impact on the study success [Ku06]. It is also important to note that the negative result was not significantly worse (question number 1), which can be interpreted as students being more challenged by the new task, but not overwhelmed.

The higher scores in the category *organization and overview* show that students were better prepared for teamwork and in particular, perceive the content of the whole week as a

useful combination. This is a considerable result that clearly illustrates the importance of the project assignment of the week and has prompted us to spend more work also in the assignments of the other degree programs.

Finally, the improvement in the category *motivation and activity* shows that students had the opportunity to incorporate also their own interests and goals into the assignment. This is important to strengthen the students' confidence in their own capabilities.

Not only the statistical analysis demonstrates the improvement, also the results of the additional question underpin this. The students rated the assignment 11 times with 5 stars and 12 times with 4 stars in y_1 which is a noteworthy improvement to 11 negative written comments regarding the project assignment in an open answer text field (for n_0=29) in y_0. Notable that there was no negative text comment regarding the assignment in y_1.

It is worth to mention that the trainer remained the same in both years. Also, the other parts of the orientation week did not change. Therefore, we can attribute the positive changes to our novel project assignment. From these results we can derive that the new assignment in combination with educational robots was a complete success for both the students as well as the teaching team.

8 Conclusion

We aimed to improve the previous project assignment for first-year CS students in the context of orientation week which was appropriate according to the evaluation results of students, trainer and didactic staff. To derive a new assignment, we first investigated necessary requirements and then showed that these requirements are all fulfilled by our new assignment. To demonstrate the enhancement of the new assignment, we compared evaluation results of two years and found significant improvements, in particular in the category *self-competence and reflection* but also in other categories such as *organization and overview* or *motivation and activity*.

Since five of the six requirements that we state are content-independent and the sixth requirement can be tailored to other subject-specific content, the requirements are adaptable for various use-cases. This means that the proposed framework can be used as a blueprint for redesigning project assignments across different subjects. In order to make the results available to other departments, the assignment was presented at our university to all study deans to serve as a catalyst for redefining the project assignments of other degree programs and potentially beyond.

Although we used the assignment in the specific context of the orientation week at our university, we anticipate a much broader use such as at the beginning of programming classes to slowly introduce the students to programming with direct experience of success or for team-building events in the context of beginner programming.

References

[Al19] Altermatt, E. R.: Academic Support From Peers as a Predictor of Academic Self-Efficacy Among College Students. Journal of College Student Retention: Research, Theory & Practice 21 (1), pp. 21–37, 2019, DOI: 10.1177/1521025116686588.

[BD14] Brooman, S.; Darwent, S.: Measuring the beginning: a quantitative study of the transition to higher education. Studies in Higher Education 39 (9), pp. 1523–1541, 2014, DOI: 10.1080/03075079.2013.801428.

[Be21] Berghoff, S.; Hachmeister, C.-D.; Huesch, M.; Thiemann, J.: Hochschulzugang und Studieneingang in Deutschland (Stand 2021). Guetersloh, CHE, 2021.

[BM19] Bosse, E.; Mergner, J.: Besser einsteigen – Analyse von Anforderungen und Gestaltungsmöglichkeiten der Studieneingangsphase. 91, pp. 71–102, 2019.

[BTK22] Biggs, J.; Tang, C.; Kennedy, G.: Teaching for quality learning at university. McGraw Hill, 2022.

[GP14] Gale, T.; Parker, S.: Navigating change: a typology of student transition in higher education. Studies in Higher Education 39 (5), pp. 734–753, 2014, DOI: 10.1080/03075079.2012.721351.

[Ha09] Hannay, J. E.; Dybå, T.; Arisholm, E.; Sjøberg, D. I.: The effectiveness of pair programming: A meta-analysis. Information and Software Technology 51 (7), Special Section: Software Engineering for Secure Systems, pp. 1110–1122, 2009, ISSN: 0950-5849.

[Ha15] Haft, A.: Heterogene Studierende - homogene Studienstrukturen. In: Herausforderung Heterogenität beim Übergang in die Hochschule. Pp. 13–28, 2015.

[HBS21] Hawlitschek, A.; Berndt, S.; Schulz, S.: Towards a Framework of Planning Collaborative Learning Scenarios in Computer Science. In: Proceedings of the 21st Koli Calling International Conference on Computing Education Research. Koli Calling '21, Association for Computing Machinery, Joensuu, Finland, 2021.

[HMS21] Heintz-Cuscianna, B.; Mayer, F.; Sigg, B.: Wegbereiter für hybride Formate in der Studieneingangsphase. Eine Evaluationsstudie zur Vorbereitungswoche an der HFT Stuttgart. In: Forschungsimpulse für die Hochschulentwicklung im Kontext hybrider Lehre. Forschung und Innovation in der Hochschulbildung, Band 11, pp. 131–146, 2021.

[Jo18] Jormanainen, I.: On computer science major students' motivation in a practically oriented robotics course. In: Proceedings of the 18th Koli Calling International Conference on Computing Education Research. Koli Calling '18, Association for Computing Machinery, Koli, Finland, 2018.

[Kö22] Körei, Attila and Szilágyi Szilvia: From Scratch to Python - Lego Robots as Motivational Tools for Coding, 2022, URL: https://doi.org/10.35925/j.multi.2022.3.22e, visited on: 02/24/2024.

[KPZ21] Kalogiannakis, M.; Papadakis, S.; Zourmpakis, A.-I.: Gamification in Science Education. A Systematic Review of the Literature. Education Sciences 11 (1), 2021.

[Ku06] Kuh, G. D.; Kinzie, J. L.; Buckley, J. A.; Bridges, B. K.; Hayek, J. C.: What matters to student success: A review of the literature. National Postsecondary Education Cooperative Washington, DC, 2006.

[La02] Lawhead, P. B.; Duncan, M. E.; Bland, C. G.; Goldweber, M.; Schep, M.; Barnes, D. J.; Hollingsworth, R. G.: A road map for teaching introductory programming using LEGO© mindstorms robots. SIGCSE Bull. 35 (2), pp. 191–201, 2002, ISSN: 0097-8418.

[Lu10] Lui, A. K.; Ng, S. C.; Cheung, Y. H. Y.; Gurung, P.: Facilitating independent learning with Lego Mindstorms robots. 1 (4), pp. 49–53, 2010, ISSN: 2153-2184.

[Lu18] Luxton-Reilly, A.; Simon; Albluwi, I.; Becker, B. A.; Giannakos, M.; Kumar, A. N.; Ott, L.; Paterson, J.; Scott, M. J.; Sheard, J.; Szabo, C.: Introductory programming: a systematic literature review. In: Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education. ITiCSE 2018 Companion, Association for Computing Machinery, Larnaca, Cyprus, pp. 55–106, 2018.

[Ly14] Lykke, M.; Coto, M.; Mora, S.; Vandel, N.; Jantzen, C.: Motivating programming students by problem based learning and LEGO robots. In: 2014 IEEE Global Engineering Education Conference (EDUCON). Pp. 544–555, 2014.

[MK10] Markham, S. A.; King, K. N.: Using personal robots in CS1: experiences, outcomes, and attitudinal influences. In. ITiCSE '10, Association for Computing Machinery, Bilkent, Ankara, Turkey, pp. 204–208, 2010.

[MO09] McWhorter, W. I.; O'Connor, B. C.: Do LEGO® Mindstorms® motivate students in CS1? In: Proceedings of the 40th ACM Technical Symposium on Computer Science Education. SIGCSE '09, Association for Computing Machinery, Chattanooga, TN, USA, pp. 438–442, 2009, DOI: 10.1145/1508865.1509019.

[NC+02] Nakamura, J.; Csikszentmihalyi, M., et al.: The concept of flow. Handbook of positive psychology 89, p. 105, 2002.

[NMM20] Noor, F. H.; Mohamad, F. S.; Minoi, J. L.: Learning Programming using Lego Mindstorms: Analysis of Learner Experiences. In: 2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC). Pp. 1–6, 2020, DOI: 10.1109/R10-HTC49770.2020.9357024.

[NS00] Norman, G. R.; Schmidt, H. G.: Effectiveness of problem-based learning curricula: theory, practice and paper darts. Medical Education 34 (9), pp. 721–728, 2000.

[Pr04] Prince, M.: Does Active Learning Work? A Review of the Research. Journal of Engineering Education 93 (3), pp. 223–231, 2004.

[RJG18] van Rooij, E.; Jansen, E.; van de Grift, W.: First-year university students' academic success: the importance of academic adjustment. European Journal of Psychology of Education 33, pp. 749–767, 2018.

[Sn20] Snijders, I.; Wijnia, L.; Rikers, R.; Loyens, S.: Building bridges in higher education: Student-faculty relationship quality, student engagement, and student loyalty. International Journal of Educational Research 100, p. 101538, 2020, DOI: 10.1016/j.ijer.2020.101538.

[Su09] Summet, J.; Kumar, D.; O'Hara, K.; Walker, D.; Ni, L.; Blank, D.; Balch, T.: Personalizing CS1 with robots. SIGCSE Bull. 41 (1), pp. 433–437, 2009.

[TB16] Trautwein, C.; Bosse, E.: The first year in higher education—critical requirements from the student perspective. Higher Education 73, 2016, DOI: 10.1007/s10734-016-0098-5.

[TW15] Tang, Y. V.; Wong, S. L.: Bridging students: successful transition from high school to college. 2015, URL: https://api.semanticscholar.org/CorpusID:55580657.

[VK21] V.K., Z.: One More Time on the Zone of Proximal Development. Cultural-Historical Psychology 17 (2), pp. 37–49, 2021.

Used materials and links

We are working with the new LEGO® Education SPIKE™ Prime basic sets, including the extension set. LEGO® Education SPIKE™ Prime is the follow-up model of the often used and well-explored LEGO® MINDSTORMS® robot. Each group had a tablet plus own devices. On the tablets the LEGO® Education SPIKE™ App was preinstalled. Students were allowed to install the app on their own devices from this link: `https://education.lego.com/de-de/downloads/spike-app/software/`.

Students were asked to build the advanced driving base and were equipped with a link to the construction manual: `https://education.lego.com/de-de/lessons/prime-competition-ready/assembling-an-advanced-driving-base`. They also could go through the material in the app to find the manual for a gripper arm (`https://education.lego.com/de-de/lessons/prime-competition-ready/time-for-an-upgrade/`) and also an idea how to code driving on a line (see e.g. `https://education.lego.com/de-de/lessons/prime-competition-ready/training-camp-3-react-to-lines/`). The app assisted the students in generating ideas for the creative topic; however, some groups chose to think completely freely and were not influenced by the given building instructions.

Correlation of Error Metrics in Python CS1 Courses

Annabell Brocker[1] and Ulrik Schroeder[1]



Abstract: Timely and effective feedback is essential for novice programmers. Despite significant advancements in Large Language Models enabling the generation of understandable feedback in CS1 courses, determining appropriate timing for delivering feedback automatically remains a persistent challenge. Compiler messages serve as a fundamental communication channel between programmers and computers, signaling syntactic and runtime errors. Various metrics associated with these messages could potentially signify the need for intervention. Hence, it is imperative to explore the correlations among these error metrics. This study conducts a comprehensive analysis of multiple public Python datasets to evaluate error metrics and explore their correlations. The findings offer insights into a potential indicator for determining appropriate timing of feedback in programming education.

Keywords: Programming, Error Quotient, Watwin, Robuste Relative, Repeated Error Density, Python

1 Motivation

Providing timely and effective feedback is crucial in facilitating the growth of novice programmers [HT07; KJH19; Na08]. Premature feedback can, for example, interrupt the learning process or foster an over-reliance on feedback [HT07]. Advancements in feedback generation, especially via LLMs [KLK23; Le23; RKJ24], have not fully addressed the challenge of automatically determining appropriate timing. Early work, such as expert-recommended intervention points [Je22], face limitations in automated implementation due to their dependence on human intuition. Compiler messages inform programmers of coding errors, and when recorded and transformed into data-driven metrics, they provide predictive insights into learner performance [PSV15]. Help-seeking strategies have a notable impact on learning behavior [Ga81]. This process comprises (a) recognizing the need for assistance, (b) deciding to seek help, (c) identifying potential helpers, and (d) utilizing strategies to request aid [VPJ14], with considerable variation among learners in disclosing difficulties and seeking aid. Metrics derived from programming interactions remain unaffected by these behaviors, making them suitable for automated determination of feedback timing.

This research aims to examine various datasets from Python programming novice courses to investigate correlations among different error metrics. The long-term goal is to explore the potential usefulness of these error metrics in automating the determination of appropriate feedback timings, as they can be monitored and computed without further effort by learners and are independent of help-seeking behaviors. A related research initiative has been proposed [Me22], but no findings have been published from this study thus far.

1 RWTH Aachen, Learning Technologies Research Group, Ahornstr. 55, 52074 Aachen, Deutschland,
a.brocker@informatik.rwth-aachen.de, https://orcid.org/0009-0007-6708-0892;
schroeder@informatik.rwth-aachen.de, https://orcid.org/0000-0002-5178-8497

doi:10.18420/delfi2024_40

2 Related Work

Detecting automatically struggling students in programming education is already researched multiple times [CAP17; Co17; MLL07; MZK19], but no consense or clear metric could be found, that can help to identify these students for offering help during the development process. Multiple methodologies for detection exist across various disciplines:

- biometric data, e.g. facial expressions, eye tracking or physiological signals [MLL07],
- demographic data [CAP17; Co17],
- interaction data such as mouse movements, key presses, temporal patterns [MZK19].

While biometric and demographic data provide detailed insights into individual characteristics, their acquisition is frequently impeded by the absence of requisite devices or legal constraints on data collection procedures. Consequently, these methods are often impractical for real-world applications. Conversely, this study utilizes data that is automatically generated by programmers during a session through regular interactions.

Error Quotient: The Error Quotient (EQ) [Ja06] is a quantitative metric that evaluates the frequency and characteristics of errors during a single programming session. It considers the number of errors, their recurrence, and their locations in subsequent compilations. EQ scores range from 0 to 1, with higher values indicating more persistent errors.

Watwin Algorithm: The Watwin algorithm [WLG13] expands upon the concept of EQ by introducing temporal considerations between events. This enhancement involves discerning error types specific to individual students in contrast to their peers, and integrating resolution time as a predictive parameter. Similar to EQ, the resulting Watwin score ranges between 0 and 1, with higher values indicative of a greater number of unresolved issues. Before computation, the compilation sequence undergoes preprocessing procedures involving various techniques, such as the removal of event pairs with identical code snapshots after comment removal from both snapshots.

Robuste Relative Algorithm: The Robuste Relative algorithm (RR) was also introduced by [Wa15] as a modification of the Watwin algorithm regarding the scoring for the correction time of an error. Instead of mean and $2MAD_e$ (comparable to standard deviation), the quantiles of the correction times are used, as these are more robust against outliers.

Repeated Error Density: The Repeated Error Density (RED) [Be16] specifically addresses the recurrence of identical errors within short programming sessions. It quantifies the frequency of repeated errors without considering their resolution time. A minimum value of 0 indicates the absence of directly repeated errors. Unlike EQ, Watwin, and RR, RED does not have an upper limit for its highest value. Becker observed that RED exhibits lower sensitivity to contextual changes compared to EQ and Watwin [Be16].

Originally, all metrics were designed to examine syntax errors in compilation sequences. However, Python code is interpreted at runtime without a compilation phase. Therefore, following the methodology in [PSV15], our analysis included all types of errors.

Coelho et al. employed Learning Analytics [SB12] techniques using interaction data during coding to forecast student performance in Python CS1 courses, generating over 30 attributes for performance prediction [Co23]. They noted that simpler metrics like EQ and Watwin exhibited *"lower accuracy rates (<66%) when using data from the entire course"*[Co23, p. 1100] compared to their methodology. However, our focus lies on predicting performance in shorter learning sessions to facilitate timely feedback and not for a whole course. In addition, the relevant attributes and the entire investigation process are not explained, so that it cannot be applied to other datasets. Rodrigo; Baker computed several metrics including average duration between compilations, aggregate count of compilations, total number of errors, and EQ [RB09]. They found that discerning average frustration levels across laboratory exercises can be achieved through factors such as *"the number of pairs of consecutive compilations with the same edit location, the number of pairs of consecutive compilations with the same error, the average time between compilations and the total number of errors."*[RB09, p. 75]. Jadud; Dorn investigated the compilation behavior of approximately 28,000 users from ten countries [JD15]. Their findings suggest that individuals with higher EQ scores may struggle more with programming tasks, leading to poorer performance on traditional assessments, while those with lower EQ scores tend to fare better. Additionally, the Watson algorithm and the percentage of lab time spent resolving errors were found to explain over 50% of final student results [WLG14]. Both observations underscores a potential correlation between compilation behavior and programming skills. Price et al. developed calculations for EQ, Watwin, and RED from code snapshots in ProgSnap2 data format and investigated their correlation on a user session level [Pr20]. In each session, any two consecutive events took place within a maximum time frame of 20 minutes. They found strong correlations among all three metrics on all datasets, with similar distributions, contrary to previous studies, and advocate for further investigations on additional datasets.

3 Methodology and Data

We started with the acquisition of publicly accessible datasets containing code snippets from Python programming novices. Datasets were chosen based on specific criteria: (1) programming language must be Python, (2) dataset must be publicly available, and (3) it must include essential information such as code, interpreter output, timestamps, tasks, and user identifiers. Unfortunately, only a restricted number of datasets fulfilled these criteria. For example, in a recent dataset authored by [De23] relevant information such as individual run outputs or the Python version used were absent, hindering the execution of the code on our end. Subsequently, a suitable dataset, namely CodeBench, was identified for analysis. CodeBench[2] is an automated evaluation tool and enables learners to code and submit solutions [Co23]. Since 2016, the authors have released all system logs in pseudonymized form[3]. In a second step, we preprocessed all data from the datasets to aggregate key elements of the ProgSnap2 format [Pr20], including user information, timestamp, task id, executed

2 CodeBench, `https://codebench.icomp.ufam.edu.br/`, last accessed: 09.03.2024

3 CodeBench Dataset, `https://codebench.icomp.ufam.edu.br/dataset/`, last accessed: 09.03.2024

code, interpreter feedback, success status, error name, and error line for each code execution, with a summary of each dataset presented in Table 1. Subsequently, the processed data was calculated according to EQ, Watwin, RR, and RED on various levels: user, user task, user session, and user task session level. At user level, data is aggregated per user. At user session level, data is aggregated across all sessions of a user. At user task level, data is aggregated across all tasks of a user. Finally, at user task session level, data is aggregated across all tasks and sessions of a user. A session was defined as any sequence of consecutive events occurring within a 20-minute interval, similar to [Pr20].

	2020-2	2021-1	2021-2	2022-1	2022-2	2023-1
# Students	203	217	224	237	198	263
# Tasks	885	1.549	2.336	2.340	2.525	2.562
# Exec. Events	176.212	188.133	225.894	209.597	162.466	243.243
# Error Events	67.366 (38,23%)	74.534 (39,62%)	91.489 (40,5%)	76.170 (36,34%)	60.658 (37,34%)	88.620 (36,43%)

Tab. 1: Basic dataset statistics with total number of students, tasks, execution events, and events that resulted in an error after execution for each dataset from CodeBench[3].

Lastly, pearson correlation coefficients were computed for each metric pair, excluding data points where both metrics had a value of 0 (no compilation problems) to focus on correlations related to compilation issues to address the subsequent research question:

> *How and on which levels do different error metrics of compilation sequences from Python CS1 code fragments correlate with each other?*

All data preprocessing and error metric calculation scripts are openly accessible [4] [5].

4 Results

Figure 1 illustrate the relationship between the granularity of levels, total amount of data points, and non-zero scores for various error metric combinations. The proportion of non-zero scores relative to the total number of data points of each level decreases progressively, with the highest mean values observed at the user level ($\bar{x}$ = 95.94%), followed by the user session level ($\bar{x}$ = 64.76%), then the user task level ($\bar{x}$ = 38.42%), and finally the user task session level ($\bar{x}$ = 35.31%). Conversely, the absolute number of non-zero values rises due to the generally higher number of data points at each subsequent level. Among these levels, RED consistently exhibits the lowest proportion of non-zero values, followed by EQ, and subsequently by Watwin/RR. This trend aligns with the inherent mathematical principles of each metric. Specifically, RED assigns error points solely for identical consecutive errors, while EQ allots points for any consecutive errors. Watwin and RR allocate points for any

4 Analysis Scripts for Error Metric Study for CodeBench Datasets, https://doi.org/10.17605/OSF.IO/52XZE, last accessed: 20.06.2024

5 pyerrormetrics, https://pypi.org/project/pyerrormetrics/, last accessed: 20.06.2024

consecutive errors and additionally for pairs of events where only the initial event contains an error. Examination of paired error metric combinations reveals a consistent presence of non-zero scores between Watwin and RR. The number of non-zero scores remains identical between EQ & Watwin and EQ & RR, as well as between RED & Watwin and RED & RR. This is attributed to RR and Watwin having an equivalent number of non-zero scores. For EQ & RED, the number of non-zero values is always equal to that of RED.

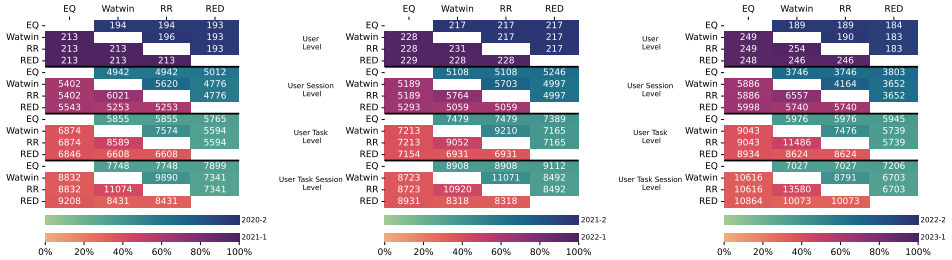


Fig. 1: Number of data points for error metric tuples with errors (non-zero scores) across all levels.

The pearson correlation coefficient does not show a consistent pattern across the different years and levels. In terms of EQ, correlations with Watwin as well as with RR are higher at user level ($min = 0.42, max = 0.73, \bar{x} = 0.57, \sigma = 0.1$) compared to user session, user task and user task session levels for each year ($min = 0.32, max = 0.6, \bar{x} = 0.46, \sigma = 0.1$). Furthermore, EQ has less correlation to Watwin ($min = 0.32, max = 0.61, \bar{x} = 0.4, \sigma = 0.07$) than to RR ($min = 0.48, max = 0.73, \bar{x} = 0.58, \sigma = 0.06$) over all years and levels. Correlations between Watwin & RR are the highest ($min = 0.5, max = 0.84, \bar{x} = 0.67, \sigma = 0.07$) compared to all other error metric combinations ($min = -0.04, max = 0.73, \bar{x} = 0.3, \sigma = 0.2$) . However, there are no clear patterns within the correlations of Watwin & RR. The correlations between RED & EQ, RED & Watwin and RED & RR reveal similar results: In general, the data from the second half of the year ($min = -0.04, max = 0.44, \bar{x} = 0.09, \sigma = 0.11$) shows less correlations compared to the data from the first half of the year ($min = 0.08, max = 0.41, \bar{x} = 0.25, \sigma = 0.09$). Particularly noticeable are the values from the second half of the year between RED & EQ at user level ($min = 0.34, max = 0.44, \bar{x} = 0.38, \sigma = 0.06$), as this is significantly higher compared to the three other levels ($min = 0.03, max = 0.11, \bar{x} = 0.06, \sigma = 0.02$).

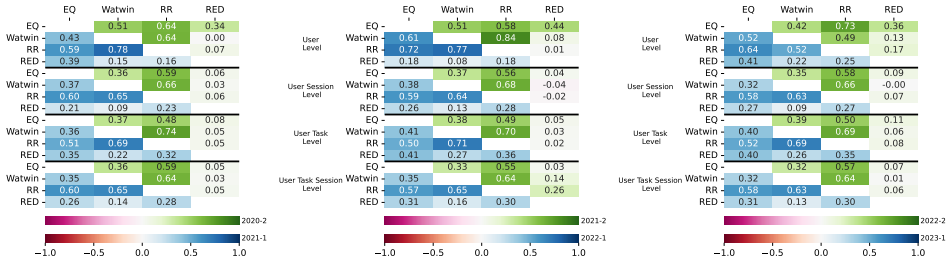


Fig. 2: Pearson correlations coefficients for all datasets over all levels.

5 Discussion

The correlation analyses show clear differences in the relationships between the error metrics derived from compilation sequences of Python CS1 code fragments. Specifically, RED consistently has nearly no correlation with the three other error metrics, diverging from prior research findings which indicated robust correlations among EQ, RED, and Watwin ($r = 0.854 - 0.988$) [Pr20]. Unlike EQ, Watwin, and RR, which attribute error points immediately upon the occurrence of distinct errors, RED assigns points only upon the recurrence of an error at least twice consecutively. Hence, the anticipated weak or negligible correlation for RED aligns with our observed results. The notable disparity in RED correlations across distinct periods (second half of the year (2020-2, 2021-2 & 2022-2) vs. the first half of the year (2021-1, 2022-1 & 2023-1)) remains unexplained. The highest correlation is observed between Watwin & RR, which aligns with expectations due to the similar calculation methods of both metrics. However, the correlation value is not consistently high enough to establish a definitive relationship between Watwin & RR. While EQ exhibits a stronger correlation with Watwin, RR, and RED at the user level compared to the three other levels, this correlation remains insufficient to reliably predict long-term associations between these metrics.

In summary, the correlations between error metrics and levels exhibit variability across the datasets, lacking consistent strong correlations. Consequently, none of the error metrics is directly predictive of any of the others. Therefore, employing a combination of error metrics may yield more targeted insights than utilizing them individually. In the context of the long-term goal to automatically identify appropriate feedback timing in Python programming novice courses, a combination of these error metrics could therefore serve as an indicator, as they reflect the development process of programming, in particular the handling of errors, and can be logged automatically without further effort by learners.

6 Conclusion and Future Work

This paper addresses the crucial aspect of providing timely feedback to novice programmers, exploring the correlation between error metrics in Python programming courses. We analyzed datasets from Python programming novices, computing error metrics across various levels to investigate their correlations. Findings indicate an increase in data points with finer granularity, alongside a relative decrease in non-zero values. The correlations between the error metrics vary between the different levels and years, with RED showing almost no correlation with other error metrics. The publicly accessible scripts[5] for error metric calculation aim to establish a standardized implementation for consistent comparison for future work. These findings establish a basis for the automated determination of feedback timing for programming novices. Future research should include conducting user studies in which learners proactively seek assistance within a programming environment. Interaction data will be logged to compute error metrics based on help request times, and thus to identify clusters indicating the need for assistance.

References

[Be16] Becker, B. A.: A New Metric to Quantify Repeated Compiler Errors for Novice Programmers. In: Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education. ITiCSE '16, Association for Computing Machinery (ACM), pp. 296–301, 2016, DOI: 10.1145/2899415.2899463.

[CAP17] Castro-Wunsch, K.; Ahadi, A.; Petersen, A.: Evaluating Neural Networks as a Method for Identifying Students in Need of Assistance. In: Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education. SIGCSE '17, Association for Computing Machinery (ACM), pp. 111–116, 2017, DOI: 10.1145/3017680.3017792.

[Co17] Costa, E. B. et al.: Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Computers in Human Behavior 73, pp. 247–256, 2017, DOI: 10.1016/j.chb.2017.01.047.

[Co23] Coelho, F. J. M. et al.: Learning Analytics in Introductory Programming Courses: a Showcase from the Federal University of Amazonas. Brazilian Journal of Computers in Education (Revista Brasileira de Informática na Educação – RBIE) 31, pp. 1089–1127, 2023, DOI: 10.5753/rbie.2023.3334.

[De23] De Freitas, A. et al.: FalconCode: A Multiyear Dataset of Python Code Samples from an Introductory Computer Science Course. In: Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1. SIGCSE '23, Association for Computing Machinery (ACM), pp. 938–944, 2023, DOI: 10.1145/3545945.3569822.

[Ga81] Gall, S. N.-L.: Help-seeking: An understudied problem-solving skill in children. Developmental Review 1 (3), pp. 224–246, 1981, DOI: 10.1016/0273-2297(81)90019-8.

[HT07] Hattie, J.; Timperley, H.: The Power of Feedback. Review of Educational Research 77 (1), pp. 81–112, 2007, DOI: 10.3102/003465430298487.

[Ja06] Jadud, M. C.: An Exploration of Novice Compilation Behaviour in BlueJ, PhD thesis, University of Kent, 2006, 268 pp., URL: https://kar.kent.ac.uk/14615/.

[JD15] Jadud, M. C.; Dorn, B.: Aggregate Compilation Behavior: Findings and Implications from 27,698 Users. In: Proceedings of the eleventh annual International Conference on International Computing Education Research. ICER '15, Association for Computing Machinery (ACM), pp. 131–139, 2015, DOI: 10.1145/2787622.2787718.

[Je22] Jeuring, J. et al.: Towards Giving Timely Formative Feedback and Hints to Novice Programmers. In: Proceedings of the 2022 Working Group Reports on Innovation and Technology in Computer Science Education. ITiCSE-WGR '22, Association for Computing Machinery (ACM), pp. 95–115, 2022, DOI: 10.1145/3571785.3574124.

[KJH19] Keuning, H.; Jeuring, J.; Heeren, B.: A Systematic Literature Review of Automated Feedback Generation for Programming Exercises. ACM Transactions on Computing Education 19 (1), pp. 1–43, 2019, DOI: 10.1145/3231711.

[KLK23] Kiesler, N.; Lohr, D.; Keuning, H.: Exploring the Potential of Large Language Models to Generate Formative Programming Feedback. In: 2023 IEEE Frontiers in Education Conference (FIE). Pp. 1–5, 2023, DOI: 10.1109/FIE58773.2023.10343457.

[Le23] Leinonen, J. et al.: Comparing Code Explanations Created by Students and Large Language Models. In: Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1. ITiCSE '23, pp. 124–130, 2023, DOI: 10.1145/3587102.3588785.

[Me22] Messer, M.: Detecting When a Learner Requires Assistance with Programming and Delivering a Useful Hint. In: Proceedings of the 15th International Conference on Educational Data Mining. International Educational Data Mining Society, pp. 778–781, 2022, DOI: 10.5281/zenodo.6852958.

[MLL07] McQuiggan, S. W.; Lee, S.; Lester, J. C.: Early Prediction of Student Frustration. In: Affective Computing and Intelligent Interaction. Springer, pp. 698–709, 2007, DOI: 10.1007/978-3-540-74889-2_61.

[MZK19] Mao, Y.; Zhi, R.; Khoshnevisan, F.: One Minute Is Enough: Early Prediction of Student Success and Event-Level Difficulty During a Novice Programming Task. In: Proceedings of the 12th International Conference on Educational Data Mining (EDM 2019). Pp. 119–128, 2019.

[Na08] Narciss, S.: Feedback Strategies for Interactive Learning Tasks. In: Handbook of Research on Educational Communications and Technology. Pp. 125–144, 2008.

[Pr20] Price, T. W. et al.: ProgSnap2: A Flexible Format for Programming Process Data. In: Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education. ITiCSE '20, Association for Computing Machinery (ACM), pp. 356–362, 2020, DOI: 10.1145/3341525.3387373.

[PSV15] Petersen, A.; Spacco, J.; Vihavainen, A.: An exploration of error quotient in multiple contexts. In: Proceedings of the 15th Koli Calling Conference on Computing Education Research. Koli Calling '15, Association for Computing Machinery (ACM), pp. 77–86, 2015, DOI: 10.1145/2828959.2828966.

[RB09] Rodrigo, M. M. T.; Baker, R. S.: Coarse-grained detection of student frustration in an introductory programming course. In: Proceedings of the fifth international workshop on Computing education research workshop. ICER '09, Association for Computing Machinery (ACM), pp. 75–80, 2009, DOI: 10.1145/1584322.1584332.

[RKJ24] Roest, L.; Keuning, H.; Jeuring, J.: Next-Step Hint Generation for Introductory Programming Using Large Language Models. In: Proceedings of the 26th Australasian Computing Education Conference. ACE '24, Association for Computing Machinery (ACM), pp. 140–153, 2024, DOI: 10.1145/3636243.3636259.

[SB12] Siemens, G.; Baker, R. S. J. d.: Learning analytics and educational data mining: towards communication and collaboration. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. LAK '12, Association for Computing Machinery (ACM), pp. 252–254, 2012, DOI: 10.1145/2330601.2330661.

[VPJ14] Vaessen, B. E.; Prins, F. J.; Jeuring, J.: University students' achievement goals and help-seeking strategies in an intelligent tutoring system. Computers & Education 72, pp. 196–208, 2014, DOI: 10.1016/j.compedu.2013.11.001.

[Wa15] Watson, C.: An Exploration of Traditional and Data Driven Predictors of Programming Performance, PhD thesis, Durham University, 2015, URL: https://etheses.dur.ac.uk/11141/.

[WLG13] Watson, C.; Li, F. W.; Godwin, J. L.: Predicting Performance in an Introductory Programming Course by Logging and Analyzing Student Programming Behavior. In: 2013 IEEE 13th International Conference on Advanced Learning Technologies. ICALT '23, pp. 319–323, 2013, DOI: 10.1109/ICALT.2013.99.

[WLG14] Watson, C.; Li, F. W.; Godwin, J. L.: No Tests Required: Comparing Traditional and Dynamic Predictors of Programming Success. In: Proceedings of the 45th ACM technical symposium on Computer science education. SIGCSE '14, Association for Computing Machinery (ACM), pp. 469–474, 2014, DOI: 10.1145/2538862.2538930.

Hackathon als Pflichtmodul im ersten Semester - Eine Fallstudie

Marc Stelzner[1], Sarah Scheer[1], Max Kaussow[1] und Stefan Fischer[1]

Abstract: Alternative Lehrformate wie Hackathons gewinnen im akademischen Lehrbetrieb zunehmend an Bedeutung, um Studierende zu erreichen und zu motivieren. Im Rahmen eines Pflichtmoduls für die Erstsemester der Informatik-Studiengänge der Universität zu Lübeck wurde ein semesterbegleitendes Projekt modernisiert und erstmalig als zweitätiger Hackathon durchgeführt, um Studierende in einem Gemeinschaftsgefühl spielerisch für das Thema Programmierung zu begeistern. In der Auswertung des Hackathons wurden dafür u.a. auch die wahrgenommene Schwierigkeit, Vorbereitung, Organisation, sowie die Auswirkung des Wettbewerbs analysiert. Auch wenn das Lehrformat dabei auf grundlegend positive Resonanz sowie den Wunsch nach Fortsetzung und Ausbau stieß, haben sich einige Fallstricke im Rahmen der angemessenen Schwierigkeit und Skalierung ergeben. Die Ergebnisse legen nahe, dass der Hackathon ein motivierendes und innovatives Werkzeug in der Lehre sein kann, jedoch eine große Planungssorgfalt erfordert, um alle Studierenden zu inkludieren.

Keywords: Fallstudie, Hackathon, Prüfung, Projekt, Alternative Lehrformate, Motivation

1 Einleitung und Motivation

In einer Welt, die zunehmend von Technologie geprägt ist, wird Programmierung zu einer essentiellen Fähigkeit für Studierende aller Fachrichtungen. Insbesondere im Bereich der Informatik stellt die Beherrschung von Programmiersprachen und algorithmischem Denken eine Grundvoraussetzung dar und wird daher bereits ab dem ersten Semester vermittelt. Dennoch gibt es viele Studierende, welche ihr Studium bereits im ersten Semester abbrechen. Auch wenn die Gründe dafür vielfältig und teilweise persönlich oder finanziell bedingt sind, so sind Gründe wie unzureichende Studienmotivation, fehlende praktische Tätigkeit sowie schlechte Studienbedingungen und -organisation [He17] Punkte, an welchen die Universität durch gut organisierten und modernen Lehrbetrieb selbst ansetzen kann.

Im Folgenden wird daher die Frage gestellt, wie der Hackathon als Pflichtformat im Rahmen des Moduls *Einführung in die Programmierung* dazu beitragen kann, Studierende des ersten Semesters für ihr Fach zu begeistern und ob dies eine zeitgemäßere Alternative zum semesterbegleitenden Projekt bietet. Dieses wurde bisher genutzt, um Studierende im Rahmen vorgegebener Entwicklungsschritte und entsprechend des Kenntnisstands zum Lehrfortschritt über Wochen hinweg ein vorgegebenes Spiel entwickeln zu lassen. Ein Hackathon, definiert als eine intensive Veranstaltung, bei der Teilnehmende innerhalb von einem

1 Universität zu Lübeck, Institut für Telematik, Ratzeburger Allee 160, 23562 Lübeck, Deutschland,
marc.stelzner@uni-luebeck.de, https://orcid.org/0009-0000-9840-2750;
sa.scheer@uni-luebeck.de, https://orcid.org/0009-0004-8441-7184;
ma.kaussow@uni-luebeck.de, https://orcid.org/0009-0003-7886-5071;
stefan.fischer@uni-luebeck.de, https://orcid.org/0000-0003-1292-8925

doi:10.18420/delfi2024_41

bis zwei Tagen ein Softwareprojekte entwickeln [Po18], bietet nun die Möglichkeit zum Abschluss praktische Erfahrungen zu sammeln, kreativ zu sein und Problemlösungsfähigkeiten zu entwickeln. Insbesondere für Studierende ohne vorherige Programmierkenntnisse kann ein Hackathon eine niederschwellige Einführung in die Welt der Softwareentwicklung bieten. Auf der anderen Seite besteht jedoch auch die Gefahr, dass er als unüberwindliche Herausforderung wahrgenommen wird, da der Kenntnisstand der Studierenden im Bereich der Programmierung gerade zu Studienbeginn stark variiert.

