GESELLSCHAFT
FÜR INFORMATIK



Veit Köppen und Sven Strickroth (Hrsg.)

23. Fachtagung Bildungstechnologien (DELFI 2025)

8.–11. September 2025
Freiberg

Gesellschaft für Informatik e.V. (GI)

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

Volume P-369

ISSN 2944-7682 (Online)

Volume Editors
Prof. Dr. Veit Köppen
Fachhochschule Potsdam
Kiepenheuerallee 5, 14469 Potsdam
veit.koeppen@fh-potsdam.de

Prof. Dr. Sven Strickroth
Ludwig-Maximilians-Universität München
Oettingenstraße 67, 80538 München
sven.strickroth@ifi.lmu.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 Frank, Universität Duisburg-Essen, Germany
Barbara Hammer, Universität Bielefeld, Germany
Falk Schreiber, Universität Konstanz, Germany
Wolfgang Karl, KIT Karlsruhe, Germany
Michael Koch, Universität der Bundeswehr München, Germany
Heiko Roßnagel, Fraunhofer IAO Stuttgart, Germany
Kurt Schneider, Universität Hannover, Germany
Andreas Thor, HFT Leipzig, Germany
Ingo Timm, Universität Trier, Germany
Karin Vosseberg, Hochschule Bremerhaven, Germany
Maria Wimmer, Universität Koblenz-Landau, Germany

Dissertations
Rüdiger Reischuk, Universität Lübeck, Germany
Thematics
Agnes Koschmider, Universität Kiel, Germany
Seminars
Judith Michael, RWTH Aachen, Germany

© Gesellschaft für Informatik, Bonn 2025



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

Vorwort

Die DELFI-Tagungsreihe der Gesellschaft für Informatik zeichnet sich durch eine engagierte Auseinandersetzung mit dem facettenreichen Themenfeld der Bildungstechnologien aus und wird besonders durch den interdisziplinären Diskurs zwischen Wissenschaftler*innen verschiedener Fachrichtungen und den Akteur*innen diverser Bildungskontexte belebt.

Die 23. Fachtagung Bildungstechnologien (DELFI 2025) an der Technischen Universität Bergakademie Freiberg trägt das Motto „Digitale Technologien für gemeinsames Lehren und Lernen“. Damit legt diese Tagung den Fokus auf die kollaborative Nutzung von Bildungstechnologien und beleuchtet deren Rolle bei der Förderung von Digitalität, Innovation und dem Einsatz offener technologischer Lösungen. Im Zentrum der Diskussion steht eine breite Palette von Themen aus der Informatik-Ausbildung, ergänzt durch die Untersuchung verschiedenster Werkzeuge, Methoden und Lösungen zur praxisnahen Kompetenzvermittlung. Darüber hinaus werden rechtliche, organisatorische und soziale Rahmenbedingungen eingehend betrachtet. Thematische Schwerpunkte bilden dieses Jahr:

- Bildungstechnologie für gemeinsame Lehre und gemeinsames Lernen
- partizipative Erstellung und Individualisierung offener Materialien
- Organisation gemeinsamer Bildungsangebote
- Adaption und Qualitätssicherung gemeinsam erstellter Bildungsinhalte
- Anrechnung und Anerkennung von kollaborativen Leistungen
- Abbildung gemeinsamer Lernprozesse in Learning Analytics

Die eingereichten wissenschaftlichen Beiträge wurden im Doppelblind-Verfahren von je drei Mitglieder*innen des Programmkomitees begutachtet. Bei einer Programmkomitee-Sitzung wurde dann über die Beiträge diskutiert, um zu einer Entscheidung durch die Programmleitung zu gelangen. Eingereicht wurden 30 Langbeiträge, von denen 9 angenommen wurden (Annahmequote 30 %) sowie 22 Kurzbeiträge, von denen 5 angenommen wurden (Annahmequote 23 %). Zudem gab es 6 Praxisbeiträge mit 3 Annahmen (Annahmequote 50 %) und 2 Einreichungen als Positionspapier, wobei beide angenommen wurden. Somit wurden von den insgesamt 60 wissenschaftlichen Einreichungen 19 Beiträge für die DELFI 2025 akzeptiert (Annahmequote 32 %).

Auch in diesem Jahr erfolgte kein Downgrade der wissenschaftlichen Beiträge, sondern die Autor*innen konnten die Hinweise der Gutachter*innen in einer zweiten Einreichungsrunde für Poster- und Demobeiträge nutzen. Hier gab es anschließend ein zweites Doppelblind-Reviewverfahren in dem letztlich 19 Poster- und 20 Demobeiträge für die DELFI 2025 in Freiberg ausgewählt wurden.

Die „Tradition“ der Open-Data-Badges aus dem letzten Jahr wurde fortgeführt. Beiträge, die Forschungsdaten oder Software öffentlich zugänglich machen, werden damit auf der jeweiligen Titelseite ausgezeichnet. Voraussetzung ist, dass Forschungsdaten oder Soft-

ware direkt im Paper referenziert sind und an einen Ort hinterlegt wurden, der den wissenschaftlichen FAIR-Ansprüchen genügt. Fünfzehn Papiere erfüllen diese Bedingungen und tragen daher das Open-Data-Badge.

Im vorliegenden Tagungsband finden Sie alle angenommenen Beiträge, zusammen mit zwei Beiträgen zu den Keynotes. Gemeinsam mit dem Programmkomitee konnten wir ein interessantes, abwechslungsreiches und qualitativ hochwertiges Programm zusammenstellen. Den Autor*innen gebührt unser herzlicher Dank für ihre wertvollen Beiträge. Wir möchten uns ausdrücklich bei den Mitglieder*innen des Programmkomitees für die ausgezeichnete und gelungene Zusammenarbeit bedanken. Diese Kooperation hebt die wissenschaftliche Qualitätssicherung, wie sie unser Tagungsmotto betont, besonders hervor.

In gewohnter Weise wird die DELFI 2025 von einem Workshoptag eingeleitet. Dabei können in zwei Ganztages- und fünf Halbtagesworkshops Ideen, Ansätze und Herausforderungen diskutiert und erarbeitet werden. Die Themen der Workshops sind dabei ebenfalls sehr vielfältig und adressieren den Einsatz von Künstlicher Intelligenz, Open Science, Extended Reality oder Bildungstechnologien in der Schule sowie Learning Analytics. Die Ergebnisse sollen in separaten Proceedings nach der Tagung veröffentlicht werden.

Neben den wissenschaftlichen Präsentationen und Diskussionen sowie den Austausch und die Vernetzung innerhalb der Community auf der DELFI 2025 bietet Freiberg und die Technische Universität Bergakademie Freiberg einen Blick über den bildungstechnologischen Tellerrand hinaus und eröffnet hoffentlich neue Einblicke, die zu interdisziplinären Ansätzen und einer fruchtvollen Zusammenarbeit über die Tagung hinaus reichen.

Wir freuen uns darüber, Sie auf der 23. Fachtagung Bildungstechnologien an der Technischen Universität Bergakademie Freiberg begrüßen zu dürfen!

Potsdam und München, im September 2025

Veit Köppen und Sven Strickroth (die Herausgeber)

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

Sebastian Zug (Gesamtleitung)
Raphael Zender (Poster Chair) und Bastian Pfleging (Demo Chair)
Anja Hawlitschek (Workshop Chair) und Birte Heinemann (Workshop Chair)

Tagungsleitung

Gesamtleitung:	Sebastian Zug, TU Bergakademie Freiberg
Leitung des Programmkomitees:	Veit Köppen, Fachhochschule Potsdam
	Sven Strickroth, LMU München
Workshop Chairs:	Anja Hawlitschek, OVGU Magdeburg
	Birte Heinemann, RWTH Aachen
Poster Chair:	Raphael Zender, Zeppelin Universität Friedrichshafen
Demo Chair:	Bastian Pfleging, TU Bergakademie Freiberg

Programmkomitee

Nadine Bergner	RWTH Aachen
Jan Bernoth	Universität Potsdam
Daniel Bodemer	Universität Duisburg-Essen
Josef Buchner	Pädagogische Hochschule St. Gallen
Julian Dehne	Gesellschaft für Informatik
André Dietrich	TU Bergakademie Freiberg
Ralf Doerner	Hochschule RheinMain
Matthias Ehlenz	RWTH Aachen
Robert Garmann	Hochschule Hannover
Paul Grimm	Hochschule Darmstadt
Volker Göhler	TU Bergakademie Freiberg
Joerg Haake	FernUniversität in Hagen
Andreas Harrer	Fachhochschule Dortmund
Jasmin Hartanto	RWTH Aachen
Anja Hawlitschek	Otto-von-Guericke-Universität Magdeburg
Peter Henning	Hochschule Karlsruhe
Claudia Hildebrandt	Pädagogische Hochschule Heidelberg
Sebastian Hobert	Technische Hochschule Lübeck
Marco Kalz	Pädagogische Hochschule Heidelberg
Paul-Thomas Kandzia	Duale Hochschule Baden-Württemberg Lörrach
Natalie Kiesler	Technische Hochschule Nürnberg Georg Simon Ohm
Johannes Konert	Hochschule Fulda
Thomas Kosch	Humboldt-Universität zu Berlin
Bastian Küppers	Maastricht University
Roland Küstermann	Deutsche Hochschule für angewandte Wissenschaften 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
Agathe Merceron	Berliner Hochschule für Technik

Tobias Moebert	Universität Potsdam
Miriam Mulders	Universität Duisburg-Essen
Andreas Mühling	Christian-Albrechts-Universität zu Kiel
Svenja Noichl	RWTH Aachen
Benjamin Paaßen	Universität Bielefeld
Vivien Petras	Humboldt-Universität zu Berlin
Niels Pinkwart	Humboldt-Universität zu Berlin
René Röpke	Technische Universität Wien
Guido Rößling	Technische Universität Darmstadt
Maren Scheffel	Ruhr-Universität Bochum
Daniel Schiffner	DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Jan Schneider	DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Ulrik Schroeder	RWTH Aachen
Sandra Schulz	Universität Hamburg
Clara Schumacher	Universität Potsdam
Andreas Schwill	Universität Potsdam
Niels Seidel	FernUniversität in Hagen
Katharina Simbeck	Hochschule für Technik und Wirtschaft Berlin
Heinrich Söbke	Hochschule Weserbergland und Bauhaus-Universität Weimar
Patrick Stalljohann	FH Münster
Michael Striewe	Hochschule Trier
Julia Maria Struß	Fachhochschule Potsdam
Veronika Thurner	Hochschule München
Alexander Tillmann	Goethe-Universität Frankfurt
Martin Wessner	Hochschule Darmstadt
Arno Wilhelm-Weidner	VDI/VDE-IT
Raphael Zender	Zeppelin Universität Friedrichshafen

Organisationsteam

Sebastian Zug	TU Bergakademie Freiberg
Ines Aubel	TU Bergakademie Freiberg
Josephine Mäke	TU Bergakademie Freiberg

Herzlich Willkommen in der Silberstadt!

Die Stadt Freiberg blickt auf eine mehr als 850-jährige Geschichte zurück. Ihre Wurzeln liegen im Silberbergbau: Seit dem 12. Jahrhundert prägten Erzfunde und Montanwesen die Entwicklung der „Silberstadt“ und machten sie über Jahrhunderte zu einem bedeutenden Wirtschafts- und Wissenschaftsstandort. Zeugnisse dieser Epoche – die historische Altstadt, das Besucherbergwerk „Reiche Zeche“ oder die Mineraliensammlung „terra mineralia“ – lassen den Reichtum dieser Tradition bis heute lebendig werden. Gleichzeitig ist mit der 1765 gegründeten TU Bergakademie Freiberg die älteste montanwissenschaftliche Hochschule der Welt in der Stadt beheimatet. Damit repräsentiert Freiberg den Wandel vom Rohstoffabbau hin zur Wissensproduktion – der Begriff der Ressource wird in zeitgemäßer Weise neu interpretiert.

Aber auch die Grundlage des Rohstoffs Wissen – die systematische Vermittlung und Weitergabe – unterliegt einem beständigen Wandel. Die Organisator*innen der DELFI 2025 möchten diesen Veränderungsprozess auch in der Gestaltung des Tagungsprogramms sichtbar machen. Entsprechend beginnt die Tagung am 8. September 2025 mit der Welcome Reception im historischen Wernerbau. Der Namensgeber, Abraham Gottlob Werner, unterrichtete bereits Humboldt und Novalis während ihres Studiums in Freiberg. Seit über 100 Jahren befindet sich an diesem Standort ein Teil der bedeutenden geowissenschaftlichen Sammlungen der Universität. Die Einordnung dieser Kollektion und ihre Nutzung im Lehrkanon wird im Vortrag „250 Jahre Mineralogielehre an der Bergakademie“ thematisiert. So richtet sich zu Beginn der DELFI der Blick in die Vergangenheit und auf die Frage, welche didaktischen Methoden und Konzepte sich zeitunabhängig bewährt haben.

Am 9. und 10. September 2025 liegt der Fokus auf der zukünftigen Entwicklung von Bildungstechnologien. Als Spiegelbild dieses Perspektivwechsels dient der zentrale Tagungsort der DELFI 2025: die neue Universitätsbibliothek der Bergakademie. Das Gebäude dokumentiert den Wandel in der Wissensvermittlung: Traditionelle Materialien werden mit digitalen Formaten und Angeboten kombiniert, um ein disziplinübergreifendes, kollaboratives Lernen zu ermöglichen. Studierende und Mitarbeiter*innen finden den Raum und die technischen Voraussetzungen, um in unterschiedlichsten Formaten individuell oder in der Gemeinschaft zu arbeiten. So wird das Motto der diesjährigen DELFI-Konferenz, „Digitale Technologien für gemeinsames Lehren und Lernen“, unmittelbar greifbar.

Mit diesem inhaltlichen und räumlichen Spannungsbogen bietet die TU Bergakademie Freiberg eine inspirierende Kulisse für den Austausch über die Zukunft des Lehrens und Lernens und freut sich darauf, die Teilnehmenden der DELFI 2025 in der Silberstadt willkommen zu heißen.

Glück auf!

Sebastian Zug

Freiberg, im September 2025

Inhaltsverzeichnis

Eingeladene Vorträge

Isa Jahnke
Digital Learning Experience Design and Research 3

Armin Weinberger
Towards Hybrid Intelligence in CSCL 5

Best-Paper-Kandidaten-Langbeiträge

Imen Azaiz, Sven Strickroth
Aligning Teaching Objectives with Formative Assessment to Uncover Student Learning Challenges 9

Jonathan Dyrna, Richard Joos, Marilisa Herchet, Thomas Köhler
Prior virtual reality experience and its determinants in educational settings 25

Dominik Evangelou, Maria Klar, Kristian Träg, Miriam Mulders, Melina Marnitz, Lara Rahner
GenAI-Chatbots as Debriefers: Investigating the Role Conformity and Learner Interaction in Counseling Training 41

David Mack, Ludger Schmidt
Integration realer Werkzeuge in eine VR-Trainingssoftware für komplexe manuelle Montageprozesse 57

Best-Paper-Kandidaten-Kurzbeiträge

Armin Egetenmeier, Sven Strickroth
Empowering Tutors with Digital Weekly Student Survey Results 73

Sigrid L. Klinger, Sven Strickroth
Assessing GPT-4's Reliability for Japanese Kanji Instruction: Accuracy, Prompt Control, and Contextual Sentences 81

Marie Trippel, René Röpke
Developing a Graph-based Visualization of Elective Courses to Support Course Selection in Higher Education 89

Künstliche Intelligenz

Philipp Matter, Ninja Leikert-Böhm
Adding the Human Error: Exploring the Impact of GenAI on Student Writing in Information Systems 99

Lea Cohausz, Thilo I. Dieing, Belinda Merkle, Marc Scheffler, Stefan Janke
Smart Start: A Critical Analysis of Chatbots as Tools to Support Study Orientation 113

Virtual Reality

Pascal Bolz, Dominic Fehling, Markus Zdrallek
Immersive Virtual Reality Training for Service Personnel in the Power Distribution Grid with Adaptable Training Environments 131

Birte Heinemann, Matthias Ehlenz, Ulrik Schroeder
Didaktische Konzepte fürs Onboarding und Tutorials in VR-Lernanwendungen 139

Learning Analytics/Gamification

Claudine Boyer, Tobias Berner, Thomas Schmalfeldt
Entwicklung eines Lehrkräfte-Dashboards für ein sicheres und schulgerechtes soziales Netzwerk 157

Josefine Welk, Christian Roschke, Dominik Breck, Max Schlosser, Marc Ritter
Evaluierung einer gamifizierten Lern-App: Usability, User Experience und Lerneffekt in der Hochschulbildung 171

Clemens Witt, Nico Reithmeier, Thiemo Leonhardt, Mareen Grillenberger
Fostering Structured Problem Solving in Machine Learning through Collaborative Gameplay 185

Martin Breuer, Annabell Brocker, Malte Persike, Ulrik Schroeder
Examiner Perspectives on Assessment Analytics as a Foundation for Developing Quality Assurance Tools in Higher Education 193

Open Science/Software

Sandra Schulz, Daniel Schiffner, Natalie Kiesler, Katarzyna Biernacka
A Publication Framework for Research Data in the EdTec Community . . 203

Sarah Schneeweiß, Andreas Harrer
Nutzungserfahrungen und partizipative Entwicklung eines Scaffolding-Frameworks zur Unterstützung von Lernprozessen 213

Nadja Dietze, Caroline Surmann
Auf dem Weg zur Umsetzung der Open Educational Resources Strategie (OER-Strategie) 221

Sven Judel, Erik Schnell, Ulrik Schroeder
Ein Data Warehouse für die Zusammenführung von Hochschuldaten . . . 229

Poster-Beiträge

Aakash Ahmad, Thomas Schmid
Synergising Human Expertise and Generative Artificial Intelligence for Curriculum Design 239

Jan Bernoth, Birte Heinemann, Matthias Ehlenz, Martin Armbruster, Ulrike Lucke
FAIR4EDTECH: Towards an Artifact Evaluation Track 245

Andrea de Santiago, Carina Ziegler, Isa Jahnke
Prompt Engineering or Creating Custom GPTs, or both? 251

André Dietrich, Jihad Hyadi, Sebastian Zug
Server or Cloud? Who needs it! Peer-to-Peer Remote Labs with edrys-Lite. 257

Flavia Gentile, Mascha Wanke, Evi Hochuli, Wolfgang Mueller
Virtual Reality in Speech Therapy Education 263

Volker Göhler, Sebastian Zug, Hilke Domsch
Enhancing Vocational Education through AI-driven Learning and Competence Assessment: The DiAgnostiK Project 271

Jasmin Hartanto, Birte Heinemann, Ulrik Schroeder
Lernendenzentriertes Onboarding in RSE: Ein Forschungsdesign zur Persona-Entwicklung 277

Anja Hawlitschek, Alisa Poleshchuk, Nadine Wegmeyer, Marie-Louise Hilberer, Christin Deinert
Design of Interactive Learning Materials with H5P 283

Maurice Herwig, Norbert Hundeshagen, Cedric Kollenberg, Robin Weber
On AI Generated Reductions from Natural Language Descriptions 289

Katja Hornoff, Mariane Liebold, Josefin Müller, Anja Schulz, Anne Vogel
Nutzung von Fremdkonzepten in der hochschuldidaktischen Praxis – Mehrwerte und Fallstricke 295

Sigrid L. Klinger, Patrick Weber, Sven Strickroth, Michael Striewe
Stimulating Reflection on Code Quality: An Adaptive and Automated Question Generation Approach for Student Code 301

Veit Köppen, Anja Hawlitschek
Supporting Program Comprehension with Digital Learning Journals: Experiences from a Course with Non-CS Students 307

Tim Komorowski, Laura Getz, Daniel König, Lilli Heimes, Meike Weiland, Michael Heister
Gemeinsames Arbeiten mit digitalen Technologien: Positive Effekte auf intrinsische Motivation und selbsteingeschätzte digitale Kompetenzen bei Lernenden in der dualen Berufsausbildung 313

Martin Sboron, Oliver Thews, Thomas Schmid
ViPATalk meets AI: Advancing Multimodal Virtual Anamnesis Simulation using LLM Fine-tuning 319

Juliane Scherlitzki, Axel Wiepke
Vergleich von Stresssituationen im simulierten Schulunterricht durch Biofeedback 325

Doerthe Vieten, Jörn Hees
Woran denken Sie, wenn Sie ‚Informatik' hören? 331

Frank Wehrmann, Raphael Zender
Modulares Open-Source-Framework für inklusive Multi-User-VR-Lehr-Lernanwendungen 337

Patrick Witthaus, Alina Abé, Samuel Tomczyk, Raphael Zender
Digital gestützte Interventionsentwicklung für Studierende mit Lernproblemen 343

Lucas Yerinç
Generative KI in Bildungskontexten: Eine Analyse auftretender Diskriminierung und ihrer Implikationen am Beispiel von ChatGPT . . . 349

Demo-Beiträge

Firas Al Laban, Jan Bernoth, Safial Islam Ayon, Axel Wiepke, Ulrike Lucke
RDMC: A Tool to Support Authors in Preparing Artifacts for Evaluation . 357

David Baberowski, Thiemo Leonhardt, René Röpke, Sila Irmak
Pack-a-Capsule – Extending a VR Serious Game on Network Communications to Enable Collaborative Gameplay 363

Volker Bach, Aloys Krieg, Ruedi Seiler
Hochschulübergreifende Kooperation am Beispiel der Mathematikkurse OMB+ und HM4MINT 369

Annabell Brockner, Ulrik Schroeder
Integrating LLM-Generated Feedback in JupyterLab to Enhance Programming Education 373

Jakob Czekansky, Justin Sauer, Diethelm Bienhaus
Shared Sessions in MICRO: Enabling Collaborative Remote Experimentation in Embedded Systems Education 379

Louis Dunkel, Imran Hossain, Patrick Sacher, Hasan Abu-Rasheed
LearnMate 385

Nghia Duong-Trung, Xia Wang, Rahul Rajkumar Bhoyar, Angelin Mary Jose, Silke Elisabeth Wrede, Lars van Rijn, Theresa Panse, Claudia De Witt, Niels Pinkwart
Kompagnon für das Verfassen von Hausarbeiten 391

Andreas Dymek, Frederic Salmen, Ulrik Schroeder
Web-Based Prototype of a Branching Scenario Editor for Adaptive Learning 395

Christopher Eck, Sven Judel, Tobias Brachmann, Ulrik Schroeder, Martin Frenz
Konzeption eines Moodle-Plugins zur Selbst- und Fremdeinschätzung von Sozialkompetenzen in Projektarbeiten 401

Rahel Natalie Engel, Hannah Lenk, André Greubel, Dorian Weber
Cats, Code, and Concepts: Learn to Program with Catpie 407

Burhan Karababa, Armin Egetenmeier, Sven Strickroth
UniGoal: Ein Unterstützungstool für SMARTe Lernziele 413

Ulrike Mascher, David Fernes, Alexander Tillmann
VRsim_EU – AI Act: Exploring Educational Simulation Gaming in Virtual Reality 419

Erik Morawetz, Roy Meissner, Nadine Hahm, Andreas Thor
FeeDi: Typunabhängiges Diagramm-Assessment durch deklarative Diagrammspezifikation 425

Anna Pakosch, Dennis Allerkamp, Anne Nadolny, Arne Koschel, Monika Stöhr, Christin Schulze
Collect Skills – Tool zur Selbsteinschätzung des individuellen Kompetenzerwerbs 431

Marc Schlicker, Astrid Schmidt, René Wilzeck, Britta Fischer, Thilo Kleickmann, Raphael Zender
Teach-R Sporthalle: Erweiterung des VR-Klassenzimmers um Lernszenarien zur Professionalisierung von Sportlehrkräften 437

Thomas Schmalfeldt, Tobias Berner, Claudine Boyer
Entwicklung einer sicheren Plattform zur explorativen Nutzung sozialer Medien im Unterricht 443

Konrad Schöbel, Jochen Merker, Paul Brassel, Moritz Zöllner, Heike Hain
PyRope – Ein codebasierter Ansatz für E-Assessment in MINT-Fächern . 449

Niels Seidel, Pio Virzi, Marie-Theres Niazi-Shahabi
Advancing Study Planning with Learning Analytics 453

Lea Sigethy, Sarah Aragon-Hahner, Sophia Sakel, Sven Strickroth
StudyOrientation: Eine mobile App zur Unterstützung der Selbstreflexion bei der Studienwahl im Fach Informatik 459

Raphael Wimmer, Leonie Schrod, Alexander Weichart
Demonstrating GraphIT - a Platform for Working with Dependency Graphs of Learning Topics 465

Autorenverzeichnis 471

Eingeladene Vorträge

Digital Learning Experience Design and Research

Isa Jahnke [1]

Abstract: Unter Verwendung des neuen Forschungsfeldes „Learning Experience Design (LXD)“ werden Designs von digitalem Lernen – „positive or memorable experiences“ (unvergessliche Lernerfahrungen) – anhand empirischer Studien vorgestellt, z. B. in den Kontexten von Mobile-Microlearning, Games for learning, und Online-Learning. Methoden umfassen unter anderem Usability-Tests, User Experience Studies, Pre-/Post-Tests, und Follow-up Interviews. Diese werden zur Gestaltung und Erfassung von effektiven, effizienten und ansprechenden digitalen Lerndesigns verwendet. Das Design von digital learning experiences umfasst drei Dimensionen: die Interaktion von Lernenden mit den Technologien (technische Dimension), die Interaktion mit anderen Lernenden und Lehrenden vermittelt durch die Lernplattformen (soziale Dimension), und die Interaktion der Lernenden mit didaktischen Elementen eingebettet in technischen Tools (pädagogische Dimension). Es werden Erkenntnisse und Implikationen für eine „Forschung zur Verbesserung“ (Research to Improve) versus „Forschung zum Beweis“ (Research to Prove) diskutiert.

[1] University of Technology Nuremberg, Learning Technologies Group, Dr.-Luise-Herzberg-Straße 4, 90461 Nürnberg, Germany, isa.jahnke@utn.de, https://orcid.org/0000-0001-6255-6416

doi:10.18420/delfi2025_01

Towards Hybrid Intelligence in CSCL

Armin Weinberger [1]

Hybride Intelligenz kombiniert idealerweise die Stärken menschlicher mit künstlicher Intelligenz [De21]. Dies ermöglicht ein Szenario computerunterstützten kooperativen Lernens (CSCL), in dem mehrere Lernende komplexe Aufgaben miteinander diskutieren und dabei von einem KI-basierten Agenten adaptiv unterstützt werden [Te16; Th24].

CSCL profitiert nachweislich erheblich von zusätzlicher Unterstützung von Kooperationsprozessen wie z. B. Kooperationsskripts [Vo16], Group Awareness Tools (GATs) [SB22] und KI-Agenten [Ar23]. Skripte spezifizieren herausfordernde Aufgaben und Aktivitäten (Feedforward), GATs geben Feedback zu Eigenschaften oder Aktivitäten der Lernenden. Diese Ansätze bauen auf Fähigkeiten zur Selbstregulierung auf, sind jedoch wenig anpassungsfähig und flexibel. KI-Agenten können Feedforward und Feedback miteinander verbinden und Lernprozesse adaptiv unterstützen. KI-Agenten fehlt aber Intentionalität, was die kritische Auseinandersetzung mit Lernenden und deren etwaigen Fehlkonzepten erschwert. Ein zentrales Problem des „AISCL“ (AI-Supported Collaborative Learning) ist vielmehr, Aufgabenlösungen vorwegzunehmen und dadurch die mentale Anstrengung Lernender zu reduzieren. Lernunterstützung sollte jedoch mitunter die Lernaufgabe in geeigneter Weise schwieriger machen [Re18].

Im Sinn hybrider Intelligenz soll hier daher ein Modell entwickelt werden, das Mensch-Mensch-Interaktion ideal mit Mensch-Maschinen-Interaktion ergänzt. Zentrale Komponenten dieses Modells sind neben der mentalen Lernanstrengung die psychologische Sicherheit sowie epistemische Wachsamkeit. Zum einen zeigen erste Studien, dass Lernende gegenüber KI-Agenten ein höheres Maß an psychologischer Sicherheit empfinden als gegenüber menschlichen Kooperationspartnern [AB15]. Dies deutet darauf hin, dass Lernende KI-Unterstützung in hohem Maß akzeptieren während Mensch-Mensch-Interaktion als herausfordernder wahrgenommen werden kann. Zum anderen kommt es in Systemen mit mehreren menschlichen und künstlichen Agenten darauf an, ein Maß an epistemischer Offenheit bzw. Wachsamkeit gegenüber den Beiträgen anderer Lernpartner sowie KI-Agenten aufzubringen, was insbesondere von transaktivem Wissen über die Interaktionspartner abhängt [No13].

Im Rahmen dieses Modells zeigen erste Studien deutlich, inwieweit KI-Agenten CSCL-Prozesse unterbrechen und kooperative Lernanstrengungen reduzieren können [Th24], aber auch auf höhere Akzeptanz stoßen können und CSCL-Prozesse lernwirksam unterstützen

[1] Saarland University, Germany,
a.weinberger@edutech.uni-saarland.de, https://orcid.org/0000-0002-5837-0819

doi:10.18420/delfi2025_02

können, wenn sie nur moderierend eingreifen, thematische Breite (vs. Tiefe) fördern und Erklärungen elizitieren anstatt sie zu geben [PWG25].

Literaturverzeichnis

[AB15] Asterhan, C. S. C.; Babichenko, M.: The social dimension of learning through argumentation: Effects of human presence and discourse style. Journal of Educational Psychology 107 (3), S. 740–755, 2015, DOI: 10.1037/edu0000014.

[Ar23] de Araujo, A. et al.: Supporting Collaborative Online Science Education With a Transferable and Configurable Conversational Agent. In: Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning - CSCL 2023. International Society of the Learning Sciences, S. 416–419, 2023.

[De21] Dellermann, D. et al.: The future of human-AI collaboration: a taxonomy of design knowledge for hybrid intelligence systems. 2021, arXiv: 2105.03354.

[No13] Noroozi, O. et al.: Scripting for construction of a transactive memory system in multidisciplinary CSCL environments. Learning and Instruction 25, S. 1–12, 2013, DOI: 10.1016/j.learninstruc.2012.10.002.

[PWG25] Pfannenstiel, W.; Weinberger, A.; Grajcevci, A.: Supporting Transactive Dialogue in Collaborative Learning Scenarios: When 'less' may be 'more' – The role of teaching and teachable GPT-based agents in CSCL, Symposium at EARLI 2025, Graz, 2025.

[Re18] Reiser, B. J.: Scaffolding Complex Learning: The Mechanisms of Structuring and Problematizing Student Work. In: The Journal of the Learning Sciences. Psychology Press, S. 273–304, 2018, DOI: 10.4324/9780203764411-2.

[SB22] Schnaubert, L.; Bodemer, D.: Group awareness and regulation in computer-supported collaborative learning. International Journal of Computer-Supported Collaborative Learning 17 (1), S. 11–38, 2022, DOI: 10.1007/s11412-022-09361-1.

[Te16] Tegos, S. et al.: Conversational agents for academically productive talk: a comparison of directed and undirected agent interventions. International Journal of Computer-Supported Collaborative Learning 11 (4), S. 417–440, 2016, DOI: 10.1007/s11412-016-9246-2.

[Th24] Thierfelder, B. et al.: Adverse effects of intelligent support of CSCL—the ethics of conversational agents, en, 2024, DOI: 10.22028/D291-46088.

[Vo16] Vogel, F. et al.: Socio-Cognitive Scaffolding with Computer-Supported Collaboration Scripts: a Meta-Analysis. Educational Psychology Review 29 (3), S. 477–511, 2016, DOI: 10.1007/s10648-016-9361-7.

Best-Paper-Kandidaten-Langbeiträge

Aligning Teaching Objectives with Formative Assessment to Uncover Student Learning Challenges

Imen Azaiz [1] and Sven Strickroth [1]

Abstract: In higher education, especially in introductory computer science courses, student diversity challenges instructors to design student-centered curricula that address students' actual learning needs. Although e-assessment and feedback systems play a critical role in supporting students' acquisition of programming skills by delivering formative feedback, they often fail to integrate explicit teaching objectives into their evaluation and analysis strategies. To address this gap, we propose a rubric-based approach that allows educators to systematically evaluate student performance in relation to the defined teaching objectives through tutor-provided formative feedback. Systematic alignment of learning objectives, teaching methods and assessment data helps educators identify students' difficulties, and reveals discrepancies between intended teaching objectives and actual assessment outcomes. This fosters a more student-centered environment, and offers instructors and students transparent insights into the achievement of teaching objectives.