2 Verwandte Arbeiten

Laut Porras et al.[Po18] werden Hackathons schon seit den frühen 2000ern nicht nur von Unternehmen organisiert, um hochwertige Lösungen für Probleme entwickeln zu lassen und Nachwuchskräfte zu finden, sondern auch seit längerem in der Lehre genutzt. Die Autoren bieten eine Übersicht über die verschiedenen Ausprägungen von Hackathons, sowie deren Erfahrung in Organisation von sowie Teilnahme in Hackthons. So werden beispielsweise *Code Camps* dazu verwendet, um Studierende bereits früh im Studium im Rahmen einer freiwilligen Blockveranstaltung über Tage oder Wochen hinweg für kollaboratives Lernen und Entwickeln zusammenzubringen. Das Ziel ist dabei weniger die Leistung oder das Endprodukt, als mehr der Weg dorthin. An der LUT in Finnland wurde dies bereits in der ersten Woche des Studiums durchgeführt, um die Studierenden zu vernetzen und sie anhand erster Programmiererfahrung in einem kompetetiven Rahmen für das kommende Studium zu motivieren. Im Resümee heben sie als die größten Vorteile von Hackathon-Veranstaltungen hervor, dass neben dem Aufbau neuer Kontakte vor allem auch die Freiheiten in der Problemlösung Kreativität und neue Fähigkeiten zum Vorschein kommen lassen oder weiterentwickeln.

Steglich et al.[St20] haben die Frage gestellt, warum Studierende an einem Hackathon teilnehmen, was sie am Hackathon motiviert, was sie von diesem erwarten und was sie dort letztlich lernen. Sie stießen dabei auf ähnliche Erkenntnisse: Das Lernen von und mit Komilitonen motiviert, während dabei auch Kontakte zu den Aufgabenstellern geknüpft werden. Durch die Veranstaltung wird das Teamwork gefördert und die Ergebnisse und eventuellen Auswirkungen des Hackathon-Endprodukts motivieren. Es hilft ihnen ihre Fähigkeiten zu verbessern und sich auf den Arbeitsmarkt vorzubereiten. Die Studierenden äußern jedoch auch Sorgen, dass sie sich aufgrund des Wettbewerbs im Hackathon im Vorfeld unbehaglich fühlen, auch wenn es im Nachhinein meist nicht als so schlimm empfunden wurde. Dadurch, dass die Studierenden während der Veranstaltung von anderen Studierenden lernen, statt -wie zuvor- nur in Vorlesungen und Übungen, entwickeln sie neue Wege und Techniken Probleme anzugehen und zu lösen.

3 Aufbau des Hackathons

Im Einsatz an der Universität zu Lübeck bringt das Lehrmodul *Einführung in die Programmierung* eine Kombination besonderer Anforderungen mit sich:

- Der Hackathon soll im Laufe des ersten Semesters an Stelle eines zuvor semesterbegleitenden Projekts stattfinden. Die Programmierkenntnisse der Studierenden variieren aufgrund unterschiedlichen Vorwissens teilweise stark.
- Der Veranstaltung ist für alle vier beteiligten Studiengänge (Informatik, IT-Sicherheit, Medieninformatik sowie Robotik und Autonome Systeme) eine Pflichtveranstaltung. Das Modul wird somit nicht nur von im Vorfeld motivierten Studierenden gewählt.
- Da die Veranstaltung prüfungsrelevant ist, muss am Ende auch eine faire und individuelle Leistungsbewertung vorgenommen werden können. Die Schwierigkeit muss den zu erwerbenden Kompetenzen des Moduls entsprechen. Diese Anforderung kann dadurch auch leicht mit dem Ziel einer motivierenden Veranstaltung kollidieren.

Um das grundlegende Ziel einer gewissen Leistungsüberprüfung beizubehalten, wird eine Durchführung des Hackathons zum Ende des Semesters anvisiert. Da es sich um eine nicht freiwillig gewählte Veranstaltung handelt, müssen zudem zumutbare Zeiten für einen zwei Tage dauernden Hackathon gefunden werden. Zur individuellen Bewertung wird im Nachgang mit jeder Gruppe eine halbstünde Code-Abnahme durchgeführt, um zu überprüfen, ob alle Gruppenteilnehmer mit dem erreichten Ergebnis ausreichend vertraut sind. Da dem gesamten Projektteil ein Aufwand von zwei Kreditpunkten zugeordnet ist, wird die restliche Zeit von ungefähr 40 Stunden genutzt die Studierenden durch fünf zweiwöchentliche Übungsaufgaben auf den Hackathon vorzubereiten. Inhalt der Aufgaben ist die schrittweise Entwicklung eines Schachprogramms, jedoch vorgegeben in festen Entwicklungsschritten. Der Hackathon soll sich ebenfalls mit einem Spiel beschäftigen, um das Thema selbst nahbar und entsprechend spielerisch zu halten. Die Studierenden haben so die Möglichkeit eigentlich abstrakte Problemstellungen, wie beispielsweise die Lösung eines Schachrätsels, leicht zu verstehen, in der Realität nachzustellen und Lösungen dafür zu entwickeln. Da die Studierenden im ersten Semester jedoch noch nicht lernen, wie größere Systemarchitekturen entstehen, wird die Spieleentwicklung noch stark angeleitet und und der Hackathon in zwei Abschnitte eingeteilt. Im ersten Abschnitt setzen die Studierenden Schritt für Schritt komplexere Methoden ähnlich denen der Vorbereitungsaufgaben um, um die Funktionalität des Spiels zu gewährleisten. Aufbauend auf dieser Grundlage folgt der kreative zweite Abschnitt. Den Studierenden wird eine Vielzahl unterschiedlicher Spielsituationen in einem gegebenen Regelwerk präsentiert, deren Lösung sich zunehmend komplexer gestaltet, sowohl in Rätsellogik als auch Implementierungskomplexität. Löst eine Gruppe die Aufgabe, so erhält sie eine Punktzahl entsprechend der Qualität ihrer Lösung, welche durch die Anzahl benötigter Züge oder die Anzahl gelöster Varianten mit Zufallsfaktoren bestimmt wird. Die Punktzahl dient den Studierenden einerseits als minimale Punkteschwelle um den Hackathon zu bestehen und andererseits auch als nach oben offene Skala für den kompetitiven Anteil der Veranstaltung.

Zur technischen Umsetzung des Hackathons wurde das *VPL (Virtual Programming Lab)* Plugin des Lernmanagementsystems Moodle als gut skalierbare, kooperative und einheitliche Lösung verwendet. Aufgabenteile können mit vorgefertigten Programmcode und Testfällen hinterlegt werden, wobei diese offen verfügbare oder auch versteckte Bestandteile beinhalten könnnen. Den Studierenden wird auf diese Weise sowohl eine einfache und durch das Studium bereits vertraute Entwicklungsumgebung zur Verfügung gestellt, als auch ein durchgängig verfügbarer *Judge* -ein englischer Begriff für die automatische Prüfung von Code gegen meist geheime Testfälle- für die Bewertung der Implementierungen. Die resultierenden Punktzahlen werden von Moodle im ersten Schritt selbst akkumuliert, danach durch ein Skript als .csv ausgelesen und im Anschluss in eine Web-Applikation basierend auf Angular überführt, welche die Bestenliste in regelmäßigen Zeitabständen aktualisiert.

4 Durchführung

Im Wintersemester 2023/24 fand der Hackathon das erste Mal statt. Nach einer Auftaktveranstaltung, in der die Aufgaben vorgestellt wurden, begaben sich die Studierenden am ersten Tag in die Arbeitsräume und arbeiteten sechs Stunden lang an den Aufgaben des ersten Abschnitts zur Umsetzung der Spielfunktionalität des Brettspiels Stratego. Zu diesem Zweck wurden acht Aufgaben als konsekutive Entwicklungsschritte nacheinander zur Verfügung gestellt. Für jeden erfolgreich absolvierten Teil gab es 10 Punkte. Am zweiten Tag wurde die Veranstaltung für neun Stunden fortgesetzt. Alle Gruppen, welche den ersten Abschnitt abgeschlossen hatten, konnten nun den zweiten Abschnitt beginnen, in welchem es bis zu 31 unterschiedliche Spielsituationen in beliebiger Reihenfolge zu lösen bzw. zu gewinnen galt. Für jedes Szenario gab es weitere 10 Punkte, sodass am Ende bis zu 390 Punkte erreicht werden konnten. Während des Hackathons konnte der aktuelle Punktestand jederzeit mit Hilfe eines vorher gewählten Gruppennamens auf der Bestenliste verfolgt werden. Pausen konnten jederzeit selbstständig gemacht werden, Speisen und Getränke standen durchgehend kostenfrei zur Verfügung. Aufgrund einiger Schwierigkeiten bei einer Aufgabe des ersten Abschnitts, wurde die Bestehensgrenze von 130 Punkte auf 90 Punkte heruntergesetzt. In einer Abschlussveranstaltung wurden die drei Gruppen mit den höchsten Punktzahlen im Rahmen einer Siegerehrung geküht und die Evaluation durchgeführt.

5 Evaluation

An der abschließenden Evaluation nahmen 129 Personen teil von denen 73 männlich, 25 weiblich und 6 divers waren. Der Altersdurchschnitt lag bei 20,43 Jahren (SD = 5,31) In der Auswertung wurde der Hackathon nach verschiedenen Aspekten bewertet. Hierbei wurde eine Likert-Skala mit sechs Items (absolut ja, meist ja, eher ja, eher nein, meist nein, absolut nein) verwendet, um die Zustimmung mit einer Aussage mit ausreichender Präzision zu bestimmen [NB14]. Die Ergebnisübersicht aller Fragen ist in Tab. 1 zu sehen. Die Durchschnittsbewertung resultiert aus der Zuordnung von 1 (absolut ja) bis 6 (absolut nein).

Im Anschluss an die Skala-Bewertung erfolgte die Möglichkeit in Form von Freitexten positive und negative Aspekte der Veranstaltung hervorzuheben.

Evaluationsfrage	∅
Die Schwierigkeit der Aufgaben zur Zulassung für den Hackathon war angemessen	3,17
Der Umfang der Aufgaben zur Zulassung für den Hackathon war angemessen	2,7
Die Übungen vor dem Hackathon haben mich angemessen auf die Veranstaltung vorbereitet	3,28
Der kompetitive Anteil (das Leaderboard) hat mich während der Veranstaltung motiviert	3,71
Die Schwierigkeit der Aufgaben zum Erlangen der Mindestpunktzahl im Hackathon selbst war angemessen	3,69
Der Umfang der Aufgaben zum Erlangen der Mindestpunktzahl im Hackathon selbst war angemessen	3,39
Die Hilfestellung durch Bilder, Texte und Mitarbeiter während des Hackathons war angemessen	2,79
Die Hackathonaufgaben als Team zu lösen, fand ich angenehm	2,4
Die gebündelte Veranstaltung des Hackathons an einem Wochenende durchzuführen, empfand ich als zumutbar	2,83
Ich habe mich im Vorfeld auf das gemeinsame Event gefreut	2,95
Das Event hat mir Freude bereitet	3,33
Ich hätte statt des Hackathons lieber ein Semester-begleitendes komplexes Projekt bevorzugt	4,14

Tab. 1: Evaluationsergebnisse mit durchschnittlicher Bewertung (∅)

Die ersten Fragen zielten auf die Vorbereitungsaufgaben und deren Bewertung ab. Sowohl die Schwierigkeit als auch der Umfang wurde als angemessen empfunden. Weiterhin haben sich die Studierenden durch die Aufgaben gut auf die Veranstaltung vorbereitet gefühlt. Bei dem Hackathon selbst wurde der Umfang, genau wie die Hilfestellung durch Bilder, Texte und Mitarbeitende als angemessen bewertet. Die Schwierigkeit wurde mit einem Durchschnittswert von 3,69 als eher zu hoch wahrgenommen. Aus den Freitexten lässt sich dies hauptsächlich auf eine konkrete Aufgabe zurückführen, die für viele Gruppen ein großes oder kaum lösbares Hindernis darstellte. Die Überprüfung der korrekten Bewegung einer Figur, vergleichbar mit dem Turm aus einem Schachspiel, war für viele Teilnehmende schwierig. Die Überprüfung bestand aus vielen Zwischenschritten der Überprüfung, ob das Zielfeld noch im Spielfeld lag, ob das Zielfeld zu betreten war, ob Hindernisse wie andere Figuren auf dem Weg standen und ob die zu ziehende Figur überhaupt dem richtigen Spieler gehört. Weiterhin waren die Testfälle aus Sicht der Studierenden teilweise nicht klar genug, um genügend Anhaltspunkte für Lösungen von Seiten der Studierenden zu bieten. Dies machte eine gesteigerte Hilfestellung vor Ort nötig, um die Studierenden bei Ihrer Fehlersuche zu unterstützen, was die vorhandenen Kapazitäten stark beanspruchte.

Aus der Evaluation lässt sich ebenfalls schließen, dass die Veranstaltung an einem Wochenende als gebündeltes Event zumutbar war. Die Bestenliste mit der Live-Punktzahl wurde jedoch von vielen als nicht motivierend wahrgenommen (3,71), da diese eher dauerhaften Druck ausübte.

Die Evaluation ergibt, dass die Studierenden gerne im Team gearbeitet haben und das

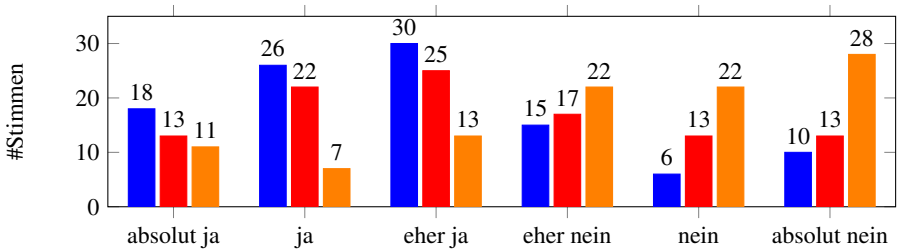


Abb. 1: Detailierte Ergebnisse von drei Fragen aus der Evaluation: Blau: *Ich habe mich im Vorfeld auf das gemeinsame Event gefreut.* Rot: *Das Event hat mir Freude bereitet.* Orange: *Ich hätte statt des Hackathons lieber ein Semester-begleitendes komplexes Projekt bevorzugt.*

Lösen von Aufgaben in Zweiergruppen als angenehm empfunden haben. Beim Vergleich der Vorfreude auf die Veranstaltung und dem Eindruck danach, lässt sich eine geringe Verschlechterung erkennen. Lag der Ausgangswert vor dem Event bei 2,95 und wurde somit mit eher ja bewertet, so hat er sich nach dem Hackathon um 0,38 verschlechtert, wurde aber immer noch als eher ja bewertet (siehe Abb. 1). Grundsätzlich wurde die Veranstaltung jedoch positiv angenommen. Bei der Frage, ob die Studierenden statt eines Hackathons lieber ein semesterbegleitendes Projekt bevorzugt hätten, kam es mit einem Wert von 4,14 zu einer Verneinung. Die meisten Befragten fanden den Hackathon dementsprechend besser. Dies zeigt sich auch in einer Frage, welche die Studierenden nach einer Note für die Veranstaltung fragt. Hier wurde eine 2,77 vergeben, wobei das semesterbegleitende Projekt im Vorjahr noch mit einer 3,26 bewertet wurde.

6 Zusammenfassung und Ausblick

Die Idee eines Hackathons, Organisation, technische Umsetzung sowie Unterstützung durch das Personal wurden in der Evaluation durch die Studierenden nahezu durchweg lobend erwähnt. Die bessere Note im Vergleich zum Vorjahr zeigt auch, dass das Projekt als Hackathon durchgeführt auf mehr Zuspruch durch die Studierenden trifft.
Dass viele Gruppen während des Hackathons ohne ausreichende Unterstützung durch die offiziellen Helfer für einige Stunden an der Lösung einer spezifischen Aufgabe scheiterten, zeigt, wie schnell die Motivation zu Frustration wechseln kann. Dieses Problem kann zukünftig durch unterschiedliche Maßnahmen verhindert werden: Eine Aufteilung von Aufgaben in atomarere Bestandteile birgt zwar die Gefahr den Blick für größere Zusammenhänge zu verlieren, führt jedoch häufiger zu Punkten und Erfolgserlebnissen. Differenziertere und ausführlicher kommentierte Rückmeldungen können die Studierenden bei der Fehlersuche besser unterstützen. An dieser Stelle kann auch die Bereitstellung einer Schritt für Schritt Kurzanleitung zur richtigen Fehlersuche unterstützen, damit Studierende sich besser selbst zu helfen wissen und das Personal entlastet wird. Eine Ersetzung des Reihenfolgezwangs bei der Bearbeitung der Aufgaben des ersten Abschnitts durch eine Empfehlung löst das Problem des Flaschenhalses und gibt den Studierenden mehr Kontrolle über ihren Fortschritt.

Literaturverzeichnis

[He17] Heublein, U.; Ebert, J.; Hutzsch, C.; Isleib, S.; König, R.; Richter, J.; Woisch, A.: Zwischen studienerwartungen und studienwirklichkeit. Ursachen des Studienabbruchs, beruflicher Verbleib der Studienabbrecherinnen und Studienabbrecher und Entwicklung der Studienabbruchquote an deutschen Hochschulen 1, 2017.

[NB14] Nemoto, T.; Beglar, D.: Likert-scale questionnaires. In: JALT 2013 conference proceedings. S. 1–8, 2014.

[Po18] Porras, J.; Khakurel, J.; Ikonen, J.; Happonen, A.; Knutas, A.; Herala, A.; Drögehorn, O.: Hackathons in software engineering education: lessons learned from a decade of events. In: Proceedings of the 2nd international workshop on software engineering education for Millennials. S. 40–47, 2018.

[St20] Steglich, C.; Salerno, L.; Fernandes, T.; Marczak, S.; Dutra, A.; Bacelo, A. P.; Trindade, C.: Hackathons as a pedagogical strategy to engage students to learn and to adopt software engineering practices. In: Proceedings of the XXXIV Brazilian Symposium on Software Engineering. S. 670–679, 2020.

Enhancing Parallel Programming Education with High-Performance Clusters Utilizing Performance Analysis

Anna-Lena Roth [1], David James [1], Michael Kuhn [2], and Johannes Konert [1]



Abstract: The use of high-performance clusters in parallel programming education is not trivial. Not only working on a cluster can be a challenge. Using clusters requires the development of complex parallel programs that are executed with many processes. This increases the necessity for performance analysis to determine bottlenecks, identify optimization opportunities, and evaluate scalability. For learners, performance analysis is a crucial aspect of understanding parallel programs and systems. However, due to its complexity and the prior knowledge required, learners are often overwhelmed by the analysis and interpretation of performance metrics. We conducted a usability test to identify barriers and issues of using a high-performance cluster in parallel programming education in conjunction with professional performance analysis using the tools Score-P, Scalasca, and Cube. At the same time, we tested EduMPI, a novel learning support tool that simplifies the process of executing parallel programs and automates performance analysis at runtime.

Keywords: HPC, MPI, Education, Parallel Programming, Cluster, Performance Analysis

1 Introduction

The ever-increasing demand for computing power in the scientific and technological environment is leading to a growing significance of High-Performance Computing (HPC). It is therefore becoming increasingly important to teach distributed and parallel programming and, if feasible, to integrate corresponding courses into undergraduate degree programs [Jo13; Pr23]. However, integrating HPC into the curriculum also challenges universities and teachers. To understand the possibilities and advantages of HPC and parallel programming, students must be given access to high-performance clusters. Some universities do not have the necessary resources, so students are expected to develop parallel code on simple computers or laptops, while knowledge about HPC systems can only be taught theoretically. Students develop only "small toy programs" in exercises [Be05]. At Fulda University of Applied Sciences, this was also the case, which meant that students had no realistic, practical experience after completing the course.
To address this issue, a local HPC cluster must be integrated into the learning environment. While allowing students to gain practical experience with HPC systems, employing a cluster in education introduces additional challenges. Using a cluster remotely adds a layer of

1 Hoschschule Fulda, University of Applied Sciences, Applied Computer Science, Leipziger Str. 123, 36037 Fulda, Germany, anna-lena.roth@cs.hs-fulda.de, https://orcid.org/0000-0002-6463-3486; david.james@cs.hs-fulda.de, https://orcid.org/0009-0009-0625-6297; johannes.konert@cs.hs-fulda.de, https://orcid.org/0000-0003-0022-535X
2 Otto von Guericke University Magdeburg, Faculty of Computer Science, Universitaetsplatz 2, 39106 Magdeburg, Germany, michael.kuhn@ovgu.de, https://orcid.org/0000-0001-8167-8574

doi:10.18420/delfi2024_42

complexity for learners. In addition, learning parallel programming does not mean simply parallelizing serial programs familiar to students [Si20]. Instead, it is necessary to develop particular parallel strategies and algorithms and to *think in parallel*, which many students find difficult [KBS13]. Special attention must be paid not only to error-free code but also to minimizing bottlenecks and optimizing runtime or resource usage, as algorithms may perform differently on various systems. Performance analysis tools are, therefore, frequently used instruments that help professional developers optimize their programs and avoid errors. In our previous work, professional performance analysis tools for parallel Message Passing Interface (MPI) programs were examined, and their use in education was discussed [RS23]. This paper explains why visualizing performance metrics, in particular, can contribute significantly to learners' motivation and better understanding. However, considerable prior knowledge of their operation and concepts such as instrumentation, measurement, profiling, and tracing is required, which makes them difficult to use in education as they can quickly overwhelm students [RS23]. Studies have documented the integration of such professional tools in educational settings [DM10; Ma19]. Due to their complexity, these tools are only used at the end of a course when students have the necessary prior knowledge. Malakar reports problematic waiting times during the execution and analysis of programs, which were criticized by students during the subsequent evaluation [Ma19]. Zhang et al. recognized the difficulty of working on a supercomputing system to execute parallel programs and, therefore, developed a corresponding online tool that automatically executes the program and provides features for subsequent performance analysis [Zh18]. In general, however, compared to tools that support the teaching of serial programming, there are hardly any tools that have been developed, especially for learning parallel programming [Ra20].

We aim to make it easier for students to run parallel programs on real high-performance clusters. They should possess a clear understanding of the processes that take place within clusters during program execution, and always keep performance in mind when developing code. In addition to learning MPI as a standard for distributed memory parallelization, performance analysis should become an integral part of development from the first MPI program. To this end, this study evaluates the operation of Fulda University's cluster via SSH in conjunction with a professional performance analysis using Score-P [Kn12], Scalasca [Ge10] and Cube [Sa15]. This is in contrast to EduMPI, which we are currently developing. It enables program execution and performance analysis entirely through a user interface.

1.1 EduMPI

EduMPI is a fork of Open MPI [Ga04], a commonly used open-source MPI implementation. In addition to the functionality provided by Open MPI, EduMPI automatically measures various performance data and metrics at runtime and sends them to a time series database. To provide a more intuitive user experience, EduMPI also includes a Graphical User Interface (GUI) where MPI programs can be executed on a cluster without further knowledge. Within

this GUI, various performance metrics are then visualized at runtime using the data collected in the database. With EduMPI, students no longer have to carry out separate instrumentation of source code and measurement of metrics. In this way, it differs significantly from many well-known performance analysis tools, which, with a few exceptions, are based on post-mortem visualization and do not support real-time analyses.

2 User tests to evaluate the usability and workflow of running and analyzing parallel programs on a high-performance cluster

To integrate the high-performance cluster as an inherent part of parallel programming education and, if possible, to use it from the first MPI application programmed, it is necessary to check how well and intuitively students can carry out program execution and performance analysis and what barriers arise here. In this context, we are investigating whether using EduMPI can improve the user experience and achieve the same goal. Two different usability tests were carried out for this purpose. Test 1 examines the usability of the conventional use of a cluster with subsequent performance analysis using Score-P, Scalasca, and Cube. Test 2 checks the usability of an alpha version of EduMPI. Our primary focus was on the usability of the tasks required to execute a program on a high-performance cluster and obtain visual analysis results, particularly the independent execution of instrumentation and measurement. The results and questionnaires of the usability tests are published and freely accessible [Ro24].

2.1 Requirements and preparations for usability tests

A total of 26 students took part in the tests. Both test versions were run a total of seven times (six teams of two and one individual). The students were randomly allocated to the two test versions. Students in teams were encouraged to collaborate on the given tasks. The usability test took place in the winter semester 2023/24 at the end of the course Parallel Programming [Fu20], which was attended by all participants, ensuring that they were familiar with MPI at a basic level and were given an introduction to performance analysis with Tuning and Analysis Utilities (TAU) [SM06]. So far, they had only run parallel MPI programs on single computers in the university laboratory (maximum of eight cores) and had not yet utilized high-performance clusters. All students were provided with an MPI program that they had previously encountered during the lecture. This program should be executed on the cluster, possibly several times, using 400 parallel processes, each on a separate core. In each test, a performance analysis should be carried out using a tool previously unknown to them. To ensure comparability, the participants of both test versions were given identical questionnaires to complete before and after the test [Ro24]. The pre-test questionnaire comprised two closed-ended questions and three Likert scale questions, which asked about the participant's level of experience, interest, and respective self-assessment with the topics of parallel programming and performance analysis. The post-test questionnaire consisted of

six free-text questions, including three additional Likert scale questions used to evaluate the previously conducted usability test. The participants were asked to compare the work on the cluster and the performance analysis with the workflow they were familiar with from the course and to rate what they found easy or challenging during the respective user test.

2.2 Test 1: Conventional cluster connection and performance analysis with Score-P, Scalasca, and Cube

An SSH connection must be established via a terminal to access the cluster conventionally. Participants were given information about the connection and the corresponding access. The program to be executed was located on the participant's local computer and had to be copied to the cluster accordingly. Steps such as compiling and executing the program had to be carried out manually using commands, as did the performance analysis, consisting of instrumentation and measurement with Score-P and Scalasca. All tools necessary to perform these tasks were pre-installed on the respective systems, and guidance materials were provided. The corresponding profiling files had to be copied from the cluster back to the local computer and visualized via Cube to obtain a performance report on the measured metrics. This local use of Cube follows the officially recommended remote scenario [Mo21]. In the first step, students had to solve the tasks independently by applying the knowledge they had acquired during the course and using the provided guidance material. If the students took a completely wrong approach or could not solve a task independently, the moderator intervened with help and hints so that the test could be continued. On such occasions, students were also allowed to use Google Search. If this also did not help, the moderator provided the students with so-called *help cards*, which contained explicit commands needed to complete a task.

2.3 Test 2: Cluster connection and performance analysis via EduMPI

In Test 2, the connection and control on the cluster and the performance analysis were carried out purely using an alpha version of EduMPI. In advance, participants received an info sheet with instructions on accessing the cluster and the database used in EduMPI. All tasks could be carried out within the GUI of EduMPI. Users initially had to establish a connection to the database, where performance data and metrics were stored at runtime. Participants then generated a bash script within the GUI by entering the student's credentials for the cluster and selecting a locally stored program. The script facilitates students to select a varying number of processes and execute the program on the cluster. In contrast to Test 1, performance analysis was carried out automatically, metrics were stored in the database, and once data was available there, real-time visualizations were presented. The participants were required to use the GUI intuitively. Similar to Test 1, the moderator intervened in the form of hints when students could not solve tasks independently. Help cards and Google Search were unnecessary because there was no need to use a console or write commands.

2.4 Results of the usability test

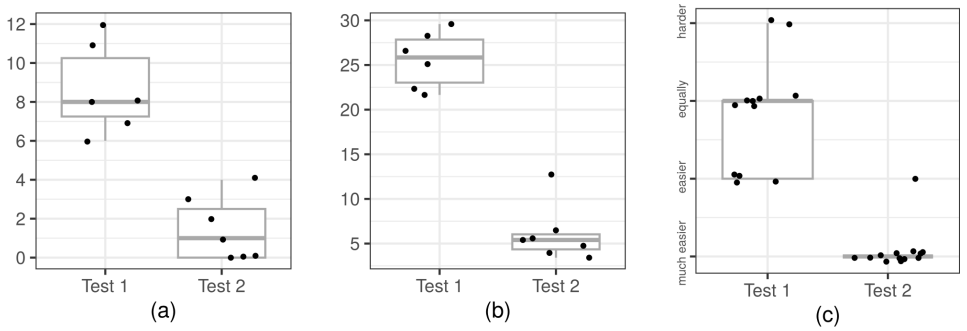


Fig. 1: (a) number of assistance instances, (b) time until first visualizations (in minutes), (c) difficulty

Test 1 had to be aborted once at a participant's request. In both test variants, each assistance was provided by the moderator via the help cards, or hints to the next step, and the number of Google searches was counted as assistance instances until a performance visualization was produced. In addition, the time it took until the first visualization of performance metrics was measured. Figure 1(a) shows a combined box and scatter plot of the number of these assistance instances grouped by test variant. During Test 1, assistance was required a median of 8 times, with a maximum of 12 and a minimum of 6 times. In contrast, in Test 2, the moderator intervened a median of only 1 time and a maximum of 4 times, while three groups did not need any assistance at all. This is also reflected in the time each group of students took to produce the first visualization, depicted in a similar plot in Figure 1(b). In Test 1, it took a median of 25.8 minutes to produce the first visualization, with a minimum of 21.7 minutes and a maximum of 29.6 minutes. In contrast, it only took a median of 5.4 minutes with a minimum of 3.4 minutes and a maximum of 12.7 minutes in Test 2. Fig. 1(c) shows the result of one Likert scale question from the questionnaire following the test, where students were asked to assess how difficult or easy they found running the program on the cluster compared to running programs on a single computer in practical sessions. In Test 1, the individual students' answers range from easier to harder, where most of the students assess working with the cluster as equally difficult. In Test 2, twelve students feel that working with the cluster is much easier, while one assesses it as easier.

2.5 Discussion

Performance analysis was introduced and explained to the students at the end of the course Parallel Programming. However, during the usability test, the students had different perceptions about it. As part of the post-test questionnaire, the students were asked if they would like to use performance analysis for developing parallel programs in the future and if the topic should have been introduced earlier in the course. Some students recognized the potential of performance analysis and felt it improved their understanding of executing

programs on the cluster. On the other hand, six students felt that performance analysis was unnecessary or too difficult to understand and suggested that it be covered as a bonus topic, if at all, or only at the end of the course. This indicates that not all students recognize the importance of performance analysis in developing parallel programs. During Test 1, there were frequent issues with the commands being applied. Some students had trouble compiling and executing MPI programs, but the real issues arose with the commands for instrumentation and measurement, which required extensive analysis as outlined in the guidance material and file transfers to and from the cluster. Particularly when commands caused errors, the students were often unable to solve the problems independently, so the moderator had to intervene by providing the correct solution via a help card. In Test 2, participants did not need to write any commands themselves. Everything could be operated and controlled through EduMPI, resulting in fewer inquiries, and the moderator only had to intervene rarely. When extrapolating this test to a general teaching scenario, it is important to consider that the participants in this test had prior knowledge of the workflow and most of the commands. In practical sessions with beginners, it is expected that more questions will be asked due to their lack of experience. In Test 1, most participants found the execution of a program equally difficult/easy as on a single computer. This is because the mere compilation and execution of programs on local systems hardly differs from the workflow on the cluster. In contrast, participants in Test 2 found running programs on the cluster much easier. Executing a program via a GUI seems to simplify the workflow immensely. In Test 2, the participants proved to be keen to experiment and usually ran the program several times in succession with a different number of processes and drew comparisons between the visualizations. In Test 1, on the other hand, the complete execution and analysis process was only run through once by all students due to time constraints, and thus, only one visualization was generated. Quick visual feedback can increase students' motivation and interest. EduMPI saves students the lengthy process of instrumentation and measurement. In addition, data is displayed at runtime and not only post-mortem. In practical sessions, students with EduMPI potentially have fewer problems and queries, so less time is needed.

3 Conclusion and future work

Overall, the usability tests indicate that the conventional use of clusters alongside professional performance analysis in education is feasible, but raises many questions and requires assistance, particularly in relation to instrumentation and measurement. This can mean that learners often have to wait a long time for help from teachers in practical sessions. Not only was less help required when using EduMPI, but the automated performance analysis and visualization also provided learners with faster feedback, potentially sparking their interest in running a program multiple times with a different number of processes and comparing the results. EduMPI introduces a new concept for parallel programming education by incorporating performance analysis right from the initial development of an MPI program. Additionally, it simplifies the usual workflow for running and analyzing a program on a high-performance cluster.

References

[Be05] Bernreuther, M.; Brenk, M.; Bungartz, H.; Mundani, R.; Muntean, I. L.: Teaching High-Performance Computing on a High-Performance Cluster. In (Sunderam, V. S.; van Albada, G. D.; Sloot, P. M. A.; Dongarra, J. J., eds.): Computational Science - ICCS 2005, 5th International Conference, Atlanta, GA, USA, May 22-25, 2005, Proceedings, Part II. Vol. 3515. Lecture Notes in Computer Science, Springer, pp. 1–9, 2005, DOI: 10.1007/11428848_1.

[DM10] Delistavrou, C. T.; Margaritis, K. G.: Survey of Software Environments for Parallel Distributed Processing: Parallel Programming Education on Real Life Target Systems Using Production Oriented Software Tools. In: 14th Panhellenic Conference on Informatics, PCI 2010, Tripoli, Greece, September 10-12, 2010. IEEE Computer Society, pp. 231–236, 2010, DOI: 10.1109/PCI.2010.26.

[Fu20] Fulda, University of Applied Sciences: Parallel Programming (AI5085), Examination Regulations, M. Sc. Global Software Development, 2020, URL: https://www.hs-fulda.de/fileadmin/user_upload/Unsere_Hochschule/Hochschulrecht/Pruefungsordnungen/Angewandte_Informatik/MSc_GSD_2020_Ae2022-23_LF.pdf, visited on: 06/19/2024.

[Ga04] Gabriel, E.; Fagg, G. E.; Bosilca, G.; Angskun, T.; Dongarra, J. J.; Squyres, J. M.; Sahay, V.; Kambadur, P.; Barrett, B.; Lumsdaine, A.; Castain, R. H.; Daniel, D. J.; Graham, R. L.; Woodall, T. S.: Open MPI: Goals, Concept, and Design of a Next Generation MPI Implementation. In (Kranzlmüller, D.; Kacsuk, P.; Dongarra, J. J., eds.): Recent Advances in Parallel Virtual Machine and Message Passing Interface, 11th European PVM/MPI Users' Group Meeting, Budapest, Hungary, September 19-22, 2004, Proceedings. Vol. 3241. Lecture Notes in Computer Science, Springer, pp. 97–104, 2004, DOI: 10.1007/978-3-540-30218-6_19.

[Ge10] Geimer, M.; Wolf, F.; Wylie, B. J. N.; Ábrahám, E.; Becker, D.; Mohr, B.: The Scalasca performance toolset architecture. Concurr. Comput. Pract. Exp. 22 (6), pp. 702–719, 2010, DOI: 10.1002/cpe.1556.

[Jo13] Joint Task Force on Computing Curricula, Association for Computing Machinery (ACM) and IEEE Computer Society: Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science. Association for Computing Machinery, New York, NY, USA, 2013, ISBN: 9781450323093.

[KBS13] Ko, Y.; Burgstaller, B.; Scholz, B.: Parallel from the beginning: the case for multicore programming in thecomputer science undergraduate curriculum. In (Camp, T.; Tymann, P. T.; Dougherty, J. D.; Nagel, K., eds.): The 44th ACM Technical Symposium on Computer Science Education, SIGCSE 2013, Denver, CO, USA, March 6-9, 2013. ACM, pp. 415–420, 2013, DOI: 10.1145/2445196.2445320.

[Kn12] Knüpfer, A.; Rössel, C.; Mey, D. a.; Biersdorff, S.; Diethelm, K.; Eschweiler, D.; Geimer, M.; Gerndt, M.; Lorenz, D.; Malony, A.; Nagel, W. E.; Oleynik, Y.; Philippen, P.; Saviankou, P.; Schmidl, D.; Shende, S.; Tschüter, R.; Wagner, M.; Wesarg, B.; Wolf, F.: Score-P: A Joint Performance Measurement Run-Time Infrastructure for Periscope,Scalasca, TAU, and Vampir. In (Brunst, H.; Müller, M. S.; Nagel, W. E.; Resch, M. M., eds.): Tools for High Performance Computing 2011. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 79–91, 2012, DOI: 10.1007/978-3-642-31476-6_7.

[Ma19] Malakar, P.: Experiences of Teaching Parallel Computing to Undergraduates and Post-Graduates. In: 26th International Conference on High Performance Computing, Data and Analytics Workshop, HiPCW 2019, Hyderabad, India, December 17-20, 2019. IEEE, pp. 40–47, 2019, DOI: 10.1109/HiPCW.2019.00020.

[Mo21] Mohr, B.: Installing POP Tools: Score-P, Scalasca, Cube, 2021, URL: https://pop-coe.eu/sites/default/files/pop_files/pop-onlinetraining-install-jsc-tools.pdf, visited on: 04/02/2024.

[Pr23] Prasad, S. K.; Weems, C. C.; Sussman, A.; Gupta, A.; Estrada, T.; Vaidyanathan, R.; Ghafoor, S. K.; Kant, K.; Stunkel, C. B.: NSF/IEEE-TCPP Curriculum on Parallel and Distributed Computing for Undergraduates - Version II - Big Data, Energy, and Distributed Computing. In (Doyle, M.; Stephenson, B.; Dorn, B.; Soh, L.; Battestilli, L., eds.): Proceedings of the 54th ACM Technical Symposium on Computer Science Education, Volume 2, SIGCSE 2023, Toronto, ON, Canada, March 15-18, 2023. ACM, pp. 1220–1221, 2023, DOI: 10.1145/3545947.3569594.

[Ra20] Raj, R. K.; Romanowski, C. J.; Impagliazzo, J.; Aly, S. G.; Becker, B. A.; Chen, J.; Ghafoor, S.; Giacaman, N.; Gordon, S. I.; Izu, C.; Rahimi, S.; Robson, M. P.; Thota, N.: High Performance Computing Education: Current Challenges and Future Directions. In: Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education. ITiCSE-WGR '20, Association for Computing Machinery, Trondheim, Norway, pp. 51–74, 2020, DOI: 10.1145/3437800.3439203.