Keywords: Competency Modeling, Constructive Alignment, E-Assessment, Formative Feedback, Instructional Design, Learning Analytics, Generative AI, Rubric for Assessment

1 Introduction

Summative assessment often relies on point-based or graded evaluations at the end of a course. This offers limited insight into students' areas of difficulty during the term, as well as their achievement of teaching objectives [BT11]. In contrast, formative assessment provides ongoing, qualitative feedback geared toward continual improvement. However, many e-assessment and feedback systems do not adequately support analysis and interpretation of these qualitative data sources. Specifically, links to measurable competency indicators and clearly defined teaching objectives are often missing [ADA22; Sc20]. Moreover, instructors and tutors lack time, capacity, or expertise to transform assessment data into actionable information [Sc20]. This limits their ability to provide targeted support [HV20]. Yet, such in-depth insights are crucial: they help educators refine their course design and identify the learning challenges students face. The terms teaching objectives and learning objectives are largely used interchangeably. In this paper, we use the term *teaching objectives*, which denotes the competencies and proficiency levels defined by the instructors (cf. [Ni08]). To support instructors in leveraging their teaching objectives, to design assignments that better shape their teaching and provide in-depth insight into student needs, we require a solution based on formative assessment and feedback aligned with these objectives. Rubrics

[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/delfi2025_03

(grading criteria) are an assessment tool that break down an assignment into its components and objectives, providing clear criteria for different levels of performance for each part (cf. [WSL23]. They have proven to be an effective assessment method, as they articulate performance expectations and align assignments with the intended teaching objectives [WSL23]. They enable transparent evaluation, efficient grading, and meaningful feedback that supports self-regulated learning [Ba15; WSL23]. Instructors found rubrics not only useful for grading assessments but also for conveying assessment expectations to students [Ma10]. Despite their potential for data-driven decision making and teaching improvement, rubrics are still underutilized – especially in e-assessment and feedback systems, where their systematic integration to support learning outcomes has yet to be realized. To fill this gap, our competence-oriented approach demonstrates how educational technologies can be used to systematically integrate rubrics into the formative assessment and feedback process.

There are two major shortcomings in current practice: (1) Teaching objectives are often only presented in the lecture and kept in mind by instructors, but they are rarely embedded in e-assessment systems or explicitly linked to student performance. (2) Instructors only have a comprehensive insight into student learning after final assessment, limiting timely formative interventions during the learning process and detailed competency level analysis.

These challenges highlight the need for a formative approach that systematically incorporates teaching objectives into assessment. This research aims to address the following research questions to enhance the formative assessment process with educational technologies:

- *RQ1*: How can teaching objectives be systematically used in formative assessment to evaluate student performance outcomes in an e-assessment and feedback system?
- *RQ2*: How can rubric-based assessment results be visualized to support educators in identifying student learning gaps and aligning instruction with teaching objectives?
- *RQ3*: How do educators perceive the proposed prototype for formative assessment and feedback?

To effectively support student-centered course design and curriculum development, assessment methods must go beyond outcome evaluation to actively reflect and reinforce teaching objectives. This requires instructional and assessment strategies that are closely aligned, enabling educators not only to measure what students achieve but also to monitor whether the course structure effectively fosters the intended learning outcomes [BT11].

2 Related Work

The Moodle learning management system provides a plug-in, named *differentiator* [Mo23], which helps instructors design learning objectives according to Bloom's taxonomy [Bl56]. However, these created learning objectives are only used to tag contents such as assignments, without adapting and integrating them in the assessment and/or feedback. One widely adopted approach in competency-based learning and assessment is the use of e-portfolios,

which supports reflective practice and the documentation of learning progress [SBD14; YWL17]. Gantikow et al. [Ga24] presented an AI-supported system for the automated assessment of student e-portfolios. Their method operationalizes predefined criteria into measurable indicators, such as the presence or absence of relevant topics and applying text metrics to estimate explanation quality. While their approach effectively visualizes content breadth and depth and assesses the concept model, it does neither explicitly assess the achievement of task-specific learning objectives, nor does it incorporate a structured competency model, which is a central focus of our work. Another example of alignment is presented by Kruse et al. [Kr24] who investigated the use of automated term extraction techniques to construct domain models from instructional materials. These models can later support the adaptation of learner models in personalized learning systems. Although this supports conceptual alignment between instructional content and domain knowledge, the method operates solely at the terminological level and does not incorporate learning objectives into the constructed domain model in a way that enables competency-oriented instructional design or assessment of student performance later on. Moreover, curriculum analysis tools such as *BuddyAnalytics* [GRS24] provide information on student progression and curriculum structure through study program analysis and cohort monitoring to support study curriculum designers. Although *BuddyAnalytics* does not consider course objectives or learning outcomes with student performance assessment. In contrast, our work focuses on competency-oriented assessment by aligning teaching objectives with student performance assessment to identify student difficulties.

Tools for instructional planning have shown that automated feedback can effectively support educators in structuring lessons, promoting reflection, and ensuring alignment with curriculum standards and didactic principles [SP17; St16]. These systems focus on the planning phase and provide value by guiding lesson design, yet they offer limited support for evaluating their planning based on student feedback or student performance assessment. To support computer science educators in the design of programming classes and assessments, Kiesler [Ki20] proposed a competency model based on Bloom's Revised Taxonomy [AKB00]. For the modeling of learning objectives in our approach, we adopt the extended competency model introduced by Kiesler and Pfülb [KP23], which systematically differentiates between technical, methodological, subject-specific, and social competencies. This model serves as the foundation for our analysis. Nevertheless, many courses still lack the tools necessary to assess how these modeled learning objectives are addressed and achieved [AS24].

Recent research by Pathak et al. [Pa25] also highlights the potential of question-specific rubrics with large language models enhancing logical assessment of student code in undergraduate computer science courses. Their analysis shows that such rubrics enable more precise feedback aligned with instructional goals, going beyond syntactic correctness.

3 Core Concept and Research Methodology

The primary goal of our approach is to provide a technological solution that enables educators to continuously monitor how their predefined teaching objectives are implemented and reflected in students' learning outcomes during the teaching process by leveraging e-assessment and feedback systems. This could help students acquire the intended competencies and minimize dropout.

The approach is based on three processes: *Design*, *Apply*, and *Evaluate*, whereas our work focuses on the latter two phases. The *Design* process includes tasks such as defining course goals and learning objectives, structuring learning objectives for topics and assignments, and instructional planning. This process typically takes place before the start of the course and is considered static, because the course goals will generally not be changed during the term. This phase results in the set of defined *teaching objectives*. This process is outside the scope of our research. The following two processes, *Apply* and *Evaluate*, are interwoven rather than sequential and are supported by educational technology: The *Apply* process encompasses the implementation of teaching and learning methods. Students work on assignments, while teaching staff observe, accompany, and provide formative feedback throughout the learning process. The *Evaluate* process involves gathering assessment results and feedback, checking the achievement of learning objectives, and identifying student difficulties. *Apply* and *Evaluate* together form a dynamic process that adapts to students' performance outcomes, ensuring alignment with the teaching objectives.

To guide the proposed approach, we applied the concept of constructive alignment [Bi03], which emphasizes aligning learning objectives, teaching methods, and assessment strategies. Our methodology focuses on three aspects:

- *Intended Learning Outcomes:* Integrating clearly defined teaching objectives into the e-assessment system to support systematic evaluation of student performance.
- *Methods and Feedback:* Utilizing annotated assignments and rubric-based formative feedback for competency-oriented assessment.
- *Teaching Analytics:* Analyzing e-assessment data by examining formative and summative assessment outcomes to systematically identify student difficulties and visualize alignment with the development of teaching objectives.

Our methodology follows a design-based research approach combining iterative design and research to gain new knowledge and improve practice. It involves cycles of developing a possible solution based on existing theories, evaluating it, and iteratively refining it with gained knowledge [Re22]. Therefore, we adopt the generic model for educational design-based research as proposed by McKenney and Reeves [MR12], which is particularly suitable for context-sensitive and iterative research in real educational technology-based settings. The aim of this approach is not only to solve a didactic issue, but to contribute to a broader understanding of how learning objectives, teaching methods, and assessment practices can be better aligned in higher education and supported by educational technologies.

Context: The study was conducted within the course *Introduction to Computer Science: Programming and Software Development* for computer science minors at LMU Munich, Germany (cf. [AS24] for course structure and content), across two cohorts (winter term 2023/24 and 2024/25). Each cohort comprised approximately 400 students who consented to the use of their anonymized data and submissions. The course consists of 15 lectures and 14/15 (first/second cohort) corresponding assignment sheets, where tasks are discussed in weekly tutorials. The student submissions were assessed semi-automatically by tutors combining personalized formative feedback and automated tests. All assignments and feedback were managed using the GATE e-assessment and feedback system [SOP11], which was introduced for the first time in this course in winter term 2023/24 [AS24]. The course was supported by two experienced lecturers (both involved in the design process and interviewed during both iterations), five student tutors per cohort (with varying levels of experience in providing feedback and familiarity with GATE), and one research assistant. In total, 65/74 individual tasks with 1672/2653 submissions were managed and evaluated.

Evaluation Design: Educational technology-based systems must be evaluated from the perspectives of all involved stakeholders – educators, tutors, and students – to ensure that the design, implementation, and learning impact are coherently aligned [BT11]. To evaluate the iterative design process, a mixed-methods approach was adopted across two iterations, each representing a separate winter term. We combined Likert-scale surveys and open-ended questions with tutors and students, qualitative expert interviews with instructors, group discussions with tutors and instructors, and a quantitative analysis of e-assessment data to gain a multi-perspective understanding of the implementation. All open-ended survey responses and interview transcripts were analyzed using qualitative content analysis based on Mayring's inductive category development method [Ma01]. This multi-perspective evaluation aligns with the core processes of design, apply, and evaluate. From the *instructor perspective*, the focus lies on the use of teaching objectives to continuously monitor their achievement and alignment with student performance assessment outcomes. From the *tutor perspective*, the formative assessment method based on rubric and reusable feedback phrasing was assessed with regard to its ability to improve feedback quality, provide diagnostic support, and to enable consistent assessment practices. From the *student perspective*, the focus was placed on how learning-objective-aligned feedback was perceived in terms of clarity, usefulness for self-reflection, and progress tracking. This triadic evaluation supports the continuous alignment of teaching objectives, assessment methods, and actual learning outcomes.

4 Evaluation and Results of the Iterative Design Process

The concept delineated in section 3 is implemented over the course of two distinct terms, each term presenting a successive iteration of design and evaluation.

4.1 First iteration: Design, Implementation and Evaluation

Design: The developed prototype was implemented as an extension to the existing GATE system. The newly introduced analysis unit was added into GATE's backend to evaluate annotated assignments with respect to defined teaching objectives. The annotation scheme introduced in [AS24] was adopted for mapping submissions to specific learning objectives. To assess the achievement of these objectives, student learning outcomes are assessed based on qualitative criteria (*rubric*) aligned with the teaching objectives, thereby maintaining a coherent alignment between teaching objectives, assessment criteria, and teaching methods [Bi03]. The rubrics are structured according to the defined teaching objectives and the tutors marked the student submissions as *fulfilled*, *partially fulfilled* or *not fulfilled*. The results were visible to students accompanied by formative tutor-feedback.

Evaluation: Building on the mixed-method, multi-perspective evaluation design described in section 3, we evaluated the approach and didactic value from the perspectives of tutors, students, and instructors.

Tutor perspective: The tutor survey included closed items capturing satisfaction, ease of use, clarity of feedback for different task types (PDF uploads, multiple-choice, code with function tests), and the perceived difficulty of assessing learning objectives using a 5-point Likert scale (1 = negative to 5 = positive). Open-ended questions explored challenges encountered during correction, suggestions for improvement, and detailed experiences with rubric-based evaluation and automated assessment practices. The results from the first iteration showed that the rubric-based assessment of teaching objectives was perceived as rather difficult (mean $\overline{x} = 3.6/5$), while the automated evaluation of multiple-choice tasks was rated as easy ($\overline{x} = 4.8$) and well-supported by automatic functional tests ($\overline{x} = 4.2$). The subsequent qualitative analysis revealed five main categories. The first category *rubric-based learning objective evaluation* was mentioned by all five tutors. Subcategories were defined to describe how tutors experienced the use of rubric-based assessment: subcategory (1) *clarity of learning objectives* refers to how precisely and measurably the learning objectives were formulated. Three out of five tutors found the objectives overly vague or broad, such as "The learning objectives were too simple or general". This lack of specificity hindered clear assessment of student achievement. Subcategory (2) *granularity of the rating scale* addresses structural limitation of the three-level rubric ("*fulfilled*", "*partially fulfilled*", "*not fulfilled*"). Two tutors argued that the current system lacks nuance to differentiate performances adequately and proposed two additional intermediate scores. Subcategory (3) *scoring transparency* highlights ambiguities in the clarity of the criteria used to assign each rubric level. Two tutors expressed uncertainty when deciding between levels, citing unclear definitions. Hence, clear guidelines on how to apply the rubric are needed. Together, these subcategories underscore the key practical challenges in ensuring consistent, fair, and objective-aligned rubric-based evaluations. The second category, *function test issues*, emerged in four responses: three tutors reported false negatives in valid solutions, while one highlighted a mismatch between test outcomes and learning objectives. For *usability*

and interface (third category), three tutors critiqued the cumbersome switching between feedback view and PDFs, advocating for direct PDF annotation tools to streamline their workflow. *Organizational aspects* were mentioned once and concerned the lack of automated or consistent submission assignment to tutors. Notably, all five tutors contributed to the *suggestions for improvement* category. Key requests included embedding model solutions, expanding the scoring scale, and to explicit feedback criteria (each twice).

Student perspective: Students who were enrolled in the course were surveyed, but there were only very few participants, so the results may be unrepresentative. In the survey students were asked to rate the clarity and usefulness of learning objectives in the context of programming exercises. Specifically, they responded to the following statements: (1) "The learning objectives on the lecture slides and exercises in GATE were clear and understandable" with 6 out of 8 students rating it positively and one strongly disagreeing, and (2) "The learning objectives helped me understand the meaning of learning content" with all strongly agreeing. Statement (3) "The learning objectives in the exercises helped me understand the requirements better", and (4) "The learning objectives integrated into the feedback helped me assess my learning progress" with all students strongly agreeing and one student strongly disagreeing. These findings suggest that learning objectives should be consistently maintained across lecture slides, GATE exercises, the e-assessment system, and feedback. Their presence provides students with a structured overview and supports the continuous tracking of their own learning goals. As the results indicate, students perceived the learning objectives as clear and helpful for understanding task requirements, and valuable for reflecting on their individual learning progress.

Figure 1 illustrates how teaching objectives were assessed by tutors. Assignments 5 to 14 had mostly achieved good coverage and achievement of the intended learning objectives. Figure 1 serves as a visualization of the rubric based assessment in response to *RQ2*. While the rubric categories *fulfilled*, *partially fulfilled*, and *not fulfilled* provide a broad indication of learning objective attainment, they do not reveal the underlying causes of unmet objectives on the student side. In particular, *partially fulfilled* remains a subjective classification dependent on individual tutor judgment, which may impact the reliability of evaluation.

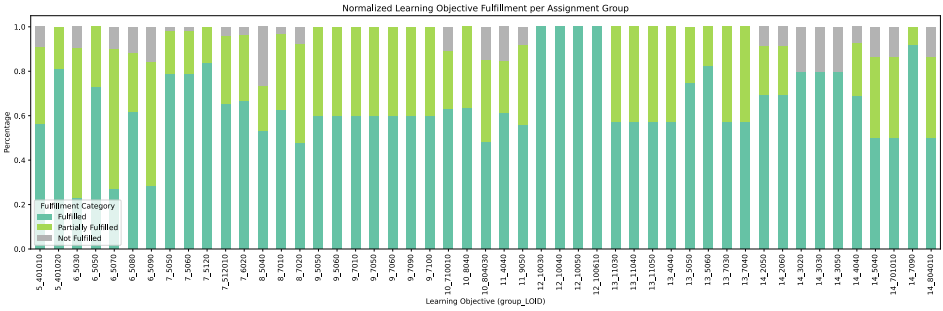


Fig. 1: Fulfillment of Teaching Objectives

4.2 Second Iteration: Redesign, Implementation, and Final Evaluation

Design: Based on the evaluation results from the first iteration, the rubric-based assessment framework was substantially revised to better align with intended learning outcomes and improve the overall assessment process. Tutors had reported challenges related to vague learning objectives and insufficient differentiated three-level scoring scheme. In response, structured rubrics were introduced that decompose each learning objective into clearly defined sub-goals and partial aspects contributing to the achievement of a teaching objective and the successful task completion. These rubrics served as operational tools, helping tutors assess the extent to which a student's submission fulfilled the core assignment requirements. To promote clarity and consistency, each rubric item was accompanied by a detailed description and example feedback snippets. These reusable and adaptable feedback snippets aim to support tutors in providing targeted, constructive feedback aligned with common student errors identified in earlier iterations and anticipated in advance. Frequently occurring mistakes were categorized and embedded into the rubric structure, facilitating diagnostic and formative assessment. Each rubric item was implemented as a checkbox in the e-assessment and feedback system (GATE), offering tutors a transparent overview of relevant criteria and facilitating consistent evaluations. Simultaneously, students could use the rubrics as a checklist to identify their mistakes and reflect on their understanding. All rubrics underwent a double-review process by educators prior to integration to ensure alignment with the assignment goals. As part of the formative assessment design, tutors received weekly correction guidelines tailored to each assignment. These guidelines included predefined error categories, learning-objective-aligned rubrics, and reusable feedback phrases. This support structure was implemented among five tutors and evaluated during the second iteration to support consistent, clear, and pedagogically meaningful feedback.

Assessment Results Overview: Tab. 1 summarizes key performance trends. Significant improvement was observed in object-oriented programming. Data structures showed consistently strong results. Logical structures and code tracing had low error rates, whereas operator precedence, memory representation, and complexity proved more challenging. Syntax errors no longer occurred from the exercise sheet of the tenth week onward. Assessment outcomes were visualized as error counts (misses) per rubric item within the system, providing a structured overview of student learning gaps, thus directly supporting the analysis in response to *RQ2*. Examples of these rubrics can be found in the Table 1.

Tutor perspective: A survey among all tutors (n = 5) was conducted followed by a group discussion. A quantitative analysis of the closed questions revealed a strong agreement on the pedagogical relevance and practical utility of the approach. All participants strongly agreed that individualized feedback is essential for the learning process ($\overline{x} = 5$). Furthermore, the correction aid, comprising structured error categories, checkboxes, and reusable feedback snippets were positively rated. Their contribution to a more systematic correction process ($\overline{x} = 4.4$), the clarity of error categories ($\overline{x} = 4.4$) were especially valued. Tutors also acknowledged that the subdivision of error types supported quicker recognition of common

Rubric Category	Initial	Follow-up	Trend and Criterion Example
EBNF Derivation	2	A	Remains weak; e. g., "Derivation steps are correct"
Operator Precedence	3	B	Remains problematic; e. g., `bday - tday / 30` correctly evaluated
Heap/Stack Representation	4	8	No improvement; e. g., heap after lines 1–11 fully and correctly drawn
Classes & Methods OOP	6	7	Strong improvement; e. g., methods consistently implemented
Data Structures	4	D	Consistently strong; e. g., `event_planung()` correctly implemented
Complexity (Time/Space)	3	6, 10	Still difficult; e. g., space complexity part b correctly explained

Tab. 1: Performance trends across tasks: Numbers refer to task groups 2-10, letters to specific midterm tasks A-D. The column "Follow-up Task(s)" indicates where the same teaching objectives were reassessed to monitor progress.

mistakes ($\overline{x}$ = 4.2), and that the rubric-based approach helped them provide more structured ($\overline{x}$ = 4.0), specific ($\overline{x}$ = 3.8), and explanatory feedback ($\overline{x}$ = 3.6) through the tool. The tool was neither perceived as an increasing workload ($\overline{x}$ = 2.4) nor as irrelevant ($\overline{x}$ = 1.6). In addition, a qualitative content analysis was performed on open responses. Three central categories emerged: (1) *Rubrics as orientation tools for feedback* (n=5): All respondents highlighted the clarity and helpful structure of the correction aid and emphasized that clearly structured rubrics aligned with learning objectives improved the transparency and coherence of the assessment. They reported that the error categories provided orientation during assessment and helped standardize tutor feedback. One tutor said: "The predefined categories gave me a clearer view of what to look for. (2) *Support for feedback quality* (n=4): the predefined feedback snippets were highlighted as an efficient way to ensure clarity and motivation in tutor responses. These snippets allowed tutors to focus on diagnosis and support rather than formulation effort, strengthening the formative function of the feedback. (3) *Integration with formative assessment principles* (n=5): tutors described that the rubric-based approach allowed them to identify common misunderstandings, address them with targeted guidance, and can support self-reflection among students. Several tutors appreciated the balance between standardization and flexibility and adapted the predefined feedback snippets. While generally positive, some tutors suggested further refinement of category descriptions for the assessment and the addition of more differentiated snippets for advanced cases. One participant commented that he rarely used the snippets because of his long experience on the course, preferring to formulate his own feedback; nevertheless, he used the feedback phrasing as a guide and orientation.

Student perspective: To assess the quality of the formative assessment and feedback mechanism used within the e-assessment and feedback system from the student perspective, an evaluation was conducted at the end of the course. The evaluation consisted of 9 closed

and open-ended questions in total, which focused on the effectiveness and clarity of the feedback, as well as the structure and usefulness of the rubric-based evaluation. Five closed questions were presented using a five point Likert-scale (Strongly disagree, Rather disagree, Rather agree, Strongly agree, and Cannot judge) to collect targeted quantitative assessments. Only n=9 participants took part in the survey. Six participants agreed that the tutor feedback was detailed, understandable, and referred to specific areas of their work, while three rather disagreed. Constructive suggestions for improvement were made by six participants; two rather disagreed and one strongly disagreed. Positive and motivating comments were acknowledged by all participants. The checklists were perceived as helpful for identifying gaps and working on them in a targeted way by four participants, while two rather disagreed and two could not judge. The automated tests in GATE were rated as helpful by seven participants, with two expressing slight disagreement. The suggestion to receive a performance overview was rated as helpful by six participants, while three responded neutrally. No negative responses were given. In the open-ended question on what participants particularly appreciated about the tutor feedback with checklist, several respondents highlighted the checklists themselves as especially valuable. It was seen as helpful for identifying what worked well and what still needed improvement, providing a clear structure for self-correction, and offering a quick overview of one's own performance. Participants also appreciated being able to see concrete evaluation details, such as syntax and variable naming. Concrete explanations of mistakes were positively mentioned as well. In addition, automated tests and interim feedback were perceived as practical and supportive. In the open-ended question on potential improvements to the tutor feedback, participants wished for more detailed feedback, especially when their solutions were incorrect. The checklists were described as clear but did not show how to proceed. Some missed explanations on what exactly was wrong and why, instead of general references to the sample solution. Short or unclear tutor comments (e. g., "does not run in Thonny [IDE]" were criticized. One participant suggested more precise feedback, particularly for complex concepts, clarifying why something was incorrect, such as in "O-notation" topic. One participant emphasized that online feedback is always anonymous and that feedback in person is more effective, though the "GATE system approach" was considered successful for large student groups.

Instructor perspective: To address the third research question (RQ3), how is the tool perceived by the course instructors, qualitative expert interviews were conducted with both instructors. The instructors evaluated the prototype's feature of linking their teaching objectives to assignments as highly beneficial. They highlighted its potential for systematically tracking objective coverage across exercises and verifying alignment during exam preparation. This connection, according to them, supports re-usability of tasks, especially when objectives recur across chapters, and allows them to monitor which goals have been addressed or require reinforcement. One key insight was the ability to map student performance directly to learning objectives, which enables more precise identification of difficulties, such as persistent weaknesses in loop-related tasks. One instructor suggests that students could also benefit by selecting tasks filtered by specific objectives, supporting targeted practice. Based on their own teaching experience, the two experts found that explicitly formulated objectives

were well received by students, increased transparency, and provided a helpful reference point in discussions. However, they also noted the importance of consistency in formulation and assignment, as partial or unclear mappings may limit the effectiveness of the approach. The interview emphasized that feedback templates and rubrics were a crucial support for tutors, particularly those with limited programming experience. Predefined examples of common errors and offering strategies helped students understand where and how to improve, thus fostering formative learning. One instructor reported that with around 80 % of the feedback predefined, tutors only needed to add about 20 % manually – enabling personalized, high-quality feedback with minimal effort. The error categorization was recognized as resembling established *comment categorization* used in professional code review, and was seen as key to ensuring clarity and consistency in feedback. One of the instructors stressed the importance of constructive, motivating feedback that highlights both strengths and weaknesses. Given that tutors' feedback quality and tone can vary, especially with high turnover and limited training, standardized templates were seen as essential for ensuring consistent and supportive feedback. Templates help to structure feedback, reduce effort, and compensate for a lack of experience ("a consistent tone and clear guidance are hard to achieve without them"). On expert emphasized that educators must be supported in defining measurable teaching objectives. Templates for learning objectives – adjustable but consistently applied – would help ensure alignment across courses. A fine-grained structuring of objectives is essential, though detailed rubrics are especially valuable for tutors and students, who engage directly with individual performances. For instructors, aggregated overviews at topic or group level in dashboards are seen as sufficient. Interest was also expressed in exploring cross-course correlations of objectives, identifying student difficulties early, and enabling preventive measures against dropout. One instructor argued that if our approach would be widely adopted, such systems could support curriculum-wide analyses, uncover dependencies between courses, and contribute to a more objective, student-centered quality enhancement in teaching through technological support. One exptert highlighted that the insights from the assessment dashboard were particularly valuable, as they "remarkably mirrored our summative exam results". It was seen as compelling that the patterns observed aligned so closely with end-of-semester performance, offering a meaningful basis for course reflection and design. In contrast, commonly used in-class tools such as audience response systems or hand-raising were seen as offering only a rough indication of conceptual understanding. The visualization of the assessment analysis, however, enables a fine-grained analysis of applied competencies, which those classroom tools cannot capture. Both approaches were described as complementary and helpful, though the dashboard insights were considered "far more precise and actionable".

5 Discussion

The approach was evaluated from three perspectives – students, tutors, and instructors – highlighting its broad applicability. Tutors, particularly those with limited experience, reported that rubrics and predefined feedback provided valuable orientation, increased

consistency, and reduced effort. These findings confirm the positive impact of predefined feedback snippets [SH22] and underline the dual function of rubrics as both assessment instruments and pedagogical scaffolds. Students perceived the feedback as motivational and useful for self-regulated learning. Challenges regarding the granularity and operational clarity of teaching objectives were addressed in the second iteration through fine-grained rubrics and categorized error types, reinforcing the need for ongoing system refinement. Continuous tutor involvement in rubric development is critical to maintain alignment and to foster long-term adoption. The findings suggest that such tools not only support feedback quality but also tutor professionalization and sustainable assessment practice. The alignment between rubric-based feedback and summative exam outcomes, as noted by instructors, further confirms the diagnostic accuracy of the assessment system and its potential to inform both ongoing and final evaluations. Moreover, the analysis of e-assessment data based on teaching objectives revealed patterns in student learning behavior, which promoted self-reflection for both students and educators. While topics like object-oriented programming improved with formative feedback and follow-up tasks, others (EBNF derivation, heap/stack representation, and complexity) remained persistently challenging. This aligns with research on cognitive load in computer education [Tu00] suggesting that abstract concepts may require additional visual and interactive scaffolding or peer review/discussion. These insights emphasize the potential of teaching analytics to enhance transparency, support formative learning processes, and inform pedagogical decisions during the course. Although both instructors were experienced, only one felt confident in defining fine-grained and measurable teaching objectives, highlighting the need for tools to support such tasks in competency-based teaching, possibly supported by GenAI.

6 Summary and outlook

This paper presents a novel approach that systematically aligns teaching objectives with formative assessment and feedback in programming education. By embedding rubric-based student performance assessment and predefined, adjustable feedback components into an e-assessment and feedback system, the approach enables competency-oriented, constructive, and consistent feedback, while supporting the identification of student learning challenges. The visualization of the assessment analysis enabled instructors to gain deeper insights into student performance, reflect on their course design, and offer targeted support throughout the semester.

Future developments should include editable teaching objective templates to help instructors define measurable goals, and explore AI-generated rubrics based on these objectives to streamline assessment design and reduce manual effort. Additionally, the system should support cross-course analysis to detect structural learning difficulties and enable preventive measures such as dropout risk detection. Lastly, teaching analytics should be leveraged to foster curriculum coherence and student-centered teaching, and to derive data-driven recommendations for adaptive instructional strategies and targeted learning interventions.

Acknowledgements: The authors thank Prof. Marie-Christine Jakobs and Dr. Philipp Wendler for their valuable support, and the tutors for their assistance. 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

[ADA22] Assami, S.; Daoudi, N.; Ajhoun, R.: Implementation of a Machine Learning-Based MOOC Recommender System Using Learner Motivation Prediction. International Journal of Engineering Pedagogy (iJEP) 12 (5), pp. 68–85, 2022, DOI: 10.3991/ijep.v12i5.30523.

[AKB00] Anderson, L.; Krathwohl, D.; Bloom, B.: A Taxonomy for Learning, Teaching, and Assessing: a Revision of Bloom's Taxonomy of Educational Objectives. http://lst-iiep.iiep-unesco.org/cgi-bin/wwwi32.exe/[in=epidoc1.in]/?t2000=015556/(100), 2000.

[AS24] Azaiz, I.; Strickroth, S.: Supporting Constructive Alignment: Visualization of the Alignment of Teaching Goals and Developed Assignments, Proceedings of DELFI 2024, 2024, DOI: 10.18420/delfi2024_34.

[Ba15] Brookhart, S. M.; and, F. C.: The quality and effectiveness of descriptive rubrics. Educational Review 67 (3), pp. 343–368, 2015, DOI: 10.1080/00131911.2014.929565.

[Bi03] Biggs, J.: Aligning Teaching for Constructing Learning. The Higher Education Academie, 2003, https://www.researchgate.net/publication/255583992_Aligning_Teaching_for_Constructing_Learning.

[Bl56] Bloom, B. S. et al.: Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain. David McKay Company, New York, 1956.

[BT11] Biggs, J.; Tang, C.: Teaching For Quality Learning At University. McGraw-Hill Education, 2011.

[Ga24] Gantikow, A. et al.: KI-basierte Analyse von E-Portfolios, Proceedings of DELFI 2024, 2024, DOI: 10.18420/delfi2024_31.

[GRS24] Görzen, S.; Röpke, R.; Schroeder, U.: BuddyAnalytics: A dashboard and reporting tool for study program analysis and student cohort monitoring, Proceedings of DELFI 2024, 2024, DOI: 10.18420/delfi2024_53.

[HV20] Hoyos, A. A. C.; Velásquez, J. D.: Teaching Analytics: Current Challenges and Future Development. IEEE Revista Iberoamericana de Tecnologias del Aprendizaje 15 (1), pp. 1–9, 2020, DOI: 10.1109/RITA.2020.2979245.

[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, Association for Computing Machinery, Trondheim, Norway, pp. 459–465, 2020, DOI: 10.1145/3341525.3387419.

[KP23] Kiesler, N.; Pfülb, B.: Higher Education Programming Competencies: A Novel Dataset. In (Iliadis, L. et al., eds.): Artificial Neural Networks and Machine Learning – ICANN 2023. Springer Nature Switzerland, Cham, pp. 319–330, 2023.

[Kr24] Kruse, T. et al.: Term Extraction for Domain Modeling, Proceedings of DELFI 2024, 2024, DOI: 10.18420/delfi2024_33.

[Ma01] Mayring, P.: Combination and Integration of Qualitative and Quantitative Analysis. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research 2 (1), 2001, DOI: 10.17169/FQS-2.1.967.

[Ma10] Maxwell, S.: Using Rubrics to Support Graded Assessment in a Competency-based Environment. National Centre for Vocational Education Research, Acelaide, 2010.

[Mo23] Moodle contributors: Differentiator Plugin, A Moodle plugin to support teachers in formulating learning objectives based on Bloom's taxonomy, 2023, https://moodle.org/plugins/local_differentiator, accessed: 04/08/2025.

[MR12] McKenney, S.; Reeves, T. C.: Conducting educational design research. Routledge, London, England, 2012.

[Ni08] Niegemann, H. M.: Kompendium Multimediales Lernen. Springer, Berlin, Germany, 2008.

[Pa25] Pathak, A. et al.: Rubric Is All You Need: Enhancing LLM-based Code Evaluation With Question-Specific Rubrics, 2025, arXiv: 2503.23989 [cs.SE].