[Ro24] Roth, A.-L.: Usability Test Data, conventional Performance Analysis vs. EduMPI, 2024, URL: https://anonymous.4open.science/r/Usability_Test_Data_Open_Access-7160/, visited on: 04/09/2024.

[RS23] Roth, A.-L.; Süß, T.: Performance Analysis Tools for MPI Applications and their Use in Programming Education. In: Companion of the 2023 ACM/SPEC International Conference on Performance Engineering. ICPE '23 Companion, Association for Computing Machinery, Coimbra, Portugal, pp. 361–368, 2023, DOI: 10.1145/3578245.3584358.

[Sa15] Saviankou, P.; Knobloch, M.; Visser, A.; Mohr, B.: Cube v4: From Performance Report Explorer to Performance Analysis Tool. Procedia Computer Science 51, International Conference On Computational Science, ICCS 2015, pp. 1343–1352, 2015, DOI: 10.1016/j.procs.2015.05.320.

[Si20] Sitsylitsyn, Y.: Methods and tools for teaching parallel and distributed computing in universities: a systematic review of the literature. SHS Web of Conferences 75, p. 04017, 2020, DOI: 10.1051/shsconf/20207504017.

[SM06] Shende, S.; Malony, A. D.: The Tau Parallel Performance System. Int. J. High Perform. Comput. Appl. 20 (2), pp. 287–311, 2006, DOI: 10.1177/1094342006064482.

[Zh18] Zhang, Y.; Li, J.; Wu, D.; Du, Y.: Improving Student Skills on Parallel Programming via Code Evaluation and Feedback Debugging. In: IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2018, Wollongong, Australia, December 4-7, 2018. IEEE, pp. 1069–1073, 2018, DOI: 10.1109/TALE.2018.8615351.

Poster-Beiträge

KI-basierte Unterstützung zum Abbau sprachlicher Barrieren für Kinder mit nichtdeutscher Herkunftssprache

Kensuke Akao[1,2]

Abstract: Seit einigen Jahren steigt der Anteil der Kinder und Jugendlichen mit nichtdeutscher Herkunftssprache in Deutschland weiter an. Damit sie in einer deutschen Regelschule gut integriert werden können, ist der Abbau sprachlicher Barrieren für das Lernen eine zentrale Aufgabe. Dabei ist es wünschenswert, dass die Chancengleichheit zur Unterstützung unabhängig von der Herkunft aller Kinder gewährleistet werden kann. Der Einsatz von der KI-Übersetzung wäre ein möglicher Ansatz, dies zu erreichen. Jedoch sollte die Qualität der Übersetzungen zu pädagogischen Zwecken und die Auswirkung des zweisprachigen Zugangs zu Fachinhalten auf das Lernen von Kindern noch sorgfältig untersucht werden. Unsere Ansätze sind: 1) die Genauigkeit der von unterschiedlichen KI-Diensten generierten Übersetzungen eines kompletten Informatiklehrbuchs in mehrere Sprachen zu überprüfen; 2) eine wöchentliche Nachhilfe für Informatik unter Verwendung der mittels *Post-Editing* korrigierten Übersetzungen an Kooperationsschulen umzusetzen, um ihre Nützlichkeit langfristig zu analysieren.

Keywords: Inklusion, mehrsprachiger Zugang, Schulinformatik, Künstliche Intelligenz, Post-Editing

1 Einleitung

Stand 2022 leben ca. 13,4 Millionen Ausländer:innen in Deutschland und diese Anzahl ist fast um das 1,5-fache höher als im Jahr 2015 [St23a]. In diesem Zusammenhang haben viele Kinder und Jugendliche mit Migrationshintergrund an deutschen Regelschulen – in Baden-Württemberg (BW) sind es 15,3 % [St23b] – eine nichtdeutsche Herkunftssprache. Nach Cummins lässt sich der Erwerb von Sprachkenntnissen in die Kommunikationsfähigkeit in alltäglichen Unterhaltungen (BICS[3]) und die Kognitionsfähigkeit zur Interpretation der Lerninhalte (CALP[4]) unterteilen; im Allgemeinen erwerben Kinder und Jugendliche innerhalb von 18 Monaten bis zwei Jahren nach ihrer Ankunft im Zielland zwar ein ausreichendes BICS; es dauert aber durchschnittlich fünf bis sieben Jahre, damit sie eine altersgemäße CALP erreichen [Cu81; Cu84]. Die Kinder und Jugendlichen mit nichtdeutscher Herkunftssprache werden in BW in eine reguläre Klasse integriert, nachdem sie ca. 2 Jahre an einem DaF-/DaZ-Kurs (Deutsch als Fremd-/Zweitsprache) z. B. in Vorbereitungsklassen (VKL) teilgenommen haben [Mi17]. Das heißt, sie sind normalerweise im Vergleich zu Kindern mit deutscher Herkunftssprache beim Lernen benachteiligt, wenn sie in den VKL nur BICS und nur bedingt CALP erworben haben.

1 Pädagogische Hochschule Ludwigsburg, Institut für Informatik, Reuteallee 46, 71634 Ludwigsburg, kensuke.akao@ph-ludwigsburg.de,
2 Universität Stuttgart, Institut für Software Engineering, Softwarequalität und -architektur, Universitätsstraße 38, 70569 Stuttgart / Institut für Erziehungswissenschaft, Lehren und Lernen mit intelligenten Systemen, Geschwister-Scholl-Str 24D, 70174 Stuttgart, kensuke.akao@iste.uni-stuttgart.de,
3 Basic Interpersonal Communicative Skills
4 Cognitive Academic Language Proficiency

doi:10.18420/delfi2024_43

In diesem Zusammenhang erwartet die aktuelle Bildungspolitik beispielsweise, dass eine offensive Einstellung von Lehrkräften mit Migrationshintergrund zur Förderung dieser Kinder (und/oder Jugendlichen) beitragen könnte [Sy19]. Diese freiwillige und personenabhängige Unterstützung kann jedoch leider nicht für Kinder aller Nationalitäten in allen Schulen gleichermaßen gewährleistet werden. Damit wir eine von der Herkunft der Kinder unabhängigen Unterstützung beim Lernen schaffen, planen wir, menschliche Unterstützung durch Künstliche Intelligenz (KI) zu ergänzen. Die Weiterentwicklung der KI als Chance zur Erfüllung der inklusiven und gerechten Bildung wurde auf dem Treffen der G7-Bildungsminister:innen am 13.05.2023 anerkannt [ME23]. KI kann mit automatisch erstellten Übersetzungen dazu beitragen, die Sprachbarrieren für Menschen zu beseitigen, die die deutsche Sprache noch nicht (ausreichend) beherrschen [DH21]. Zum Beispiel veröffentlicht das *Deutsche Schulportal* der Robert Bosch Stiftung[5] Tipps zum Einsatz der KI-Übersetzung am Beispiel der Aufnahme ukrainischer Kinder in deutschen Regelschulen.

Dennoch ist bekannt, dass die heutige KI auf Basis der neuronalen maschinellen Übersetzung aufgrund verschiedener ungelöster Probleme auch Fehlübersetzungen von langen Fachtexten generiert [KK17; ZN17]. Problematisch ist, dass deutsche Lehrkräfte höchstwahrscheinlich keine Möglichkeit haben, die Genauigkeit der von KI in verschiedenen Sprachen erzeugten Übersetzungen zu überprüfen. Hinzu kommt, dass den Ergebnissen der PISA-Studie 2022 zufolge die Lesekompetenz in den Ländern, aus denen die meisten Kinder mit ausländischer Staatsangehörigkeit in BW kommen [St23c], in vielen Fällen unter dem OECD-Durchschnitt (476 Punkte) liegt [OE23]. Wir müssen also auch davon ausgehen, dass einige Kinder die in ihre Herkunftssprache übersetzten Texte nicht im Detail verstehen könnten. Deshalb ist es noch nicht klar, ob der Einsatz einer KI-Übersetzung beim Lernen für Kinder mit nichtdeutscher Herkunftssprache vorteilhaft oder schädlich ist. Wir bringen dafür das neue Projekt *I-KASA* (Informatikunterricht mithilfe von KI-Anwendungen für Schulkinder aus dem Ausland) auf den Weg.

2 Projekt I-KASA

2.1 Forschungsziel, Forschungsfragen und Forschungsansätze

Wir zielen auf die Schaffung einer von der Herkunft unabhängigen Chancengleichheit beim Lernen in deutschen Regelschulen durch zweisprachigen Zugang zu Lehrmaterialien mithilfe von KI ab. Dies erfordert die wissenschaftliche Analyse über das Potenzial und die Grenzen der KI-Übersetzung bei einer solchen Anwendung, um eine hohe Bildungsqualität zu gewährleisten. Zu diesem Zweck beschäftigen wir uns am Beispiel des *Aufbaukurses Informatik* in Klasse 7 in BW mit folgenden Forschungsfragen:

1. Kann eine KI die Fachinhalte der Schulinformatik in verschiedenen Sprachen mit einer für den Bildungszweck ausreichenden Qualität übersetzen?

5 vgl.: https://deutsches-schulportal.de/schule-im-umfeld/praktische-tipps-fuer-schulen-die-ukrainische-kinder-aufnehmen/ (Stand: 22.03.2024)

2. Wie wirkt sich eine genaue Übersetzung des Lehrmaterials auf das Lernen von Kindern mit nichtdeutscher Herkunftssprache im Rahmen des Informatikunterrichts aus?
3. Ist eine KI für Kinder mit nichtdeutscher Herkunftssprache vorteilhaft, damit sie das gemäß den Lernzielen erwartete Fachwissen und die angestrebten Kompetenzen erwerben?

Um diese Fragen zu klären, planen wir die folgenden zwei Punkte zu überprüfen: 1) die Genauigkeit der KI-Übersetzung von schulfachspezifischen Beschreibungen der Informatik und 2) die Nutzbarkeit von einem zweisprachigen Zugang zu Fachinhalten für das Informatiklernen von Kindern. Das Projekt verfolgt einen halb-quantitativen Forschungsansatz[6] basierend auf dem *Post-Positivismus* [Gu90] mit den in den folgenden Abschnitten beschriebenen zwei Forschungsmethoden.

2.2 Forschungsmethode zur Überprüfung der Genauigkeit der KI-Übersetzung

Im ersten Teil der Studie werden wir die Genauigkeit der KI-Übersetzung eines kompletten Informatiklehrbuchs in mindestens drei Sprachen durch verschiedene KI-Dienste testen. Zuerst wird eine Referenzübersetzung des Lehrbuchs erstellt, indem die durch einen KI-Dienst generierte Übersetzung mittels *Post-Editing* [Al03] von bilingualen Personen[7], die Deutsch und die übersetzte Sprache beherrschen, korrigiert wird. Danach vergleichen wir die Genauigkeit der Übersetzungen desselben Lehrbuchs von verschiedenen KI-Diensten mittels der Referenzübersetzung. Dabei kombinieren wir die automatische Bewertung durch die Software mit der manuellen Bewertung durch zwei Muttersprachler:innen[8], wie es auch andere Projekte mit der maschinellen Übersetzung machen [Bi20]. Wir verwenden als Maßstäbe den *BLEU-Score* [Pa02] für die automatische Bewertung und eine 5-stufige Likert-Skala für die Geläufigkeit und Angemessenheit bei jedem Satz und jedem Abschnitt zur manuellen Bewertung. Nach der manuellen Bewertung führen wir mit den einzelnen Bewertenden ein Interview, um qualitative Einblicke in die Schwächen der KI-Übersetzung und Hypothesen für Verbesserungen zu gewinnen.

2.3 Forschungsmethode zur Überprüfung der Nutzbarkeit von zweisprachigem Zugang zu Fachinhalten

In dem zweiten Teil der Studie werden wir die im Abs. 2.2 erstellten Referenzübersetzungen in der Schulpraxis weiterverwenden. Zur Kooperation suchen wir mindestens zwei Regel-

6 Da die Probanden bei einem Teil unseres Projektes Schulkinder sind, werden wir unsere Erhebungen nach der erforderlichen Beratung von der Ethikkommission unserer Hochschule und mit der Genehmigung des Bildungsministeriums durchführen.

7 Der derzeitige Plan sieht vor, dass bilinguale Informatik(-lehramts-)studierende als studentische Hilfskräfte diese Aufgabe übernehmen.

8 Für diese Bewertung benötigen sie keine Deutschkenntnisse oder Informatik-Fachkenntnisse, da eine Referenzübersetzung vorliegt.

schulen, die aktiv Kinder mit nichtdeutscher Herkunftssprache aufnehmen, dann bieten wir die Übersetzungen und eine wöchentliche Nachhilfe nach dem regulären Unterricht an[9]. Wir schicken unsere ausländischen Studierenden als Tutor:innen an die Schulen, um eine Gruppe von Kindern bei der Wiederholung der auf Deutsch gelernten Inhalte in ihrer jeweiligen Herkunftssprache zu unterstützen. Allerdings ist es im Rahmen unseres Projekts nicht möglich, die Übersetzung und Nachhilfe in allen Sprachen bereitzustellen. Steht einer Gruppe von Kindern keine Übersetzung bzw. Hilfskraft in ihrer Herkunftssprache zur Verfügung, erhalten sie Nachhilfe von deutschen Studierenden. Die Nachhilfe auf Deutsch wird auch für deutsche Kinder als dritte Gruppe angeboten, um ihre Gesamtlernzeit den anderen Gruppen anzugleichen. Die Teilnahme an der Nachhilfe ist freiwillig und wird von den Eltern der Kinder nach Wunsch des Kindes und der Empfehlung der Informatiklehrkraft festgelegt. Zur Evaluation werden sowohl die Leistungsvergleiche zwischen verschiedenen Gruppen mit pseudonymisierten Datensatz als auch Interviews mit Lehrkräften sowie Tutor:innen über die Auswirkungen unseres Angebots auf den Lernfortschritt der Kinder verwendet.

3 Ausblick

Indem wir die Möglichkeiten und Grenzen der Anwendung von der KI-Übersetzung im Bildungsbereich aufzeigen, kann erfasst werden, was bei der Umsetzung dieses Ansatzes in der Schulpraxis derzeit beachtet werden sollte. Darüber hinaus können wir im Laufe der Forschung neue bilinguale Datensätze (*Parallelkorpus* [ZN17]) gewinnen, die aus dem deutschen Originaltext, Referenzübersetzungen und Beispielen falscher Übersetzungen bestehen. Wir sind der Ansicht, dass sie als überwachte Trainingsdaten zur Weiterentwicklung der für den Informatikunterricht spezifischen maschinellen Übersetzung und des maschinellen Post-Editing durch KI verwendet werden können. Alternativ kann ein Teil dieser Datensätze beim *Few-Shot Prompting* [Br20] zur Verfeinerung des Übersetzungsergebnisses von allgemeinen KI-Diensten verwendet werden. Außerdem ist es auch in unserem Interesse, dass unsere Tutor:innen, die das Nachhilfeangebot durchführen, irgendwann durch KI ersetzt werden können, um schlussendlich ein von der Herkunft der Kinder unabhängiges Unterstützungsangebot zu schaffen. Dabei könnten die Erkenntnisse, die sich aus unserer Nachhilfe ergeben, ein wichtiger Baustein sein, um die jeweilige Bedarfslage an Schulen mithilfe von E-Tutoring zielführend zu adressieren.

Derzeit wird eine Umfrage unter Schulleitungen und Fachleitungen der Informatik an Regelschulen in BW als Vorstudie durchgeführt. Wir werden auf Basis dieser Ergebnisse die Lehrbücher, Sprachen und KI-Dienste für unsere Studie auswählen. Dann beginnen wir mit der Erstellung und der Bewertung von KI-Übersetzungen. Ferner wird ab dem Schuljahr 2025/26 auch die Nachhilfe unter Verwendung der Übersetzung für Schulen angeboten werden.

9 Unser Ziel ist die Stärkung der schulischen Inklusion, daher soll der reguläre Unterricht nicht durch die Herkunftssprache des Kindes ersetzt werden. Es ist für uns wichtig, dass unser Angebot nur eine zusätzliche Maßnahme zur Anpassung sein darf.

Literaturverzeichnis

[Al03] Allen, J. H.: 16. Post-editing. In (Somers, H., Hrsg.): Computers and Translation - A translator's guide. Bd. 35, John Benjamins Publishing Company, Amsterdam, S. 297–317, 2003.

[Bi20] Bie, L.; Cerda-i-Cuco, A.; Degroote, H.; Estela, A.; Garcıa-Martınez, M.; Herranz, M.; Kohan, A.; Melero, M.; O'Dowd, T.; O'Gorman, S.; Pinnis, M.; Rozis, R.; Superbo, R.; Vasilevskis, A.: Neural Translation for the European Union (NTEU) Project. In: Proceedings of the 22nd Annual Conference of the European Association for Machine Translation. S. 477–478, 2020.

[Br20] Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; Amodei, D.: Language Models are Few-Shot Learners. In (Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M. F.; Lin, H., Hrsg.): Advances in Neural Information Processing Systems. Bd. 33, Curran Associates, Red Hook, NY, S. 1877–1901, 2020.

[Cu81] Cummins, J.: Age on Arrival and Immigrant Second Language Learning in Canada: A Reassessment. Applied Linguistics 2 (2), S. 132–149, 1981.

[Cu84] Cummins, J.: Wanted: A theoretical framework for relating language proficiency to academic achievement among b. gual students. In (Rivera, c., Hrsg.): Language Proficiency and Academic Achievement. Mulitlingual Matters, Clevedon, S. 2–19, 1984.

[DH21] Djeffal, C.; Horst, A.: Übersetzung und künstliche Intelligenz in der öffentlichen Verwaltung, de, Techn. Ber. Nr. 17, Berlin: Nationales E-Government Kompetenzzentrum e. V., 2021.

[Gu90] Guba, E. G.: The Alternative Paradigm Dialog. In (Guba, E. G., Hrsg.): The Paradigm Dialog. Sage Publications, Newbury Park, Calif, S. 17–27, 1990.

[KK17] Koehn, P.; Knowles, R.: Six Challenges for Neural Machine Translation. In: Proceedings of the First Workshop on Neural Machine Translation. Association for Computational Linguistics, Vancouver, S. 28–39, 2017.

[ME23] MEXT (Ministry of Education, Culture, Sports, Science and Technology Japan): Toyama-Kanazawa Declaration by G7 Education Ministers' Meeting, Outcome Documents, 2023, URL: https://www.mext.go.jp/en/content/20230516_mxt-kouhou02-000028270_2.pdf.

[Mi17] Ministerium für Kultus, Jugend und Sport Baden-Württemberg: Verwaltungsvorschrift des Kultusministeriums über die Grundsätze zum Unterricht für Kinder und Jugendliche mit nichtdeutscher Herkunftssprache und geringen Deutschkenntnissen an allgemein bildenden und beruflichen Schulen(Az.: 31-6640.0/908), 2017, URL: https://www.landesrecht-bw.de/bsbw/document/VVBW-VVBW000026191.

[OE23] OECD, Hrsg.: PISA 2022 Ergebnisse: Lernstände und Bildungsgerechtigkeit. wbv Media, 2023.

[Pa02] Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.-J.: BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL). Association for Computational Linguistics, Philadelphia, Pennsylvania, S. 311–318, 2002.

[St23a] Statistisches Bundesamt: Ausländische Bevölkerung nach Bundesländern und Jahren (Stand: 22.03.2024). 2023, URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Tabellen/auslaendische-bevoelkerung-bundeslaender-jahre.html.

[St23b] Statistisches Landesamt Baden-Württemberg: Ausländeranteil an allgemeinbildenden Schulen (Stand: 22.03.2024). 2023, URL: https://www.statistik-bw.de/BildungKultur/SchulenAllgem/AS_auslaender.jsp.

[St23c] Statistisches Landesamt Baden-Württemberg: Zwei Millionen Ausländerinnen und Ausländer leben in Baden-Württemberg Höchster Ausländeranteil in Pforzheim mit 30 %. Pressemitteilung 117/2023, 2023, URL: http://www.statistik-bw.de/Presse/Pressemitteilungen/2023117.

[Sy19] Syring, M.; Merk, S.; Cramer, C.; Topalak, C.; Bohl, T.: Der Migrationshintergrund Lehramtsstudierender als Prädiktor ihrer Einstellungen zu heterogenen Lerngruppen. Zeitschrift für Bildungsforschung 9 (2), S. 201–219, 2019.

[ZN17] Zakir, H.; Nagoor, M.: A Brief Study of Challenges in Machine Translation. International Journal of Computer Science Issues 14 (2), S. 54–57, 2017.

Forschungsdaten in der Bildungstechnologie: Worüber sprechen wir und was ist wann relevant?

Michael Striewe [1]

Abstract: Die Notwendigkeit der guten Dokumentation und auffindbaren Publikation von Forschungsdaten ist auch in der Bildungstechnologie geläufig. Doch was genau sind alles Forschungsdaten in der Bildungstechnologie? Und hängt die Relevanz der Daten von der Forschungsfrage ab? Dieser Posterbeitrag versucht, eine praktisch motivierte Kategorisierung von Forschungsdaten vorzunehmen, die bei der Erkennung und Benennung relevanter Daten unterstützt. Diese kann perspektivisch auch als Grundlage dienen, um Daten in einer Forschungsdateninfrastruktur wiederverwendbar abzulegen und semantisch zu beschrieben.

Keywords: Forschungsdaten, Forschungssoftware, Open Science, Metadaten, NFDIxCS

1 Hintergrund

Die FAIR-Prinzipien [Wi16] bieten Leitlinien, wie Forschungsdaten auffindbar publiziert und wiederverwendbar dokumentiert werden können. Auch in der Bildungstechnologie sind diese Prinzipien bekannt und grundsätzlich anerkannt (vertreten z. B. durch den AK Open Science der FG Bildungstechnologien in der GI [2]). Es mangelt jedoch sowohl an der breiten Publikation von Forschungsdaten im Allgemeinen [JS23] als auch an der Anerkennung für die Publikation von Software-Artefakten im Besonderen [KS22]. Ausgehend von der Annahme, dass Forschungsdaten und -software nicht vorsätzlich zurückgehalten werden, sondern vielmehr die Relevanz dieser Artefakte nicht ausreichend beachtet oder gewürdigt wird, gibt dieser Beitrag einen Überblick über potenziell relevante Artefakte im Kontext bildungstechnischer Forschung. Dies soll die systematische Erkennung, Dokumentation, Publikation und Wiederverwendung erleichtern. Im vorliegenden Status beschreibt der Beitrag ein initiales, vereinfachtes, gedankliches Framework, das als Diskussionsgrundlage dienen und in zukünftigen Arbeiten systematisch validiert und verfeinert werden soll.

2 Art und Herkunft von Artefakten in der Forschung

Artefakte entstehen in der Forschung nicht in einer bestimmten Situation oder zu einem bestimmten Zeitpunkt. Abb. 1 gibt einen Überblick über verschiedene Entitäten im Kontext bildungstechnischer Forschung, die Artefakte erzeugen. Die grundsätzliche Annahme dabei

1 Universität Duisburg-Essen, Fakultät für Informatik, Gerlingstraße 16, 45127 Essen, Deutschland, michael.striewe@paluno.uni-due.de, https://orcid.org/0000-0001-8866-6971

2 https://fg-bildungstechnologien.gi.de/fachgruppe/arbeitskreise/ak-open-science

doi:10.18420/delfi2024_44

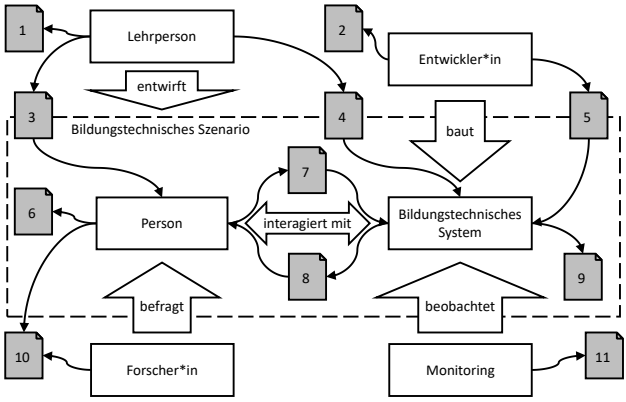


Abb. 1: Übersicht über Artefakte im Kontext bildungstechnischer Forschung in einfachen Szenarien. Pfeile und graue Kästchen geben an, wo Artefakte entstehen oder verwendet werden, die als Forschungsdaten relevant sein können. Die Nummern der Kästchen werden im Text erläutert.

ist, dass (einfache) bildungstechnische Szenarien existieren, in denen eine einzelne Person individuell mit einem bildungstechnischen System interagiert. Die Person, die das Szenario entwirft, wird im Folgenden als Lehrperson bezeichnet, während die Person, die das bildungstechnische System entwickelt, als Entwickler*in bezeichnet wird. Beide Rollen können von derselben Person eingenommen werden und existieren unabhängig davon, ob Forschung stattfindet. Falls Forschung stattfindet, kommt die Rolle eines Forschers bzw. einer Forscherin hinzu und es können zusätzliche, technische Systeme zum Monitoring eingesetzt werden. Nicht dargestellt ist in der Abbildung, dass Forschende alle Art von entstandenen Artefakten betrachten und (ggf. unter Einsatz weiterer Forschungssoftware) weiterverarbeiten können. Ebenfalls nicht dargestellt sind komplexere Szenarien, in denen es zusätzlich zu einer Interaktion zwischen Personen oder zwischen der Lehrperson und dem System kommt oder in denen es mehrere Ebenen der Interaktion gibt. Eine entsprechende Erweiterung des Modells ist für zukünftige Arbeiten vorgesehen.

Die in der Abbildung dargestellten Artefakte (graue, nummerierte Kästchen) lassen sich in ihrer Art und Herkunft wie folgt beschreiben:

1. Bei der Entwicklung eines bildungstechnischen Szenarios erzeugt eine Lehrperson Daten, die z. B. den didaktischen Zweck des Szenarios oder das erwartete Vorwissen der teilnehmenden Personen beschreiben.
2. Bei der Entwicklung eines bildungstechnischen Systems erzeugen Entwicklerinnen und Entwickler Artefakte (z. B. Anforderungen, UI Mockups oder Modelle), die das System in seinem technischen Aufbau beschreiben. Auch das System selbst ist als (Forschungs-)Software ein solches Artefakt.

3. In einem Szenario geben Lehrpersonen Daten an die teilnehmenden Personen, z. B. in Form von Arbeitsmaterial oder -anweisungen.
4. In einem Szenario geben Lehrpersonen Daten in das bildungstechnische System ein, z. B. in Form von Lerninhalten oder Bewertungsregeln für Aufgaben.
5. Entwicklerinnen und Entwickler können ebenfalls Daten (z. B. ein vorab trainiertes KI-Modell) in das bildungstechnische System eingeben.
6. Eine Person kann zusätzlich zur Interaktion mit dem System weitere Daten erzeugen, z. B. in Form von handschriftlichen Notizen oder Interaktionen mit anderen Systemen.
7. Eine Person gibt Daten in das bildungstechnische System ein, z. B. explizit durch Antworten auf eine Aufgabe oder implizit durch die Verweildauer auf einem Dokument.
8. Das bildungstechnische System erzeugt Daten, die zur weiteren Nutzung durch die Person gedacht sind, z. B. explizites Feedback als Reaktion auf eine Eingabe oder implizite Adaption aufgrund gesammelter Informationen.
9. Das bildungstechnische System erzeugt weitere Daten, die anderen Zwecken dienen, z. B. Logdaten.
10. Forschende können Personen befragen oder beobachten und dadurch weitere Daten erzeugen, die normalerweise nicht durch das Szenario entstehen würden, z. B. Fragebögen oder Pre-/Post-Tests und die zugehörigen Antworten.
11. Zusätzliches Monitoring kann eingesetzt werden, um weitere Daten über ein System zu erzeugen, z. B. durch die Messung der CPU-Auslastung während der Nutzung.

Alle diese Artefakte können potenziell als Forschungsdaten relevant sein. Ob sie es tatsächlich sind, hängt von der betrachteten, konkreten Forschungsfrage ab (siehe nächster Abschnitt). Die Artefakte weisen insbesondere aber auch sehr verschiedene Datentypen auf, folgen keiner einheitlichen Struktur, unterliegen partiell den Regeln des Datenschutzes und werden in Teilen typischerweise nicht systematisch aufgezeichnet. Software als besonderes Artefakt erfordert überdies eine andere Handhabung als reine Daten [La20].

3 Abhängigkeiten zur Forschungsfrage

Ob die oben betrachteten Artefakte im Kontext einer Forschungsfrage relevant sind und welche Rolle sie zur ihrer Beantwortung oder zur Dokumentation der Untersuchung spielen, hängt vom Inhalt der Forschungsfrage ab. Drei Beispiele sollen dies exemplarisch illustrieren:

- Ist der Forschungsgegenstand der Lernprozess, der aus der Interaktion mit dem System entsteht, sind primär die mit der Person verknüpften Daten relevant. Insbesondere sind die Daten 3 und 8 aus der Abbildung als Eingaben in den Lernprozess relevant, während die Daten 6, 7 und 10 potenziell Produkte des Lernprozesses beschrieben, die diesen sicht- und messbar machen. Die Daten 1, 2, 4 und 5 sind potenziell relevant,

um die Studie unter vergleichbaren Bedingungen zu wiederholen, während die Daten 9 und 11 weder für die Forschungsfrage noch den Kontext unmittelbar relevant sind.

- Ist der Forschungsgegenstand dagegen das Verhalten des Systems, sind primär die damit verknüpften Daten relevant. Somit beschreiben nun die Daten 4, 5 und 7 die Eingaben, während die Produkte des Verhaltens die Daten 8, 9 und mittelbar auch 11 sind, die zur Beantwortung der Forschungsfrage herangezogen werden müssen. Die Daten 1, 2 und 3 müssen möglicherweise herangezogen werden, wenn die Studie wiederholt werden soll, während diesmal die Daten 6 und 10 ohne unmittelbare Relevanz für Forschungsfrage oder Kontext sind.
- Ist schließlich die Interaktion der Person mit dem System der Forschungsgegenstand, sind die Daten 3, 4 und 5 die Eingaben, aus denen eine Interaktion resultiert, die wiederum die Daten 6 bis 11 produziert und somit durch diese beobachtet werden kann. Für eine Reproduktion müssen wiederum die Daten 1 und 2 herangezogen werden, so dass in diesem Fall alle der oben genannten Daten von unmittelbarer Relevanz für Forschungsfrage oder Kontext sind.

4 Konsequenzen und zukünftige Arbeiten

Die Beispiele zeigen, dass dieselben Artefakte in unterschiedlichen Rollen mal unmittelbar zur Beantwortung der Forschungsfrage dienen und mal den Kontext beschreiben. Daraus folgt, dass eine systematische Erfassung und Dokumentation aller Artefakte (z. B. im Rahmen der Nationalen Forschungsdateninfrastruktur [3]) sinnvoll ist, da sie spätestens durch die Wiederverwendung in einem anderen Forschungsvorhaben Relevanz bekommen können, auch wenn sie in ihrem ursprünglichen Kontext irrelevant erscheinen. Dazu müssen jedoch auch z. B. gesetzliche Regeln für die langfristige Speicherung personenbezogener Daten beachtet werden und es müssen (Beschreibungs-)Standards gefunden werden, um Artefakte aus verschiedenen Forschungsvorhaben verknüpfen zu können. Dieser Beitrag liefert dazu ein initiales Modell, das zukünftig validiert, erweitert und verfeinert werden soll. Dazu ist insbesondere geplant, das Modell auf komplexere Szenarien in der Bildungstechnologie zu erweitern und somit komplexere Interaktionen, aber auch z. B. KI-generierte Artefakte mit zu berücksichtigen. Ferner sollen konkrete Artefakte aus realen Forschungsprojekten aus dem In- und Ausland auf das Modell abgebildet werden, um die Vollständigkeit des Modells zu überprüfen. Umgekehrt soll überprüft werden, ob existierende Standards und Ontologien für Metadaten aus dem Bereich der Daten-Management-Pläne (DMP) bzw. Software-Management-Pläne (SMP) (z. B. [Ca24; MWN20]) genutzt werden können, um das bisher informelle Modell in eine formalisiertere Fassung zu überführen.

Danksagung: Die Arbeit in diesem Beitrag wurde partiell durch die Deutsche Forschungsgemeinschaft (DFG) unter der Projektnummer 501930651 gefördert.

3 https://nfdixcs.org/

Literaturverzeichnis

[Ca24] Castro, J. L.; Giraldo, O.; Geist, L.; Quiñones, N.; Solanki, D.; Rebholz-Schuhmann, D.: machine-actionable Software Management Plan Ontology (maSMP Ontology), 2024, DOI: 10.5281/zenodo.10582073.

[JS23] Jate, W. M.; Striewe, M.: Umgang der DELFI-Community mit Forschungsdaten und Softwareartefakten. In: DELFI 2023, Die 21. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V. S. 167–172, 2023, DOI: 10.18420/delfi2023-27.

[KS22] Kiesler, N.; Schiffner, D.: On the Lack of Recognition of Software Artifacts and IT Infrastructure in Educational Technology Research. In: DELFI 2022, Die 20. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V. S. 201–206, 2022, DOI: 10.18420/DELFI2022-034.

[La20] Lamprecht, A.-L.; Garcia, L.; Kuzak, M.; et al.: Towards FAIR principles for research software. Data Science 3 (1), S. 37–59, 2020, URL: https://content.iospress.com/articles/data-science/ds190026.

[MWN20] Miksa, T.; Walk, P.; Neis, P.: RDA DMP Common Standard for Machine-actionable Data Management Plans, 2020, DOI: 10.15497/rda00039.

[Wi16] Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1), 2016, DOI: 10.1038/sdata.2016.18.

Towards Computer-Aided Teaching of Reductions in Theoretical Computer Science

Maurice Herwig [1], Norbert Hundeshagen [1], Marit Kastaun [2], and Cedric Kollenberg [3]

Abstract: Reductions play a crucial role in the theory of computer science, aiding in the identification of computationally unsolvable or intractable problems. Despite their significance, mastering reductions remains challenging for students due to their high level of abstraction. In this work we report on an educational approach to learn reductions in a more practical way as a programming exercise. Through a pilot study ($n = 41$) with three measurement points, insights were gathered on the usage of a prototype learning tool for reductions, leveraging Python as the main computational model. Initial findings highlight further enhancements of computer-aided learning and teaching of reductions, such as incorporating mathematical foundations in a tools feedback, visualizing and generically generating problem instances, as well as improving extensibility by simplifying the creation of exercises.

Keywords: complexity, computability, learning tool, NP-completeness, reductions

1 Introduction

One of the fundaments of computer science is the existence of computationally unsolvable or, because of resource constraints, intractable problems. In this realm reductions are a major method to prove, if a given problem is one of the latter. Therefore, understanding and performing reductions is an important part of every undergraduate course on computability and/or complexity. In contrast to its relevance, the latter topic is named as one of the "fundamental ideas of computer science" (cf., [Sc93]), performing reductions is rather challenging for students (e. g. in [AG06; AGT05; Ar04; Ar09; En14; GAS21; NA23]). This is mainly due to the "essentially abstract" (cf., [GT16], p.6) nature of the formal concept itself. A condensed version of the definition of a *many-to-one* reduction, as used in our courses, is the following: *A problem A is called reducible onto B (typically written as $A \leq B$) if there is a computable function f such that, $x \in A$ iff $f(x) \in B$.* Somehow contradictory to the abstractness and perceived difficulty, the basic principle itself is widespread and can be interpreted as the ability to use a given solution for a problem B to solve a problem A. Thus, reducing A onto B essentially requires students to write an algorithm that translates instances of problem A to instances of problem B. The latter perspective immediately hints

1 University of Kassel, Germany, Theoretical Computer Science / Formal Methods,
maurice.herwig@uni-kassel.de, https://orcid.org/0009-0005-5519-3043;
hundeshagen@uni-kassel.de, https://orcid.org/0009-0003-5900-7821
2 University of Kassel, Germany, Didactics of Biology,
m.kastaun@uni-kassel.de, https://orcid.org/0000-0001-9122-0815
3 Hochschule Fulda – University of Applied Sciences, Germany, Applied Computer Science,
cedric.kollenberg@informatik.hs-fulda.de, https://orcid.org/0009-0008-8205-0277

doi:10.18420/delfi2024_45

how learning of reductions can be supported by educational tools. There are already existing tools aimed at supporting the teaching and learning of reductions (e. g. in [CEK13; Cr10; Ma15; Pa97; Pa98; ZHD22]). Most notable, the application RACSO [CFG14] should be mentioned here that provides several given reduction tasks between NP-complete problems. The solution to these problems is automatically tested for correctness by integrating an SAT solver. In the event of an error, the failed test case is provided to the user. A different approach is taken by Grange et al. [Gr23]. Their focus is on reductions restricted to a decidable correctness check, which can be described graphically. Finally, the GraphBench application [Br06] focuses on an intuitive and interactive representation of reductions for selected NP-complete problems. However, this implementation does not offer students the opportunity to write reductions themselves. In our project, we are aiming for a combination of the existing technical concepts in order to maximize the potential of testing and using a common programming language and to minimize difficulties in learning reductions. In the long term, two questions will be investigated with the objective of promoting the teaching/learning process of reductions in line with the development and optimization of a teaching/learning tool: How can the abstraction level of reductions be minimized by testing worked examples (e. g. in [Re14; Re17; SAK11]) and using *Python*[4]? How and to what extent should automated feedback be provided to support learners in their learning process of reductions?

2 Procedure & Method

The Course on Computability and Complexity and the Protoype ReTrain Reductions are a central part of a 2nd year course in the BSc computer science curriculum at the University of Kassel. Lecture topics include models of computation, decidability, and intractability, standard in such courses (e.g. in [Sc08]). Typically, many-to-one reductions are first introduced as the major method to prove undecidability, then adapted to polynomial-time reductions for NP-hardness. Recently, the course has been rearranged to present reductions in a more practical way using Python as a computational model, similar to the approach in [Ma18]. This offers the possibility to actually implement them, thus, reducing abstractness by presenting reductions as "[...] procedures that use 'functionally specified' subroutines [...]" (cf., [Go06], p.10) for a problem B to solve another problem A. For the latter, we use a prototype of a learning tool (called ReTrain) implemented in two bachelor theses [Le22; We23]. For a reduction task in form of two decision procedures given by a lecturer, students can implement their solution in ReTrain. Furthermore, they can use self-selected test cases to verify the correctness of their reduction. Fig.1 exemplifies this for the reduction 3SAT $\leq$ IndependentSet.[5]. Such a test for an instance (grey dotted box) for

4 `www.python.org`

5 The 3SAT problem: Given a formula φ in conjunctive normal form with at most 3 literals in a clause, is φ satisfiable?
The IndependentSet problem: Given a graph $\mathcal{G}$ and a number k, does $\mathcal{G}$ contain k vertices that are pairwise not connected by edges?

the reduction code (green dotted box) involves the following two steps: (1) translate the instance of 3SAT by the reduction code into an instance of IndependentSet (yellow box), (2) and use the lecturer code for IndependentSet to compute the result (turquoise box), which should be true, if the given test case is a satisfiable 3CNF-formula and false otherwise.

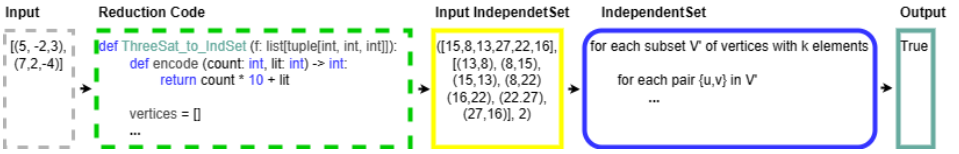


Fig. 1: Simplified view of ReTrain frontend

Evaluation setup (design & participants) ReTrain was used for the first time in the course "Computability and Complexity" in the winter term 23/24 ($n = 41$; 1st-3rd semester = 47.1%; 4th-7th semester = 26.5%; >7th semester = 26.5%). An evaluation study with three measurement points (M1-M3) was conducted to get a first impression of the usefulness of ReTrain. The specific attitudes, knowledge and perceptions of the students in relation to reductions and ReTrain were measured by using qualitative/quantitative instruments. At M1, in addition to the socio-demographic data, prior knowledge of sets and functions (3 items: 4-point Likert scale from (1) uncertain to (4) very certain), as well as the use of programming languages (in particular Python; 2 items: 4-point Likert scale from (1) uncertain to (4) very certain) was collected by self-assessments. Before using ReTrain (M2), the students' knowledge of the mathematical basics of sets and functions has been measured reduction-specifically. For 6 descriptions of functions we asked, if these are *well-defined*, and for 6 pairs of sets A and B the question was, if $A \in B$, $A \subseteq B$ or none of both. Then, the students were given two tasks to solve with the help of ReTrain. First, a given reduction should be corrected by answering a series of questions, and second, a polynomial-time reduction between two given problems should be implemented. Furthermore, usage data, like time spent and click events, was recorded. Finally, after using the tool (M3), the students were asked to what extent they considered ReTrain to be helpful. Using open and closed items, students were invited to assess their individual experiences with ReTrain.

3 First Results and Future Work

From the initial use of ReTrain and the data derived from the accompanying study, we were able to identify first implications for further developments of a digital learning tool for reductions. More specifically, the adjustments should take two sides into account. On the one hand, the focus is on the students as direct users of the learning application. On the other hand, the application should also be used by lecturers to manage the exercises.

Importance of Mathematical Foundations Because of the abstract nature of reductions, this topic requires "significant mathematical skills" (cf., [GT16] p.5), the survey incorporated 12 mathematical questions. Although the students assessed their own abilities in dealing

with functions ($M = 2.82; SD = 0.52$), sets ($M = 2.91; SD = 0.76$), and their domains and ranges ($M = 2.79; SD = 0.64$) highly, they scored very differently in exercises on functions ($M = 1.67; SD = 0.13$) and sets ($M = 1.50; SD = 0.23$), where 1 encodes an incorrect answer and 2 a correct answer. Particularly for set problems, one of the two relationships ($A \in B$ or $A \subseteq B$) could be directly excluded due to the definitions. We refer to such errors as type mismatches, which accounted for over 35% of the errors. Since there is a strong connection between the set theory and programming, as shown in [COP13], we suspect that type mismatches also affect the programming of reductions. Therefore, the future feedback system should incorporate type feedback beyond that given by an IDE.

Python as a Programming Language for Reductions Using Python as the computational model for implementing reductions is beneficial for several reasons. If Python is already part of the curriculum, students do not have to master an additional syntactical level imposed by an artificial model of computation, where typically a kind of pseudo-code is introduced (e. g. in [Sc08]). This also ensures that the reduction is computable. The data shows that the processing time due to programming the reduction is significant. In an example, in which a given reduction had to be corrected, the average processing time was approximately 20 minutes. A possible reason was given by a student (1st-3rd semester): "However, for some problems it is quite complicated to specify the reduction, as many transformations have to be made due to the encoding." Nevertheless, the problem of encoding can also be used to shed light on certain aspects of the runtime complexity. Therefore, in conjunction with the usage data and the qualitative feedback from students, there is future need for assistance in supporting students in the use of Python. However, the possibility to implement reductions in ReTrain was also rated positively: "I think it is very good that there is the possibility of testing reductions using pre-defined queries." (student, 1st-3rd semester).

Usability The problem of encoding mentioned above can be addressed using an additional graphical form to represent instances of the problem, as for instance done with the GraphBench application [Br06]. Another crucial aspect for usability and therefore learning success through the application is the expansion of the instance tester. Due to the current manual entry of instances, students must have a sufficient overview of the task in order to be able to specifically test their reduction. It turns out that most of students solely use the given example instances to test their reduction. Only two students created their own instances. The relevance of finding good examples is confirmed by the fact that edge cases are often ignored and therefore, the reduction was incorrect. By generating random instances, the understanding of an incorrect reduction can be fostered.

Extensibility Another important aspect of the application should be extensibility and flexibility. This mainly includes a generic way for a lecturer to introduce new decision problems and accompanying data structures quickly. A possibility to ease this process in comparison to implement them directly is shown in [CFG14], where instances of decision problems are translated into SAT-instances and data structures are represented by an abstract language. Finally, we believe that a useful learning tool should also deal with reductions between undecidable problems. As there are no decision procedures, simple testing is not working. Therefore, such reductions can be limited to semi-decidable problems, where testing needs to be bounded by a timeout.

Acknowledgements: The authors would like to thank Kathrin Lehman, Clemens Weiße, and Marco Sälzer for their work in developing ReTrain. Furthermore, the work is partially funded by *Stiftung Innovation in der Hochschullehre* within the project *UKS_digi*.

References

[AG06] Armoni, M.; Gal-Ezer, J.: Reduction – an Abstract Thinking Pattern: The Case of the Computational Models Course. In: Proc. of the 37th SIGCSE Technical Symposium on Computer Science Education. Pp. 389–393, 2006, DOI: 10.1145/1124706.1121461.

[AGT05] Armoni, M.; Gal-Ezer, J.; Tirosh, D.: Solving problems reductively. Journal of Educational Computing Research 32 (2), pp. 113–129, 2005, DOI: 10.2190/6PCM-447V-WF7B-QEUF.

[Ar04] Armoni, M.: Reductive Thinking: How to Teach It? In: Proc. of the 9th annual SIGCSE Conf. on Innovation and Technology in Computer Science Education. P. 280, 2004, DOI: 10.1145/1026487.1008118.

[Ar09] Armoni, M.: Reduction in CS: A (Mostly) Quantitative Analysis of Reductive Solutions to Algorithmic Problems. ACM Journal of Educational Resources in Computing 8, 11:1–30, 2009, DOI: 10.1145/1482348.1482350.

[Br06] Brändle, M.: GraphBench: Exploring the Limits of Complexity with Educational Software, PhD dissertation, ETH Zürich, 2006, DOI: 10.3929/ETHZ-A-005128663.

[CEK13] Crescenzi, P.; Enström, E.; Kann, V.: From Theory to Practice: NP-completeness for Every CS Student. In: Proc. of the 18th ACM Conf. on Innovation and Technology in Computer Science Education. Pp. 16–21, 2013, DOI: 10.1145/2462476.2465582.

[CFG14] Creus, C.; Fernández, P.; Godoy, G.: Automatic Evaluation of Reductions between NP-Complete Problems. In: Proc. of the 17th Int. Conf. on Theory and Applications of Satisfiability Testing. Springer, Cham, pp. 415–421, 2014.

[COP13] Cantone, D.; Omodeo, E.; Policriti, A.: Set Theory for Computing - From Decision Procedures to Declarative Programming with Sets. Springer, 2013.

[Cr10] Crescenzi, P.: Using AVs to Explain NP-completeness. In: Proc. of the 15th Conf. on Innovation and Technology in Computer Science Education. P. 299, 2010, DOI: 10.1145/1822090.1822175.

[En14] Enström, E.: On Difficult Topics in Theoretical Computer Science Education, PhD thesis, KTH Royal Institute of Technology, 2014.

[GAS21] Gaber, I.; Armoni, M.; Statter, D.: Teaching Reduction as an Algorithmic Problem-Solving Strategy. In: Proc. of the 3rd International Conf. on Computer Science and Technologies in Education. IEEE, pp. 19–26, 2021, DOI: 10.1109/CSTE53634.2021.00012.

[Go06] Goldreich, O.: On Teaching the Basics of Complexity Theory. In: Theoretical Computer Science: Essays in Memory of Shimon Even. Springer, pp. 348–374, 2006, DOI: 10.1007/11685654_15.

[Gr23] Grange, J.; Velken, F.; Vortmeier, N.; Zeume, T.: Formal Foundations of Teaching Reductions, Talk at Highlights of Logic, Games and Automata (Highlights 2023), Available at url=https://highlights-conference.org/2023/program. Accessed 21 June 2024., Kassel, 2023.

[GT16] Gal-Ezer, J.; Trakhtenbrot, M.: Identification and addressing reduction-related misconceptions. Computer Science Education 26 (2-3), pp. 89–103, 2016, DOI: 10.1080/08993408.2016.1171470.

[Le22] Lehmann, K.: Toolgestütztes Lernen im Kontext der Berechenbarkeitstheorie: ein Backend-Prototyp zur automatisierten Überprüfung von Reduktionen, BSc Thesis, Universität Kassel, 2022.

[Ma15] Maji, N.: An Interactive Tutorial for NP-completeness, MSc Thesis, Virginia Tech, 2015.

[Ma18] Maccormick, J.: What Can Be Computed?: A Practical Guide to the Theory of Computation. Princeton University Press, 2018.

[NA23] Nakar, L.; Armoni, M.: On Teaching Abstraction in Computer Science: Secondary-School Teachers' Perceptions vs. Practices. In: Proc. of the 2023 Conf. on UK & Ireland Computing Education Research. Pp. 1–7, 2023, DOI: 10.1145/3610969.3611124.

[Pa97] Pape, C.: Using Interactive Visualization for Teaching the Theory of NP-completeness. Universität Karlsruhe, 1997, DOI: 10.5445/IR/65597.

[Pa98] Pape, C.: Teaching the Reduction Technique with Interactive Visualizations: Report on a Simple Automatic Tutor. Universität Karlsruhe, 1998, DOI: 10.5445/IR/19898.

[Re14] Renkl, A.: The Worked Examples Principle in Multimedia Learning. In (Mayer, R. E., ed.): The Cambridge Handbook of Multimedia Learning. Cambridge Handbooks in Psychology, Cambridge University Press, pp. 391–412, 2014, DOI: 10.1017/CBO9781139547369.020.

[Re17] Renkl, A.: Learning from worked-examples in mathematics: Students relate procedures to principles. Zentralblatt für Didaktik der Mathematik 49 (4), pp. 571–584, 2017.

[SAK11] Sweller, J.; Ayres, P.; Kalyuga, S.: The Worked Example and Problem Completion Effects. Cognitive load theory, pp. 99–109, 2011, DOI: 10.1007/978-1-4419-8126-4_8.

[Sc08] Schöning, U.: Theoretische Informatik - kurz gefasst. Spektrum Akademischer Verlag, 2008.

[Sc93] Schwill, A.: Fundamentale Ideen der Informatik. Zentralblatt für Didaktik der Mathematik 25 (1), pp. 20–31, 1993.

[We23] Weiße, C.: Frontend-Entwicklung eines Reduktions-Trainers für Studierende, BSc Thesis, Universität Kassel, 2023.

[ZHD22] Zhang, C.; Hartline, J. D.; Dimoulas, C.: Karp: A Language for NP Reductions. In: Proc. of the 43rd ACM SIGPLAN International Conf. on Programming Language Design and Implementation. ACM, San Diego, CA, USA, pp. 762–776, 2022, DOI: 10.1145/3519939.3523732.

Relation between struggle and learning personality in programming exercises

Alina Deriyeva [1], Jesper Dannath [1] und Benjamin Paaßen [1]

Abstract:

Personality-related characteristics can have an impact on learning experiences and learning outcomes. Moreover, understanding learning approaches of students can help to make personalized pedagogical decisions. Particularly, this has a potential to improve learning outcomes and mitigate user attrition in digital learning environments (DLEs). We hypothesize that persistent individual characteristics may influence a learners' tendency to struggle during programming exercises. In a study within a digital learning environment for Python programming (N=55), we evaluated the relation between learners' self-reported personality characteristics and their tendency to struggle during the exercises. We find that, in our sample, some problem solving approaches and all learning preferences are related to struggle indicators. These results indicate that it may be helpful to include personality-related features, like problem solving strategies, as context information for pedagogical modeling in DLEs.

Keywords: Programming Education, Struggle, Learning Personality, Learner Modeling

1 Motivation and Setup

Investigating struggle is one of the ways to determine success in learning. In programming tasks, struggle can indicate students in need of assistance [Do21] and can be used as a basis for learner modelling. While struggling can be caused by the ability of the learner being lower than the difficulty of the task, we hypothesize that there might be additional individual aspects that impact the tendency to struggle. If this is the case, taking into account such aspects will be important to drive personalized pedagogical decisions in digital learning environments, such as the timing, amount, or type of hints or the selection of learning tasks with the goal of mitigating attrition in online courses [KH15] and enhancing learning impact.

However, including personal characteristics in learner modeling is complicated [Le23], and, to the best of our knowledge, there is no unified model that would describe these traits. We distinguish three broad sub-features of learning personality characteristics: personality traits (which remain more of less constant over the time), learning preferences (which can undergo changes, but very slowly), and problem solving strategies (that can be learned within the course of several months, and sometimes can be treated as a meta-skill). Some of these

1 University of Bielefeld, Faculty of Technology, Inspiration 1, 33619 Bielefeld, Germany,
aderiyeva@techfak.uni-bielefeld.de, https://orcid.org/0009-0003-6757-5722;
jdannath@techfak.uni-bielefeld.de, https://orcid.org/0009-0008-5174-8365;
bpaassen@techfak.uni-bielefeld.de, https://orcid.org/0000-0002-3899-2450

doi:10.18420/delfi2024_46

aspects have already been associated with learning outcomes. For example, many studies investigated the impact of the Big 5 Traits (personality Traits) onto academic performance, and have found that, for example, the Conscientiousness factor can work as a strong predictor for task completion [Ma22; PHD02; Ve14].

To investigate the relation between struggle and learning characteristics, we conducted a classroom study with a DLE for programming. Figure 1 shows the timeline shared by all experiment sessions. First, we introduced the students to the system and provided instructions on how to operate it. The participants were also instructed that they could opt out of the experiments at any point. Then, we asked the students to fill the pre-questionnaire, including a shortened assessment scale for the Big 5 Traits[RJ07], questions about problem solving strategies and overall learning preferences. After that, the participants worked in the DLE for 65 minutes. Lastly, we asked them to complete the post-questionnaire (feedback questionnaire). Sharing system interaction data was voluntary, as well as participation in the surveys. Also, because the participants were allowed to leave at any point of the experiment, the timeline in the Fig. 1 remained approximate and was not adhered to by all the participants.

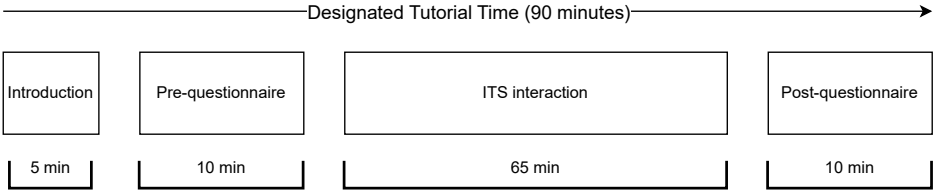


Fig. 1: The approximate timeline of the user study.

We used two struggle indicators from Dannath et al. (see [DDP24] in this volume) to broaden our perspective of the task solving behaviour. The first struggle indicator was the time until the first submission of the correct solution to the task. For times longer than the median of all learners on all tasks (02:01 minutes) we assumed struggle. We compared this measure with progress-based struggle indicators. Therefore, we introduced a progress LLM-based struggle indicator, where we calculated embeddings of the student code-states on the text-level, using PromptEOL on a quantized CodeLlama-7B model [Ji23; Ro23]. The distance between embeddings states did then serve as a signal of progress, and insignificant progress over a certain time indicates struggle.

Using these struggle indicators, we fit a 2-parameter Item Response Theory model to assess the ability of learners to complete tasks without struggling. We then looked for the correlation between these results and the individual learning characteristics collected through the survey during the experiment.

2 Results

Overall, 97 students participated in the experimental sessions. From all the participants, 84 consented for their data to be used in research, and 55 completed the pre-questionnaire.

Table 1 presents the correlation between ability to avoid struggle and the preferred strategies in solving complex problems. It is apparent that both chosen measures display similar correlations with the survey questions in terms of the direction of the relation. Still, it is interesting to observe that only few problem solving strategies are related to struggle-based ability estimation, for example, the „Trial and Error“ one. Using chatbots, on the other hand, has an opposite effect. This aligns with previous results of help-seeking behaviour being negatively correlated with proficiency [Al16]. All of the learning preference dimensions are significant for at least one of the introduced measures, indicating that they all are related to struggle. Moreover, both measures exhibit significant correlation with the question „I get easily frustrated, if I have to attempt solving a problem several times before success“, which can indicate that susceptibility to frustration can significantly impact overall learning experience. Generally, the time-based measure significantly correlates only with learning preferences (but with all of them), whereas the progress-based measure is also related to some problem solving strategies.

Problem Solving	time-based	progress-based
Using Chatbots	-0.18	-0.28*
Sub-problem decomposition	0.05	0.07
Using a visual debugger	0.15	0.22
Trial and Error	0.25	0.36**
Revisiting learning material	-0.16	-0.22
Revisiting online resources	0.03	0.01
Conceptualizing solutions through visualization	-0.1	-0.03
Learning Preferences	**time-based**	**progress-based**
Acquiring only the minimum skill	-0.31*	-0.22
I like solving bonus tasks	0.28*	0.33*
Challenging tasks increase my interest	0.29*	0.33*
I get easily frustrated	-0.41**	-0.28*
I tend to quit the course if it becomes too complicated	-0.33*	-0.17

Tab. 1: Correlation of struggle-based ability measures with problem-solving strategies and learning preferences (* $p < 0.05$, ** $p < 0.01$)

Additionally, frustration sensitivity rendered significant ($p \leq 0.05$) negative moderate correlation with number of tasks completed, number of submissions, and number of attempts, reaching -0.44, -0.36, and -0.41 respectively. This indicates that struggling while solving a task has an impact onto the overall performance of the learners, as well as on the early learner attrition in our system. Moreover, we computed Pearson's correlation between these metrics and struggle-based ability estimation with the Big 5 personality traits. Counter to our expectations, no significant correlation was detected.

3 Conclusion and Future Work

Our results indicate that there is indeed a correlation between struggling and individual learning characteristics of the learners. Particularly, the time-based struggle measure is related to learning preferences, and the progress-based struggle measure is related to some of the problem-solving strategies. While we cannot state for sure the causality of this relation, it indicates that including such characteristics into modeling users can be beneficial for supporting pedagogical decisions. Surprisingly, in our experimental setup, the effect of the Big 5 personality traits onto learning progress was not found. This may be because the relation between personality traits and struggle may be mediated by other variables, such as frustration sensitivity.

Future direction would be to investigate the struggle within tasks: for example, currently we propose two struggle indicators, which in our opinion should reflect the struggle within the tasks to some extent, but do not necessarily correspond to the subjective experience of the students (which in following experiments can be collected through self-assessment or observations). Moreover, we would like to investigate the ways how the information regarding individual learning characteristics can be integrated into the pedagogical decision mechanisms of digital learning environments.

4 Acknowledgements

This work is part of the "SAIL: SustAInable Lifecycle of Intelligent Socio-Technical Systems" project. SAIL is funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia (grant no. NW21-059A). We gratefully acknowledge this support.

Literaturverzeichnis

[Al16] Aleven, V.; Roll, I.; McLaren, B. M.; Koedinger, K. R.: Help Helps, But Only So Much: Research on Help Seeking with Intelligent Tutoring Systems. en, International Journal of Artificial Intelligence in Education 26 (1), S. 205–223, 2016, ISSN: 1560-4292, 1560-4306, DOI: 10.1007/s40593-015-0089-1.

[DDP24] Dannath, J.; Deriyeva, A.; Paaßen, B.: Evaluating Task-Level Struggle Detection Methods in Intelligent Tutoring Systems for Programming. In Proceedings of the 22th Fachtagung Bildungstechnologien (DELFI), 2024.

[Do21] Dong, Y.; Samiha, M.; Preya, S.; Price, T.; Barnes, T.: Using Student Trace Logs To Determine Meaningful Progress and Struggle During Programming Problem Solving. In: Proceedings of the 14th International Conference on Educational Data Mining. Paris, France, S. 439–445, 2021.

[Ji23] Jiang, T.; Huang, S.; Luan, Z.; Wang, D.; Zhuang, F.: Scaling Sentence Embeddings with Large Language Models, en, arXiv:2307.16645 [cs], 2023.

[KH15] Kizilcec, R. F.; Halawa, S.: Attrition and Achievement Gaps in Online Learning. In: Proceedings of the Second (2015) ACM Conference on Learning @ Scale. ACM, Vancouver BC Canada, S. 57–66, 2015, ISBN: 978-1-4503-3411-2, DOI: 10.1145/2724660.2724680.

[Le23] Lester, J. et al.: LEARNER MODELING IN INTELLIGENT TUTORING SYSTEMS SWOTANALYSIS. In: Design Recommendations for Intelligent Tutoring Systems. U.S. Army Research Laboratory, S. 43–52, 2023.

[Ma22] Mammadov, S.: Big Five personality traits and academic performance: A meta-analysis. Journal of Personality 90 (2), S. 222–255, 2022, DOI: 10.1111/jopy.12663.

[PHD02] PytlikZillig, L.; Hemenover, S.; Dienstbier, R.: What Do We Assess When We Assess a Big 5 Trait? : A Content Analysis of the Affective, Behavioral, and Cognitive Processes Represented in Big 5 Personality Inventories. Personality and Social Psychology Bulletin 28, S. 847–858, 2002, DOI: 10.1177/0146167202289013.

[RJ07] Rammstedt, B.; John, O. P.: Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality 41 (1), S. 203–212, 2007, ISSN: 0092-6566, DOI: 10.1016/j.jrp.2006.02.001.

[Ro23] Rozière, B. et al.: Code Llama: Open Foundation Models for Code, en, arXiv:2308.12950 [cs], 2023.

[Ve14] Vedel, A.: The Big Five and tertiary academic performance: A systematic review and meta-analysis. Personality and Individual Differences 71, S. 66–76, 2014, ISSN: 0191-8869, DOI: 10.1016/j.paid.2014.07.011.

Demo-Beiträge

Neugestaltung eines Mixed Reality Prototypen für das SchülerlARbor Chemie

Felix Meinhardt[1], Sergej Görzen[1], Richard Werkes[1], Philipp Pelzer[2], Daniel Krause[2], Michael Schroeder[2] und Ulrik Schroeder[1]



Abstract: Augmented Reality (AR) stellt großes, noch unausgeschöpftes Potenzial für Lernanwendungen dar. Im Chemieunterricht bietet sich etwa das Visualisieren von mikroskopischen Prozessen oder Modellen mithilfe von AR an. Diese Demo stellt einen quelloffenen AR Prototypen für die Microsoft HoloLens vor, welcher Begleitmaterialien für verschiedene Module im Schülerlabor für Chemie der RWTH Aachen bietet. Die Themen der aktuellen Version behandeln Piezokristalle, den Aufbau und die Funktionsweise von Lithium-Ionen-Akkus sowie virtuelle Experimente zu den Konfigurationsmöglichkeiten dieser Akkus. Zusätzlich ist eine Einführung in HoloLens-Interaktionen enthalten, um die Einstiegsbarriere für diese moderne Technologie zu mindern. Evaluiert wurde zunächst die Bedienbarkeit der Applikation.

Keywords: Augmented Reality, Mixed Reality, Chemieunterricht, SchülerlARbor

1 Motivation

Der Einsatz von Augmented Reality (AR) bietet erhebliches Potenzial für den Einsatz in Lernanwendungen aufgrund der nachgewiesenen Effekte wie gesteigerter Motivation und Freude am Lernen, verbessertem Lernerfolg und der erleichterten Vermittlung abstrakter Konzepte [Av21; Ma22; Wu13]. Im Fachbereich der Chemie ist die Visualisierung mikroskopischer Prozesse oder Modelle ein idealer Anwendungsbereich für diese Technologie, da solche Prozesse in schulgeeigneten Experimenten oft nur schwer oder gar nicht beobachtbar sind. Viele Studien zu "AR in der Chemieausbildung" setzen für ihre AR-Anwendungen Smartphones oder Tablets zur Darstellung der digitalen Inhalte ein [Ma22].

Im Schülerlabor für Chemie CheERs! von der Didaktik für Chemie an der RWTH Aachen University sollten Experimente durch AR Technologie unterstützt werden. Ziel war es, sowohl Einblicke in chemische Prozesse zu ermöglichen, die im Real-Experiment nur schwer beobachtbar wären, als auch den Schüler*innen während der Real-Experimente zusätzliche Informationen bereitzustellen. Hierfür wurde eine AR-Applikation für Tablets entwickelt und im Schülerlabor eingesetzt. Dabei wurden didaktisch aufbereitete Arbeitsschritte mithilfe von physischen Arbeitsblättern angeleitet. Auf diesen befanden sich zudem Marker (z.B. QR Codes) um weitere Inhalte in der AR-Applikation abzurufen. Leider waren die positiven

1 RWTH Aachen University, Learning Technologies, Aachen, Germany, felix.meinhardt@rwth-aachen.de; goerzen@cs.rwth-aachen.de, https://orcid.org/0000-0003-3853-2435; werkes@cs.rwth-aachen.de; schroeder@cs.rwth-aachen.de, https://orcid.org/0000-0002-5178-8497

2 RWTH Aachen University, Didaktik der Chemie, Aachen, Germany, pelzer@ddc.rwth-aachen.de; daniel.krause@rwth-aachen.de; schroeder@rwth-aachen.de, https://orcid.org/0000-0003-2559-7838

doi:10.18420/delfi2024_47

Effekte des Einsatzes der AR-Applikation auf Tablets nicht so hoch, wie erwartet. Es wurde beobachtet, dass die Schüler*innen das Tablet und die dazugehörigen Arbeitsblätter während der Experimente beiseite legen mussten, was zu einer Unterbrechung der AR-Erfahrung führte. Obwohl AR Technologien auf mobilen Endgeräten viel Potenzial haben, war das Tablet im direkten Einsatz bei den Experimenten eher hinderlich.

AR als unterstützendes Werkzeug (insbesondere bei Chemie-Experimente) verspricht positive Lerneffekte [CWC14]. Um eine fließende AR Erfahrung während der Experimente im Schülerlabor zu ermöglichen, bedarf es einer direkten Integration beim Bearbeiten der Arbeitsblätter - ohne einen Medienbruch. In Virtual Reality gibt es bereits erfolgreiche Ansätze für den Einsatz in Chemie Experimenten (z.B. [Du20; Ha17; Le22]). Da jedoch im Schülerlabor die Experimente im realen Raum stattfinden, ist Virtual Reality eher ungeeignet. Stattdessen eignet sich die hybride AR-Technologie "Mixed Reality (MR)"[Al22]. Durch das reale Setup bietet es einen hohen Grad an Immersion und die Hände bleiben während der Nutzung frei. Das ermöglicht eine fließende Bedienbarkeit parallel zum Experiment, ohne auf virtueller Hinweise und Modelldarstellungen verzichten zu müssen.

2 Umsetzung

Die Entwicklung des MR-Prototyps für den Einsatz im Chemie Schülerlabor wurde in Zusammenarbeit mit dem Lehrstuhl für Lerntechnologien und mit der Didaktik der Chemie der RWTH Aachen University [Me23] auf Basis der vorhandenen Lernmaterialien des Schülerlabors für Chemie CheERs! entwickelt [Pe23].

Als Technologie zeigte die Microsoft HoloLens bereits positive Ergebnisse bezüglich Motivation, Bedienbarkeit und Lerneffekt [PBC21; Wy21]. Da diese Hardware bereits zur Verfügung stand und alle Bedingungen für das Experiment-Setup im Schülerlabor erfüllte, wurde die vorhandene mobile AR-Applikation darauf umfunktioniert und erweitert. Für die Entwicklung wurde Unity mit dem Mixed Reality Toolkit (MRTK 2) [Mi22] verwendet. Der Quellcode ist Open-Source auf OSF und GitLab [Me24a; Me24b] verfügbar. Die konzeptionelle Designphase der Entwicklung orientierte sich an einem iterativen Designansatz, wie in [Je16] beschrieben. Dabei wurden unter anderem User Stories, Personas, Storyboards etc. erstellt (siehe [Me23]).

Die realisierten Lerneinheiten sind in drei Themenbereiche und Lernziele unterteilbar. Diese wurden modular umgesetzt, wobei die angewandten Interaktionsmethoden und Darstellungsformen variieren. Jede Einheit soll Schritt für Schritt entweder in einem Anweisungsfenster (siehe links in (2) aus Abb. 1) oder Arbeitsblatt angeleitet werden. So können Schüler*innen selbst bestimmen, in welchem Tempo sie lernen.

Um den Einstieg in die Lernapplikation zu vereinfachen, wurde ein Tutorial implementiert, in dem die möglichen Interaktionen erklärt werden. Dies soll eine anfängliche Überforderung durch das System vorbeugen und den Fokus auf die Lerninhalte fördern. Anschließend

kann zwischen den drei Chemie-Themenbereichen gewählt werden: versuchsbegleitende Erklärungen zu Aufbau und Funktionsweise von Lithium-Ionen-Akkus, Simulation eines materiell aufwendigen Experiments zu verschiedenen Lithium-Ionen-Akkukompositionen und Betrachtung eines Piezokristallmodells mit Simulation des piezoelektrischen Effekts. Ein an die Lerneinheiten anschließendes Quiz soll eine Echtzeit-Überprüfung über das Gelernte bieten. Die Abb. 1 bietet ein Einblick in die umgesetzten Szenarien, die nachfolgend genauer erläutert werden.

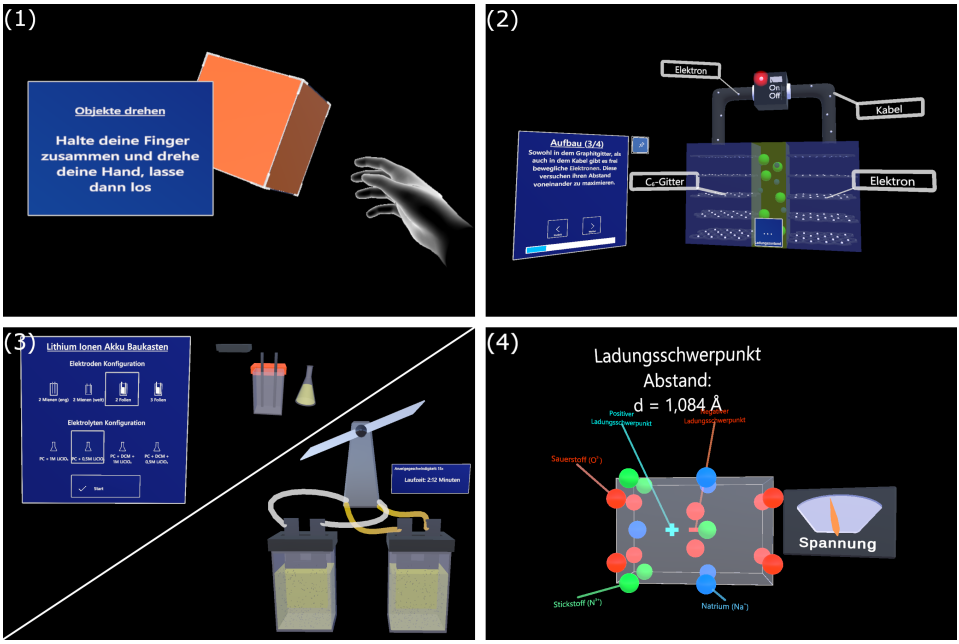


Abb. 1: Die vier Module des Prototypen: (1) Tutorial-Szene als Einführung in die erforderten HoloLens Interaktionen, (2) Versuchsbegleitende Erklärungen beim Lithium-Ionen-Akku Aufbau, (3) Simulation eines Experiments, (4) Betrachtung eines Piezokristallmodells

Das Interaktionstutorial (1) führt Nutzer*innen Schritt für Schritt durch die für die Verwendung der Applikation benötigten Interaktionstechniken, wie etwa Menübedienung oder Objektmanipulation. Nach Abschluss haben die Nutzer*innen die Möglichkeit, eines der Versuchsszenarien in beliebiger Reihenfolge auszuwählen.

In Modul (2) ist das Lernziel die Funktionsweise und der Aufbau von Lithium-Ionen-Akkus. Schüler*innen können ein animiertes und von Infotexten begleitetes Modell eines solchen Akkus beim Lade- und Entladeprozess sowie im geladenen und entladenen Zustand betrachten. Um eine Verbindung vom Experiment zu dem Modell zu schaffen, besteht die Möglichkeit, einen QR-Code zu platzieren, der von der MR-Brille erkannt wird. Das abstrakte Modell wird dadurch auf das tatsächliche Experiment projiziert.

Das Modul zur Experiment-Simulation (3) dient zur Exploration der Auswirkungen verschiedener Akku-Konfigurationen durch einen konstruktivistischen Ansatz. Die Schüler*innen müssen selbständig aus einem Menü verschiedene Elektroden und Elektrolyten wählen, die sie testen wollen. Anschließend stellen sie virtuell durch das Zusammenziehen der Teile den Versuchsaufbau nach. Danach wird in erhöhter Geschwindigkeit der Lade- und Entladeprozess simuliert und dabei die Laufzeit eines angeschlossenen Verbrauchers gemessen. Das selbe Experiment wurde bereits als interaktives Lernvideo umgesetzt [Pe23]. Schüler*innen können durch wiederholte Durchführung der Experimente mit verschiedenen Konfigurationen dann selbstständig erarbeiten, welchen Effekt verschiedene Entscheidungen bei der Konfiguration auf die Akkulaufzeit haben.

Das Piezo-Modul (4) stellt Nutzer*innen das Modell eines Piezokristalls ($NaNO_2$) sowie ein virtuelles Spannungsmessgerät zur Verfügung. Schüler*innen können an den Seiten des Kristalls ziehen, um zu beobachten, wie die Verformung des Kristalls eine messbare Spannung erzeugt. Dadurch soll die elektrische Auswirkung von mechanischer Kraft, wie z.B. einem Zusammenstoß mit dem Kristall, simuliert werden.

3 Evaluation

Eine erste Pilot-Studie fokussierte sich auf die Bedienbarkeit (Usability), da dies ein wichtiger Faktor bei einer Lernerfahrung ist [Ro17]. Evaluiert wurden sechs Teilnehmer*innen mit Chemie-Hintergrund. Während der AR-Erfahrung wurde die Think-Aloud-Methode angewandt. Anschließend wurde mithilfe der "System Usability Scale" die Nutzer*innenerfahrung eingestuft. Weitere Details sind in [Me23] zu finden. Der durchschnittliche System Usability Scale Score lag bei 83, was, um diesen anschaulicher zu machen, nach [SL16] auf der konventionellen A-F Notenskala der Note A entsprechen würde. Obwohl das Ergebnis insgesamt positiv bezüglich der Usability bewertet wurde, war eine weitere persönliche Anleitung (trotz des Tutorials) weiterhin notwendig.

4 Ausblick

Die entwickelte AR-Applikation wird derzeit in das Lehr- und Lernangebot des chemischen Schülerlabors der RWTH Aachen University eingeführt. Bei der Durchführung ergeben sich weitere Anforderungen zur Bedienung oder zu den Inhalten. Neben weiteren Verbesserung der Nutzer*innenerfahrung ist eine Evaluation des didaktischen Mehrwerts interessant. Da der Lehrinhalt beider Applikationen nahezu gleich ist, bietet sich der Vergleich zwischen der mobilen AR- und der MR-Variante an. Weitere Features wie ein Kollaborationsmodus (kollaboratives Lernen), in dem mehrere Nutzer*innen gleichzeitig dasselbe Objekt betrachten und manipulieren können, wäre ebenfalls interessant für die Experimentdurchführung. Um mehr Einblicke auf Basis von quantitativen Daten zu erhalten, ist das Sammeln der Interaktionsdaten geplant. Das würde Learning Analytics ermöglichen. Dafür ist bereits der Einsatz von xAPI mithilfe des OmiLAXR Frameworks geplant [GHS24].

Literaturverzeichnis

[Al22] Alnagrat, A.; Che Ismail, R.; Syed Idrus, S. Z.; Abdulhafith Alfaqi, R. M.: A Review of Extended Reality (XR) Technologies in the Future of Human Education: Current Trend and Future Opportunity. Journal of Human Centered Technology 1 (2), S. 81–96, 2022, ISSN: 2821-3467, DOI: 10.11113/humentech.v1n2.27, Stand: 03. 03. 2024.