[Re22] Reinmann, G.: Was macht Design-Based Research zu Forschung? Die Debatte um Standards und die vernachlässigte Rolle des Designs: Die Debatte um Standards und die vernachlässigte Rolle des Designs. EDeR. Educational Design Research 6 (2), 2022, DOI: 10.15460/eder.6.2.1909.

[SBD14] Struyven, K.; Blieck, Y.; De Roeck, V.: The electronic portfolio as a tool to develop and assess pre-service student teaching competences: Challenges for quality. Studies in Educational Evaluation 43, pp. 40–54, 2014, DOI: 10.1016/j.stueduc.2014.06.001.

[Sc20] Schildkamp, K. et al.: Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. International Journal of Educational Research 103, p. 101602, 2020, DOI: 10.1016/j.ijer.2020.101602.

[SH22] Strickroth, S.; Holzinger, F.: Supporting the Semi-Automatic Feedback Provisioning on Programming Assignments. In (Temperini, M. et al., eds.): Methodologies and Intelligent Systems for Technology Enhanced Learning, 12th International Conference. Springer International Publishing, Cham, pp. 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., eds.): Tagungsband der 9. e-Learning Fachtagung Informatik (DeLFI). Gesellschaft für Informatik e.V., Bonn, Germany, pp. 115–126, 2011, https://dl.gi.de/handle/20.500.12116/4740.

[SP17] Strickroth, S.; Pinkwart, N.: Planung von Schulunterricht: Automatisches Feedback zur Reflexionsanregung über eigene Unterrichtsentwürfe, Bildungsräume 2017, 2017.

[St16] Strickroth, S.: Unterstützungsmöglichkeiten für die computerbasierte Planung von Unterricht – ein graphischer, zeitbasierter Ansatz mit automatischem Feedback, PhD Thesis, Humboldt-Universität zu Berlin, 2016, DOI: 10.18452/17613.

[Tu00] Tuovinen, J. E.: Optimising student cognitive load in computer education. In: Proceedings of the Australasian Conference on Computing Education. ACSE '00, Association for Computing Machinery, Melbourne, Australia, pp. 235–241, 2000, DOI: 10.1145/359369.359405.

[WSL23] Walvoord, B. E.; Stevens, D. D.; Levi, A. J.: Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback, and Promote Student Learning. Routledge, 2023.

[YWL17] Yang, M.; Wang, T.; Lim, C. P.: E-Portfolios as Digital Assessment Tools in Higher Education. In (Spector, M. J.; Lockee, B. B.; Childress, M. D., eds.): Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy. Springer International Publishing, Cham, pp. 1–23, 2017, DOI: 10.1007/978-3-319-17727-4_83-1.

Prior virtual reality experience and its determinants in educational settings

Insights from secondary data analysis of three previous studies from 2021 to 2025

Jonathan Dyrna [1], Richard Joos[2], Marilisa Herchet[3] and Thomas Köhler[4]

Abstract: The effectiveness of virtual reality (VR) to enhance learning has been repeatedly demonstrated empirically. However, the technology is still rarely adopted in educational practice – possibly due to a lack of competencies and operational experience among instructional participants. Since little is known about this issue, a secondary data analysis of three studies ($N = 220$) conducted between 2021 and 2025 was performed to examine the current level of experience with VR technology among instructional participants along potential determinants. The sample included instructors, facilitators, learners, and participants of VR-based trainings and scientific workshops, of which almost half were from the real estate sector. Results indicate that, on average, instructional participants had limited prior VR experience. Its level was significantly associated with their trait technology acceptance and literacy beliefs, but not with age, gender, control beliefs, or year of survey. These findings underscore the need to foster technology readiness among instructional participants to support the adoption of innovative technologies like VR.

Keywords: Virtual Reality, VR, education, learning, boundary conditions, individual differences, prior VR experience, determinants, technology readiness, sociodemographics

1 Introduction

The effectiveness of using virtual reality (VR) to enrich educational settings like classroom teaching and vocational training has been empirically proven in numerous studies and meta-analyses (e. g., [CBG22; HGJ21; Pa21; Qu24; WYG20]). For instance, a recent meta-analysis of 15 randomized controlled trials or quasi-experimental studies indicates that the use of VR exerts a large effect on student engagement in learning, in particular with regard to procedural knowledge acquisition [Ch24]. According to a further meta-analysis of 43 studies, the application of VR technologies in learning settings leads to greater interest in the learning object, higher motivation, presence and satisfaction, and fosters competence acquisition [Yu21]. In spite of these favorable empirical findings and the growing emergence of solutions to current challenges like demands on learners limited

[1] Technische Universität (TU) Dresden, Center of Interdisciplinary Digital Sciences (CIDS), Strehlener Straße 22/24, 01069 Dresden, jonathan.dyrna@tu-dresden.de, https://orcid.org/0000-0002-3718-0645
[2] Technische Dresden, CIDS, Strehlener Straße 22/24, 01069 Dresden, richard.joos@tu-dresden.de
[3] Technische Dresden, CIDS, Strehlener Straße 22/24, 01069 Dresden, marilisa.herchet@tu-dresden.de
[4] Technische Dresden, CIDS, Strehlener Straße 22/24, 01069 Dresden, thomas.koehler@tu-dresden.de

doi:10.18420/delfi2025_04

cognitive capacity and VR sickness through appropriate design measures (e. g., [MMP23; Sa24]), the technology is still scarcely used in educational practice [Fi21]. This is particularly surprising given that, at the same time, VR hardware is increasing in performance and affordability, the number of technical platforms for VR is growing, a wide range of development tools are available [Al24], and revenues for virtual reality technology and content have been rising for years and are forecast to continue climbing in future [AR24]. Research models concerning the implementation of VR in education suggest that the latter is inhibited by further factors like the stakeholders' states and traits. For instance, teachers' and learners' competencies gained from prior experience with VR technologies and applications is likely to play a role in this regard, but we know little about how it is developed among stakeholders. In the present article, we initially refer to viable frameworks for successful implementation of VR in (formal) education. We then review current empirical findings with regard to impact factors on the intention to use VR to derive sound hypotheses about the influence of potential determinants on present VR experience. Subsequently, we report results of a secondary data analysis that addressed the question of how familiar instructional participants – for example teachers, instructors and facilitators of learning, as well as learners and participants of VR-based lessons, trainings, and workshops – in Germany currently are with VR technologies and what individual differences determine it. Finally, we discuss our findings concerning limitations and implications for future research and practice.

2 Theoretical Framework

2.1 Frameworks for successful implementation of VR in education

When talking about VR, we refer to entirely computer-generated worlds that engage as many human sensory modalities as possible and can thus be perceived by users as an illusion of a physical environment as well as (co-)designed and modified with the help of input devices to provide an interactive and immersive experience [Ni06]. VR distinguishes from reality-enhancing technologies like augmented and extended reality (see reality-virtuality continuum by [MK94]) due to the full computer simulation of its representation, and can be depicted on two-dimensional (2D) displays (called 'desktop VR') or head-mounted displays (called 'immersive VR'). Several frameworks can be considered to obtain insights regarding factors that determine the successful implementation of VR technologies in education.

One example is the Framework for the Implementation of VR in Formal Education (FIVFE; [Fi21]), that builds on the work of Euler and Seufert [ES07] regarding the sustainable implementation of technology enhanced learning innovation in higher education (see Fig. 1). In a nutshell, it proposes five areas of action for educational managers to take and apply purposeful decisions to ensure sustainable implementation: didactics (e. g., development of sound didactic concepts for the use of VR), technology (e. g.,

procurement of appropriate VR hardware), economics (e. g., ensuring an appropriate cost-benefit ratio of VR enhanced education programs), organization (embedding into and possible expand of institutional structures) and socio-culture (e. g., measures to involve all stakeholders in implementation activities to increase acceptance and use of VR). As illustrated by the examples above, the development of implementation strategies takes place within the scope of three parameters, whose conditions must be taken into account: the education system with its regulations for structures and processes, the implementing institution along its present facilities, strategies, cultural values and support offers, and the involved individuals (i. e., stakeholders) who possess certain experiences and expectations [Fi21].

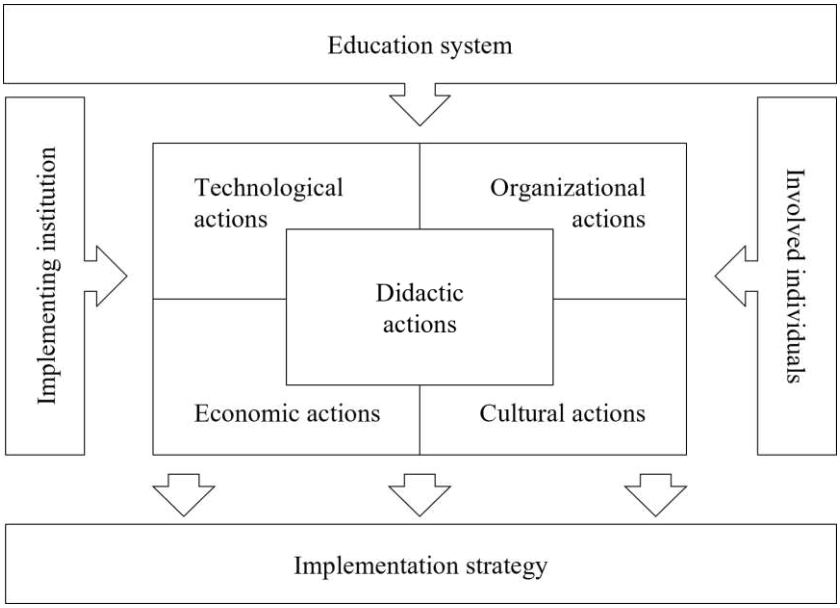


Fig. 1: The Framework for the Implementation of VR in Formal Education (Fischer et al., 2021)

More specific indications can be derived from socio-psychological models to predict the acceptance and use of emerging technologies, in particular the Technology Acceptance Model in its third edition (TAM3; [VB08]) and the Unified Theory of Acceptance and Use of Technology (UTAUT; [Ve03]). Briefly summarized, they state that individual expectations (like output quality) and characteristics (like computer self-efficacy), social and organizational influences, and technology-related factors (like objective usability and facilitating conditions), as well as moderating factors predict how useful and easy-to-use users perceive a technology and to what extent they (relatedly) will use it. Besides age, gender and the voluntariness of use, the moderating variables also encompass the users' prior experience with the respective technologies. More specifically, the application of the TAM to the context of e-learning, called the General Extended Technology Acceptance

Model for E-Learning (GETAMEL; [AW16]), reveals that learners' experience significantly predicts both the perceived usefulness and the perceived ease of use of learning technologies to forecast their prospective use. In order to determine how VR technologies can be better integrated into future educational practice, it consequently seems appropriate to take a closer look at teachers' and learners' previous experience with this technology and to identify specific factors that affect it.

2.2 Factors to determine individuals' prior VR experience

Prior experience with VR refers to individuals' previous exposure to and interaction with VR systems and applications depicted by the frequency and duration of its usage, familiarity with VR interfaces, and the types of activities performed within VR environments. It is to be distinguished from prior knowledge as the sum of all subject-related information on the learning topic already known by the learner (see also [Mu23]). Although according to the frameworks outlined above, its importance for the successful implementation of VR technologies in education seems to be crucial, there has been little research to date on the contemporary extent and determinants of teachers' and learners' experiences with these technologies. We therefore aim to address the research question (RQ) of how prior experience with VR is currently developed among instructional participants and which factors drive it.

Due to a lack of research with particular respect to prior VR experience, we build our hypotheses on more fundamental theories and empirical results on how individual factors determine the use of new technologies, as their degree of use by instructional participants is indicative of the level of present experience. One of the few studies in this area, surveying a total of 94 junior and senior students majoring in Business Administration, indicates that acceptance and use of VR technologies are significantly higher among males than females [LD22]. In contrast, a further study with 58 preservice teachers in rural China found no influence of gender on the acceptance of a specific VR technology after use [LLC23]. In addition, age only exerted a significant effect until additional impact factors like prior technology acceptance or presence experience were included into the model. This contradicts the results of a survey of 283 participants, indicating that younger age is linked to a higher perceived usefulness, which in turn leads to a more positive attitude towards using and intention to use VR hardware [MC19]. Furthermore, several studies demonstrate a positive impact of self-efficacy (with computers or IT) on the acceptance and intention to use VR technologies, suggesting that competency beliefs also represent a key predictor of VR experience [Al23; CH12; LD22]. In addition, two studies with 120 and 300 Korean second and third graders indicate that VR technology acceptance is higher for learners with lower levels of computer anxiety, thus highlighting the role of control beliefs in this regard [CPS23]. Last but not least, it is likely that the proliferation of VR hardware and applications in recent years [Al24] has resulted in instructional participants increasingly being exposed to VR in their professional or private activities and consequently possessing greater prior experience with VR than a few years ago.

From these assumptions and results, which are partially inconsistent, possibly due to varying sample characteristics, we derive the following hypotheses (H) for our study:

- H1: The younger instructional participants are, the greater their prior experience with VR.
- H2: Male instructional participants possess higher prior VR experience than female instructional participants.
- H3: The more recent the year of the survey, the greater the reported prior VR experience of instructional participants.
- H4: The higher instructional participants' overall technology acceptance, the greater their prior VR experience.
- H5: The higher instructional participants' technology competency beliefs, the greater their prior VR experience.
- H6: The higher instructional participants' technology control beliefs, the greater their prior VR experience.

To examine our research question and hypotheses, we conducted a secondary data analysis of three studies on the perception and effectiveness of VR-supported educational programs. The method and results of the analysis are described below.

3 Method

3.1 Research design

In our previous studies, we measured socio-demographic variables like age, gender, technology readiness and previous experience with VR according to common practice to precisely describe our samples and take possible covariates into account. For sustainability reasons, we therefore decided to perform a secondary data analysis as a basis for examining our research questions and hypotheses rather than collecting new data. The data was obtained from three studies. The first study aimed to evaluate a VR-supported training for real estate professionals on how to properly inspect apartments involving a role play between vocational school teachers and learners, which was conducted from 2021 to 2022. The second study, carried out between 2023 and 2025, comprised four trials on vocational education and training learners' experience of a virtual 360-degree panorama-based interactive learning space for conducting meetings of apartment owners The third study focused on examining the fit of carefully designed social VR environments for conducting virtual workshops for academic staff like professors and lecturers, practitioners, and students. It comprised a total of six sub-studies implemented from 2023 to 2024. As a result, the studies addressed a wide range of participants, from higher education instructors and school teachers to university students and vocational school pupils. As two of the three

studies focused on topics from the real estate sector, almost half of the participants had a professional background in this area. To assess individual characteristics, all studies used the same standardized online survey, which is further described below.

3.2 Measures

The used questionnaire contained four questions to measure socio-demographic data, prior VR experience and technology readiness. The year in which the participants answered the survey, which is to be taken into account as a predictor in the analysis (see H3), was automatically recorded by the survey tool LimeSurvey.

Sociodemographics. Participants' age was assessed by the use of an open numeric question ("How old are you?") asking them to type their current age. Perceived gender was measured via a multiple-choice question ("Which gender do you feel yourself to belong to?") offering four options: "female", "male", "diverse", and "other". Furthermore, the qualification of the participants was assessed by a single-choice question ("What is your highest educational qualification?)" with a total of seven items ("no degree", "secondary school degree", "degree in vocational training", "(technical) A-levels", "bachelor's degree", "Diploma or state examination or magister or master's degree", and "doctorate or habilitation".

Prior VR experience. Since to date there is no validated questionnaire for measuring prior experience with VR technology, an ad-hoc item was created to evaluate this parameter. Participants were asked "How often have you already used virtual reality (VR) technology?" to be reported on a 5-point Likert scale ranging from "very rarely" to "very often".

Technology readiness. To measure learners' technology readiness (Cronbach's $\alpha = .81$), we used the Technology Commitment Short Scales [NFG12]. They contain a total of twelve items including four items to assess each of the three subscales that measure participants' technology acceptance ($\alpha = .78$; e. g., "I am very curious about new technical developments."), technology literacy beliefs ($\alpha = .80$; e. g., "For me, dealing with technical innovations is usually too much of a challenge."), and technology control beliefs ($\alpha = .74$; e. g., "It mainly depends on me whether or not I am successful in using modern technology."). All items were to be rated on a 5-point Likert scale ranging from "does not apply at all" to "fully applies".