[Av21] Avila-Garzon, C.; Bacca-Acosta, J.; Kinshuk, 0.; Duarte, J.; Betancourt, J.: Augmented Reality in Education: An Overview of Twenty-Five Years of Research. Contemporary Educational Technology 13 (3), ep302, 2021, ISSN: 1309517X, DOI: 10.30935/cedtech/10865.

[CWC14] Cai, S.; Wang, X.; Chiang, F.-K.: A case study of Augmented Reality simulation system application in a chemistry course. Comput. Hum. Behav. 37, S. 31–40, 2014, DOI: 10.1016/j.chb.2014.04.018.

[Du20] Duan, X.; Kang, S.-J.; Choi, J. I.; and, S. K. K.: Mixed Reality System for Virtual Chemistry Lab. KSII Transactions on Internet and Information Systems 14 (4), S. 1673–1688, 2020, DOI: 10.3837/tiis.2020.04.014.

[GHS24] Görzen, S.; Heinemann, B.; Schroeder, U.: Towards Using the xAPI Specification for Learning Analytics in Virtual Reality. In: Learning Analytics from Virtual Reality (LAVR) Workshop at the 14th International Conference on Learning Analytics and Knowledge (LAK24). Kyoto, Japan, 2024.

[Ha17] Han, J.; Tian, Y.; Song, W.; Fong, S.: An Implementation of VR Chemistry Experiment System. In: Proceedings of the International Conference on Big Data and Internet of Thing. BDIOT '17, Association for Computing Machinery, London, United Kingdom, S. 205–208, 2017, ISBN: 9781450354301, DOI: 10.1145/3175684.3175708, URL: https://doi.org/10.1145/3175684.3175708.

[Je16] Jerald, J.: The VR book: human-centered design for virtual reality. Association for computing machinery Morgan & Claypool publishers, New York] [San Rafael, California, 2016, ISBN: 978-1-970001-12-9.

[Le22] Le, H.-M.; Nguyen, G.-H.; Huynh, V.-T.; Le, M.-K.; Tran, M.-T.; Nguyen, T. V.; Tran, T. N.-D.: Chemisim: A Web-based VR Simulator for Chemistry Experiments. In: 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). S. 850–854, 2022, DOI: 10.1109/ISMAR-Adjunct57072.2022.00183.

[Ma22] Mazzuco, A.; Krassmann, A. L.; Reategui, E.; Gomes, R. S.: A Systematic Review of Augmented Reality in Chemistry Education. Review of Education 10 (1), 2022, ISSN: 2049-6613, 2049-6613, DOI: 10.1002/rev3.3325.

[Me23] Meinhardt, F.: Redesigning "SchülerlARbor Chemie"for Microsoft HoloLens, Bachelorthesis, 2023.

[Me24a] Meinhardt, F.: GitLab Release von 'Neugestaltung eines Mixed Reality Prototypen für das SchülerlARbor Chemie', en-us, 2024, URL: https://gitlab.com/chemielarbor/hololens-client/-/releases/DELFI24-Demo, Stand: 19. 06. 2024.

[Me24b] Meinhardt, F.; Görzen, S.; Werkes, R.; Pelzer, P.; Krause, D.; Schroeder, M.; Schroeder, U.: Quellcode zu "Neugestaltung eines Mixed Reality Prototypen für das SchülerlARbor Chemie", 2024, DOI: 10.17605/OSF.IO/6PZ85, URL: https://osf.io/6pz85.

[Mi22] Microsoft-Learn: MRTK2-Unity Developer Documentation - MRTK 2, en-us, 2022, URL: https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/?view=mrtkunity-2022-05, Stand: 19. 06. 2024.

[PBC21] Park, S.; Bokijonov, S.; Choi, Y.: Review of Microsoft HoloLens Applications over the Past Five Years. Applied Sciences, 2021, DOI: 10.3390/app11167259.

[Pe23] Pelzer, Philipp: Ein interaktives Video mit multiplen Lernpfaden zum Thema Öptimierung einer Lithium-Ionen-Batterie (unveröffentlicht)", Masterthesis, RWTH Aachen University, 2023.

[Ro17] Rodríguez, J. J.; Pérez, A. S.; Costa, S. M.; Nolasco, A. S.: Usability and design for augmented reality learning interfaces. Digital Education Review, S. 61–79, 2017, DOI: 10.1344/der.2017.31.61-79.

[SL16] Sauro, J.; Lewis, J. R.: Chapter 8 - Standardized Usability Questionnaires. In (Sauro, J.; Lewis, J. R., Hrsg.): Quantifying the User Experience (Second Edition). Morgan Kaufmann, Boston, S. 185–248, 2016, ISBN: 978-0-12-802308-2, DOI: 10.1016/B978-0-12-802308-2.00008-4.

[Wu13] Wu, H.-K.; Lee, S. W.-Y.; Chang, H.-Y.; Liang, J.-C.: Current Status, Opportunities and Challenges of Augmented Reality in Education. Computers & Education 62, S. 41–49, 2013, ISSN: 03601315, DOI: 10.1016/j.compedu.2012.10.024.

[Wy21] Wyss, C.; Bührer, W.; Furrer, F.; Degonda, A.; Hiss, J. A.: Innovative Teacher Education with the Augmented Reality Device Microsoft HoloLens - Results of an Exploratory Study and Pedagogical Considerations. Multimodal Technol. Interact. 5, S. 45, 2021, DOI: 10.3390/mti5080045.

Web-based prototype of a visual and interactive deep learning simulation

Christian Koch[1], Frederic Salmen [2], and Ulrik Schroeder [2]

Abstract: Due to its rapid development and growing relevance in society, the research field of artificial intelligence (AI) is finding its way into more and more schools. One subsection of artificial intelligence, deep learning, is especially relevant since most of the rapid development of the study field in recent years can be attributed to it. Because of its complex nature, deep learning is very suitable to be taught in a visual and interactive way, which can benefit learning outcomes. Existing solutions can convey basic knowledge about deep learning, but none of them are well suited to customization or facilitating deep understanding. We introduce a visual and interactive deep learning simulation with rich possibilities for personalization by teachers and students, usable in any web context.

Keywords: Explorables, neural networks, simulations

1 Motivation & Related Work

How can today's deep learning and large language models (LLMs) be understood? That it is necessary to understand them seems out of question to us: Through transformer-based models [Va23] such as GPT [Ra18], Gemini [VP23], and LlaMA [To23], large language models have permeated the digital world, and the topic is discussed across the research community [Ph23; Ra23; VP23]. Yet, there are fundamental challenges in understanding those models completely [Te20]: They are brittle (can be derailed by small changes in input data), threaten privacy (since they require massive amounts of data), biased (biases in training data produce similarly biased models) and they are opaque (it is difficult to reason why they behave the way they do). Understanding these attributes of LLMs which are caused by their data-driven, rather than algorithmic, nature requires new mental models and competencies, which can be summarized as "Computational Thinking 2.0"[TDT21]. Learning technologies may be able to support educational approaches: Simulations are already well-investigated and successful for teaching complex topics [WAP08]. They could be helpful to teach LLMs, as well. For these reasons, a visual and interactive deep learning simulation is worth prototyping.

Apart from traditional learning materials like textbooks and collections of worksheets about artificial intelligence, more modern approaches exist in the form of interactive simulations. Different key strategies to effectively visualize deep learning in such a simulation were

1 Unaffiliated Author, mail@christian-koch.eu
2 RWTH Aachen University, Learning Technologies, Ahornstr. 55, 52074 Aachen, Germany, salmen@cs.rwth-aachen.de, https://orcid.org/0009-0007-5206-1146; schroeder@cs.rwth-aachen.de, https://orcid.org/0000-0002-5178-8497

doi:10.18420/delfi2024_48

proposed in the community, among them visualizing a neural network through an (animated) node-link-diagram, plotting input data and training metrics, and providing interactivity up to a degree that lets users customize parameters and the network architecture [Ho18].
A popular simulation making use of these techniques is *TensorFlow Playground* [Sm17][3] which is a web application that visualizes a neural network and offers rich options to customize the network and corresponding training parameters. However, because of the complexity of the tool, it provides novice users not with much more than a basic intuition of what deep learning and neural networks are. Furthermore, tools like *Machine Learning for Kids*[4] and *Teachable Machine* [Ca20][5] by Google explicitly target students. While the tools introduced do a great job at introducing the topic and giving first insights, they do not go in-depth or allow customization by teachers or students.

2 The Simulation

Based on the previously introduced different strategies to visualize neural networks effectively and the features existing simulations offer or do not offer, a new visual and interactive deep learning simulation was designed and developed (see Fig. 1). A major focus was put on offering teachers the greatest possible flexibility in customizing the experience for students while making the simulation accessible and easy-to-understand for students in each possible configuration. This flexibility is expressed in the opportunities for teachers to disable and hide certain features such as advanced training parameters all together or to restrict certain editing options for students in a separate settings pane, only accessible in a *teacher mode*. This is valuable to not confuse students with too many options that might even not be part of the curriculum. At the same time, care was taken to ensure that advanced options are still explained in simple terms when chosen to enable them. The simulation can also run fully offline (see subsection 'Technology' for details).

Canvas. The centrepiece of the simulation is a canvas filling the main part of the viewport. On that canvas, users can gradually assemble a neural network by dragging *layers* from a sidebar on the right onto the canvas. There are a few basic types of layers available with the option to add new advanced types of layers with future updates without introducing any breaking changes for existing simulations. When placed on the canvas, a selection of a layer (or a *neuron* inside it) spawns a new pane in the right sidebar that explains the layer (resp. neuron) and offers additional configuration options (if not restricted or hidden by the teacher). These options include customizing the number of neurons in a layer or changing a layer's activation function. Also, layers can be freely connected with each other, offering the possibility to not only create simple feed-forward networks but also more complex setups.

Data Input. A second focus is feeding the neural network with well-formatted data. The simulation offers a few data sets by default. Users can inspect the raw data itself and view

3 `https://playground.tensorflow.org/`, accessed April 1, 2024
4 `https://machinelearningforkids.co.uk`, accessed April 1, 2024
5 `https://teachablemachine.withgoogle.com/`, accessed April 1, 2024

descriptions for its *features* and *labels*. In addition to that, the simulation enables users to create and upload their own data sets. This is facilitated by a multi-step wizard that guides the user through the process of first selecting the right parameters for the data set and then writing or pasting the data (in a standard, comma-separated format) into a text field. This way, a teacher can, for example, create a humorous example that involves predicting a students' month of birth based on his/her height. It also enables the teacher (or the students) to do web searches for other deep learning data sets and use them right within the simulation with minimal additional configuration effort.

Training. With a correctly designed neural network and a suitable selected data set, the user can customize training parameters and start the training in another pane. During the training, statistics are shown and the network visualization is constantly updated.

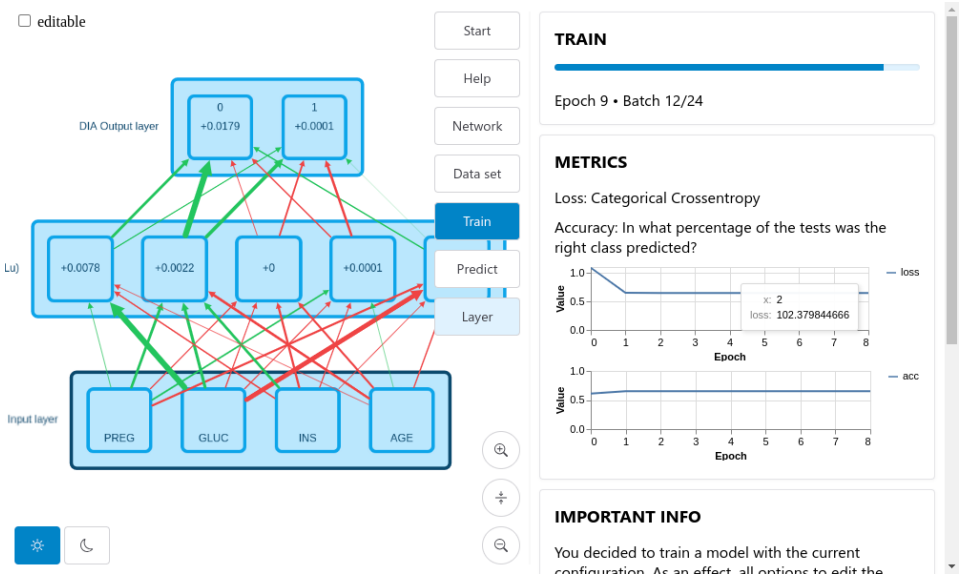


Fig. 1: The simulation while the training is in progress

Technology. The simulation is implemented as a Web Component with all functionality running in the browser, distributed as a Node Package `@webwriter/neural-network` on NPM[6]. It is written in TypeScript[7], with major technology choices including the library Cytoscape[8] to control the canvas, and Tensorflow.js[9] to handle training. This structure attains high standards of FAIRness in line with the expectations of the research community [Ch22; KS22], especially in terms of providing a long-lived, properly versioned and uniquely identified artifact (as a package on NPM) that is interoperable with any and

6 https://www.npmjs.com/, accessed April 3, 2024
7 https://www.typescriptlang.org/, accessed April 3, 2024
8 https://cytoscape.org/, accessed April 3, 2024
9 https://www.tensorflow.org/js, accessed April 3, 2024

all web frameworks (as a web component). Additional benefits include the ability to use the simulation offline. The simulation also satisfies the so-called widget interface of the Explorable authoring tool WebWriter, enabling the seamless use and editing of the simulation directly within the authoring tool.

Use Case. The simulation integrates into systems such as WebWriter [SRS23]. It may be used as part of an Explorable, a kind of digital worksheet that mixes traditional content like text with rich multimedia content like images, video and audio, and even interactive elements called widgets [Ho20; Vi11]. For example: An educator creates an Explorable for a first lesson on neural networks. They introduce the topic of neural networks with a video, then define some terms and give some examples with text. Each example could already be illustrated with an interactive simulation where they choose to disable most options to avoid distractions at this stage. For the main part of the lesson, the Explorable includes exercises to be solved with the simulation, this time enabling more options for the students, allowing them to explore the simulation and to answer questions using quiz inputs like multiple choice. Finally, students can be encouraged to reflect at the end of the lesson with a simple prompt and a free text input asking them what they have learned about neural networks. While this example is very simplified in terms of didactics, it showcases the potential of the simulation. Due to the interoperability of the simulation, other uses are possible, such as embedding it in a digital textbook or integrating it into a larger learning application.

3 Conclusion

The visual and interactive simulation that was just introduced provides a fresh take on visualizing deep learning, putting the needs of teachers and students front and center. It provides some unique value compared to the state of the art, such as the many customization options and the fully locally executed training on user-provided data. The interoperable and FAIR choices of technology may also enable future researchers to use the same simulation to address different research questions. The most crucial limitation of the simulation currently is its less than optimal performance, especially while the training is in progress. Since larger network configurations result in more and more nodes and links between nodes being displayed and these are re-rendered after each single training step, the application can get into an unresponsive state quickly. Therefore, improvements to the architecture need to be made. Future work should additionally focus on evaluating the simulation, both in terms of usability and of the value for education. This could include evaluating the usability and educational value from the perspective of teachers, but also testing the simulation with students. It is worth investigating which age groups the simulation could be used for. While it is reasonable to assume that younger students would struggle more with the innate complexity of the topic, efforts by other researchers have shown that machine learning fundamentals can be made accessible even for young students [Po23]. With some further research, this simulation could support the wider efforts of the research community to address the novel topics of deep learning and LLMs in education.

Acknowledgments We were supported by the research project NERD II (005-2201-0014), sponsored by the state of North Rhine-Westphalia.

References

[Ca20] Carney, M.; Webster, B.; Alvarado, I.; Phillips, K.; Howell, N.; Griffith, J.; Jongejan, J.; Pitaru, A.; Chen, A.: Teachable Machine: Approachable Web-Based Tool for Exploring Machine Learning Classification. In. Association for Computing Machinery, 2020, ISBN: 9781450368193, DOI: 10.1145/3334480.3382839, URL: https://doi.org/10.1145/3334480.3382839.

[Ch22] Chue Hong, N. P.; Katz, D. S.; Barker, M.; Lamprecht, A.-L.; Martinez, C.; Psomopoulos, F. E.; Harrow, J.; Castro, L. J.; Gruenpeter, M.; Martinez, P. A.: FAIR Principles for Research Software (FAIR4RS Principles). Zenodo, 2022.

[Ho18] Hohman, F.; Kahng, M.; Pienta, R.; Chau, D. H.: Visual analytics in deep learning: An interrogative survey for the next frontiers. IEEE transactions on visualization and computer graphics 25 (8), pp. 2674–2693, 2018.

[Ho20] Hohman, F.; Conlen, M.; Heer, J.; Chau, D. H. (: Communicating with Interactive Articles. Distill 5 (9), e28, 2020, ISSN: 2476-0757, DOI: 10.23915/distill.00028.

[KS22] Kiesler, N.; Schiffner, D.: On the Lack of Recognition of Software Artifacts and IT Infrastructure in Educational Technology Research. In: 20. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., pp. 201–206, 2022, ISBN: 978-3-88579-716-6.

[Ph23] Philbin, C. A.: Impact of Generative AI on K-12 Students' Perceptions of Computing: A Research Proposal. In: Proceedings of the 18th WiPSCE Conference on Primary and Secondary Computing Education Research. WiPSCE '23, Association for Computing Machinery, New York, NY, USA, pp. 1–2, 2023, ISBN: 9798400708510, DOI: 10.1145/3605468.3609775.

[Po23] Pope, N.; Kahila, J.; Vartiainen, H.; Tedre, M.: Children's AI Design Platform for Making and Deploying ML-Driven Apps, 2023, DOI: 10.36227/techrxiv.24320794.v1, visited on: 04/09/2024.

[Ra18] Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I.: Improving Language Understanding by Generative Pre-Training. 2018.

[Ra23] Rabe, A.; Kreinsen, M.; Grospietsch, F.; Schulz, S.: How to Deal with Students' Conceptions: Conceptual Change Texts on Artificial Intelligence. In: Proceedings of the 18th WiPSCE Conference on Primary and Secondary Computing Education Research. WiPSCE '23, Association for Computing Machinery, New York, NY, USA, pp. 1–2, 2023, ISBN: 9798400708510, DOI: 10.1145/3605468.3609781.

[Sm17] Smilkov, D.; Carter, S.; Sculley, D.; Viégas, F. B.; Wattenberg, M.: Direct-Manipulation Visualization of Deep Networks, 2017, arXiv: 1708.03788 [cs.LG].

[SRS23] Salmen, F.; Roepke, R.; Schroeder, U.: WebWriter: A System to Author and Remix Explorables–Requirements & First Prototype. 2023.

[TDT21] Tedre, M.; Denning, P.; Toivonen, T.: CT 2.0. In. Pp. 1–8, 2021, DOI: 10.1145/3488042.3488053.

[Te20] Tedre, M.; Vartiainen, H.; Kahila, J.; Toivonen, T.; Jormanainen, I.; Valtonen, T.: Machine Learning Introduces New Perspectives to Data Agency in K—12 Computing Education. In: 2020 IEEE Frontiers in Education Conference (FIE). IEEE, pp. 1–8, 2020.

[To23] Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; Rodriguez, A.; Joulin, A.; Grave, E.; Lample, G.: LLaMA: Open and Efficient Foundation Language Models, 2023, DOI: 10.48550/arXiv.2302.13971, arXiv: 2302.13971 [cs].

[Va23] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I.: Attention Is All You Need, 2023, DOI: 10.48550/arXiv.1706.03762, arXiv: 1706.03762 [cs].

[Vi11] Victor, B.: Explorable Explanations, http://worrydream.com/ExplorableExplanations/, 2011, visited on: 08/23/2022.

[VP23] Vo, G. M.; Pancratz, N.: AI Education in German K-10 Computer Science Curricula. In: Proceedings of the 18th WiPSCE Conference on Primary and Secondary Computing Education Research. WiPSCE '23, Association for Computing Machinery, New York, NY, USA, pp. 1–4, 2023, ISBN: 9798400708510, DOI: 10.1145/3605468.3605471.

[WAP08] Wieman, C. E.; Adams, W. K.; Perkins, K. K.: PhET: Simulations That Enhance Learning. Science 322 (5902), pp. 682–683, 2008, DOI: 10.1126/science.1161948.

Paare finden: Ein konfigurierbares Lernspiel mit Analyse-Dashboard



Michael Striewe [1]

Abstract: Lernspiele, bei denen Kartenpaare mit zueinandergehörigen Abbildungen oder Begriffen aufgedeckt werden müssen, können in vielen verschiedenen Lernszenarien sinnvoll eingesetzt werden. Diese Demo-Beitrag stellt ein web-basiertes System vor, mit dem solche Spiele angeboten werden können. Das System enthält verschiedene Konfigurationsoptionen für die zielgerichtete, aber zufällige Zusammenstellung von Kartenpaare sowie ein Analyse-Dashboard, mit dem schwierige Kartenpaare und häufige Fehlgriffe analysiert werden können.

Keywords: Lernspiel, Paare finden, Dashboard

1 Hintergrund und Anforderungen

Es gibt viele, verschiedene Lernszenarien, in denen Spiele eingesetzt werden können, bei denen zueinandergehörige Kartenpaare aufgedeckt werden müssen. Dies fördert allgemein die Gedächtnisleistung und unterstützt das Lernen von Fakten, beispielsweise bei der Zuordnung von römischen Kaisern zu ihrer Regierungszeit und bekannten Portrait-Büsten oder der Zuordnung von Bundesländern zu ihren Wappen und ihren Hauptstädten. Der Einsatz ist jedoch nicht auf das Training derartiger Kompetenzen begrenzt, sondern kann beispielsweise auch erfolgen, um Studierende die Übersetzung zwischen verschiedenen Notationen wie zwischen Strukturformeln und Skelettformel in der organischen Chemie trainieren zu lassen. Dazu ist es jedoch notwendig, dass die zwei Karten eines Paares nicht denselben Inhalt zeigen und es kann notwendig sein, Karten aus einer größeren Menge zufällig ziehen zu müssen, um ungewollte Erinnerungseffekte auszuschließen. Anders als existierende Literatur, die oft einzelne Spiele und deren Wirkung vorstellt (z. B. [M 23; ZOP20]), stellt dieser Beitrag ein Framework vor, mit dem solche Spiele als Einzelspieler-Spiel für Computer oder Mobilgeräte umgesetzt werden können. Anders als ein bereits existierendes Framework [EJG20] erlaubt es zudem die Nutzung externer Datenquellen und integriert ein Analyse-Dashboard. Der Quellcode des Frameworks ist unter `https://gitlab.com/michael.striewe/paare-finden` verfügbar.

Die Idee für die Entwicklung entsprang einem konkreten Bedarf in der Lehre, bei dem eine bisher papierbaiserte Lösung digital umgesetzt werden sollte. Das Framework sollte dazu in der Lage sein, sowohl Bild- als auch Textinhalte auf den Karten darstellen können und Spiele mit unterschiedlich viele Kartenpaare anzubieten. Um eine hohe Nutzbarkeit auch auf kleinen Displays (z. B. auf Smartphones) sowie eine gute Lesbarkeit von längeren

1 Universität Duisburg-Essen, Fakultät für Informatik, Gerlingstraße 16, 45127 Essen, Deutschland, michael.striewe@paluno.uni-due.de, https://orcid.org/0000-0001-8866-6971

doi:10.18420/delfi2024_49

Begriffen gewährleisten zu können, sollte die Zahl der im Spiel befindlichen Kartenpaare allerdings begrenzt sein, damit die einzelnen Karten nicht zu klein werden. Daher sollte das Framework über Funktionen verfügen, durch die die im Spiel befindlichen Kartenpaare nach bestimmten Vorgaben aus einer größeren Menge von Karten zufällig ausgewählt werden, um auch umfangreichere Themen in Spielen abbilden zu können, ohne dass zu viele Karten gleichzeitig auf dem Bildschirm dargestellt werden müssen.

2 Spieloberfläche und -funktionen

Das Framework stellt eine browserbasierte Spieloberfläche (siehe Abb. 1) zur Verfügung, die mit der Maus oder auf einem Touchscreen bedient werden kann. Im oberen Bereich der Spieloberfläche kann ein Anleitungstext mit Beispielen für gültige oder ungültige Kartenpaare angezeigt werden. Darunter folgt das eigentliche Spielfeld, das immer so skaliert wird, dass es die Bildschirmfläche voll ausfüllt. Auf dem Spielfeld liegen verdeckte Karten, die durch Anklicken bzw. Antippen einzeln umgedreht werden können. Werden zwei zueinandergehörige Karten aufgedeckt, werden diese mit einem kleinen, grünen Haken markiert und bleiben für den Rest des Spieldurchgangs aufgedeckt. Andernfalls werden beide Karten nach einer kurzen Wartezeit wieder umgedreht.

Das Framework kann mit Einwilligung der Nutzenden anonymisiert Logdaten aufzeichnen, aus denen die Reihenfolge der aufgedeckten Karten und die Zeitabstände zwischen jeder Aktion hervorgehen. Dazu erfolgt vor dem Start des Spiels eine Abfrage zur Zustimmung für die Datenerhebung. Logdaten werden nur für vollständig abgeschlossene Spiele an den Server gesendet. Nachdem ein Spieldurchgang abgeschlossen wurde besteht die Möglichkeit, sofort einen neuen Durchgang zu starten, für den die Karten neu gemischt werden.

3 Konfiguration von Lernspielen

Das Framework stellt seine Funktionen unabhängig von einem konkreten Einsatzzweck zur verfügung, der erst durch das Anlegen der jeweiligen Lernspiele definiert wird. Die Konfiguration einzelner Lernspiele erfolgt dazu jeweils anhand einer JSON-Datei. In dieser können beschreibende Informationen (Name des Spiels, Anleitungstext, Credits), die Hintergrundfarbe des Spielfelds, die Zahl der Kartenpaare sowie eine beliebige Menge von Beispielen für korrekte und falsche Paare hinterlegt werden. Außerdem kann angegeben werden, ob anonymisierte Logdaten erhoben werden sollen.

Der wichtigste Inhalt der Konfiguration ist jedoch die Angabe der Kartenpaare. Dabei kann grundsätzlich zwischen der Angabe einer festen Liste bzw. Tabelle von Elementen (Texte oder URLs) und der Angabe einer SPARQL-Query gewählt werden. Letztere muss so gestaltet sein, dass sie wiederum eine Liste bzw. Tabelle zurückgibt. Die Tabellen dürfen mehr Einträge enthalten, als benötigt werden. Dann werden für jedes Spiel zufällig Zeilen ausgewählt.

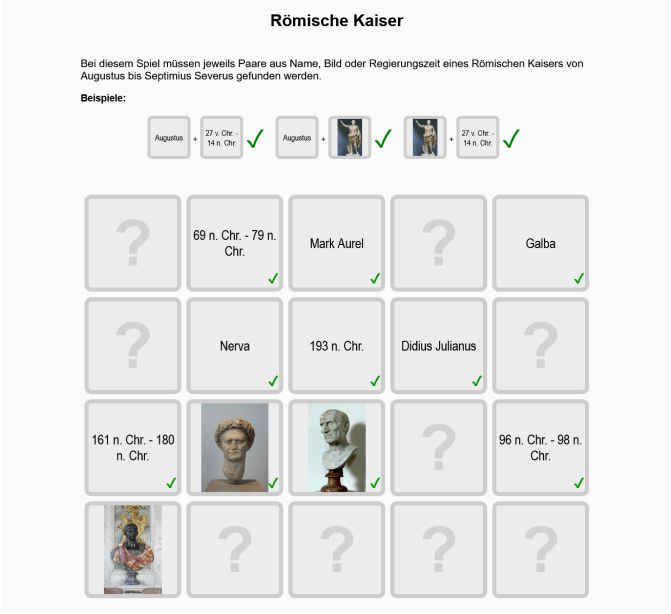


Abb. 1: Ansicht des Spielfelds mit einigen aufgedeckten Karten. Die Karte links unten wurde zuletzt aufgedeckt. Die anderen, aufgedeckten Karten gehören zu erfolgreich identifizierten Kartenpaaren.

Enthält die angegebene bzw. über die Query erzeugte Tabelle nur eine Spalte, werden Kartenpaare erzeugt, bei denen beide Karten denselben Inhalt haben. In diesem Fall kann mit dem Spiel eine fachunabhängige Gedächtnisleistung trainiert werden. Enthält die Tabelle zwei Spalten, werden Kartenpaare erzeugt, bei denen die eine Karte aus der ersten Spalte und die zweite Karte aus der zweiten Spalte gebildet wird. Enthält die Tabelle drei Spalten, werden für jedes Kartenpaar zufällig zwei der drei Spalten ausgewählt. Mit einer zusätzlichen Option kann erzwungen werden, dass immer die erste Spalte verwendet wird und nur für die zweite Karte eine zufällige Wahl erfolgt. Mit diesen Optionen können Spiele zum Training fachspezifischer Erinnerungsleistungen (z. B. Bundesländer und ihre Hauptstätte) oder Analysefähigkeiten (z. B. Erkennung identischer Moleküle in unterschiedlichen Notationen) erstellt werden.

Das Framework ist in der Lage, mit Sondersituationen in der Tabelle umzugehen: Insbesondere bei der Erzeugung der Tabelle über eine SPARQL-Query kann es vorkommen, dass einzelne Zellen leer bleiben. Beispielsweise haben die Stadtstaaten unter den deutschen Bundesländern keine separate Hauptstadt wie die anderen Bundesländer. In diesem Fall werden vom Framework nur Kartenpaare aus dem Namen des Bundeslandes und dem Wappen gebildet und keine leeren Karten erzeugt. Ebenso kann es passieren, dass es doppelte Einträge in der Tabelle gibt, die zu uneindeutigen Zuordnungen führen würden. Beispielsweise gibt es zwei römische Kaiser, die ausschließlich im Jahr 69 n. Chr. regierten.

In diesem Fall stellt das Framework sicher, dass in jedem Spieldurchgang maximal einer dieser Kaiser vorkommt.

4 Analyse-Dashboard

Im Framework ist ein Analyse-Dashboard integriert (siehe Abb. 2), mit dem die anonymisiert erhobenen Logdaten jedes Spiels ausgewertet werden können. Das Dashboard gibt die durchschnittliche Anzahl an Aktionen pro Spieldurchgang an, um so einen Vergleich der Schwierigkeit verschiedener Spiele zu ermöglichen.

Ferner gibt es für jede in der Konfiguration definierte Karte an, wie häufig diese im Spiel vorgekommen ist. Für jede Karte wird zudem angegeben, wie oft diese falsch zugeordnet wurde. Dabei wird zwischen verschiedenen Arten von Fehlern unterschieden, z. B. dem Aufdecken einer falschen, schon einmal aufgedeckten Karte und dem Aufdecken einer bisher unbekannten Karte, obwohl die gesuchte Karte ebenfalls schon bekannt war.

Alle Daten werden pro Kartenset aggregiert und es wird berechnet, ob es bei einzelnen Karten oder Kartensets signifikante Abweichungen von der durchschnittlichen Zahl falscher Zuordnungen gibt. So können Karten identifiziert werden, die besonders häufig oder selten verwechselt werden, was Rückschlüsse auf die fachliche Schwierigkeit der dargestellten Inhalte ermöglicht. Weitergehende Analysen der Daten (insbesondere zum zeitlichen Verlauf der Spieldurchgänge) sind möglich (vgl. [OAG22] für weitere Auswertungsmöglichkeiten derartiger Spiele), aber bisher nicht implementiert. Dabei ist insbesondere zu berücksichtigen, dass je nach Gestaltung und Ziel des Spiels unterschiedliche Analysen sinnvoll sind.

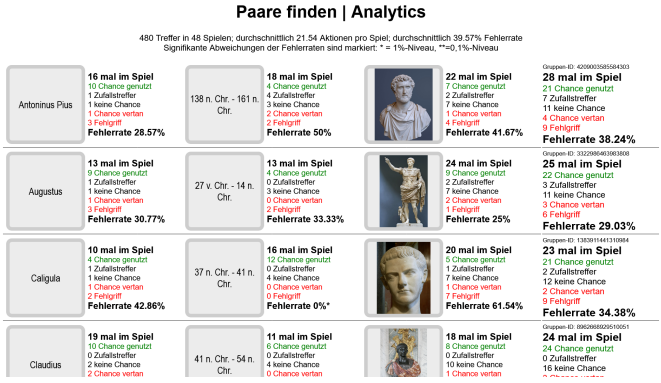


Abb. 2: Ausschnitt des Analyse-Dashboards mit statistischen Informationen über die Gesamtzahl der Spiele (oben), einzelne Karten (links) und Kartensets (rechts).

Literaturverzeichnis

[EJG20] de Estrella Hernández-Rivera, L.; Jaimez-González, C. R.; García-Mendoza, B.: Web Platform to Create Memory Games for Educational Purposes. Universal Journal of Educational Research 8 (12), S. 6498–6508, 2020, DOI: 10.13189/ujer.2020.081215.

[M 23] M R, A.; Vengali, A.; Jain, A.; Parashar, A.; Jatti, A.; Chankhore, N.: Card Blitz: Memory Retention Football Card Matching Game. In: 2023 3rd International Conference on Intelligent Technologies (CONIT). IEEE, S. 1–6, 2023, DOI: 10.1109/conit59222.2023.10205763.

[OAG22] Orellana, M.; Acosta-Urigüen, M.-I.; García, R. R.: Implementation of Clustering Techniques to Data Obtained from a Memory Match Game Oriented to the Cognitive Function of Attention. In: Information and Communication Technologies. S. 201–216, 2022, DOI: 10.1007/978-3-031-18272-3_14.

[ZOP20] Zhunio, C. S.; Orellana, P. C.; Patiño, A. V.: A Memory Game for Elderly People: Development and Evaluation. In: Seventh International Conference on eDemocracy & eGovernment (ICEDEG). S. 248–252, 2020, DOI: 10.1109/ICEDEG48599.2020.9096862.

Sicherer Einsatz von xAI in der Bildung: Erkennung von LLM-Halluzinationen bei der Generierung von Lehr- und Lernmaterialien

Benjamin Ledel [1] und Tabea Schwarz [2]

Abstract: Diese Demonstration bietet einen praktischen Einblick in eine neu entwickelte generative xAI (explainable AI), die im Bildungskontext eingesetzt werden kann, da sie eine strikte Einhaltung der DSGVO gewährleistet. Es handelt sich hierbei um ein für den Bildungsbereich optimiertes Large Language Model (LLM) - ohne Bindung an Open AI o.ä. - das in Deutschland betrieben und gehostet wird. Der Schwerpunkt der Demonstration liegt auf der Veranschaulichung der Funktionsweise dieser xAI, die neben der Textgenerierung auch Video-, Bild- und Audiogenerierung umfasst. Es wird gezeigt, wie durch die Struktur der xAI im Gegensatz zu herkömmlichen KIs Halluzinationen erkannt werden können. Darüber hinaus wird einerseits demonstriert, wie die Schnittstelle zwischen dieser künstlichen Intelligenz und H5P es Lehrenden ermöglicht, interaktives Lehrmaterial zu erstellen. Des Weiteren wird gezeigt, wie die xAI auf Basis des Lehrmaterials Fragen beantworten und somit als virtueller Tutor fungieren kann.

Keywords: Hochschulentwicklung, Technik, Datenschutz, Generative Künstliche Intelligenz, Dialogprozesse, Large Language Model, Generierung von Lehrmaterialien, explainable AI, Halluzinationen

1 Einleitung

Künstliche Intelligenz (KI) und maschinelles Lernen (ML) weisen eine Vielzahl revolutionärer und vielfältiger Anwendungsmöglichkeiten auf. Allerdings ist das Vertrauen der Nutzer:innen gegenüber diesen Tools oft begrenzt. Ein Teil dieses Misstrauens entsteht durch die Intransparenz herkömmlicher KI-Systeme. Es ist nicht nachvollziehbar, auf welcher Basis Ergebnisse erstellt wurden und welchem Entscheidungsprozess sie zugrunde liegen. Diese Intransparenz wird oft auch als 'Black Box' bezeichnet [An21], [Ru19], [Pa15], [LBH15]. Die fehlende Transparenz bedingt zudem zwei Probleme. Zum einen ist unklar, wie Daten verarbeitet werden, wodurch diese Systeme nicht DSGVO-konform sind. Zum anderen ermöglicht es unkontrollierte Halluzinationen. Dies bedeutet im Bereich der generativen KI, dass Ergebnisse entstehen, die auf den ersten Blick richtig und überzeugend wirken, da sie scheinbar durch die Trainingsdaten gestützt werden, jedoch objektiv falsch sind [Ji23]. Die Überprüfung dieser falschen Ergebnisse gestaltet sich aufgrund des verborgenen Entscheidungsprozesses von KI-Systemen als schwierig und ist in einigen Fällen ohne Zugang zu den Originalquellen nicht möglich.

1 Digital Learning GmbH, Geschäftsführer, Am Sportplatz 4, 37115 Duderstadt, Deutschland, benjamin.ledel@digitallearning.gmbh, https://orcid.org/https://orcid.org/0000-0002-6776-0472
2 Digital Learning GmbH, Projektmanagerin, Am Sportplatz 4, 37115 Duderstadt, Deutschland, tabea.schwarz@digitallearning.gmbh, https://orcid.org/https://orcid.org/0009-0000-6966-8912

doi:10.18420/delfi2024_50

Die Herausforderungen, denen Anwender:innen herkömmlicher KI-Systeme gegenüberstehen, sind besonders bedeutend in sensiblen Sektoren wie dem Gesundheitswesen sowie anderen Bereichen, die das menschliche Leben, individuelle Rechte, Finanzen und die Privatsphäre betreffen [An21]. Dies schließt den Bildungsbereich mit ein. Um jedoch die Potenziale von KI-Systemen im Bildungsbereich nutzen zu können, bedarf es einer Lösung, die diese Probleme entweder behebt oder zumindest kontrollierbar macht. Eine solche Lösung wird als 'erklärbare KI' (xAI) bezeichnet. xAI ermöglicht es Menschen, die Gründe für maschinelle Entscheidungen zu verstehen und ihre Vertrauenswürdigkeit zu bewerten [An21]. Diese Kontrolle und Transparenz würden auch den Einsatz von KI in Bildungseinrichtungen fördern. Lehrkräfte und Lernende könnten die Antworten der KI-Systeme verstehen und überprüfen.
Insgesamt bietet die Integration generativer xAI-Technologien in das Bildungswesen ein vielversprechendes Potenzial, um das Vertrauen der Nutzer:innen zu stärken, die Transparenz von KI-Systemen zu erhöhen und die Bildungsqualität zu verbessern. Es ist jedoch entscheidend, diese Technologien sorgfältig zu integrieren und sicherzustellen, dass sie ethischen und pädagogischen Standards entsprechen. Dieser Artikel stellt eine solche Lösung vor: educa AI. Zunächst werden die Herausforderungen traditioneller KI und die Definition von xAI vorgestellt. Anschließend wird erläutert, wie educa AI entwickelt wurde, um diesen Herausforderungen zu begegnen.

2 Probleme durch Halluzinationen von KIs und xAI als Lösung

Die Zukunft der Bildung mit generativen KI-Modellen birgt sowohl Chancen als auch Risiken. Halluzinationen, also falsche Inhalte, die richtig erscheinen, stellen eine Gefahr dar [Ji23],[MP13],[Fi09], die zu Fehlinformationen bei Lernenden und falschen Inhalten bei Lehrenden führen können. Intransparente Informationsquellen und Verarbeitungsprozesse erhöhen dieses Risiko. Halluzinationen können die Leistung beeinträchtigen, Sicherheitsbedenken aufwerfen und die Privatsphäre gefährden, was Bildungseinrichtungen vor großen Herausforderungen stellt.
Daher ist es entscheidend, Halluzinationen von KI-Systemen zu adressieren. Der Einsatz von xAI bietet die Möglichkeit, Halluzinationen zu erkennen und entsprechend im Bildungskontext zu reagieren, wo eine präzise Informationsverarbeitung von großer Bedeutung ist. Die Notwendigkeit der Erklärbarkeit für die Einführung von KI-Systemen ist schon seit längerer Zeit in verschiedenen Bereichen deutlich geworden. Diskussionen über den Einsatz in autonomen Fahrzeugen, in der medizinischen Diagnostik und bei Finanzdienstleistungen zeigen ihre Bedeutung aus praktischen, sozialen und rechtlichen Gründen. Verbraucher haben gemäß der DSGVO das Recht auf Informationen über automatisierte Entscheidungen, weshalb Erklärungen das Systemmodell verständlich machen und Vertrauen stärken sollten [Co21]. Doch trotz jahrelanger Forschung besteht immer noch Uneinigkeit über das allgemeine Verständnis und die Definition(en) von Erklärbarkeit [Co21]. [Co21] definieren sieben Kriterien für xAI, die menschliche Bedürfnisse im Umgang mit KI berücksichtigen und eine gemeinsame Basis zwischen Mensch und KI-Systemen schaffen sollen: Kausalität,

Kontextualisierung, Sozialisation, Selektivität, Transparenz, Semantik und Interaktivität (eigene Übersetzung von [Co21]).
Es können also durch den Einsatz solcher xAI-Systeme im Bildungsbereich Halluzinationen erkannt und die Qualität der Ergebnisse verbessert werden, indem sichergestellt wird, dass sie den inhaltlichen und pädagogischen Anforderungen entsprechen. Dies wird am Beispiel von educa AI veranschaulicht.

3 educa AI: Ein xAI Tool für den Bildungsbereich

educa AI ist ein speziell in Deutschland entwickeltes Large Language Model (LLM), welches den sicheren Einsatz von KI im Bildungsbereich mit entscheidenden Vorteilen gegenüber herkömmlichen KI-Systemen ermöglicht. Es erfüllt dabei die für xAI essentiellen Eigenschaften gemäß den Studien von [Co21], wodurch das Risiko unkontrollierter Halluzinationen, wie sie bei herkömmlichen LLMs [Ji23] auftreten können, deutlich reduziert wird. Die Lösung von educa AI adressiert die Herausforderungen bei der Implementierung von KI in Bildungseinrichtungen, indem sie Transparenz, Anpassungsfähigkeit und Vielseitigkeit bietet und so das Vertrauen in KI stärkt und die Unterstützung von Bildungseinrichtungen reformiert. Nutzer:innen können jederzeit den Entscheidungsprozess und die verwendeten Quellen nachvollziehen. educa AI unterstützt Bildungseinrichtungen mit einer Vielzahl von Funktionen, darunter die Beantwortung von Wissensfragen, automatisierte Recherche, Erstellung von Inhalten und die nahtlose Integration von xAI mit H5P für interaktive Bildungsinhalte. Lehrende können maßgeschneiderte Inhalte erstellen, während H5P diese Inhalte auf moderne und interaktive Weise präsentiert, um den Lehr- und Lernprozess zu unterstützen. Darüber hinaus unterstützt educa AI das kontinuierliche Lernen durch die unbegrenzte Generierung von Übungsaufgaben, die es den Lernenden ermöglichen, Inhalte gezielt zu wiederholen.
Um die Anforderungen an die Transparenz und Erklärbarkeit zu erfüllen, wurde ein Agent-Pattern eingesetzt. Dies bedeutet, dass das trainierte LLM nur für das Verständnis der Sprache und nicht als eigentliche Wissensquelle verwendet wird. Die programmierten Agents, nutzen das Prinzip des Function Calling, um zunächst einen Plan zu entwickeln, wie eine bestimmte Aufgabe gelöst werden soll. Beispielsweise wird dem System eine Wissensfrage gestellt: Wann findet die Delfi 2024 statt? Der Agent, Knowledge AI, erstellt daraus einen entsprechenden Plan, um die Frage zu beantworten. Je nach Konfiguration kann der Plan variieren, beispielsweise kann er vorrangig oder ausschließlich lokales Wissen nutzen oder auch Suchmaschinen wie Google, DuckDuckGo oder ähnliche verwenden. In diesem Schritt werden also anfragerelevante Quellen gesucht. Dem Agent stehen aber auch andere Tools zur Verfügung, wie beispielsweise Tools zum Lesen von Webseiten, zum Zusammenfassen von Texten, zum Vergleichen von Texten, zum Bewerten von Relevanzen etc. Je nach Plan werden dann unterschiedliche Tools ausgewählt, um die Anfrage zu lösen. Das Besondere an educa AI ist, dass der Pfad der Quellen während des gesamten Prozesses erhalten bleibt. Dabei werden den Nutzer:innen sowohl der Plan des Agents als auch die Zwischenergebnisse zur Verfügung gestellt. Ziel ist es, dass der Prozess entsprechend

überprüft und nachvollzogen werden kann. Im Fall der oben geschilderten Wissensfrage können beispielsweise die ersten Treffer von Google verwendet werden. Anschließend werden die Seiten gelesen, zusammengefasst und abgeglichen.
In Kombination mit der niedrigen Temperatur des LLM und der Nachvollziehbarkeit des Quellenflusses verwenden wir eine xAI-typische Methode, um unser Ergebnis für die endgültigen Erklärungen der Ausgabe des LLM (insbesondere für die Zusammenfassung von Texten) zu validieren. Ähnlich zu Prinzipien wie SHAP und LIME (vgl. [Mo22] oder [Zh23]) blenden wir einzelne Treffer aus und untersuchen anschließend, wie sich das Ergebnis des LLMs von der Ausgabe unterscheidet. Da LLMs nicht deterministisch sind, wird dieser Vorgang mehrmals wiederholt. Innerhalb von educa AI kann so eine Distanz bestimmt werden, die es erlaubt, die Relevanz einer Quelle zum LLM-Output zu bewerten. Ein Beispiel für einen solchen Vorgang ist, wenn sich die Ausgabe eines LLMs nicht ändert, obwohl die Quelle nicht mehr in das LLM eingegeben wurde. In diesem Fall ist es naheliegend, dass die Ausgabe des LLMs auf dem "Wissen" im Modell selbst basiert. Wenn sich die Ausgabe jedoch unterscheidet, dann ist es naheliegend, dass die Quelle einen relevanten Einfluss auf die Ausgabe des LLMs hat.
In Kombination mit der Nachvollziehbarkeit der Quellen können wir so eine Art Mapping berechnen, welches dann den Nutzer:innen zusammen mit dem Plan angezeigt wird. Auf diese Weise können die Nutzer:innen sehen, wie der Agent vorgegangen ist, welche Quellen verwendet wurden und welchen Einfluss diese auf das Ergebnis haben - ganz im Sinne von xAI.
Durch ihre Struktur bietet educa AI eine für den Bildungsbereich zugeschnittene Lösung, die auf Transparenz, Sicherheit und Anpassbarkeit basiert und somit eine effektive und vertrauenswürdige Unterstützung für Lehrende und Lernende darstellt.

4 Konklusion

Die Integration von KI-Systemen in den Bildungsbereich verspricht, die Unterstützung von Lehr- und Lernprozessen zu revolutionieren. Das Vertrauen in herkömmliche KI-Systeme ist jedoch oft begrenzt, insbesondere aufgrund ihrer mangelnden Transparenz und der damit verbundenen Gefahr von Halluzinationen. Diese Herausforderungen werden durch xAI adressiert, welches ermöglicht, die Entscheidungsprozesse von KI-Systemen nachzuvollziehen und zu kontrollieren.
educa AI stellt dabei eine auf den Bildungsbereich zugeschnittene xAI-Lösung dar. Gegenüber herkömmlichen KI-Systemen bietet sie entscheidende Vorteile wie erhöhte Transparenz, kontrollierbare Halluzinationen und Anpassbarkeit an die spezifischen Bedürfnisse von Bildungseinrichtungen. Darüber hinaus wird durch die Integration von educa AI in die Open Source Plattform H5P der Lehrprozess nicht nur effizienter, sondern auch offener gestaltet. educa AI ist somit eine wertvolle Entwicklung für den Bereich der xAI-Systeme für den Bildungsbereich. Es ist geplant, weitere Tests in weiteren Bildungseinrichtungen durchzuführen und damit die Auswirkungen dieses Systems auf das Vertrauen und die Usability durch eine Studie wissenschaftlich zu untersuchen.

Literaturverzeichnis

[An21] Angelov, P.; Soares, E.; Jiang, R.; Arnold, N.; Atkinson, P.: Explainable artificial intelligence: an analytical review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 11 (5), e1424, 2021.

[Co21] Confalonieri, R.; Coba, L.; Wagner, B.; Besold, T.: A historical perspective of explainable Artificial Intelligence. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 11 (1), e1391, 2021.

[Fi09] Fish, W.: Perception, hallucination, and illusion. OUP USA, 2009.

[Ji23] Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y.; Madotto, A.; Fung, P.: Survey of hallucination in natural language generation. ACM Computing Surveys 55 (12), S. 1–38, 2023.

[LBH15] LeCun, Y.; Bengio, Y.; Hinton, G.: Deep learning. nature 521 (7553), S. 436–444, 2015.

[Mo22] Mosca, E.; Szigeti, F.; Tragianni, S.; Gallagher, D.; Groh, G.: SHAP-based explanation methods: a review for NLP interpretability. In: Proceedings of the 29th international conference on computational linguistics. S. 4593–4603, 2022.

[MP13] Macpherson, F.; Platchias, D.: Hallucination: Philosophy and psychology. MIT Press, 2013.

[Pa15] Pasquale, F.: The black box society: The secret algorithms that control money and information. Harvard University Press, 2015.

[Ru19] Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence 1 (5), S. 206–215, 2019.

[Zh23] Zhang, J.; Ma, X.; Zhang, J.; Sun, D.; Zhou, X.; Mi, C.; Wen, H.: Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. Journal of environmental management 332, S. 117357, 2023.

easyProtocol: Ein digitales Unterstützungstool für mündliche Prüfungen

Zoe Jebing[1], Armin Egetenmeier[1] und Sven Strickroth[1]

Abstract: Protokolle stellen eine notwendige Dokumentation für spezielle Prüfungsformate wie mündlichen Prüfungen dar. Das Protokoll erfüllt dabei primär die rechtlichen Anforderungen eines Nachweises des korrekten Ablaufs, wobei dessen Inhalte oft nicht weiter genutzt werden. In diesem Demo-Beitrag wird ein Prototyp eines Unterstützungstools für mündliche Prüfungen vorgestellt. Das Tool unterstützt zum einen den Protokollführenden, den Prüfungsablauf strukturiert zu dokumentieren und ein rechtskonformes Protokoll zu erstellen. Zum anderen werden die eingegebenen Daten aber auch analysiert, um z. B. die Bewertung der Prüfung durch eine übersichtliche Zusammenfassung zu unterstützen. Dadurch sollen u. a. die Prüfungsbewertungen transparenter und vergleichbarer werden.

Keywords: Protokoll, mündliche Prüfung, Technology-Enhanced Learning, Assessment Analytics

1 Motivation

Prüfungen sind ein essentieller Bestandteil im Bildungsbereich als summative Bewertung des Lernergebnisses einzelner Lehrveranstaltungen [KP08]. Dieser Nachweis kann durch verschiedene Prüfungsformen erfolgen, wie beispielsweise schriftliche Klausuren, studentische Ausarbeitungen (in Form von Portfolios oder Seminararbeiten), Vorträge oder mündliche Prüfungen. Ziel der Prüfungen ist dabei, das Wissen der Lernenden bzw. das Verständnis bezüglich eines Themas festzustellen und zu bewerten. Neben einer Benotung muss jede Prüfung entsprechend dokumentiert (und archiviert) werden. Dies ist bei den meisten schriftlichen Prüfungsformaten einfach und ermöglicht mit überschaubarem Aufwand auch einen Vergleich zwischen unterschiedlichen abgelegten Prüfungen.

Allgemein ist die Durchführung von Prüfungen für Lehrende mit einem nicht zu unterschätzenden Aufwand verbunden. Daher wurden zum Teil spezielle Unterstützungssysteme für Prüfungen entwickelt, die vor allem in Großveranstaltungen einen effizienten Ablauf ermöglichen [SB22]. Etablierte Systeme, die schriftliche Prüfungen auf verschiedenen Ebenen unterstützen können, sind z. B. EvaExam (für sowohl rein digitale als auch scanbare Papier-Klausuren) oder spezialisierte Systeme wie JACK [GS09] oder GATE [SOP11] für Programmieraufgaben. Diese Systeme unterstützen die Erstellung von Klausuren, die digitale Durchführung inkl. Protokollierung der Prüfung, oder die (semi-)automatische Bewertung [SH22]. Teilweise werden studentische Ausarbeitungen auch über LMS wie Moodle oder per E-Mail eingereicht, um den (organisatorischen) Aufwand der Lehrenden zu

1 Ludwig-Maximilians-Universität München, Institut für Informatik, Oettingenstraße 67, 80538 München, Z.Jebing@campus.lmu.de; armin.egetenmeier@ifi-lmu.de, https://orcid.org/0009-0001-3944-3277; sven.strickroth@ifi.lmu.de, https://orcid.org/0000-0002-9647-300X

doi:10.18420/delfi2024_51

mindern. Den Autoren sind bisher keine Systeme bekannt, welche sich auf die Unterstützung der Lehrpersonen in mündlichen Prüfungen beziehen.

In diesem Beitrag wird ein spezielles Unterstützungstool für mündliche Prüfungen präsentiert. Im Folgenden wird zuerst der spezielle Kontext von mündlichen Prüfungen vorgestellt und schließlich auf die Konzeption und Umsetzung des easyProtocol-Tools eingegangen. Im Ausblick werden Features für die nächste Weiterentwicklung des Tools dargestellt.

2 Mündliche Prüfungen

Akimov et al. beschreiben den aktuellen Forschungsstand im Bereich mündliche Prüfungen mit Blick auf Vor- und Nachteile und zeigen in ihrer Fallstudie, dass dies ein qualitativ hochwertiges Prüfungsformat ist [AM20]. Dabei werden die Herausforderungen der Studierenden (Prüfungsangst), der Hochschulen (zeitlicher Aufwand in Vorbereitung und Durchführung) und des Prüfungsformats aus testtheoretischer Sicht (Objektivität, Reliabilität und Validität) diskutiert. Auf die rechtlichen Aspekte der Dokumentationspflicht in Form eines Protokolls wird nicht eingegangen.

Mündliche Prüfungen stellen eine dynamische (auch hektische) Prüfungsform dar, die sich in der Praxis durch eine Vielzahl an möglichen Ausprägungen (hinsichtlich Ablauf, inhaltlicher Gestaltung, Dauer, etc.) auszeichnet [AM20; Jo98; St02; Wi18]. Die Vergleichbarkeit einzelner Prüfungen dieses Formats ist daher komplex, insbesondere wenn sich der Prüfungszeitraum über mehrere Tage oder Wochen erstreckt. Auch objektive Bewertungskriterien und zugehörige Fragen zu finden, kann sich als schwierig herausstellen. Dies wird meist nur durch die Expertise und Erfahrung des Prüfenden sichergestellt. Dabei orientieren sich diese Personen oft z. B. an einer Liste der Themengebiete oder vorbereiteter, spezifischer Fragen. Aus rechtlichen Gründen müssen alle wesentlichen Aspekte der Prüfung in einem Ergebnisprotokoll dokumentiert werden. Dieses Protokoll wird durch eine sachkundige Person (meist wissenschaftliches Personal) während der Prüfung erstellt – oft handschriftlich oder mit Hilfe einer einfachen Textverarbeitungssoftware – und ist eigenhändig zu unterschreiben. Weiterhin wird das Protokoll oft auch zur Rekapitulation der Prüfung bei der Notenfindung herangezogen. Die Qualität des Protokolls hängt hierbei von der Erfahrung, den Fertigkeiten und der Verfassung (z. B. Konzentration, Müdigkeit, Aufmerksamkeit) des Protokollierenden ab. Durch eine technische Unterstützung kann dies vereinfacht und eine bessere Vergleichbarkeit mündlicher Prüfungen (zur gleichen Lehrveranstaltung) angestrebt werden.

Dies bildet die Grundidee für ein (digitales) Unterstützungstool dieses Prüfungsformates. Das Tool soll einerseits die Erstellung „guter“ Protokolle vereinfachen und anderseits Prüfenden weitere Einblicke ermöglichen. Dabei sollen insbesondere Protokollierende mit wenig Erfahrung unterstützt werden. Im Folgenden werden einige Features des easyProtocol-Tools beschrieben, das aktuell noch weiterentwickelt wird.

3 Konzeption, Umsetzung & Features

Basierend auf einer Interviewstudie (Paper in Vorbereitung), wurden die Anforderungen an das Tool abgeleitet. Neben rechtlichen Aspekten (vgl. [FJD22] bzw. Prüfungsordnungen), wurde insbesondere die bestehende Praxis und Arbeitsweise beachtet. Anhand von einzelnen Fallbeispielen in Form typischer mündlicher Prüfungsformaten wurde das Tool konzipiert.

Der Prototyp ist als webbasierte Anwendung umgesetzt und fokussiert sich auf das grundlegende Design für die Erstellung eines Protokolls aus Sicht des Beisitzers. Das Tool ermöglicht die Eingabe der gestellten Fragen sowie der gegebenen, zugehörigen Antworten (als Freitext oder durch Auswahl von gegebenen Schlüsselbegriffen oder Schlagwörtern) zur Dokumentation der Prüfung (siehe Abb. 1). Neben dem eigentlichen Inhalt des Ergebnisprotokolls (d. h. Prüfungsfragen und Antworten) kann auch eine zugehörige Bewertung (gemäß einer wählbaren Skala) eingetragen werden oder mittels Buttons zur vereinfachten Eingabe von wiederkehrenden Ereignissen erhoben werden (z. B. Nichtbeantwortung einer Frage). Auch weitere Aspekte, die sich auf die Notengebung auswirken könnten (wie Hilfestellung des Prüfenden), können über einen Mausklick dokumentiert werden. Vor der eigentlichen Prüfung werden grundlegende (Meta-)Daten (wie Fach, Namen der beteiligten Personen, Datum, Prüfungsfähigkeit, . . .) strukturiert abgefragt. Der Start und das Ende der Prüfung sowie der einzelnen Fragen werden automatisch erfasst und die zeitliche Dauer angezeigt. Nach der Prüfung kann die Note der Prüfung eingetragen werden. Als Ergebnis wird ein einheitliches Protokoll erstellt, das ausgedruckt, unterschrieben und archiviert werden kann.

Ein spezielles Feature ist eine optische Aufbereitung des Prüfungsablaufs zur Rekapitulation für die Notenfindung. Hierbei wird auch die Dauer jedes einzelnen Frage-Antwort Artefakts der Prüfung dargestellt. Die chronologische Darstellung bietet einen schnellen Einblick, wobei nennenswerte Ereignisse wie zusätzliche Erläuterungen durch den Prüfenden im Nachgang an die Prüfung schnell erfasst werden können. Dies bildet die Grundlage für weitere Features, die Auswertungen im Sinne von Assessment Analytics [El17] ermöglichen.

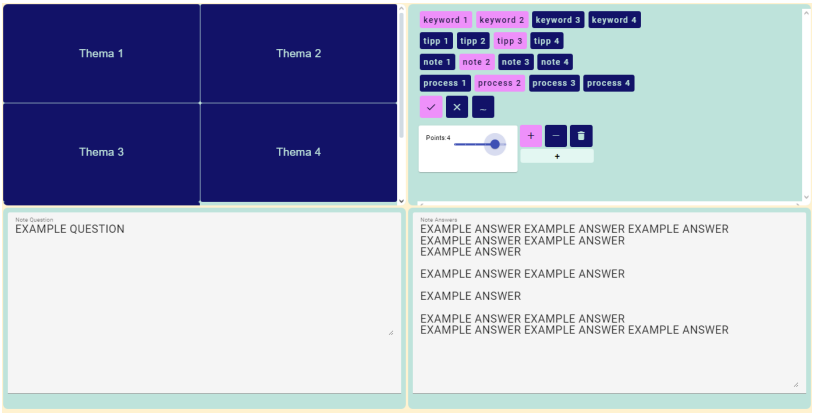


Abb. 1: Eingabemaske für das digitale Prüfungsprotokoll

4 Ausblick

Das easyProtokoll-Tool dient als konzeptioneller und explorativer Prototyp, welcher anhand spezifischer Fallbeispiele entwickelt wurde. Der Ansatz eignet sich daher für viele unterschiedliche Ausprägungen von mündlichen Prüfungen, auch außerhalb der Informatik. Eine Evaluation des Prototyps z. B. hinsichtlich Usability steht zur Zeit noch aus. Es ist geplant, den Prototypen auch für weitere Anwendungsfälle weiterzuentwickeln (z. B. zur Dokumentationen von Verteidigungen und Programmierpraktika).

Ein weiterer zentraler Aspekt des Konzepts ist die Verwendung der erhobenen Daten für Assessment Analytics. Durch die komplett digitale und strukturierte Erfassung des Prüfungsprozesses kann direkt in oder nach der Prüfung bzw. im Nachgang einer ganzen Prüfungsphase eine Unterstützung der Lehrenden erfolgen. In der Prüfung könnte das Tool adaptive Informationen und Rückmeldungen für den Prüfenden bereitstellen (z. B. eine Echtzeit-Analyse für den Lehrenden als Recommender-System, das auf seltenere Themen oder Fragen hinweist, oder Hinweise auf Abweichungen zu ähnlichen, bereits abgeschlossenen Prüfungen). Im Anschluss an eine Prüfung oder eine Prüfungsphase können zudem detailliertere Auswertung erfolgen. Bereits die Auswertungen der Metadaten, wie Anzahl Fragen/Themen pro Prüfung oder Dauer einzelner Themen in der Prüfung (z. B. durchschnittliche Zeit pro Frage-Antwort Artefakt) können zur Notenfindung herangezogen werden oder auch auf Prüfungen hinweisen, die vom angestrebten/typischen Muster abweichen. Die Zusammenfassungen zu den einzelnen Themengebieten können den Prüfenden auf schwierige Themen aus Studierendensicht hinweisen und auch häufig bzw. selten abgefragte Themengebiete werden damit sichtbar. In einer Weiterentwicklung könnte beispielsweise eine Auswertung der typischen Fragen als Zusammenfassung ausgegeben werden, die als Grundlage für künftige Prüfungen dienen kann. Somit kann sichergestellt werden, dass (immer) alle prüfungsrelevanten Themen angesprochen wurden. Diese Auflistung stellt eine Kohärenz zwischen der vergebenen Prüfungsnote und der Antworten auf die Fragen her. Dies ermöglicht es dem Prüfenden, einzelne Prüfungen besser zu reflektieren, deren Schwierigkeitsgrad abzuschätzen und bei Bedarf Änderungen am „typischen“ Ablauf weiterer Prüfungen vorzunehmen. Zur Unterstützung des Lehrenden ist auch eine einfache, teilautomatisierte Einordnung der Fragen gemäß der (überarbeiteten) Bloom'schen Taxonomie [AK01] vorgesehen. Diese Vorauswahl kann im Nachgang der Prüfung durch den Nutzer bestätigt bzw. angepasst werden. Auf ähnliche Weise könnten auch Freitextfragen im Nachgang einzelnen Themengebieten zugeordnet werden, um einen besseren Überblick zu erhalten und um Doppelungen in Fragen/Themen zu vermeiden. Dies soll ebenso zum besseren Verständnis des Prüfungsgeschehens und der daraus resultierenden Bewertung beitragen. Insgesamt soll das Tool so einen explorativen Einblick in dieses dynamische Prüfungsformat bieten. Für die genannten Features ist die Einbindung von KI-Methoden in einer Weiterentwicklung des easyProtokoll-Tools angedacht bzw. zu prüfen. Dabei geht es auch um die Unterstützung von eher formalen Aspekten, wie Formulierungshilfen.

Danksagung: Teile dieser Forschung wurden vom Bundesministerium für Bildung und Forschung (BMBF) im Projekt AIM@LMU unter dem Kennzeichen 16DHBKI013 gefördert.

Literatur

[AK01] Anderson, L. W.; Krathwohl, D. R.: A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman, New York, 2001.

[AM20] Akimov, A.; Malin, M.: When old becomes new: a case study of oral examination as an online assessment tool. Assessment & Evaluation in Higher Education 45 (8), S. 1205–1221, 2020, DOI: 10.1080/02602938.2020.1730301.

[El17] Ellis, C.: The Importance of E-Portfolios for Effective Student-Facing Learning Analytics. In (Chaudhuri, T.; Cabau, B., Hrsg.): E-Portfolios in Higher Education. Springer Singapore, Singapore, S. 35–49, 2017, DOI: 10.1007/978-981-10-3803-7_3.

[FJD22] Fischer, E.; Jeremias, C.; Dieterich, P.: Prüfungsrecht. C.H. Beck, München, 2022, ISBN: 978-3-406-77900-8.

[GS09] Goedicke, M.; Striewe, M.: A Flexible and Modular Software Architecture for Computer Aided Assessments and Automated Marking. In: Proceedings of the First International Conference on Computer Supported Education. SciTePress - Science, 2009, DOI: 10.5220/0001966900540061.

[Jo98] Joughin, G.: Dimensions of Oral Assessment. Assessment & Evaluation in Higher Education 23 (4), S. 367–378, 1998, DOI: 10.1080/0260293980230404.

[KP08] Kirkwood, A.; Price, L.: Assessment and student learning: a fundamental relationship and the role of information and communication technologies. Open Learning: The Journal of Open, Distance and e-Learning 23 (1), S. 5–16, 2008, DOI: 10.1080/02680510701815160.

[SB22] Strickroth, S.; Bry, F.: The Future of Higher Education is Social and Personalized! Experience Report and Perspectives. In: Proceedings of the 14th International Conference on Computer Supported Education - Volume 1: CSEDU. Bd. 1, INSTICC, SciTePress, S. 389–396, 2022, DOI: 10.5220/0011087700003182.

[SH22] Strickroth, S.; Holzinger, F.: Supporting the Semi-Automatic Feedback Provisioning on Programming Assignments. In: Methodologies and Intelligent Systems for Technology Enhanced Learning, 12th International Conference. Springer International Publishing, Cham, S. 13–19, 2022, DOI: 10.1007/978-3-031-20617-7_3.

[SOP11] Strickroth, S.; Olivier, H.; Pinkwart, N.: Das GATE-System: Qualitätssteigerung durch Selbsttests für Studenten bei der Onlineabgabe von Übungsaufgaben? In (Rohland, H.; Kienle, A.; Friedrich, S., Hrsg.): DeLFI 2011. GI-Edition Proceedings, Ges. für Informatik, Bonn, S. 115–126, 2011.

[St02] Stary, J.: „Doch nicht durch Worte nur allein. . . " – Die mündliche Prüfung. Neues Handbuch Hochschullehre 1 00 02 01 (H 2.1) (1), S. 1–30, 2002.

[Wi18] Willert, M.: Mündliche Prüfung: Handreichung der Prüfungswerkstatt, hrsg. von Zentrum für Qualitätssicherung und -entwicklung Mainz, Mainz, 2018.

Augmented Reality an der Schnittstelle von Hochschuldidaktik und Technologie: Integration von realen Objekten für medial-gestütztes Lernen

Marcella Haller [1] und Alexander Ast [1]

Abstract: Das Toolkit für Augmented Reality (bestehend aus den Applikationen CYCLP und ARtemis) unterstützt Lehrende und Studierende dabei, digitalisierte Objekte in eine AR-Umgebung niedrigschwellig zu integrieren, um exploratives Lernen anzuregen und interaktive sowie immersive Lernerfahrungen zu ermöglichen. CYCLP als Photogrammetrie-Software ermöglicht auf einfache Weise durch Fotoaufnahmen Objekte zu digitalisieren, als 3D-Objekt in die ARtemis-App (kurz für: Augmented Reality Toolkit for Educational Methodology and Interactive Simulation) zu importieren und darzustellen. In Verbindung mit einem Weiterbildungsangebot rund um Augmented Reality und das Toolkit bilden sie für Einsteiger:innen einen ersten Schritt zum Einsatz von Augmented Reality in der universitären Lehre. Die genannten Komponenten stellen zusammen ein Ökosystem dar, welches um den Reality Composer von Apple aufgebaut wurde.

Keywords: Augmented Reality, Digitale Lernumgebung, E-Learning, Hochschuldidaktik, Interaktion, Objekte digitalisieren, Photogrammetrie, Weiterbildung

1 Ausgangsüberlegungen

Das Toolkit für Augmented Reality (AR) der Technischen Universität Darmstadt wurde durch die E-Learning Arbeitsgruppe vor dem Hintergrund mediendidaktischer und –technischer Konzeptionierung und Umsetzung von Lehr- und Lernszenarien mittels AR-Technologie entwickelt. Dieses Toolkit dient als niedrigschwelliger Einstieg für den Einsatz von AR in der Lehre.

Die Integration von AR in der Universität stellt eine Möglichkeit dar, die Lehre interaktiver zu gestalten. Trotz ihres Potenzials zur Steigerung der Lerneffizienz und zur Motivation [Wy22] stößt jedoch die Implementierung von AR-Technologien auf Herausforderungen, wie etwa zum einen die technische Komplexität von AR-Systemen, 3D-Objektgenerierung, Game-Engines und Programmiersprachen. Zum anderen müssen Lehrende nicht nur in der Bedienung dieser Technologie geschult werden, sondern auch in der pädagogischen Integration in die Lehre. Weiterhin können Bedenken hinsichtlich der Datenschutz- und Sicherheitsaspekte bei der Nutzung von AR entstehen [Wy22]. Zusammengefasst lässt sich daraus schließen, dass die Generierung von AR-Lerncontent spezifische Anforderungen, wie Zeit, Ressourcen und Kompetenzen erfordert.

1 Technische Universität Darmstadt, Hochschuldidaktische Arbeitsstelle (HDA) , Hochschulstraße 1, 64289 Darmstadt, Germany, marcella.haller@tu-darmstadt.de, https://orcid.org/0000-0002-1712-390X; alexander.ast@tu-darmstadt.de, https://orcid.org/0009-0005-6293-564X

doi:10.18420/delfi2024_52

2 Konzeption und technische Umsetzung

Vor dem Hintergrund der Ausgangsüberlegungen wurde sich bei der Konzeption des Toolkits an den mediendidaktischen Überlegungen nach Kerres [Ke18] orientiert. So adressiert in Anlehnung an Letzteren das Toolkit primär Lehrende für den Einsatz in der akademischen Lehre, aber auch Studierende und alle anderen Personen, die sich potenziell in pädagogischen Kontexten vor den Ansätzen explorativen oder kollaborativen Lernens mit der AR-Technologie auseinandersetzen, um ein Lernen mit und/oder über Medien [Ru14] anzuregen (bspw. in der Schule). Des Weiteren wurde die Endbenutzer-Entwicklung (EUD) [Li06] bei der technischen Gestaltung im Hinblick auf die bereits genannten Herausforderungen herangezogen. Nach umfassender technischer Recherche stellt der Reality Composer[2] von Apple ein gutes Beispiel für die Anwendung der EUD-Prinzipien dar. Das Tool ermöglicht AR-Erlebnisse auf intuitive Weise ohne notwendige Programmierkenntnisse zu erstellen, kreativ umzusetzen und zu bearbeiten, wodurch der Einstieg zur Erstellung von AR-Inhalten deutlich vereinfacht werden kann. Daraus ergab sich die Entwicklung eines AR-Ökosystems. Dieses ist um den Reality Composer herum aufgebaut und wurde durch die proprietäre Photogrammetrie-Software CYCLP sowie die proprietäre ARKit-App ARtemis[3] ergänzt. Lehrende können ihre Projekte einfach mit Hilfe von iPads und Macs erstellen und mit den Studierenden teilen. Der Workflow folgt einem dreistufigen Prozess: (1) Generieren, (2) Komponieren und (3) Implementieren.

1. **Generieren:** Der Prozess beginnt mit der Generierung von 3D-Modellen realer Objekte (z. B. historische Artefakte, Präparate, Maschinen), die als Grundlage dienen. Die Photogrammetrie-Software CYCLP ermöglicht, aus einer Serie von Fotos ein digitales 3D-Modell eines physischen Objekts zu erstellen. Lehr-Lernmaterialien aus der realen Welt können digital so erfasst und in ein Format konvertiert werden, welches für AR-Anwendungen geeignet ist.
2. **Komponieren:** Im nächsten Schritt werden die generierten 3D-Modelle in den Reality Composer importiert. Hier können diese in interaktive AR-Szenen integriert werden, mit Animationen versehen, mit digitalem Content versetzt oder mit Touch-Gesten interaktiv erlebbar gemacht werden. Bspw. kann das 3D-Modell eines Elektromotors per Click zu einer Explosionsdarstellung verändert werden.
3. **Implementieren:** Der letzte Schritt beinhaltet die Nutzung der ARtemis-App, um die erstellten AR-Szenen zu erleben. Durch die Verwendung können Studierende mit den AR-Szenen interagieren, indem sie über die Verwendung eines mobilen Endgerätes diese in ihrer physischen Umgebung platzieren und aus verschiedenen Winkeln betrachten. Studierende könnten so bspw. die AR-Szene des menschlichen Herzens in ihrem eigenen Lernumfeld platzieren, um die Struktur und Funktion des Herzens aus nächster Nähe zu untersuchen und so ein besseres Verständnis der Herzfunktion zu vermitteln.

2 https://developer.apple.com/documentation/realitykit/creating-3d-content-with-reality-composer, letzter Zugriff am 28.03.2024

3 ARtemis im App Store: https://apps.apple.com/de/app/artemis-tuda/id6471366440

Hieraus ergibt sich insgesamt folgender Überblick für die Arbeit mit dem Ökosystem Abb. 1:



Abb. 1: Skizzenhafte Darstellung des Workflows im AR-Ökosystem

Die technische Umsetzung von CYCLP und ARtemis wurde mittels Xcode[4] und SwiftUI[5] ermöglicht. Xcode erlaubt eine native Entwicklung, die speziell für die Zielplattformen MacOS und iOS optimiert ist, was zu einer höheren Leistung und einer besseren Integration mit dem Betriebssystem führen kann. SwiftUI hingegen wurde als modernes UI-Framework genutzt und ermöglicht die Erstellung von benutzerfreundlichen Oberflächen. Es zielt darauf ab, das UI-Design durch effizienten Code zu vereinfachen und verbessert dabei auch die Wartbarkeit und Erweiterbarkeit der App-Entwicklung. CYCLP als Photogrammetrie-Software wurde für macOS mittels SwiftUI und dem RealityKit realisiert und erforderte eine Kombination aus Swift-Programmierung, Kenntnissen in der Verarbeitung und Optimierung von 3D-Daten sowie einem grundlegenden Verständnis von Computer Vision. Apple's Object Capture API ist ein Bestandteil des RealityKit-Frameworks und erlaubt die Generierung von 3D-Modellen aus einer Sequenz von 2D-Bildern. Es wurden hierfür Verknüpfungen der Funktionen, wie Bild-Upload sowie die Anzeige des Verarbeitungsfortschrittes, mit der Benutzeroberfläche umgesetzt. Des Weiteren wurde eine Funktion zum Exportieren der erstellten 3D-Modelle in gängige Formate wie USDZ implementiert, um eine AR-fähige Kompatibilität zu gewährleisten. Hierfür wurden Optimierungstechniken zur Verbesserung der Qualität und Reduzierung der Verarbeitungszeit der 3D-Modelle für Benutzer:innen als Hintergrundprozess angelegt. Die Herausforderung an ARtemis bestand zudem darin, interaktive 3D-Objekte in einer Weise zu implementieren, die sowohl leistungsfähig als auch benutzerfreundlich sind, insbesondere unter Berücksichtigung von EUD. Um EUD-Prinzipien zu integrieren, musste die ARtemis-App Benutzer:innen die Möglichkeit bieten, ohne umfangreiche technische Kenntnisse mit den AR-Inhalten zu interagieren, diese zu importieren, anzupassen und darzustellen. Die Entwicklung einer intuitiven Benutzeroberfläche, die die Dateiauswahl, Szeneanzeige und Touch-Gesten-Interaktion unterstützt, wurde unter Verwendung von SwiftUI in Xcode umgesetzt. Durch das Iterieren der .usdz- und .reality Dateien wurde die Dateiauswahl als Funktion auf zwei Buttons reduziert. Somit wird Benutzer:innen ermöglicht, Dateien zur Laufzeit in die Szene zu integrieren. Um Probleme zu vermeiden, die beim Dateiimport hervorgerufen werden, wurde das sichere Teilen von Dateien von Anfang an beim Programmieren berücksichtigt. Es wurden Funktionen zum Teilen von Dateien via Kabel, Server oder WiFi implementiert. Darüber hinaus bietet die App Optionen zur lokalen Speicherung auf dem Benutzer:innen-iPad (z. B. iTunes Shared Folder, Files App), um die Kontrolle über die eigenen Daten zu stärken.