3.3 Procedure

At the start of each study, participants were asked to complete the preliminary questionnaire containing the measurements described above. Next, they completed a trial with an educational VR environment and answered a subsequent questionnaire, whose data was not included in the secondary data analysis. To carry out the latter, all data records were first checked, cleaned (e. g., for incomplete records and duplicates) and merged. The

final dataset was analyzed in a semi-automated manner using the associated features for descriptive and regression analyses of IBM SPSS Statistics 30.0.

3.4 Participants

A total of 220 instructional participants from multiple vocational schools and universities answered the voluntary survey. It is important to note that in the studies whose data were reanalyzed as part of the present secondary analysis, the role of the subjects was not clearly captured. This was partially related to the design of the investigated educational settings. In the role-playing activities related to apartment inspections, for example, vocational school teachers primarily assumed the role of learning facilitators while some of temporarily acted as learners during initial trial runs for testing purposes. The virtual academic workshops on the other hand were attended by academic staff, practitioners and university students. Here, some participants assumed instructional roles (e. g., as lecturers or chairs) in one session, while taking on learning roles (e. g., as auditors) in others. As a result, and also taking into account that the boundary between teaching and learning is increasingly blurring, particularly with regard to the education of (young) adults, it was not feasible to count and consider instructional participants' specific roles individually.

The majority of participants (55.9 %) perceived themselves as female (44.1 % male; .0 % diverse; .0 % other). They varied in age from 17 and 71 years (M = 31.34 years, SD = 12.67). Of the participants, 21 persons (9.5 %) were younger than 20, 106 persons (48.2 %) were between 20 and 29 years old, 29 persons (13.2 %) were between 30 to 39, 42 persons (19.1 %) were aged 40 to 49, and 22 persons (10.0 %) were more than 50 years old. Among the subjects, 18 persons (8.2 %) reported the secondary school degree as their highest qualification, while 12 persons (5.5 %) held a degree in vocational training, and 57 persons (26.0 %) successfully finished their (technical) A-levels. Another 47 persons (21.5 %) held a bachelor's degree, while 53 persons (24.2 %) graduated with a diploma, state examination, magister, or master's degree, and 32 persons (14.6 %) with a doctorate or habilitation. Regarding their technology readiness, participants reported medium to high overall technology acceptance (M = 3.72 out of 5; SD = .81), high technology literacy beliefs (M = 4.29 out of 5; SD = .72), and medium to high technology control beliefs (M = 3.56 out of 5; SD = .76), resulting in a rather high total readiness (M = 3.83 out of 5; SD = .56). Seventeen subjects (7.7 %) answered the survey in 2021, 55 (25.0 %) in 2022, 65 (29.5 %) in 2023, and another 64 (29.1 %) in 2024, while a total of 19 persons (8.6 %) participated in 2025.

4 Results

To address our main RQ, we first analyzed the prior experience of all participants with VR technologies using descriptive statistics. Results show that, on average, the instructional participants surveyed reported little prior experience with VR (M = 1.89 out of 5;

SD = 1.25). A closer look reveals that more than half of the respondents (126 to be precise, that is 57.3 %) possessed very little prior experience with VR, whereas a total of 63 people could draw on at least a low (*N* = 38; 17.3 %) to medium (*N* = 25; 11.4 %) level of previous experience with this technology (see Fig. 2). Only 31 of the respondents possessed a high (*N* = 17; 7,7 %) or very high (*N* = 14; 6.4 %) degree of prior experience with VR.

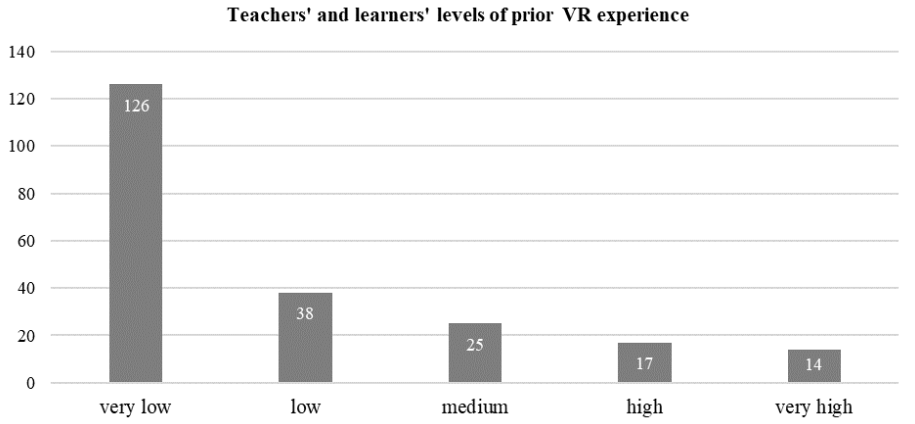


Fig. 2: Self-reported prior VR experience scores of instructional participants surveyed

To test H1 to H6, a hierarchical multiple regression analysis was conducted to identify the determinants of teachers' and learners' prior VR experience. In the first step, sociodemographic variables (age and gender) and year of survey (2021 to 2025) were entered into the model. Results indicate that this model did not significantly predict prior VR experience ($R^2 = .17$, $R^2_{adj} = .03$, $F\ (3, 216) = 2.02$, $p = .113$). Gender was a significant negative predictor ($\beta = -.16$, $p = .023$), indicating that male instructional participants reported slightly higher prior VR experience, whereas age ($\beta = .05$, $p = .445$) and year of survey ($\beta = .02$, $p = .768$) did not significantly contribute to the model. In the second step, individual technology readiness characteristics (technology acceptance, literacy beliefs, and control beliefs) were added. Their inclusion significantly improved the model, explaining an additional 6 % of the variance in prior VR experience ($\Delta R^2 = .06$, $\Delta F\ (3, 213) = 4.89$, $p = .003$). Individuals' fundamental technology acceptance ($\beta = .15$, $p = .044$) and technology literacy beliefs ($\beta = .20$, $p = .007$) emerged as positive predictors, suggesting that higher prior VR experience is associated with stronger technology acceptance and literacy beliefs. In contrast, teachers' and learners' technology control beliefs did not significantly predict their prior experience with VR technologies ($\beta = -.11$, $p = .130$). After adding the technology-readiness-related factors, the previously significant effect of gender was no longer statistically meaningful ($\beta = -.09$, $p = .199$), indicating that individuals' technology acceptance and literacy beliefs play a more dominant role regarding their prior experience with VR technologies than gender alone. The results of the conducted hierarchical multiple regression analysis support H4 and H5 but not H1 to H3 and H6.

Overall, the final model (see Fig. 3) explained 9% of the variance in prior VR experience (R^2 = .09, R^2_{adj} = .06, F (6, 219) = 3.51, p = .003).

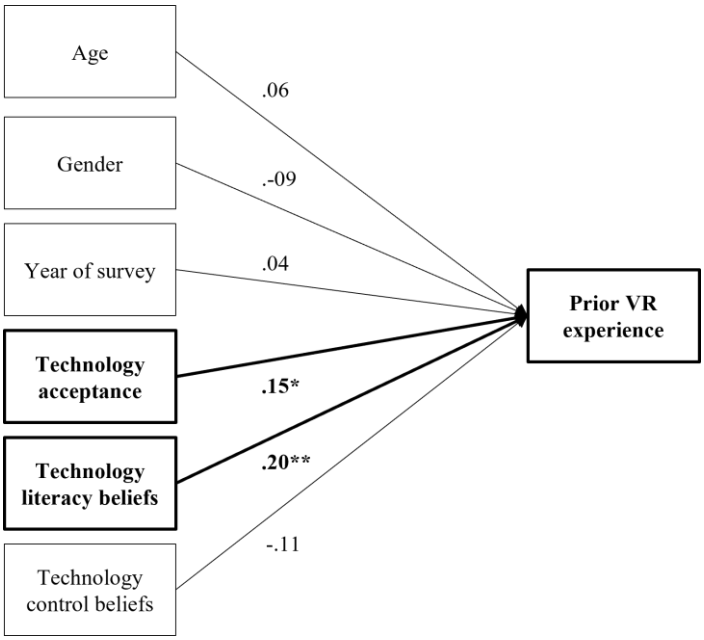


Fig. 3: Final regression model concerning the individual determinants of prior VR experience

Subsequently, we carried out a descriptive analysis to determine whether the experience of instructional participants with VR has at least tended to increase in recent years. However, the results did not show an upward trend. While the subjects' prior VR experience reported in 2022 (M = 1.93; SD = 1.03), 2023 (M = 2.00; SD = 1.30) and 2024 (M = 1.91; SD = 1.44) was at least slightly higher compared to 2021 (M = 1.53; SD = .87, the participants surveyed in 2025 in turn stated slightly lower values (M = 1.63; SD = 1.26). As the results of the regression analysis already indicated, the scores for all five years ranged very closely, suggesting no chronological increase in teachers' and learners' VR experience.

5 Discussion, implications, limitations, and future directions

The results of the secondary analysis of five years of data highlight that, despite their great potential for educational purposes as demonstrated in a series of meta-analyses (e. g., [Yu21]), instructional participants still lack experience with VR technologies like head-mounted-displays and 360-degree-applications. The dearth of experience may represent one reason why these technologies have hardly been implemented in educational practice

to date. Vice versa, a shortage of integration of VR in educational programs naturally means that instructional participants can hardly gather any (positive) experience with it. Accordingly, educational science and policy are required to jointly develop goal-oriented concepts for systematic and beneficial use of VR technologies in general and vocational schools as well as universities. Only if primary stakeholders possess sufficient experience and competencies to quickly adapt and apply VR-supported educational programs, the potential of the numerous promising didactic concepts developed in educational science over more than a decade (e. g., [KMS13; Ze22]) can be exploited to full extent.

Furthermore, our findings do not support the widespread assumption that (previous) use of VR technologies is determined by socio-demographic characteristics like age and gender. Instead, it appears to be significantly influenced by individuals' level of technological readiness, in particular their technology acceptance and literacy beliefs. These results are comparable to the findings of a similar study, in which the impact of age and gender on the future intention to use a specific VR application for education is also substantially reduced when more technology-related factors, in particular participants' fundamental intention to use VR technology, are included in the analysis [LLC23]. In addition, they are consistent with the results of more general studies on factors influencing the acceptance and use of e-learning technologies, which also found no significant impact of respondents' age and gender in this regard (e. g., [FBN17; Sh13]). Accordingly, future research is required to challenge and replicate studies that (at least partly) identified this kind of impact with reference to e-learning in general (e. g., [Re08; Sh13]) or VR-supported learning in particular [Di19; MC19; LD22] by taking into account possible moderator and mediator variables like individuals' readiness to use technology. In this regard, it may also be worth examining whether the relationship between prior VR experience and technology readiness is of unidirectional or bidirectional nature. Considering the rather low degree of reported experience with VR technologies in the present, it is likely that subjects' readiness for technology has been shaped by their own use and social influences with regard to more widespread technologies like notebooks and smartphones. However, if it turns out that especially negative experiences with VR also reduce instructional participants' general technology readiness, it is of even more importance to ensure that in particular less experienced educational stakeholders gain positive initial experiences with VR through suitable measures (e. g., further training for teachers or the use of VR in lessons for learners).

Given that, to our best knowledge, there have been no empirical findings on the experience with VR among the population to date, not to mention instructional participants, the results of our study offer a valuable initial insight in this regard. Nevertheless, it is important to note that our approach of conducting a secondary data analysis of our own studies was largely driven by concerns of sustainability and efficiency, primarily in order to make some initial progress in this respect. As a result, it must be acknowledged that a potential selection bias cannot be ruled out due to the non-random composition of the sample. Moreover, although the sample examined is fairly large and diverse in terms of its individual characteristics, it is not representative and almost half of it comes from the real estate sector due to the focus of the former primary analyses. This group may have had

occasional exposure to virtual 360-degree property presentations or 3D visualization tools during training or practice. In consequence, such experiences might have slightly inflated the reported average, suggesting that a more diverse sample could reveal even lower levels of prior VR experience. As a first exploratory approach, this study relied on a consistent secondary dataset to examine individual determinants of prior VR experience across several studies. This focus ensured methodological rigor and internal comparability within the analysis. However, future research could benefit from combining such data with external sources like open datasets – although a brief search on established open dataset repositories using the terms ('virtual reality' AND ('prior experience' OR 'previous experience')) did not yield any relevant datasets. Expanding the dataset in this way may improve generalizability, support more diverse comparative analyses, and offer deeper interdisciplinary insights into instructional participants' VR experience and its determinants.

Furthermore, for methodological reasons outlined above, it was not possible to examine the previous VR experience of instructional participants in teaching roles (e. g., academic lecturers, workshop instructors, and school teachers) and learning roles (e. g., university students or pupils), separately. Combining both groups may have masked systematic differences in VR experience, as instructors typically have greater access to technology in educational settings, while learners may accumulate experience through curricular or extracurricular exposure. In order to investigate the degree of experience with VR technologies among educational stakeholders, in particular teachers and learners, in more depth, future studies should choose a within-subject design with a representative sample carefully distinguishing between both groups rather than a mixed-sample between-subject design in order to be able to validly analyze the development of the proliferation and use of VR over several years. Separate findings on the prior experience of teachers and learners provide fertile ground for follow-up research on whether the acceptance and experience of teachers or that of learners is crucial for the successful implementation of VR in educational practice. In advance, the lack of a standardized assessment tool for people's prior experience of VR should be addressed, which necessitated the study described above resorting to an ad hoc item by developing and validating a scale that enables more reliable measurements in this respect. With regard to research sustainability, further secondary data analyses of previously conducted studies are a common alternative option, especially since prior experience with VR has been surveyed in quite a few studies (e. g., [LLC23; Mu23]). Here, however, a broader data basis should be included compared to our study. In the present study, the investigated characteristics of socio-demographics and technology readiness of instructional participants only accounted for 9 % of the variance. It suggests that additional factors not examined in this study may influence instructional participants' prior experience with VR. These may include access to VR hardware, the availability of VR-based learning opportunities within educational institutions, workplaces, or public events, as well as the specific vocational field participants are engaged in. Furthermore, social influences – such as peers or role models with prior VR experience – may also play a role by shaping interest and access. Consequently, we encourage follow-up research to include further potential impact factors on their (prior)

experience with VR based on substantiated theories. They contribute further insights into which variables determine the previous experience of stakeholders and thus the successful implementation of VR in educational practice.

Future work on teachers' and learners' prior experience with VR technologies is also encouraged since, in addition to successful and sustainable implementation, this factor furthermore appears to affect the user experience and learning outcomes using VR. For instance, a recent study involving 110 adult participants demonstrates that a higher level of prior experience with VR can contribute to learners experiencing less VR sickness [Jo23], which may in turn impede learning. A further study, in which psychology and engineering students worked on a multi-step assembly task in a VR environment, showed that participants who reported a higher level of prior VR experience rated their own performance more positively than their peers with less prior experience [Sa20]. Yet these results contradict another study of 132 middle school students using a VR application to learn about a history topic that revealed no significant moderating effects of prior VR experience (here referred to as prior technical knowledge) on learning outcomes [Mu23]. These findings and their partial inconsistency emphasize the need for follow-up studies that look in depth and systematically at the role of involved individuals' prior experience with VR technologies in educational settings in order to shed light on this so far insufficiently investigated issue.

For educational practice, our findings suggest that the trait technology readiness of instructional participants – particularly their technology acceptance and literacy beliefs – is a crucial step toward enabling the effective use of VR in teaching and learning. To achieve this, professional development programs for teachers should explicitly address the pedagogical potential and limitations of VR, include hands-on sessions with the technology, and offer opportunities to reflect on its meaningful integration into curricula. or learners, early exposure to immersive media through low-threshold, guided VR experiences (e. g., virtual lab simulations, language immersion environments, or interactive history reconstructions) could foster familiarity and reduce hesitation. Moreover, pilot projects or institutional test phases may help educational institutions identify best practices while avoiding premature policy restrictions or prohibitions. Overall, promoting constructive, competence-oriented engagement with VR technologies across all levels of education may help bridge the current gap between empirical potential and practical adoption.