4 https://developer.apple.com/xcode/, letzter Zugriff am 28.03.2024
5 https://developer.apple.com/xcode/swiftui/, letzter Zugriff am 28.03.2024

3 Aktueller Stand und Zukunftsperspektiven

Für den Workflow wurde ein Workshopangebot als Unterstützung und Weiterbildung der Lehrpersonen im Umfang von zwei Stunden konzipiert, mit den Zielen der Medienkompetenzförderung sowie der Vermittlung didaktischer und technischer Grundlagen im praktischen Einsatz und Umgang von AR und dem Toolkit. Der Workshop ist in drei Phasen gegliedert, bestehend aus einer Theoriephase, einer Praxisphase mit kooperativen Übungen und einer Reflexionsphase. Aktuell ist die Anpassung der ARtemis-App an weitere Zielgruppen, insbesondere ältere Personen und Menschen mit Behinderungen, ein zentraler Aspekt. Die Anpassung umfasst die Integration von Design- und Funktionsmerkmalen (z. B. die Einbindung von Game Controller sowie Access-Controller), die auf die spezifischen Bedürfnisse und Präferenzen dieser Gruppen abgestimmt sind, um Inklusion in der Lehre oder auch in unterschiedlichen Schulformen zu fördern. Die Zukunft liegt in der kontinuierlichen Optimierung und Weiterentwicklung der Nutzererfahrung. Diese wird maßgeblich von der Evaluation des Workshops sowie den generierten Anwendungsfällen (Use Cases) von Lehrenden unterschiedlicher Fachbereiche bestimmt. Die Relevanz dieser Anwendungsfälle bietet im Entwicklungsprozess eine solide Grundlage für eine benutzerzentrierte Gestaltung, die sich unmittelbar auf die Akzeptanz und den Erfolg des Toolkits einerseits und damit andererseits auf den Lernerfolg auswirkt. Ein intensiver Austausch mit den Lehrenden ist hier von wesentlicher Bedeutung. Zuletzt wird durch CYCLP im Aufbau einer umfangreichen 3D-Bibliothek und deren Zugänglichmachung von 3D-Objekten im universitären Bereich perspektivisch ein weiterer Schritt gesehen. Die Bibliothek bietet nicht nur wertvolle Ressourcen für die akademische Lehre und Forschung, sondern soll die Kollaboration über Fachbereiche hinweg fördern.

Literaturverzeichnis

[Ke18] Kerres, Michael: Mediendidaktik: Konzeption und Entwicklung digitaler Lernangebote 5. Auflage. DE GRUYTER, 2018.

[Li06] Lieberman, Henry; Paternò, Fabio; Klann, Markus; Wulf, Volker: End-user development: An emerging paradigm. S. 1–8. Springer, 2006.

[Ru14] Ruge, Wolfgang B: Pädagogik* Medien–Eine Suchanfrage. Perspektiven der Medienbildung, S. 187–207, 2014.

[Wy22] Wyss, Corinne; Furrer, Florian; Degonda, Adrian; Bührer, Wolfgang: Augmented Reality in der Hochschullehre: Überlegungen zu einer zukunftsweisenden Ausbildung von Lehrpersonen. MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung, 47(09):118–137, 2022.

BuddyAnalytics: A dashboard and reporting tool for study program analysis and student cohort monitoring

Sergej Görzen [1], René Röpke [2], and Ulrik Schroeder [1]

Abstract: With students leaving digital traces in Campus Management Systems when registering and completing courses and exams, there is a growing interest in data-driven study program analysis and student cohort monitoring. To support study program designers with tasks, such as planning courses and exams, creating (re-)accreditation reports, and improving curricula and study plans, insights into the students' behavior throughout their studies present valuable input and may allow for evidence-based curriculum development. This demo presents BuddyAnalytics, a web-based tool providing dashboards and analysis reports for study program analysis and cohort monitoring. It enables study program designers to review various metrics and indicators relevant to understanding students' behavior and potential issues of study programs. Developed using user-centered design methodology, the tool is closely tailored to users' needs and requirements. Future evaluation with the target group will assess its suitability and provide valuable feedback for improving the tool over time.

Keywords: Cohort Monitoring, Study Program Analysis, Curriculum Analytics, Dashboard

1 Introduction

With the advance of Campus Management Systems (CMS) as large-scale administrative software solutions to support student life cycles in higher education institutions (HEIs), many study-related processes of an organizational nature are captured through digital traces of students (e.g., events for (de-)registering to courses and exams, receiving exam results and updates on students' (re-)enrolment status). Researchers, administrative staff, and instructors understandably show a growing interest in analyzing student lifecycle data. Data-driven insights gained by applying methods of *educational data mining* and *curriculum analytics* (CA) yield potential in evidence-based curriculum development and student cohort monitoring. Different stakeholders of HEIs may use analytics results to identify bottlenecks and issues in curricula, as well as thriving and beneficial study paths suitable for academic advisory services, in supporting students and providing guidance and advice regarding individual study planning. Further use cases include creating analytics reports for (re-)accreditation processes and quality assurance matters.

Insights into the students' behavior throughout their studies may provide valuable input to support the different stakeholders in their tasks. However, a diverse set of requirements must

1 RWTH Aachen University, Learning Technologies, Aachen, Germany,
goerzen@informatik.rwth-aachen.de, https://orcid.org/0000-0003-3853-2435;
schroeder@informatik.rwth-aachen.de, https://orcid.org/0000-0002-5178-8497
2 TU Wien, Learning Technologies and eDidactics, Vienna, Austria,
rene.roepke@tuwien.ac.at, https://orcid.org/0000-0003-0250-8521

doi:10.18420/delfi2024_53

be considered due to the different stakeholders' tasks and responsibilities. As such, we use the term *study program designer* to summarize the heterogeneous set of stakeholders with their respective tasks while still allowing us to describe distinct use cases and application scenarios.

In this demo paper, we introduce BuddyAnalytics, a web-based tool providing dashboards and analysis reports for study program analysis and cohort monitoring. The tool enables study program designers to review various metrics and indicators relevant to understanding students' behavior and potential issues of study programs. It is being developed in the context of the AIStudyBuddy project. Within the project, the two paradigms of rule-based, symbolic artificial intelligence (AI) and process mining (PM) are combined to support individual study planning and student cohort monitoring.

2 Related Work

The analysis and visualization of curriculum and student cohort data are domains of curriculum analytics (CA), a sub-domain of learning analytics. CA utilizes artificial intelligence and data science methods to support and guide evidence-based curriculum development [DH14; Hi20; Hi22]. It can be beneficial in identifying students' needs, monitoring student cohorts, and reducing dropout rates [Da15], and thus is particular interest for study program designers. Various research projects have focused on CA and the analysis of student lifecycle data (e.g., [BY15; Gu20; PMY17; Ta22]). Also, methods of educational process mining allow the analysis events in student lifecycle data to identify high-failure rate courses leading to late dropouts [Sa21] or to answer students' questions regarding successful pathways to their degree [Sc17]. Further work is summarized in [BCR18].

Suitable applications to make CA results accessible to study program desigerns are not yet widely available. With our work on BuddyAnalytics, we address this gap with a modular, extensible web-based tool providing dashboards and analysis reports for study program analysis and cohort monitoring.

3 Technological Background and Concept

We first identified stakeholders' requirements and needs during a workshop as part of an iterative user-centered design methodology. These requirements were organized into user stories, resulting in smaller, feasible tasks and features. This paper presents some of the main requirements for BuddyAnalytics, which mainly influence our concept and implementation.

The tool's foundation is being developed using Angular for the frontend and NodeJS for the backend. The backend communicates with an instance of ExcaliburLA [JS22], an extensible learning analytics infrastructure supporting data analysis from CMS and other systems in HEI. The project utilizes data from CMS of three German university. A general data model

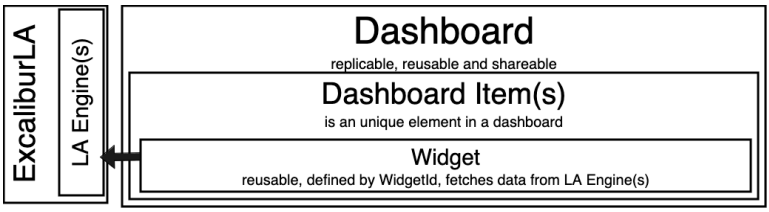


Fig. 1: A concept of a replicable dashboard of holding items that manage their own widgets and data.

for CMS data is in development, so far a first approach of a preliminary general data model was presented in [Wa23]. By providing decoupled, batch-based analysis and storing results in a result store, applications like BuddyAnalytics can request pre-computed data efficiently without considering differences between the origin CMSs.

Roles and Permissions. Study program designers are a heterogeneous group with varied roles and functions in HEI. Therefore, a scalable yet flexible approach for permissions and roles is necessary to ensure fine-granular, secure data access and compliance with data protection regulations. While it might be possible to match the permissions and roles to existing concepts in a single university, a tool designed for multiple institutions requires a more adaptable and customizable system. We employ the classical concept of *user accounts* to which multiple roles can be assigned for an indefinite or a limited period and scope (e.g., for a specific study program for x months). Each role compromises a list of *permissions* to restrict access to specific resources. For example, the permission *canCreateUser* allows the creation of users. Access is controlled by a flat set of permissions, where the number of permissions defines possible hierarchies. This way, different access levels and hierarchies are enabled and ready for use in a university context.

Dashboard Items and Widgets. Customizability is a key requirement for supporting users in personalizing their dashboards and adapting them to specific tasks. This need was confirmed in a first-user study, where study program designers expressed strong interest in filter options and configurations. Consequently, the goal was to provide a tool supporting custom dashboards with flexible content elements, extensive filter options, and export functionalities. The dashboard is designed with a component-based flexible grid layout (see Figure 1). We distinguish between different dashboard types: *template* for fixed, pre-configured dashboards on study programs and modules and *custom* for user-created dashboards with individualized content and configurations). The grid components called *Dashboard Items*, are unique elements that refer to widgets by ID, containing scoped filters (depending on widgets) and grid options (size and position). Each widget is loaded lazily upon loading a dashboard and is responsible for fetching data from the connected LA infrastructure Excalibur LA, which provides pre-computed analysis results based on the collected user stories of study program designers. The data is received as "ready" depending on the provided widget ID and type, and the associated visualization is chosen dynamically. Through sandboxing, widgets and dashboard items can be copied and reused for across

different dashboards while maintaining a consistent design and individual filters. Figure 2 presents an exemplary study program dashboard of BuddyAnalytics in its current version. Some charts are conceptual (e.g., the Sankey diagram), and real data integration is ongoing. The final visualizations are aimed to help users to gain in-depthinsights. However, the dashboard won't generate action recommendations due to ethics and data quality concerns.

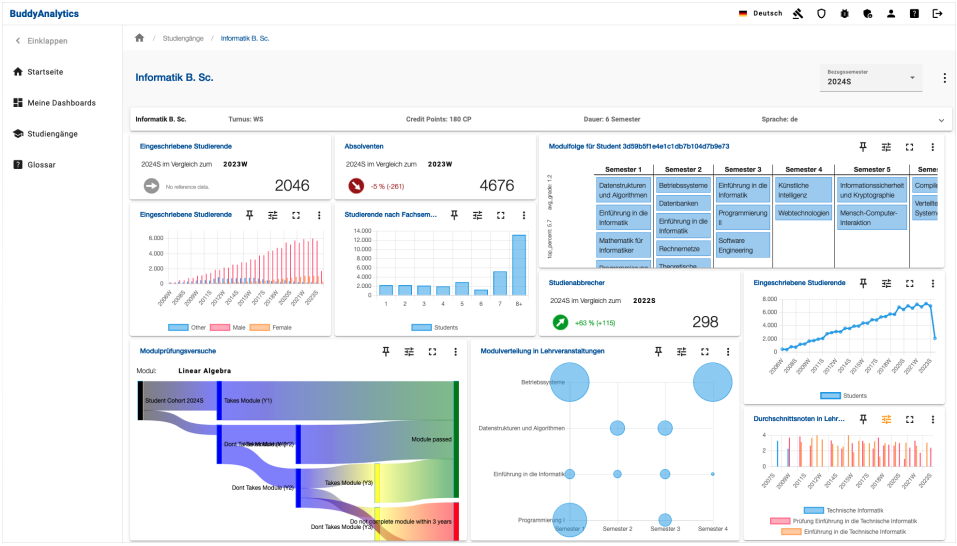


Fig. 2: Screenshot of BuddyAnalytics showing parts of a study program dashboard

4 Conclusion and Outlook

With BuddyAnalytics, we present a modular, extensible tool that supports the creation of informative template dashboards at study program and module levels and custom dashboards for tailored support of study program designers. We developed a prototype that meets current and upcoming requirements, adhering to the regulations and needs of the involved institutions and participating study program designers.

Alongside the presented key requirements for the design of BuddyAnalytics, more features and functionalities will be integrated over time. While development has primarily utilized synthetic test data, the prototype should be evaluated using real-world data to capture valuable feedback from study program designers. The extensible LA infrastructure allows data analysis results with custom filter options. We aim to incorporate additional visualizations, such as process models providing detailed insights into study path data. This will require the integration of specific process mining approaches for data analysis and custom visualizations. Finally, study program designers must assess the resulting dashboards in terms of usability and efficiency for supporting different tasks (e.g., creating accreditation reports).

References

[BCR18] Bogarín, A.; Cerezo, R.; Romero, C.: A Survey on Educational Process Mining. WIREs Data Mining and Knowledge Discovery 8 (1), e1230, 2018, DOI: 10.1002/widm.1230.

[BY15] Bendatu, L. Y.; Yahya, B. N.: Sequence Matching Analysis for Curriculum Development. Jurnal Teknik Industri 17 (1), 2015, DOI: 10.9744/jti.17.1.47-52.

[Da15] Daniel, B.: Big Data and Analytics in Higher Education: Opportunities and Challenges. British Journal of Educational Technology 46 (5), pp. 904–920, 2015, DOI: 10.1111/bjet.12230.

[DH14] Dawson, S.; Hubball, H.: Curriculum Analytics: Application of Social Network Analysis for Improving Strategic Curriculum Decision-Making in a Research-Intensive University. Teaching & Learning Inquiry: ISSOTL 2 (2), pp. 59–74, 2014, DOI: 10.2979/teachlearninqu.2.2.59.

[Gu20] Gutiérrez, F.; Seipp, K.; Ochoa, X.; Chiluiza, K.; De Laet, T.; Verbert, K.: LADA: A Learning Analytics Dashboard for Academic Advising. Computers in Human Behavior 107 (105826), 2020, DOI: 10.1016/j.chb.2018.12.004.

[Hi20] Hilliger, I.; Aguirre, C.; Miranda, C.; Celis, S.; Pérez-Sanagustín, M.: Design of a Curriculum Analytics Tool to Support Continuous Improvement Processes in Higher Education. In: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge. LAK '20, ACM, New York, pp. 181–186, 2020, DOI: 10.1145/3375462.3375489.

[Hi22] Hilliger, I.; Aguirre, C.; Miranda, C.; Celis, S.; Pérez-Sanagustín, M.: Lessons Learned from Designing a Curriculum Analytics Tool for Improving Student Learning and Program Quality. Computing in Higher Education, 2022, DOI: 10.1007/s12528-022-09315-4.

[JS22] Judel, S.; Schroeder, U.: EXCALIBUR LA - An Extendable and Scalable Infrastructure Build for Learning Analytics. In: 2022 International Conference on Advanced Learning Technologies (ICALT). IEEE, Bucharest, Romania, pp. 155–157, 2022, DOI: 10.1109/icalt55010.2022.00053.

[PMY17] Priyambada, S. A.; Mahendrawathi, E.; Yahya, B. N.: Curriculum Assessment of Higher Educational Institution Using Aggregate Profile Clustering. Procedia Computer Science 124, 2017, DOI: 10.1016/j.procs.2017.12.155.

[Sa21] Salazar-Fernandez, J. P.; Sepúlveda, M.; Munoz-Gama, J.; Nussbaum, M.: Curricular Analytics to Characterize Educational Trajectories in High-Failure Rate Courses That Lead to Late Dropout. Applied Sciences 11 (4), p. 1436, 2021, DOI: 10.3390/app11041436.

[Sc17] Schulte, J.; Fernandez de Mendonca, P.; Martinez-Maldonado, R.; Buckingham Shum, S.: Large Scale Predictive Process Mining and Analytics of University Degree Course Data. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference. LAK '17, ACM, New York, pp. 538–539, 2017, DOI: 10.1145/3027385.3029446.

[Ta22] Tatel, C. E.; Lyndgaard, S. F.; Kanfer, R.; Melkers, J. E.: Learning While Working: Course Enrollment Behaviour as a Macro-Level Indicator of Learning Management Among Adult Learners. Journal of Learning Analytics 9 (3), pp. 104–124, 2022, DOI: 10.18608/jla.2022.7625.

[Wa23] Wagner, M.; Helal, H.; Röpke, R.; Judel, S.; Doveren, J.; Görzen, S.; Soudmand, P.; Lakemeyer, G.; Schroeder, U.; Aalst, W.: A Combined Approach of Process Mining and Rule-Based AI for Study Planning and Monitoring in Higher Education. In. Pp. 513–525, 2023, ISBN: 978-3-031-27814-3, DOI: 10.1007/978-3-031-27815-0_37.

19squared - Interaktive XR-Inhalte gemeinsam selbst erstellen

Matthias Ehlenz[1] und Frederic Maquet[1]

Abstract: Die hier vorgestellte Demo präsentiert ein webbasiertes Tool, welches Lehrkräften und Schüler:innen die Erstellung eigener immersiver und interaktiver Lerninhalte für den Unterricht ermöglicht. Als zentrales Medium dienen 360°-Aufnahmen (Fotos und Videos), welche mittlerweile mit relativ geringem Aufwand mit Actioncams, Smartphone-Gimbals oder Drohnen selbst erstellt werden können. Das Open Source-Werkzeug ermöglicht die Verarbeitung der Aufnahmen zu interaktiven Touren, welche im Browser, auf dem Tablet oder in Virtual Reality erkundet werden können. Zudem bietet die Software die Möglichkeit der kollaborativen Bearbeitung, die Erstellung von virtuellen Escape Rooms sowie die Beobachtung des Lernfortschritts durch Learning Analytics-Integration. Im Rahmen der Demonstration können sowohl das Erstellen interaktiver Touren mit vorbereitetem 360°-Material als auch das Durchführen von Exkursionen, die bereits in der Lehre eingesetzt werden, erprobt werden. Zudem kann auch die eigenständige Erstellung von 360°-Medien auf verschiedene Arten getestet werden, sofern die Rahmenbedingungen es zulassen.

Keywords: 360-Grad-Medien, Open Educational Resources, Autorentools, Immersive Learning

1 Einleitung

Schulen werden zunehmend digitalisiert. Dies wird durch den Digitalpakt sowie weitere Förderungen von Bund, Ländern, Kommunen und Schulträgern vorangetrieben. In der jüngeren Vergangenheit wurde so viel Hardware in die Klassenzimmer eingeführt, auch solche, die über die klassischen Interaktionsformen mit Maus (oder Touchpad) und Tastatur hinausgehen. Die Verfügbarkeit von Tablets in Klassenstärke ist inzwischen keine Seltenheit mehr und auch VR-taugliche Hardware findet sich in ersten Schulen oder steht in den jeweils zuständigen Medienzentren zur Ausleihe bereit. Allerdings fehlt es an zugänglichen Möglichkeiten, um Inhalte für diese Medien bereitzustellen. Im Bereich der erweiterten und virtuellen Realität stellt sich dabei stets die Herausforderung, dass die Erstellung von Inhalten in der Regel hohe technische Hürden entweder im Bereich der Modellentwicklung oder der Generierung von Aufnahmen mit sich bringt. Es ist jedoch erfreulich, dass sich die Technik nicht nur im Bereich der Wiedergabemedien, sondern auch bei den Aufnahmegeräten spürbar weiterentwickelt. Inzwischen sind Actioncams mit 360°-Funktion für einen breiten Nutzerkreis erschwinglich geworden, Hobbydrohnen mit integrierten Kameras verfügen über entsprechende Funktionen und auch (Smartphone-)Gimbals sind entsprechend ausgestattet. Dadurch können technische interessierte Lehrkräfte selbst solche Kugelpanoramen, auch mit Bewegtbild, anfertigen. Im Folgenden wird ein Autorentool präsentiert, mit dessen Hilfe didaktisch wertvolle, interaktive Lernmedien erstellt werden können.

1 RWTH Aachen University, MediaLab Lehramt, Kármánstraße 17-19, 52056 Aachen, Deutschland, ehlenz@lbz.rwth-aachen.de, https://orcid.org/0000-0001-6189-6056; maquet@rwth-aachen.de

doi:10.18420/delfi2024_54

2 Verwandte Arbeiten

Die Idee Autorenwerkzeuge für (Lern-)Umgebungen in virtueller Realität zur Verfügung zu stellen ist nicht neu. Es gab bereits einige interessante Entwicklungen und Ansätze. Insbesondere im kommerziellen Bereich gibt es zahlreiche Produkte und Dienstleister, die für Bildungseinrichtungen jedoch häufig unerschwinglich sind[2]. Der Fokus in der Recherche lag daher auch auf kostenlos einsetzbaren Lösungen. Im Rahmen der Recherche wurden zum einen freie Plattformen großer Anbieter betrachtet, zum anderen freie Werkzeuge:
Auf der einen Seite gibt es immer wieder Projekte wie Google VR (eingestellt 2019) [Go24] und Mozilla Hubs (eingestellt 2024)[Mo24], die grundsätzlich auch die Erstellung von immersivem Content ohne Programmierkenntnisse ermöglichen. Allerdings sind diese zum einen aus Datenschutz-Perspektive fragwürdig und zum anderen durch die enge Kopplung an einen Anbieter wenig nachhaltig, wie das Scheitern der genannten Produkte unterstreicht. In der Open Source-Welt dagegen sind diese Probleme zwar nicht gegeben, dafür zeigen sich hier andere Herausforderungen. Es stehen mit Blender [Bl24] und Co. (in der Regel in Kombination mit Unity o. Ä.) sehr mächtige Werkzeuge zur Verfügung, mit denen ganze immersive Welten gebaut werden können. Allerdings setzen diese solide Vorkenntnisse voraus und weisen eine recht steile Lernkurve auf. Ein Beispiel für ein zugänglicheres Tool ist Marzipano [Ma24], mit dem sich interaktive Touren mit 360°-Medien erstellen lassen. Die Bedienung ist sehr nutzerfreundlich, die erstellten Inhalte sind auf vielen Medien wie VR-Headsets (auch der Cardboard-Variante), Tablets und am PC nutzbar und somit für viele Einsatzszenarien zu gebrauchen. Allerdings weist Marzipano einen entscheidenden Nachteil auf, der sich insbesondere bei der nachhaltigen Erstellung von Unterrichtsmedien bemerkbar macht. So müssen die erstellten Touren nach dem Export auf einen eigenen Server geladen werden. Zudem ist ein Wiederimport der exportierten Ergebnisse nicht vorgesehen, was die Nutzung als Open Educational Resource (OER) deutlich einschränkt.

3 Open Source für den Schuleinsatz

Das hier vorgestellte Open Source-Tool zielt darauf ab, diese Lücke zu schließen und darüber hinauszugehen. 19squared ist frei verfügbar[3] und wird aktiv weiterentwickelt. Im Folgenden wird ein Überblick über den Technologiestack, die aktuellen Features und die gegenwärtigen Weiterentwicklungen gegeben. Erste Fallstudien in fachübergreifenden Kontexten wurden bereits durchgeführt, werden an dieser Stelle jedoch nicht im Detail beschrieben.

3.1 Technische Grundlagen

Nachhaltigkeit und Offenheit sind die priorisierten Kriterien für die gewählten Technologien. Darüber hinaus soll eine möglichst breite Interoperabilität erreicht werden. Aus diesen

2 Beispiel: [Vi24], bei mehr als 8 Nutzern kostenpflichtig, >50$/Monat (abgerufen 8.4.24)
3 `https://git.rwth-aachen.de/medialab/interactive360vr`

Gründen wurde sowohl für die Produktebene der interaktiven Touren als auch für das Autorenwerkzeug selbst (browserbasierte) Webtechnologien gewählt und, wo vorhanden und sinnvoll, auf Open Source-Libraries zurückgegriffen. Das Kernstück der Anwendung stellt die Darstellung der 360°-Ansicht dar, welche plattformübergreifend mittels WebVR genutzt werden kann. Als Fallback-Lösung ist die Anwendung auf PCs mit Maus oder Touch-Display bedienbar, wobei die Immersion auf anderen Geräten spürbar steigt. Auf mobilen Endgeräten werden die Lagesensoren ausgelesen, sodass ein aktives Umschauen durch entsprechende Bewegung möglich ist. Die aktive Kopfbewegung im VR-Headset erlaubt ein noch immersiveres Erlebnis. Die Verwendung von A-Frame als Darstellungsframework für WebVR ermöglicht die Integration der interaktiven Hotspots, zudem ergeben sich Weiterentwicklungsmöglichkeiten, beispielsweise durch die Einbettung von 3D-Objekten. Das Interface zur Erstellung und Bearbeitung von interaktiven Exkursionen ist bewusst schlank gehalten. Bislang wird auf die Verwendung komplexer Webframeworks verzichtet. Lediglich jQuery als JavaScript-Convenience-Lösung und Bootstrap als CSS-Framework sind eingebunden. Auch serverseitig basiert das Tool auf offenen Technologien. Hier kommt Python mit dem Framework Flask zum Einsatz, die Datenbank wird über SQLAlchemy angebunden und kann in kleineren Deployments mit SQLite, in größeren Produktivsetups aber auch mit MariaDB oder vergleichbaren Systemen betrieben werden. Zur Erleichterung der Inbetriebnahme werden Docker-Files bereitgestellt, welche die Verwendung des Tools in einer Containerumgebung ermöglichen.

3.2 Funktionsumfang

19squared ist als Produkt, Werkzeug und Gegenstand von Lehre und Forschung eine fortlaufende Entwicklung. Dennoch befindet es sich bereits jetzt im Produktiveinsatz und wurde schon im Rahmen zweier Projekte genutzt. Zum Zeitpunkt der Einreichung befassen sich drei weitere studentische Arbeiten mit der Weiterentwicklung. Der hier beschriebene Stand beschreibt den Status der Version 2.0.6, wobei davon auszugehen ist, dass auch die im folgenden Abschnitt beschriebenen Weiterentwicklungen zum Konferenzzeitpunkt bereits getestet werden können.
Beim Aufruf der Webapplikation können Nutzer:innen sich einloggen oder einen Nutzeraccount erstellen, in dem sie dann in der Folge neue Touren erstellen können. Abbildung 1 zeigt einen Screenshot der Szenenbearbeitungsansicht des Autorenwerkzeugs. Auf der linken Seite sind die hochgeladenen Medien zu finden, in der Mitte befindet sich die 360°-Ansicht und auf der rechten Seite eine Übersicht der bereits erstellten Szenen und der zugeordneten Hotspots. Diese können frei in der Szene platziert werden und zeigen zusätzlichen Text, Bilder, Videos oder geben Audio wieder. Des Weiteren besteht die Möglichkeit, die Szenen über entsprechende Hotspots miteinander zu verknüpfen, um so ganze Rundgänge zu erstellen. Ist die Erstellung einer Tour abgeschlossen, kann ein Link generiert werden, mit dem diese ohne Benutzeraccount frei abrufbar und erlebbar ist. Zudem können erstellte Touren exportiert und durch andere Nutzer:innen oder in andere Instanzen importiert und weiterbearbeitet und somit als OER verbreitet werden.

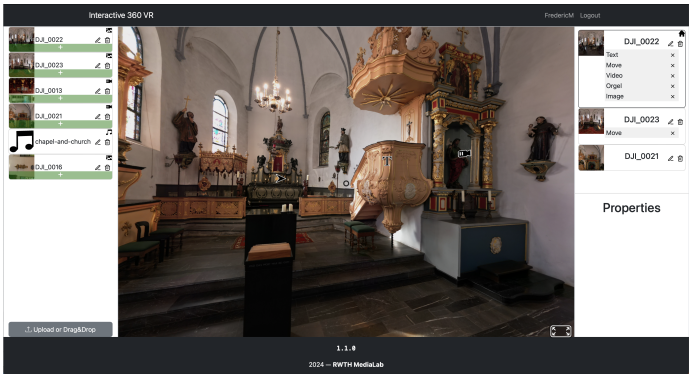


Abb. 1: Der Editor des Autorenwerkzeugs

3.3 Weiterentwicklungen

Im Rahmen der aktuellen Weiterentwicklungen erfolgt nicht nur eine Überarbeitung des User-Interface, sondern auch die Integration neuer Features, welche die didaktischen Möglichkeiten zum Schuleinsatz deutlich erweitern sollen. Hierzu gehört die Integration neuer Hotspots, die ganz ähnlich zu z. B. den interaktiven Videos in H5P noch mehr Interaktion über Quizzes und ähnliche Elemente ermöglichen sollen, sowie die Erweiterung auf bedingte Hotspots. In Kombination ermöglichen diese die Erstellung eigener Escape-Rooms. Des Weiteren wird die kollaborative Produktion virtueller Exkursionen entwickelt. Hierbei können Lehrkräfte beispielsweise die Medien bereitstellen und Schülerinnen und Schüler, auch gleichzeitig an verteilten Endgeräten, gemeinschaftlich Szenen und Rundgänge bearbeiten. Eine Learning Analytics-Schnittstelle zur Datenerhebung in xAPI, sowohl für die Erstellung als auch die Erkundung der virtuellen Rundgänge, wird parallel dazu integriert.

4 Offene Herausforderungen

Für die langfristige Nutzung sind noch Herausforderungen offen, die insbesondere die Skalierbarkeit betreffen. Konzepte hierzu sind in der Entwicklung und umfassen die Nutzung von Streaming-Servern oder verteilten Datenbanken zur Medienbereitstellung. Zunächst wird hier mit Quotas gearbeitet, bis ein größer skalierter Produktivbetrieb absehbar ist. In den nachfolgenden Schritten muss zudem eine Nutzbarmachung der erfassbaren Interaktionsdaten untersucht werden. Ziel ist es, daraus im Sinne von ächtionable"Learning Analytics nutzbringende Informationen für Lehrkräfte in einem Dashboard zusammenzutragen.
Andere Funktionen hingegen werden den kommerziellen Anbietern in diesem Feld überlassen werden, beispielsweise die Möglichkeit mehrere User gemeinsam die VR-Umgebung erkunden zu lassen, die Erweiterung um 3D-modellierte Umgebungen oder potenzielle 6-DoF-Anwendungen. Des Weiteren ist die Bereitstellung nativer Apps für spezifische Systeme derzeit nicht vorgesehen.

Literaturverzeichnis

[Bl24] Blender Foundation: blender.org - Home of the Blender project - Free and Open 3D Creation Software, en, 2024, URL: https://www.blender.org/, Stand: 12. 04. 2024.

[Go24] Google VR: 360° Media | Google VR, en, 2024, URL: https://developers.google.com/vr/discover/360-degree-media, Stand: 12. 04. 2024.

[Ma24] Marzipano: Marzipano - a 360° viewer for the modern web, 2024, URL: https://www.marzipano.net/, Stand: 12. 04. 2024.

[Mo24] Mozilla Hubs: Hubs - Private, virtual 3D worlds in your browser, 2024, URL: https://hubs.mozilla.com/, Stand: 12. 04. 2024.

[Vi24] Virbela: Frame - Immersive Meetings, Classes, Events, en, 2024, URL: https://learn.framevr.io/, Stand: 12. 04. 2024.

Produktionstechnisches Vorgehen und Demonstration der Erstellung einer Selbstlernakademie für IT-Sicherheit

Tabea Dobbrunz[1], Miriam Naß[2], Sina Warmer[1], Norbert Pohlmann[2] und Andreas Harrer[1]

Abstract: Im Rahmen des vom Ministerium für Kultur und Wissenschaft des Landes NRW und von der Digitalen Hochschule NRW (DH.NRW) geförderten Projekts SecAware.nrw wurde eine Selbstlernakademie zum Thema IT-Sicherheit entwickelt. Mit dem Ziel, Bewusstsein zu schaffen, stellt dieses Lernangebot eine Präventionsmaßnahme für die vermehrt stattfindenden Cyber-Angriffe auf Hochschulen dar. Das Angebot setzt dabei auf ein multimediales didaktisches Konzept, bestehend aus informativ aufbereiteten und zur Exploration anregenden Elementen. In diesem Beitrag wird die produktionstechnische Vorgehensweise bei der Konzeption und Realisierung des Lernangebots beschrieben und das Resultat vorgestellt.

Keywords: Selbstlernkurs, IT-Sicherheit, Multimediales Lernangebot, Produktionsprozess

1 Einleitung

Die steigende Anzahl von Cyberangriffen auf Universitäten und Hochschulen, wie durch die „Jahresberichte 2022 und 2023 des Bundesamtes für Sicherheit in der Informationstechnik (BSI)“ dokumentiert, unterstreicht die Notwendigkeit, die Cyberabwehr dieser Institutionen zu stärken. Von den 23 im Jahr 2023 registrierten Ransomware-Attacken im Bildungssektor fanden 13 auf Hochschulen und Forschungseinrichtungen statt [Bu23]. Diese Entwicklung zeigt, wie die fortschreitende Digitalisierung Sicherheitsrisiken für die akademische Gemeinschaft mit sich bringt. Zwar gibt es kommerzielle Schulungsprogramme zum Thema IT-Sicherheit, wie SoSafe [So22] oder Lucy [Lu22], diese werden jedoch primär für Mitarbeitende eingesetzt und sind nicht spezifisch auf den Hochschulkontext ausgerichtet.

Die Online-Selbstlernakademie SecAware.nrw adressiert gezielt die Schwachstelle „Mensch“ der IT-Sicherheitsarchitektur von Hochschulen. Dabei ist die Entwicklung des modularen und interaktiven Selbstlernangebots spezifisch auf die Bedürfnisse von Studierenden, wissenschaftlichen Mitarbeitenden und Professorenschaft ausgerichtet. In diesem Beitrag steht die iterative Konzeption und Realisierung im Fokus: die Produktion relevanter Inhalte und systematische Integration interner Feedbackschleifen sowie externer Überprüfung spielen eine zentrale Rolle. Damit werden Lerninhalte erstellt, die IT-Kompetenzen und Bewusstsein für Cyber-Sicherheit im akademischen Umfeld stärken.

1 Fachhochschule Dortmund, Institut für Digitalisierung von Arbeits- und Lebenswelten, Otto-Hahn-Straße 23 & 27, 44227 Dortmund, Deutschland, tabea.dobbrunz@fh-dortmund.de; sina.warmer@fh-dortmund.de; andreas.harrer@fh-dortmund.de, https://orcid.org/0009-0001-1076-7964

2 Westfälische Hochschule Gelsenkirchen, Institut für Internet-Sicherheit, Neidenburger Straße 43, 45897 Gelsenkirchen, Deutschland, miriam.nass@w-hs.de; pohlmann@internet-sicherheit.de, https://orcid.org/0009-0007-6221-7327

doi:10.18420/delfi2024_55

2 Produktionsprozess

Die Erstellung der Lernmaterialien im Rahmen der Online Selbstlernakademie SecAware.nrw erfolgt iterativ mittels Feedbackschleifen. Dazu gehört die kontinuierliche Einbindung externer Fachexperten und der Zielgruppen für alle Inhalte. Der Prozess umfasst mehrere Schlüsselschritte: Konzeption, Skripterstellung, Videoproduktion, Vertonung, Entwicklung interaktiver Elemente sowie Bereitstellung.

Konzeption: Die Konzeptionsphase dient der grundlegenden Planung der Inhalte. In dieser Phase werden aktuell relevante Kernthemen (*Module*, [Na24]) im Kontext von Cybersicherheit an Hochschulen identifiziert. In Rücksprache mit *Informationssicherheitsbeauftragten (CISOs)* von Hochschulen und Universitäten aus NRW wird ein Curriculum (fort-)entwickelt, welches die Herausforderungen von Hochschulangehörigen adressiert.

Skripterstellung: Die Erstellung der Skripte erfolgt auf Grundlage umfassender Recherchen und Analysen durch eine journalistisch ausgebildete *Redaktion*. In enger Zusammenarbeit zwischen dem Projektteam und *externen Fachexperten*, werden die Inhalte mehrfach überarbeitet, um fachliche Korrektheit sicherzustellen. Dieses Skript dient als Grundlage für weitere Versionen, die im Laufe des Prozesses spezifisch angepasst werden: Erstens für die Optimierung gemäß den Anforderungen der Videoproduktion, zweitens für die Vertonung durch *professionelle Sprechende* und drittens zur Erstellung eines begleitenden Textes, der als Alternative zum Videoangebot und aus Barrierefreiheitsgründen gelesen werden kann. Zusätzlich werden ergänzende Informationsquellen bereitgestellt, um diese der Zielgruppe in dem Abschnitt „weiterführende Informationen“ zum Nachlesen anzubieten.

Videoproduktion: Für die Videoproduktion werden die Skripte mittels eines KI-basierten Text-zu-Sprache-Tools in ein vorläufiges Voice-Over transformiert. Vektorgrafiken werden in Adobe Illustrator erstellt und anschließend in Adobe After Effects importiert und animiert. Nach mehrfacher Inspektion und Überarbeitung eines Videos erfolgt die Übergabe der finalen Videoversion und der entsprechenden Skriptversion an professionelle Sprechende.