References

[AW16] Abdullah, F.; Ward, R.: Developing a General Extended Technology Acceptance Model for E-Learning (GETAMEL) by analysing commonly used external factors. Computers in Human Behavior 56, pp. 238–256, 2016. DOI: 10.1016/j.chb.2015.11.036

[Al23] Al-Adwan, A. S.; Li, N.; Al-Adwan, A.; Abbasi, G. A.; Albelbisi, N. A.; Habibi, A.: Extending the Technology Acceptance Model (TAM) to Predict University Students'

Intentions to Use Metaverse-Based Learning Platforms. Education and Information Technologies 28, pp. 15381–15413, 2023. DOI: 10.1007/s10639-023-11816-3

[Al24] Albrecht, M.; Eder, M.; Erenli, K.; Fankhauser, L.; Mayerl, M.; Petak, L.; Schachinger, V.: The VR Chronicles: Insights from the VR Factsheet 2024. In: 2024 IEEE Gaming, Entertainment, and Media Conference. IEEE, Turin, pp. 1–6, 2024. DOI: 10.1109/GEM61861.2024.10585504

[AR24] ARtillery Intelligence: VR Global Revenue Forecast: 2023-2028, https://artilleryiq.com/reports/vr-global-revenue-forecast-2023-2028, 2025-03-24.

[CPS23] Chang, H.; Park, J.; Suh, J.: Virtual reality as a pedagogical tool: An experimental study of English learner in lower elementary grades. Education and Information Technologies 29, pp. 4809–4842, 2023. DOI: 10.1007/s10639-023-11988-y

[Ch24] Chen, J.; Fu, Z.; Liu, H.; Wang, J.: Effectiveness of Virtual Reality on Learning Engagement: A Meta-Analysis. International Journal of Web-Based Learning and Teaching Technologies 19, pp. 1–14, 2023. DOI: 10.4018/IJWLTT.334849

[Ch12] Chow, M.; Herold, David K.; Choo, T.-M.; Chan, K.: Extending the technology acceptance model to explore the intention to use Second Life for enhancing healthcare education. Computers & Education 59, pp. 1136–1144, 2012. DOI: 10.1016/j.compedu.2012.05.011

[CBG22] Coban, M.; Bolat, Y. I.; Goksu, I.: The potential of immersive virtual reality to enhance learning: A meta-analysis. Educational Research Review 36, pp. 100452, 2022. DOI: 10.1016/j.edurev.2022.100452

[ES07] Euler, D.; Seufert, S.: Change Management in der Hochschullehre: Die nachhaltige Implementierung von e-Learning-Innovationen [Change management in academic teaching: The sustainable implementation of e-learning innovations]. Zeitschrift für Hochschulentwicklung 3, pp. 3–15, 2007. DOI: 10.3217/ZFHD03/02

[Fi21] Fischer, H.; Arnold, M.; Philippe, S.; Dyrna, J.; Jung, S.: VR-Based Learning and Teaching. A Framework for Implementation of Virtual Reality in Formal Education. In (Gómez Chova, L.; López Martínez, A.; Candel Torres, I., Eds.): INTED2021 Proceedings. IATED, pp. 3304–3314, 2021. DOI: 10.21125/inted.2021.0694

[FBN17] Fleming, J.; Becker, K.; Newton, C.: Factors for successful e-learning: does age matter? Education + Training 59, pp. 76–89, 2017. DOI: 10.1108/ET-07-2015-0057

[HGJ21] Howard, M. C.; Gutworth, M. B.; Jacobs, R. R.: A meta-analysis of virtual reality training programs. Computers in Human Behavior 121, S. 106808, 2021. DOI: 10.1016/j.chb.2021.106808

[Jo23] Josupeit, J.: Does Pinocchio get Cybersickness? The Mitigating Effect of a Virtual Nose on Cybersickness. In (Ahram, T.; Karwowski, W., Eds.): Proceedings of the AHFE International Conference on Human Factors in Design, Engineering, and Computing. AHFE, Hawaii, pp. 99–109, 2023. DOI: 10.54941/ahfe1004445

[KMS13] Köhler, T.; Münster, S.; Schlenker, L.: Didaktik virtueller Realität: Ansätze für eine zielgruppengerechte Gestaltung im Kontext akademischer Bildung [Virtual reality didactics: approaches for target group-oriented design in the context of higher education]. In (Reinmann, G.; Ebner, M.; Schön, S., Eds.): Hochschuldidaktik im

Zeichen von Heterogenität und Vielfalt: Doppelfestschrift für Peter Baumgartner und Rolf Schulmeister [Higher education didactics in the light of heterogeneity and diversity: Double commemorative publication for Peter Baumgartner and Rolf Schulmeister]. BIMS e. V., Bad Reichenhall, pp. 99–112, 2013.

[LLC23] Li, Q.; Liu, Q.; Chen, Y.: Prospective Teachers' Acceptance of virtual reality technology: a mixed study in Rural China. Education and Information Technologies 28, pp. 3217–3248, 2023. DOI: 10.1007/s10639-022-11219-w

[LD22] Luo, Y.; Du, H.: Learning with desktop virtual reality: changes and interrelationship of self-efficacy, goal orientation, technology acceptance and learning behavior. Smart Learning Environments 9, pp. 1–22, 2022. DOI: 10.1186/s40561-022-00203-z

[MC19] Manis, K. T.; Choi, D.: The virtual reality hardware acceptance model (VR-HAM): Extending and individuating the technology acceptance model (TAM) for virtual reality hardware. Journal of Business Research 100, pp. 503–513, 2019. DOI: 10.1016/j.jbusres.2018.10.021

[MMP23] Mayer, R. E.; Makransky, G.; Parong, J.: The Promise and Pitfalls of Learning in Immersive Virtual Reality. International Journal of Human–Computer Interaction 39, pp. 2229–2238, 2023. DOI: 10.1080/10447318.2022.2108563

[MK94] Milgram, P.; Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Transactions on Information and Systems 77, pp. 1321–1329, 1994.

[Mu23] Mulders, M.: Learning about Victims of Holocaust in Virtual Reality: The Main, Mediating and Moderating Effects of Technology, Instructional Method, Flow, Presence, and Prior Knowledge. Multimodal Technologies and Interaction 7, p. 28, 2023. DOI: 10.3390/mti7030028

[NFG12] Neyer, F. J.; Felber, J.; Gebhardt, C.: Entwicklung und Validierung einer Kurzskala zur Erfassung von Technikbereitschaft [Development and Validation of a Short Scale to Measure Technology Commitment]. Diagnostica 58, pp. 87–99, 2012. DOI: 10.1026/0012-1924/a000067

[Ni06] Ni, T.; Schmidt, G. S.; Staadt, O. G.; Livingston, M. A.; Ball, R.; May, R.: A survey of large high-resolution display technologies, techniques, and applications. In: IEEE Virtual Reality Conference (VR 2006). IEEE, Alexandria, USA, pp. 223–236, 2006. DOI: 10.1109/VR.2006.20

[Pa21] Parong, J.: Multimedia Learning in Virtual and Mixed Reality. In (Mayer, R. E.; Fiorella, L., Eds.): The Cambridge Handbook of Multimedia Learning. Cambridge University Press, Cambridge, pp. 498–509, 2021.

[Qu24] Qiu, X. B.; Shan, C.; Yao, J.; Fu, Q. K.: The effects of virtual reality on EFL learning: A meta-analysis. Education and Information Technologies 29, pp. 1379–1405, 2024. doi: 10.1007/s10639-023-11738-0

[Re08] Rezaei, M.; Mohammadi, H. M.; Asadi, A.; Kalantary, K.: Predicting E-Learning Application in Agricultural Higher Education Using Technology Acceptance Model. Turkish Online Journal of Distance Education 98, pp. 85–95, 2008.

[Sa20] Sagnier, C.; Loup-Escande, E.; Lourdeaux, D.; Thouvenin, I.; Valléry, G.: User Acceptance of Virtual Reality: An Extended Technology Acceptance Model.

International Journal of Human–Computer Interaction 36, pp. 993–1007, 2020. doi: 10.1080/10447318.2019.1708612

[Sa24] Santiago, L. G.; Kronemberger, V. A.; Porcino, T.; Cherullo, R.; Raposo, A.: An XR Framework Proposal to Assist Designers in Minimizing Cybersickness. In: Anais Estendidos do XXVI Simpósio de Realidade Virtual e Aumentada. Sociedade Brasileira de Computação, pp. 21–22, 2024. DOI: 10.5753/svr_estendido.2024.244711

[Sh13] Shah, S. A. M.; Iqbal, N.; Janjua, S. Y.; Amjad, S.: Employee Behavior Towards Adoption of E-learning Courses: Validating Technology Acceptance Model. Mediterranean Journal of Social Sciences 14, pp. 765–774, 2013. DOI: 10.5901/mjss.2013.v4n14p765

[Ve03] Venkatesh, V.; Morris, M. G.; Davis, G. B.; Davis, F. D.: User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 27, pp. 425–478, 2003. DOI: 10.2307/30036540

[VB08] Venkatesh, V.; Bala, H.: Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences 39, pp. 273–315, 2008. DOI: 10.1111/j.1540-5915.2008.00192.x

[WYG20] Wu, B.; Yu, X.; Gu, Xiaoqing: Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. British Journal of Educational Technology 51, pp. 1991–2005, 2020. DOI: 10.1111/bjet.13023

[Yu21] Yu, Z.: A meta-analysis of the effect of virtual reality technology use in education. In: Interactive Learning Environments 31, pp. 4956–4976, 2021. DOI: 10.1080/10494820.2021.1989466

[Ze22] Zender, R.; Buchner, J.; Schäfer, C.; Wiesche, D.; Kelly, K.; Tüshaus, L.: Virtual Reality für Schüler: innen: Ein «Beipackzettel» für die Durchführung immersiver Lernszenarien im schulischen Kontext [Virtual reality for pupils: An “instruction leaflet” for the implementation of immersive learning scenarios in schools]. MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung 47, pp. 26–52, 2022. DOI: 10.21240/mpaed/47/2022.04.02.X

GenAI-Chatbots as Debriefers: Investigating the Role Conformity and Learner Interaction in Counseling Training

Dominik Evangelou [1], Maria Klar [1], Kristian Träg [1], Miriam Mulders [1], Melina Marnitz [1], and Lara Rahner [6]

Abstract: Debriefing is essential for the effectiveness of simulation-based training but is generally considered resource-intensive. Generative Artificial Intelligence (GenAI)-based debriefing can be an alternative to human debriefers. However, there is no research yet whether GenAI chatbots can take up this role and how learners react to them. This paper presents a qualitative analysis of a debriefing following a counseling training conducted in Virtual Reality (VR) with the support of a GenAI chatbot. The debriefing helped students to analyze their experiences and application of consulting techniques. The analysis of the chatlogs (n = 22) are focused on the role conformity of the bot and the students' ability to reflect their behavior within VR. The results revealed the chatbot's strong role conformity but also its tendency for overly complimentary answers. However, this bias does not seem to influence students' self-reflection. Instead, they maintained a self-critical attitude. Future research on AI-assisted debriefing could expand on these findings in related areas.

Keywords: Counseling, Virtual Reality, Training, Chatbot, AI, GenAI, Debriefing, Roleconformity, Student behavior

1 Introduction

Effective learning through simulations and serious games critically depends on the structured processing and transformation of experiences in subsequent debriefing sessions [Cr23]. The pedagogical value of even the most technologically and aesthetically advanced simulations can only be fully realized when learners are given the opportunity to reflect systematically on their actions and decisions. As argued by [Th90], such reflective engagement serves as a central mechanism for integrating experience into meaningful knowledge structures and facilitating transfer to novel contexts.

Despite its recognized importance, debriefing—particularly in Virtual Reality (VR) and simulation-based learning environments—remains a resource-intensive practice, often constrained by the limited availability of trained facilitators [MHC07]. While research on simulation-based debriefing is steadily growing, it remains fragmented across domains, with

[1] University of Duisburg-Essen, Chair of Educational Technology and Instructional Design, Universitätsstraße 2, 45117 Essen, Germany, dominik.evangelou@uni-due.de, https://orcid.org/0009-0002-7152-1594;
maria.klar@uni-due.de, https://orcid.org/0000-0002-5298-8458;
kristian.traeg@uni-due.de, https://orcid.org/0009-0004-4557-2026;
miriam.mulders@uni-due.de, https://orcid.org/0000-0003-0683-2310;
melina.marnitz@rwth-aachen.de, https://orcid.org/0009-0007-8373-4729

[6] TU Dortmund, Center for Higher Education, Hohe Straße 141, 44139 Dortmund, Germany, lara.rahner@tu-dortmund.de, https://orcid.org/0009-0000-1298-8640

doi:10.18420/delfi2025_05

the healthcare sector currently at the forefront of empirical and conceptual developments (e. g., [DY14; Fa24; Ga15; Lu21]).

Recent advancements in generative artificial intelligence (GenAI), especially in the form of large language model (LLM)-based chatbots, offer new opportunities to support reflective learning processes at scale. These systems may be capable of simulating debriefing dialogues that approximate human-facilitated reflection. The present study investigates this potential by examining the chatbot's ability to maintain role conformity in the function of a debriefer, as well as students' responses to and engagement with such AI-mediated interactions.

2 Background

2.1 Debriefing

The integration of a debriefing following the actual learning situation has long been an essential component in educational contexts, even before the availability of simulations in VR. Research on simulation games and role-playing has already shown that meaningful learning effects cannot be achieved merely by exposing learners to a model of reality that is both appropriately complex and authentic. Rather, in addition to providing complex learning scenarios, simulation game instructional frameworks – such as the design of an effective debriefing for reflection and learning transfer – are necessary to transfer the experiences from the simulation games to real-world professional practice [KS07]. The simulation game method is particularly well-suited for developing competencies, as it applies the principles of problem-based learning. Learners are confronted with complex and authentic situations that enable experiential learning. It is essential that learners engage in cooperative learning through team-based action and receive appropriate support from the instructor – including in the form of joint reflection. Notably, the aspect of collective reflection directly relates to the debriefing process. Accordingly, [KS07] emphasizes that only knowledge acquired through reflection can be transferred to unfamiliar domains. It follows that especially when learning through online courses or acquiring competencies within virtual reality simulations, the content to be learned needs to be reviewed and discussed in a debriefing afterward.

Debriefing can generally be conducted in two fundamentally different ways: as a guided reflection or as a self-debriefing. Additionally, these two methods can be varied by factors such as the number of participants, the instructions, technical aids, or even the instructor themselves (e. g. [DY14; Fa24; Lu21]). In guided debriefing, the instructor moderates the process by posing questions related to the pre-defined learning objectives. Particularly for inexperienced learner groups, facilitating the reflection process seems to be of great importance [KS07]. According to [Th11], this type of debriefing is considered the most suitable reflection method. A review of literature reveals that guided debriefing is indeed the most frequently used method (see [Bo11; Ti13; TTH15; Ve18]); however, there is no empirical evidence indicating that this method is also the most effective [DY14]. In contrast,

unguided debriefing allows learners to reflect on their learning experiences independently and self-organized, for example, within a group setting. Studies show that there is no statistically significant difference between guided debriefings and self-debriefings. Both methods lead to performance improvements [Bo11; Bo13].

In addition to comparing guided and unguided debriefings, several studies have examined variations in specific debriefing aspects. For instance, some studies compare the effectiveness of asynchronous debriefings [At21], or video-assisted debriefing [Gr10]. These studies further confirm the finding that any form of debriefing has positive effects on performance. The only factor that appears to lead to significant differences in debriefing quality is the instructor's skill as a debriefer [Ti13; TTH15]. Studies focusing on the debriefing of simulations are, as previously mentioned, primarily found in the healthcare sector. A literature review by [DY14] indicates that research on post-simulation debriefings is still in its infancy. The researchers identified only 13 articles that either compared various debriefing strategies or examined student perceptions of debriefing. The relatively limited amount of relevant research literature is further supported by the reviews of [Ga15] and [Lu21], who describe eight and seven articles, respectively, comparing different debriefing strategies or explored student perceptions of debriefing.

It can be stated that post-simulation debriefing plays a crucial role in learning. Research indicates that debriefing can be resource-intensive, depending on the format, and that no single method has been proven more effective than others to date [DY14]. This highlights the significant potential for further exploration of different debriefing methods. The following sections will explore the potential use of AI in the context of debriefing. The focus will first be on GenAI Chatbots, followed by an examination of their potential application as debriefers.

2.2 Chatbots

While scripted Chatbots have been around for decades, AI-based generative Chatbots have taken off only recently, notably with the publication of ChatGPT in late 2022 [Wa24]. These types of chatbots are large language models (LLMs), meaning they have been fed large amounts of text-data to analyze the likelihood of words occurring together, and use those likelihoods to be able to generate new coherent sentences [Na23].