Vertonung: Die Stimmen von professionellen Sprechenden wurden abwechselnd nach weiblichen und männlichen Personen gewählt. Nach Übergabe der Audiospur ersetzt die professionelle Aufnahme das vorläufige, KI-erzeugte Voice-Over, das teilweise eine unnatürliche Aussprache / Betonung aufweist. Vor dem finalen Export werden Untertitel in Englisch und Deutsch hinzugefügt, um die Zugänglichkeit (accessibility) zu verbessern.

Interaktive Elemente: Zur Aktivierung der Lernenden werden interaktive Elemente iterativ entwickelt (siehe [Na24]). In Workshops werden mittels Brainstorming pro Thema Ideen generiert, um mit den Videos inhaltlich abgestimmte interaktive Darstellungen sowie Selbstüberprüfungsaufgaben zu konzipieren, und deren Umsetzbarkeit bewertet.

Daraufhin werden mit Hilfe von *Storyline 360*[3] Prototypen entworfen (siehe Abb. 1). In der Software werden Storypfade aus Präsentationsfolien (Übersicht: linke Toolbar)

3 https://www.articulate.com/360/storyline/

erstellt, welche durch definierte Nutzende- oder Systemevents (rechte Toolbar: Trigger) beeinflusst werden können. Diese ermöglichen außerdem das Anzeigen und Manipulieren bestimmter Folien, Folienebenen oder -objekte (untere Toolbar). Mittels *xAPI* [Iee23] können Informationen an ein Lernmanagementsystem (LMS) übersendet werden.

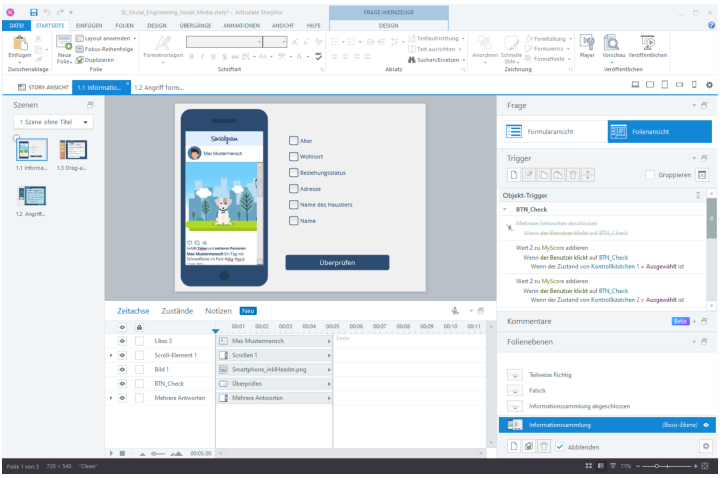


Abb. 1: Ansicht der Gestaltung eines interaktiven Elements in Storyline 360

Die aus Storyline exportierten Prototypen (siehe Abb. 2) werden in das Selbstlernangebot eingebettet und zunächst projektintern und nach Überarbeitung mit professionellen Grafiken projektextern evaluiert.

Bereitstellung: Nach der Produktion gemäß des oben beschriebenen Prozesses und der Zusammenführung der Inhalte im Autorentool *Articulate Rise 360*[4] ist die Bereitstellung der Selbstlernakademie und deren Module die abschließende Aufgabe, wobei auch hier das Vorgehen iterativ und fortlaufend ist. Nach dem Launch der ersten acht Module für die breite Öffentlichkeit im Dezember 2023 laufen zwei Prozesse parallel und kontinuierlich:

- Erweiterung: Unter Nutzung des oben beschriebenen Produktionsprozesses werden unter gemeinsamer Priorisierung von Projektteam und CISOs des Hochschulnetzwerkes NRW weitere Module entwickelt. Beispielsweise ist ein Modul zum Thema „Künstliche Intelligenz und IT-Sicherheit“ derzeit in Bearbeitung.
- Aktualisierung / Revision bisheriger Inhalte: Durch die rasanten Veränderungen im Themenbereich können Informationen schnell veralten und selbst ein Sicherheitsrisiko darstellen. Beispiele dafür sind die empfohlene Länge für starke Passwörter und die Einführung von Multifaktor-Authentifizierung an Hochschulen. Für diesen Zweck werden regelmäßige Überprüfungen und aus aktuellem Anlass erforderliche Revisionen durchgeführt, um die Qualität der veröffentlichten Inhalte zu gewährleisten.

4 https://www.articulate.com/360/rise/



Abb. 2: Interaktives Element, in dem personenbezogene Informationen aus einem Social Media Feed gesammelt und für die Formulierung eines Phishing Angriffs genutzt werden sollen

Neben der Bereitstellung der Selbstlernakademie als Webanwendung[5] steht sowohl das gesamte Angebot als auch einzelne Module für den Download als SCORM-Module zur Verfügung. Dies ermöglicht Hochschulen die spezifische Nutzung jener Inhalte und Integration in deren jeweilige Lernmanagementsysteme, wie beispielsweise ILIAS und Moodle. Aufgrund der Lizenz CreativeCommons `CC-BY` können auch Dritte sich die Ressourcen zu eigen machen, anpassen und nutzen.

3 Fazit

Der hohe Arbeitsaufwand bei der Entwicklung der Lernmaterialien und der Umsetzung der Lernplattform mit dem Faktor 1000:1 gegenüber der Konsumierungszeit ist vergleichbar mit Erfahrungen in der Literatur [Ch10]. Besonders bei der Videoproduktion mit externen Sprechenden ist zu beachten, dass ein vielschrittiger Prozess durchlaufen werden muss. Die Verwendung einer KI-basierten Text-to-Speech Stimme für die initiale Videoerstellung bis hin zur finalen Vertonung reduziert den Aufwand einer iterativen Vorgehensweise.

Der Einsatz von Autorentools für die Generierung von Modulen anstatt einer Eigenimplementierung hat sich positiv ausgewirkt: sowohl die Einhaltung von E-Learning Standards, wie SCORM, als auch eine einheitliche Gestaltung wird durch das Tool ermöglicht. Jedoch gibt es Limitationen in der Umsetzung der Selbstlernakademie, z.B. in der Gestaltung von Navigation, Farbe und Symbolik sowie Nachverfolgung von Lernaktivitäten.

Aufgrund der hohen Dynamik im Feld der IT-Sicherheit ist es nötig, die Selbstlernakademie kontinuierlich zu aktualisieren und nach Bedarf um neue Module zu erweitern. Dies wird neben der Evaluation des Angebotes in den nächsten drei Jahren Fokus vom Projekt SecAware.nrw sein.

5 https://secaware.nrw/selbstlernakademie/

Literaturverzeichnis

[Iee23] IEEE Standard for Learning Technology–JavaScript Object Notation (JSON) Data Model Format and Representational State Transfer (RESTful) Web Service for Learner Experience Data Tracking and Access. IEEE Std 9274.1.1-2023, Conference Name: IEEE Std 9274.1.1-2023, S. 1–103, 2023, DOI: 10.1109/IEEESTD.2023.10273185, URL: https://ieeexplore.ieee.org/document/10273185, Stand: 19. 06. 2024.

[Bu23] Bundesamt für Sicherheit in der Informationstechnik (BSI): Die Lage der IT-Sicherheit in Deutschland 2023, Bundesamt für Sicherheit in der Informationstechnik, 2023, URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2023.html?nn=132646, Stand: 07. 03. 2024.

[Ch10] Chapman, B.: How Long Does it Take to Create Learning? Chapman Alliance, 2010, URL: https://www.cedma-europe.org/newsletter%20articles/misc/How%20long%20does%20it%20take%20to%20develop%20training%20by%20Brian%20Chapman%20(Sep%2010).pdf.

[Lu22] Lucy Security AG / ThriveDX Lucy: LUCY Security Awareness Software, 2022, URL: https://lucysecurity.com/, Stand: 12. 06. 2024.

[Na24] Naß, M.; Dobbrunz, T.; Warmer, S.; Harrer, A.; Pohlmann, N.: Entwicklung, Evaluation und Distribution einer Selbstlernakademie zum Thema IT-Sicherheit. In. 22. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., Bonn, 2024.

[So22] SoSafe: Security Awareness und Human Risk Management, 2022, URL: https://sosafe-awareness.com/de/, Stand: 12. 06. 2024.

Enhancing Geographical Learning Through GeoVR: Immersive Exploration and Topographic Analysis Educational Application

Sam Sabah [1], Alexander Tillmann [2], David Weiß [2] und Hendrik Drachsler [3]

Abstract: GeoVR, a Virtual Reality immersive educational application, facilitates pedagogical approaches in geographical education, enabling students and educators to explore and analyze landscape features. This research examines the benefits of this educational environment, enabling students to seamlessly transition between macro and micro perspectives of terrain, facilitating a comprehensive understanding of surface structures. GeoVR aims to improve learning experiences, while simultaneously enhancing the user's understanding of topographical features and spatial relationships through scalable observation and interactive analysis tools. The application notably extends the scope of actions available to learners and educators in fields such as geography, geology, environmental studies, urban planning, and education. The precision of the representation, the interactive capabilities, the infinitely adjustable scales and the contribution to a realistic and immersive learning environment for geoinformation and processes can lead to a better understanding of the systems under investigation. Despite potential challenges, GeoVR seeks to bridge the gap between virtual and physical environments, changing how people learn and explore landscapes digitally.

Keywords: Virtual Reality, Geography, Geomorphologie, Field trip, Educational Applications

1 Motivation and Background

The implementation of GeoVR (a Virtual Reality immersive educational application) involves a comprehensive approach to integrating advanced technologies, pedagogical principles, and interactive functionalities. This section outlines the key components and strategies employed in the development and deployment of GeoVR. The use of VR technology in educational settings, particularly in the exploration and analysis of landscape features and processes, has been shown to enhance teaching and learning experiences. VR is able to create immersive and interactive environments, allowing students to explore and analyze landscapes that may not be locally accessible [Su21]. The Technology can potentially open up interesting geographical and geographical sites to students, academics and others who may not have had the opportunity to visit such sites previously. Recent

1 Goethe University of Frankfurt, studiumdigitale and Faculty of Computer Sciencee, Robert-Mayer-Str. 11-15 , 60325 Frankfurt am Main, Germany, sabah@sd.uni-frankfurt.de, https://orcid.org/0009-0009-5111-3068
2 Goethe University of Frankfurt, studiumdigitale, Eschersheimer Landstraße 155-157 , 60323 Frankfurt am Main, Germany, tillmann@sd.uni-frankfurt.de, https://orcid.org/0000-0001-7230-7042; weiss@sd.uni-frankfurt.de, https://orcid.org/0000-0002-7689-7833
3 Goethe University of Frankfurt, studiumdigitale and Faculty of Computer Sciencee and, DIPF | Leibniz Institute for Research and Information in Education , Rostocker Str. 6, 60323 Frankfurt am Main , h.drachsler@dipf.de, https://orcid.org/0000-0001-8407-5314

doi:10.18420/delfi2024_56

studies suggest that VR can serve as a potential alternative to traditional textbook-style learning, resulting in comparable performance levels and increased mood and engagement among learners[AM18].

This research aims to explicate the functionality and potential of GeoVR in augmenting spatial cognition and geographical comprehension, highlighting the importance of its educational applications. GeoVR combines immersive environments, precise 3D terrain models, and compatibility with various spatial data sources to create interactive learning experiences. The application significantly extends the scope of actions available to learners and educators in fields such as geography, geology, environmental studies, urban planning, and education.

Virtual Reality technologies have the capability to present abstract information in a concrete manner, thereby enhancing education through the integration of visual and auditory modalities, making VR-based experiments more immersive and engaging compared to traditional systems, furthermore educators need to worry less about ever complex and realistic models, and invest more into positive prior experiences of using technology [Sa02; Wr23].

Through scalable observation and interactive analysis tools, GeoVR encourages active involvement in topographic analysis and deepens comprehension. This research will discuss the methodology and principles of GeoVR and provide examples of its practical applications in different disciplines. We will also address advantages, challenges, and potential future directions for this innovative Virtual Reality application in geographical education and beyond.

2 Implementation

The implementation of GeoVR, as an immersive educational application, involves a comprehensive approach towards integrating advanced technologies, pedagogical principles, and interactive functionalities. This section outlines the key components and strategies employed in the development and deployment of GeoVR.

2.1 Development Framework and Technologies

GeoVR is built using the Unity game engine, leveraging its capabilities in rendering immersive 3D environments and supporting various VR hardware platforms. The use of Unity enables seamless integration of diverse visualization techniques, interactive elements, and analytical tools, ensuring a rich and engaging user experience. The application relies on real-world geolocation height maps to generate accurate 3D terrain models, enhancing spatial cognition and geographical comprehension. By incorporating high-resolution spatial data sources, GeoVR achieves precision in representing landscape structures and topographic features, contributing to a realistic and immersive learning environment.

2.2 User Interface and Interaction Design

The user interface of GeoVR is designed to be intuitive and user-friendly, catering to both novice users and experienced professionals in geospatial sciences. The application offers two distinct view modes – micro view and macro view (sandbox/open world) –, accommodating different learning preferences and objectives. Additionally, GeoVR provides multiple user movement methods, including short-distance teleportation and efficient navigation within the virtual environment, enhancing exploration and comparison of terrain features.

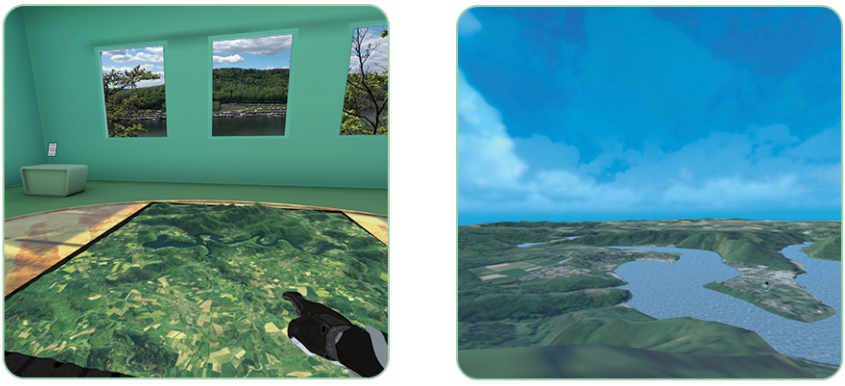


Abb. 1: micro and macro view of GeoVR

Interactive tools such as height indicators, compasses for orientation, and elevation tools empower users to conduct varied analytical tasks seamlessly using VR headset controllers. These tools facilitate activities such as identifying elevations, comparing heights, and monitoring changes in terrain over time, enriching the learning experience and promoting active engagement with geospatial data.

2.3 Scalability and Collaboration

GeoVR is designed to be scalable and adaptable to various learning scenarios and environments. The application accommodates multi-user collaboration, allowing students and professionals to interact and collaborate in real-time within the virtual landscape. This feature promotes collaborative learning experiences and facilitates knowledge sharing among users, enhancing the effectiveness and practicality of geospatial education.

3 Conclusions and Future Work

The future development of Virtual Learning Environments (VLEs) should prioritize the creation of educational experiences tailored to various levels of education [Gi24].

GeoVR demonstrates a novel approach in immersive educational technology, offering an engaging solution for geospatial learning through Virtual Reality. The development and implementation of GeoVR showcase its potential in enhancing spatial cognition, geographical comprehension, and observational skills among students and professionals in geography, geology, environmental studies, and urban planning. Through the integration of advanced technologies, pedagogical principles, and interactive functionalities, GeoVR provides a comprehensive platform for immersive exploration and topographic analysis. The application's user-friendly interface, diverse visualization techniques, and interactive tools contribute to a rich and engaging learning experience, bridging the gap between virtual and physical field trips. Regarding future work, GeoVR could explore further refinements and expansions of its functionalities, such as incorporating advanced analytical tools, enhancing realism in terrain representation, and integrating adaptive learning algorithms. Additionally, research efforts may consider evaluating GeoVR in diverse educational settings and populations, investigating its impact on learning outcomes, engagement, and retention over time.

GeoVR presents an intriguing new direction for geospatial education, providing students and professionals with an immersive and interactive platform to explore, analyze, and understand the complexities of our dynamic landscapes. As Virtual Reality technology continues to evolve, GeoVR aims to contribute to innovative pedagogical approaches, shaping the future of geographical learning and exploration.

Literaturverzeichnis

[AM18] Allcoat, D.; von Mühlenen, A.: Learning in virtual reality: Effects on performance, emotion and engagement. Research in Learning Technology 26, 2018.

[Gi24] Gielstra, D.; Moorman, L.; Kelly, J.; Schulze, U.; Resler, L. M.; Cerveny, N. V.; Gielstra, J.; Bryant, A.; Ramsey, S.; Butler, D. R.: Designing Virtual Pathways for Exploring Glacial Landscapes of Glacier National Park, Montana, USA for Physical Geography Education. Education Sciences 14 (3), S. 272, 2024.

[Sa02] Sala, N.: Virtual reality in the educational environments. In: ProceedingsIEEE International Conference on Advanced Learning Technologies ICALT. S. 417–419, 2002.

[Su21] Sun, S.; Meng, Q.; Ma, Y.; Ren, Z.: Application of virtual reality technology in landscape design. 2021.

[Wr23] Wright, P. N.; Whitworth, M.; Tibaldi, A.; Bonali, F.; Nomikou, P.; Antoniou, V.; Vitello, F.; Becciani, U.; Krokos, M.; Van Wyk de Vries, B.: Student evaluations of using virtual reality to investigate natural hazard field sites. Journal of Geography in Higher Education 47 (2), S. 311–329, 2023.

Adaptive Learning Systems in Programming Education: A Prototype for Enhanced Formative Feedback

Dominic Lohr [1], Marc Berges [1], Abhishek Chugh[2], and Michael Striewe [3]

Abstract:

Formative feedback is crucial in programming education, yet many learning systems fall short, concentrating mostly on pinpointing errors rather than guiding learners on how to resolve them. This is particularly unhelpful for novices who often lack advanced skills like debugging. Feedback is considered more valuable when it addresses error causes rather than just symptoms. However, this is challenging using only conventional methods like unit testing. Identifying error causes requires detailed information about both the error and the learner. Our proposed prototype introduces a new approach to integrating programming exercises into adaptive learning systems. It directly categorizes student code into so-called *answer classes* using a combination of static and dynamic code analysis. When integrated with data derived from a learner model, this approach enables tailored feedback that lowers the barrier to learning programming while keeping motivation high.

Keywords: programming education, adaptive feedback, CS1

1 Motivation

Programming educators in higher education are faced with the challenge of teaching diverse groups of learners whose prior knowledge of programming varies significantly [Ca15]. This discrepancy often results from the different school systems and curricula that students have passed through [Hu15]. While some learners have never written a single line of code others already have experience in developing (small) projects. Not overwhelming the former and not boring the latter is a pedagogical balancing act for educators, especially in the first few sessions of a programming course.

Adaptive learning systems (ALS) are considered as a promising approach to meet the heterogeneous needs of learners. Especially for programming novices, effective feedback is seen as crucial to support learners. However, the type of feedback required varies greatly depending on the individual learning level and type of error. Although providing elaborated feedback is crucial for learners, most learning systems offer only simple feedback focused mainly on mistakes [KJH19, Je22], without customization to individual needs. This often stems from the challenge of linking mistakes to their underlying causes, as many systems

1 FAU, Computer Science Education, Martensstraße 3, 91058 Erlangen, Germany,
dominic.lohr@fau.de, https://orcid.org/0000-0002-6330-2327;
marc.berges@fau.de, https://orcid.org/0000-0002-9982-547X
2 https://sophize.org, abc@sophice.de
3 University of Duisburg-Essen, The Ruhr Institute for Software Technology, Gerlingstraße 16, 45127 Essen, Germany, michael.striewe@paluno.uni-due.de, https://orcid.org/0000-0001-8866-6971

doi:10.18420/delfi2024_57

lack a model of the learner's unique profile, which we claim is crucial in most cases to infer from symptoms to causes – considered as more valuable feedback by learners [LB21]. The presented demo showcases a prototype that illustrates how an adaptive learning system can address these challenges. It achieves this by integrating diagnostic features with adaptive learning functionalities, merging established systems with a large language model (LLM).

2 Architecture

The architecture is composed of three integrated components: (A) the JACK system, (B) the ALEA system, and (C) GPT4 – a state-of-the-art large language model.

To evaluate the learner's program code we use JACK – an e-assessment system designed for grading programming exercises [St16]. JACK allows for both static and dynamic analysis of Java code. Based on these analyses we map the program code to specific answer classes. From a large number of existing systems for automated feedback generation for programming exercises [SS22], JACK is a suitable choice in this endeavor for two reasons: First, the architecture of JACK is specifically designed for modularity and extensibility, what makes it easy to integrate only the required components from JACK into the new prototype. Second, JACK employs a dedicated, customizable module for static code analysis [SG14], that allows the creation of exercise-specific rules and thus produces more fine-grained answer classes.

ALEA is an adaptive learning assistant for university courses that can provide customized content for learners based on semantically annotated course materials [Kr23, Be23]. In ALEA, learning objects are annotated following the Y-model framework [Lo23]. This approach makes details regarding necessary prerequisite knowledge, relevant concepts, learning objectives, and answer classes available to the system. Additionally, ALEA incorporates a learner model, capable of offering insights into the concepts that learners have successfully grasped within a specific cognitive dimension.

We use GPT4 to generate the feedback messages as the potential of the model to generate feedback has already been demonstrated in studies [KLK23, LKK24]. However, research shows that the generated feedback often lacks customization and, therefore, adaptivity. By applying retrieval augmented generation (RAG), we can dynamically add specific information to the prompt, making the feedback more tailored to the learner's needs.

2.1 System Design

Fig. 1 gives an overview of the architecture and the feedback pipeline. When feedback is requested, the code is first sent to the JACK to analyze whether it is compilable code. If the code contains syntax errors, there are three possibilities: (1) the compiler error message is sent directly to the learner (2) an explanation of the error message is generated with GPT4 and sent to the learner or (3) the program code is repaired to compilable code using GPT4

and then examined using the JACK system with regard to the answer classes. The latter variant also enables the provision of feedback on semantic or logical errors if the program code does not compile. This fundamentally expands the possibilities, as previous tools were unable to provide feedback on logical or semantic errors in non-compilable code.

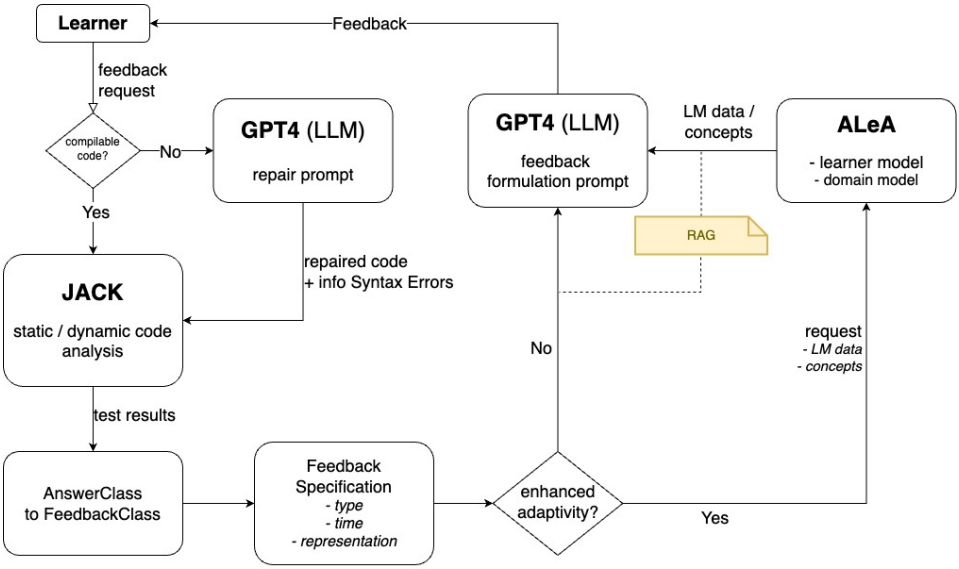


Fig. 1: Overview of system design

2.2 Functionality

Fig. 2 shows a screenshot of the prototype[4]. While working on a task, learners can use various methods to request different types of feedback, and it is up to the learner to decide when and how to ask for help. This fosters self-directed learning, as learners have different preferences. We also don't want to interrupt learners during their cognitive processes (a 'thinking step'). Therefore, we leave the timing of the feedback to the learner but implement automated support functionality when obviously struggling with the task, e. g. trying to compile five times with the same syntax error. Feedback from the compiler can be requested by using the RUN button. If there are test cases provided in the task the learner can request the test results using the TEST button. When struggling with the task the learner can request help any time using the HELP functionality.

4 Example task and code is taken from the Coding-Bat dataset [Ki22]

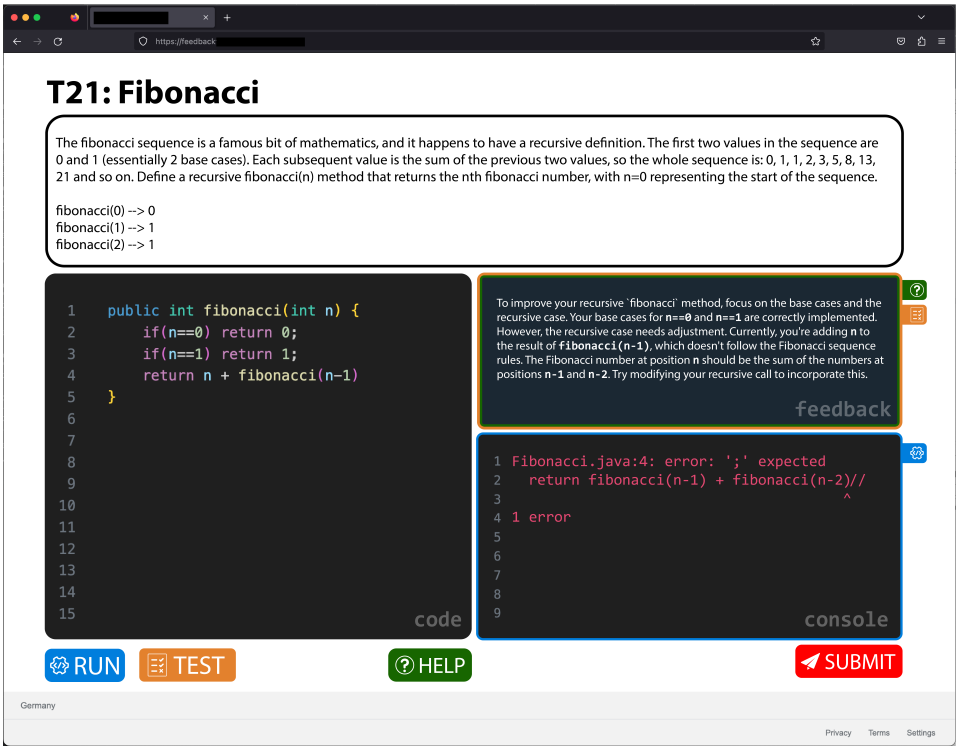


Fig. 2: Screenshot of the Prototype

3 Future Work and Vision

The prototype presented in this demo is limited to the programming language Java. One reason for this is that the JACK system currently only supports an analysis of Python and Java tasks. However, we claim that the system presented generalizes well to other programming languages – presumably with a few adaptations. Developing suitable answer classes for the programming tasks and writing suitable tests requires a considerable amount of time for programming instructors. However, some of the answer classes can be generalized across task types. It is, therefore, conceivable that in the future, the educator already has a well-stocked toolbox of answer classes with the appropriate analysis tests available for certain programming tasks.

Automated adaptive programming feedback offers a solution to the diverse needs of students in large introductory courses. Traditionally, the lack of time and resources makes it challenging for educators to provide personalized support, especially to those who need it most. Our approach demonstrates an initial step towards providing automatic, tailored feedback, addressing this critical gap.

Bibliography

[Be23] Berges, Marc; Betzendahl, Jonas; Chugh, Abhishek; Kohlhase, Michael; Lohr, Dominic; Müller, Dennis: Learning Support Systems Based on Mathematical Knowledge Management. In (Dubois, Catherine; Kerber, Manfred, eds): Intelligent Computer Mathematics, volume 14101, pp. 84–97. Springer Nature Switzerland, Cham, 2023.

[Ca15] Capovilla, Dino; Berges, Marc; Mühling, Andreas; Hubwieser, Peter: Handling Heterogeneity in Programming Courses for Freshmen. In: 3rd International Conference on Learning and Teaching in Computing and Engineering (LaTiCE). IEEE Press, Los Alamitos, pp. 197–203, 2015.

[Hu15] Hubwieser, Peter; Giannakos, Michail N.; Berges, Marc; Brinda, Torsten; Diethelm, Ira; Magenheim, Johannes; Pal, Yogendra; Jackova, Jana; Jasute, Egle: A Global Snapshot of Computer Science Education in K-12 Schools. In (Ragonis, Noa; Kinnunen, Päivi, eds): Proceedings of the 2015 ITiCSE on Working Group Reports. ACM Press, New York, pp. 65–83, 2015.

[Je22] Jeuring, Johan; Keuning, Hieke; Marwan, Samiha; Bouvier, Dennis; Izu, Cruz; Kiesler, Natalie; Lehtinen, Teemu; Lohr, Dominic; Peterson, Andrew; Sarsa, Sami: Towards Giving Timely Formative Feedback and Hints to Novice Programmers. In: Proceedings of the 2022 Working Group Reports on Innovation and Technology in Computer Science Education. ACM, Dublin Ireland, pp. 95–115, December 2022.

[Ki22] Kiesler, Natalie: Dataset: Recursive problem solving in the online learning environment CodingBat by computer science students. Online, June 2022. Datenerhebung: 2017. Version: 1.0.0. Datenpaketzugangsweg: Download-SUF. Hannover: FDZ-DZHW. Datenkuratierung: İkiz-Akıncı, Dilek.

[KJH19] Keuning, Hieke; Jeuring, Johan; Heeren, Bastiaan: A Systematic Literature Review of Automated Feedback Generation for Programming Exercises. ACM Transactions on Computing Education, 19(1):1–43, March 2019.

[KLK23] Kiesler, Natalie; Lohr, Dominic; Keuning, Hieke: Exploring the Potential of Large Language Models to Generate Formative Programming Feedback. In: 2023 IEEE Frontiers in Education Conference (FIE). IEEE, College Station, TX, USA, pp. 1–5, October 2023.

[Kr23] Kruse, Theresa; Berges, Marc; Betzendahl, Jonas; Kohlhase, Michael; Lohr, Dominic; Müller, Dennis: Learning with ALeA: Tailored Experiences through Annotated Course Material. In: INFORMATIK 2023 - Designing Futures: Zukünfte Gestalten. Gesellschaft für Informatik e.V., Bonn, pp. 395–398, 2023.

[LB21] Lohr, Dominic; Berges, Marc: Towards Criteria for Valuable Automatic Feedback in Large Programming Classes. In: Hochschuldidaktik Informatik HDI 2021. Jörg Desel, Simone Opel, Dortmund, pp. 181–186, September 2021.

[LKK24] Lohr, Dominic; Keuning, Hieke; Kiesler, Natalie: You're (Not) My Type – Can LLMs Generate Feedback of Specific Types for Introductory Programming Tasks? Journal of Computer Assisted Learning, 2024. accepted.

[Lo23] Lohr, Dominic; Berges, Marc; Kohlhase, Michael; Müller, Dennis; Rapp, Max: The Y-Model - Formalization of Computer Science Tasks in the Context of Adaptive Learning Systems. In: 2023 IEEE 2nd German Education Conference (GECon). IEEE, Berlin, Germany, pp. 1–6, August 2023.

[SG14] Striewe, Michael; Goedicke, Michael: A Review of Static Analysis Approaches for Programming Exercises. In: Proceedings of the International Conference on Computer Assisted Assessment (CAA 2014). CCIS 439, Zeist, Netherlands, pp. 100–113, 2014.

[SS22] Strickroth, Sven; Striewe, Michael: Building a Corpus of Task-Based Grading and Feedback Systems for Learning and Teaching Programming. International Journal of Engineering Pedagogy (iJEP), 12(5):26–41, Nov. 2022.

[St16] Striewe, Michael: An Architecture for Modular Grading and Feedback Generation for Complex Exercises. Science of Computer Programming, 129:35–47, November 2016.

Autorenverzeichnis

A
Akao, Kensuke, 467
Appel, Sven, 291
Ast, Alexander, 523
Aumann, Alexander, 201
Azaiz, Imen, 381

B
Baberowski, David, 115
Bauer, Katrin, 361
Baur, Melanie, 275, 425
Berges, Marc, 369, 549
Bergner, Nadine, 115
Berkhan, Chantal, 225
Bernoth, Jan, 281
Bocklet, Tobias, 133
Brettschneider, Philipp, 415
Breuer, Martin, 267
Brocker, Annabell, 441

C
Canz, Michael, 361
Christ, Paul, 331
Chugh, Abhishek, 549
Cordes, Andre, 141

D
Dahal, Prabin, 361
Dannath, Jesper, 97, 485
Deriyeva, Alina, 97, 485
Dobbrunz, Tabea, 407, 539
Drachsler, Hendrik, 545
Durski, Sara, 347

E
Egetenmeier, Armin, 397, 517
Ehlenz, Matthias, 251, 533

F
Fernes, David, 305
Finster, Rebecca, 153
Fischer, Stefan, 449
Freier, Carolin, 133
Freisinger, Steffen, 133
Fuchs, Andreas, 291
Funke, Florian, 125

G
Gantikow, Alexander, 347
Geis, Michael, 313
Georges, Munir, 133
Golovanev, Svetlana, 313
Görzen, Sergej, 267, 493, 527
Grimm, Paul, 291
Grogorick, Linda, 153

H
Haag, Maike, 171
Haake, Jörg, 331
Hacke, Alexander, 211
Haller, Marcella, 523
Harrer, Andreas, 389, 407, 539
Heinemann, Birte, 251
Herwig, Maurice, 107, 479
Hillenbrand, Gisela, 361
Hofmann, Sven, 125
Hundeshagen, Norbert, 107, 479
Hundhausen, John, 107

I
Isking, Andreas, 347

J
James, David, 457
Jebing, Zoe, 517
Judel, Sven, 81

K
Kablowski, Stefan, 107
Kapfenstein, Ann-Kathrin, 179
Kastaun, Marit, 479
Kaussow, Max, 449
Keuning, Hieke, 19
Kiesler, Natalie, 63, 251
Kießling, Peter, 125
Klees, Guido, 241
Koch, Christian, 499
Koch, Nadine Nicole, 179
Kohlhase, Michael, 369
Kollenberg, Cedric, 479
Konert, Johannes, 457
Krause, Daniel, 493
Kravcik, Milos, 171
Kreinsen, Moritz, 163
Kruse, Theresa, 369
Kuhn, Michael, 457

L
Laban, Firas Al, 281
Lange, Martin, 107
Ledel, Benjamin, 511
Leonhardt, Thiemo, 115
Libbrecht, Paul, 347
Lilienthal, Lanea, 115
Lohr, Dominic, 369, 549
Lucke, Ulrike, 211, 225

M
Mack, David, 49
Manz, Christian, 211
Maquet, Frederic, 533
Mascher, Ulrike, 305
Meinhardt, Felix, 493
Meißner, Niklas, 179
Metternich, Joachim, 313
Moghbeli, Halimeh, 369
Müller, Wolfgang, 201, 347
Munkelt, Torsten, 331
Mushoff, Maximilian, 313

N
Naß, Miriam, 407, 539
Neumann, Alexander Tobias, 171
Nowotny, Florian, 225
Nugroho, Saptadi, 361

O
Ostermann, Simon, 347

P
Paaßen, Benjamin, 97, 485
Pelzer, Philipp, 493
Pengel, Norbert, 171
Persike, Malte, 267
Pohlmann, Norbert, 407, 539

R
Ranzenberger, Thomas, 133
Rebholz, Sandra, 347
Reinartz, Lea, 389
Reinold, Luca, 133
Riedel, Christian, 281
Riedhammer, Korbinian, 133
Robra-Bissantz, Susanne, 153
Röpke, René, 251, 527
Roth, Anna-Lena, 457

S
Sabah, Sam, 545
Salmen, Frederic, 267, 499

Sänger, Volker, 361
Santelmann, Kirsten, 241
Scheer, Sarah, 449
Schiffner, Daniel, 23, 63, 251
Schlicker, Marc, 321
Schmidt, Claudia, 361
Schmidt, Ludger, 49
Schneeweiß, Sarah, 389
Schneider, Fabian, 133
Scholl, Andreas, 63
Schröder-Kroll, Roland, 313
Schroeder, Michael, 493
Schroeder, Ulrik, 81, 267, 441, 493, 527
Schulz, Sandra, 163, 251
Schütz, Marcel, 369
Schwarz, Tabea, 511
Sedlmair, Johanna, 275, 425
Sedlmeier, Teresa, 361
Simic, Christopher, 133
Simon, Claudia, 133
Soliman, Hassan, 171
Speiser, Sebastian, 275, 425
Sprenger, Sandra, 163
Stelzner, Marc, 449
Stratmann, Jörg, 201
Strickroth, Sven, 251, 381, 397, 517
Striewe, Michael, 473, 505, 549

T
Thiems, Maimon, 163
Tillmann, Alexander, 545

V
Varney, Valerie, 313

W
Wagenitz, Axel, 313
Warmer, Sina, 407, 539
Weber, Markus, 241
Wehrmann, Frank, 31
Weiß, David, 545
Weißbach, Rüdiger, 313
Weitzel, Holger, 201
Werkes, Richard, 493
Wiepke, Axel, 211, 281
Wilhelm-Weidner, Arno, 251
Wilzeck, Rene, 233
Wirzberger, Maria, 179
Wohlers, Dennis, 313
Wollersheim, Heinz-Werner, 171

Y
Yin, Yue, 171

Z
Zeise, Christopher, 81
Zender, Raphael, 31, 233, 321
Zohner, Michael, 415