Researchers from varying fields have investigated the use of GenAI-chatbots in education. [Ho23] identifies authentic language use as a key benefit that GenAI offers for language teaching, in addition to personalized tutoring via AI. [CH23] find in a survey that university students in Hong Kong hold generally positive attitudes towards GenAI in teaching and learning and would like to integrate them into their learning practices, for example for personalized feedback. However, many students seemed concerned about topics like privacy, ethics, and transparency when handling GenAI, as well as over-reliance on AI-technology and the degradation of skills accompanied by it [CH23]. [SQ24] show that texts that ChatGPT

generated for potential use in climate change education were scientifically accurate, but did not address certain economic activities and actors that contribute to increased carbon emissions. Overall, educators seem to have mixed experiences with GenAI, with tertiary education being the main focus of research up to this point [JLC23; Kh22]. The present paper attempts to expand upon this body of research by introducing a GenAI chatbot as a conversational partner to training for counsellors.

Utilizing GenAI chatbots as conversational partners in education might lead to a few problems. For educators, it might be difficult to use and teach a technology that they themselves are still unfamiliar with [Ng23]. For learners, the heavy amount of cognitive offloading afforded by assisting chatbots may lead to a lack of own contributions, possibly resulting in placebo effects when learning [Sk24]. The following section will go into more depth on using chatbots during debriefings of training interventions where these challenges need to be kept in mind.

2.3 Chatbots as Debriefers

As stated earlier, debriefer skills are essential for debriefing quality [Ti13; TTH15], so the question is whether current GenAI chatbots are performant enough to fill in the debriefer role. Debriefing requires a dynamic interaction between the debriefer and the learner that relates back to the VR simulation. The debriefer's role, therefore, requires going on tangents in a conversation and coming back to maintain the overall thread of conversation. While LLMs can be used to generative one-off texts like feedback [Da23] or test questions [Le24], it is more of a challenge to set up LLMs in such a way that they can hold up a complex conversation over a longer period [Ji24]. For AI-led debriefing the question arises whether role adherence is good enough to reach a sufficient quality.

Sycophancy, an LLM's tendency to generate output that agrees with or pleases the user, could be a further issue in the context of debriefing. This phenomenon might have adverse effects on perceived authenticity and trust [SW25]. Authenticity and trust are key factors not only in successful counselling, but also for debriefings [Ro57; Ro62; SFE16]. As debriefing requires the critical reflection of the learner's experience and actions, it would not be beneficial to the learning process to have an overly positive and uncritical chatbot debriefer [Ma24]. It is therefore vital to evaluate chatbot debriefers for this tendency.

However, even if the chatbot did not adhere to the role perfectly or displayed sycophantic or other awkward behavior (e. g., overly praising the learner or breaking the role), this would not necessarily mean it could not be used for debriefing as learners themselves could mitigate these weaknesses to some degree [Ji24]. The evidence that self-guided debriefing can be equally effective indicates that learners can take on responsibility for the debriefing process and that the quality of the debriefing does not only depend on the debriefer (human or AI) but also the learner. The responsibility for the debriefing process can be shared

between learner and AI [JNH23; Mo22]. Nonetheless, there is little research on how learners interact with GenAI chatbots and whether learners can amend chatbot weaknesses.

There is research showing that learners have difficulty to prompt chatbots effectively [Za23] and that they do not use chatbots as a dialogue partner but rather an information source [Kl24]. When students interacted with a chatbot versus a peer in problem-solving task, it was found that they interacted less dialogically and more predictably with the chatbot than with the peer, but preferred the chatbot nevertheless [So25]. Based on these findings, it could be expected that students do not easily engage in a fluent debriefing conversation and amend chatbot weaknesses.

2.4 Research Questions

In this article, we aim to investigate the extent to which GenAI chatbots can be utilized for post-simulation debriefings in educational contexts and thus represent a resource-saving but still significant alternative to classic debriefings. When analyzing the chatlogs, we focus on the aspects of debriefers' role conformity and student behavior: (1) Does a GenAI chatbot maintain its role as a debriefer? (2) How do students interact with the chatbot debriefer?

3 Methods

3.1 Procedure

The study was conducted as part of a VR training aimed at enhancing counseling skills among participants. The training session took place in a controlled lab setting, where participants engaged in a simulated counseling interaction with a standardized client (avatar) within a VR environment. Prior to the training, participants were informed about the study and asked to provide informed consent. They were then assigned a unique pseudonym code to ensure anonymization throughout the study.

After completing a pre-questionnaire, participants received instructions on operating the VR equipment and were introduced to the counseling case. They then entered the VR environment using a head-mounted display and interacted with the virtual client for approximately 20 minutes. Participants applied previously learned techniques such as summarizing and asking open-ended questions during the consultation.

Following the VR session, participants completed a second questionnaire. Subsequently, they took part in one of two types of debriefing: either a guided debriefing led by a trained facilitator or a structured, GenAI chatbot-based debriefing. As we are only referring to chatbot-based debriefing in this article, we will only describe this in the following. The chatbot's task was to guide the test participants through the debriefing. The chatbot should

follow a pre-coded protocol aligned with the three-phase debriefing model [PFS16]. To enable authentic communication with a role-conforming chatbot, we tested several system prompts in advance. In doing so, we tested many slightly varied prompts and used different LLMs. We optimized our prompt in iterative process loops with several research assistants and student participants. The system prompt employed in this study is documented in the online appendix [Ev25]. For the underlying model, we selected Meta Llama 3.1 8B Instruct, as it demonstrated robust performance despite its relatively small parameter size. Temperature and nucleus sampling were both set to 0.5 to enable sampling beyond the most common word while retaining some level of coherence in the chatbot's responses. The study concluded with a final post-questionnaire and informing the participants about the study objectives.

To evaluate the effectiveness of the training, participants' self-efficacy and counseling competence were assessed at all measurement time points. These constructs were selected as key indicators of participants' confidence in their own counseling abilities and their perceived skill development through the VR-based learning experience. Self-efficacy and counseling competence were measured using scales based on [He09]. In addition, the third measurement time point also included participants' perception of the debriefing experience, which was assessed using the Debriefing Experience Scale (DES) developed by [Re12].

The participants were able to start the debriefing with the chatbot themselves. They used a laptop provided by us for this purpose. Next to them was a written guide with instructions (e. g., "Be as detailed as possible and give specific examples."), which they had to follow during the 20-minutes conversation.[7] The study concluded with a final post-questionnaire and informing the participants about the study objectives.

To evalutae the effectivness of the training, participants' self-efficacy and counseling competence were assessed at all measurement time points. These constructs were selected as a key indicators of participants' confidence in their own counseling abilities and their perceived skill development through the VR-based learning experience. Self-efficacy and counseling competence were measured using scales based on [He09]. In addition, the thied measurement time point also included participants' perception of the debriefing experience, which was assessed using the *Debriefing Experience Scale* developed by [Re12].

3.2 Participants

Participants were students at a large university in Germany enrolled in a higher education counseling training seminar. All participants had prior exposure to theoretical knowledge of counseling techniques. The final sample consisted of 46 persons, of which 22 were randomly assigned to the chatbot condition. Of these 22 (age $M = 23.3$; $SD = 4.3$), 20 were female and 2 were male.

[7] The full guide can be accessed via [Ev25].

3.3 Data Analysis

The chatlogs of the 22 dialogs were saved and analyzed according to the principles of qualitative content analysis [KR20]. Two distinct coding schemes were developed by two experienced scientists for this purpose, one relating to research question 1, i. e., role conformity, and another one relating to research question 2, i. e., student behavior. Both coding schemes were applied using MAXQDA. The first coding scheme assessed whether and how the chatbot adhered to its predefined role as a debriefer, based on the three-phase model of debriefing by [PFS16]. Furthermore, this coding scheme captured structural elements of the chatbot's language and behavior. The second coding scheme assessed how participants reflected on their counseling experience and the debriefing process. All codes were inductively developed based on recurring themes in the participants' written responses. The tables each contain a brief description of the (sub-)code, one or more direct quotes from the chatlogs for each code, and the cumulative number of mentions across all chatlogs.[8]

After the study, two experienced research assistants independently examined the chatlogs according to the (sub)-codes contained in the coding schemes. More than a third (n = 8) were double-coded by the assistants. The interrater reliability of $\kappa = .82$ is in the acceptable range [Co68; Ku19].

We supplemented our qualitative data from chatlogs with one item of the Debriefing Scale ([Re12] from the final questionnaire, in which students were asked to rate their agreement to the statement "The chatbot provided adequate guidance during the debriefing" on a Likert-scale from 1 (*totally disagree*) to 5 (*totally agree*).

4 Results

The following results section presents the qualitative analysis of the chatlogs, focusing on the chatbot's role conformity and students' behavior. A total of eighteen categories, including subcategories, were identified for role conformity, and fourteen categories for students' behavior, using both deductive and inductive approaches. Due to the large number of categories, the complete tables containing all codes and exemplary quotes for each code are provided in the online appendix. Tables 1 and 2 are intended to illustrate exemplary and interesting results of the qualitative analysis.

4.1 Chatbot‘s Role conformity

In terms of role conformity, the chatbot adhered to most of the required steps outlined in the system prompt. In all but one chat, it introduced itself as the debriefer. Notably, in one

[8] The complete tables with codes and subcodes can be found at [Ev25].

case the participant did not enter the designated keyword *"start"* to initiate the debriefing as recommended. This likely caused the opening of the debriefing to deviate from the system prompt, as seen in the example: *"Thank you for being willing to speak with me. I hear that you are feeling [. . .] a bit overwhelmed after completing the VR training."* Interestingly, despite this deviation, the chatbot proceeded with the debriefing according to the three-phase structure. Despite the presumably less authentic communication, this student rated the chatbot-guided debriefing as highly appropriate, giving it a 5 out of 5. The other participants also provided consistently positive ratings (M = 4.33, SD = 0.58). After its introduction, the chatbot outlined the debriefing process in all cases and guided the conversation through the respective phases using appropriate transitions. Remarkably, only in transcript SD10 did it explicitly conclude the reaction phase with the statement: *"This is the conclusion of the reaction phase."* In contrast, transitions to the understanding phase ($\sum$= 21; 16/22) – the symbol $\sum$ represents the total number of mentions across all chat transcripts, while the fraction (x/22) indicates the number of chat transcripts in which this specific (sub)code appears – and summary phase ($\sum$= 14; 14/22) occurred more frequently. As the conversation progressed, the chatbot's outputs primarily focused on prompting reflection on the learning content ($\sum$= 236; 22/22). For example, it asked: *"We have learned three techniques: summarizing, paraphrasing, mirroring, I-messages, and questioning techniques. Which technique did you find most difficult or most interesting?"* In addition to reflecting on conversation techniques, applying learned content to real-world practice was a key objective of the debriefing. The chatbot facilitated this transfer ($\sum$= 86; 19/22) by posing targeted questions such as: *"Which aspects do you think are most important, and how can you integrate them into your counseling practice?"* It is also noteworthy that the chatbot actively addressed students' emotions. This was achieved, in part, through direct inquiries about their emotional state, as illustrated in table 1. In fact, all chatlogs included at least one such inquiry. Moreover, in sixteen transcripts ($\sum$= 39) the chatbot demonstrated attempts to convey understanding or empathy in response to students' input (see Table 1). In the final phase of the debriefing, students were expected to summarize their key takeaways. The chatbot frequently facilitated this process by posing targeted questions such as: *"What are the key insights you can take away from this debriefing?"* In some instances, it also generated independent summaries of students' responses (see Table 1). Across eighteen transcripts, the summarization phase was explicitly referenced 38 times.

Beyond the content-driven aspects of the chatbot debriefer, which were shaped by the system prompt, certain chatbot-specific characteristics may have contributed to a perception of inauthentic communication. For instance, twelve transcripts ($\sum$= 28) contain stage directions, such as: *"(Pause to give the student time to collect their thoughts and mention key points.)"* These instructions typically appeared after the chatbot posed a question. Authenticity in conversation can also be disrupted by the chatbot's tendency toward sycophantic behavior, as discussed in Section 2.3. This is particularly evident in its overly enthusiastic responses, which appear in every chatlog. For example, it frequently used adjectives such as interesting ($\sum$= 30; 13/22), awesome ($\sum$= 67; 19/22), fantastic ($\sum$= 5; 3/22), and very good ($\sum$= 40; 13/22). As shown in table 1, the word important appears 156 times across 21 debriefings.

Codes and Subcodes	Examples
Asks about emotions (∑ = 35; 22/22)	*Let's dive deeper into your feelings. How did you feel when you began to understand the situation better and started applying the counseling techniques? Did you feel more confident, or were there still uncertainties?*
Expresses empathy (∑ = 39; 16/22)	*It is understandable that you didn't think about taking notes in that situation. The situation was complex, and you were focused on communicating with the client.*
Summary of key findings (∑= 38; 18/22)	*Let's now summarize what we have discussed:* • *You felt comfortable while conducting the VR training.* • *[. . .]*
Uses adjective "important" (∑= 156; 21/22)	*That is an important step toward becoming a successful counselor.*

Tab. 1: Exemplary codes for role conformity (∑ = number of mentions across chatlog files; x/22 = share of chatlog files with this (sub)code)

Additionally, the chatbot exhibited repetition within individual conversations. For instance, the sentence: *"That shows that you engaged with the situation and that you were willing to open up and try new things."* appears 16 times in one case.

4.2 Student behavior

Next to the chatbot behavior, the question is how students reacted to the chatbot and whether they engaged in a reflection that fulfill the function of a debriefing. The main function of a debriefing is a critical review of the simulation to facilitate transfer to real-world scenarios. Furthermore, a debriefing should allow for expression of affective responses to the simulation, especially if the simulation was emotionally activating [Cr23].

In this study, most students indeed processed their current emotional state once or several times (see Table 2). They also evaluated their role as a counselor either positively, neutrally, or negatively at least once (see Table 2). In absolute numbers, they assessed their performance positively the least often, but in relative numbers negative self-appraisal occurred the least often. Assuming a normal distribution of actual counseling performance, these reflections can be seen as adequately balanced. There were a considerable number of statements on transfer to real-world counseling (see Table 2). Similarly, 14 of the 22 participants felt that they have learned something in the simulation (∑= 20). This is in line with a rather large number of assessments of the counseling techniques (∑= 36; 16/22), e. g., *"Yes, the*

questioning techniques helped me better understand the person's concerns and structure the conversation accordingly", as well as a comparatively large number of statements on the difficulty level (∑= 14; 10/22), e. g., *"[. . .] I concentrated more on paraphrasing and mirroring. I found the I-Messages difficult to implement. [. . .] I was unsure whether they were the right questions [. . .]."*

Next to these reflections on the content level, there were a number of reflections on the meta level. 17 participants (∑= 27) chatted about the simulation and study situation, e. g., *"I think one challenge of VR training is that facial expressions and general body language are currently difficult to represent in the digital space. [. . .] However, the VR interface allows for greater anonymity [. . .]."* More than half expressed ideas on what could help or could have helped them in their learning process: *"Observing an experienced counselor and focusing on how they apply techniques could be helpful [. . .]."* One student discussed the debriefing itself three times, e. g., by writing: *"The debriefing forces me [. . .] to reflect on my perceptions and consider whether the counseling was successful or not. It helps me clearly understand what happened in the conversation and realize if and how I applied different techniques"* Only rarely did the students express that they do not know how to answer (∑= 6; 6/22), use conversational phrases like *"thank you"* (∑= 2; 2/22), and only in one case did the student take the lead three times and moderated the conversation, e. g., *"Instead of continuing to discuss the person's situation, I would like to talk about how I can train my counseling skills in the future."*

Codes and Subcodes	**Examples**
Current emotional state of the participant (∑= 35; 20/22)	*I was a bit nervous because I had never done anything like this before. The situation was unfamiliar [. . .] I wasn't sure what to do.*
Positive reflection on own behavior (∑= 16; 13/22)	*Still, I am proud of myself for conducting a counseling conversation, and I am mostly satisfied.*
Neutral reflection of own behavior (∑= 24; 14/22)	*I mainly used open-ended questions because I wanted to learn as much as possible about the client's feelings and thoughts. I also used I-Messages to show that I appreciated the fact that the person already had certain goals in mind.*
Critical reflection on own behavior (∑= 24; 10/22)	*[. . .] I quickly chose the Tree-of-Life method, even though there might have been better alternatives.*

Tab. 2: Exemplary codes for student behavior (∑= number of mentions across chatlog files; x/22 = share of chatlog files with this (sub)code)

5 Discussion

Our findings indicate that the GenAI chatbot predominantly fulfilled its function as a debriefer in accordance with the structured three-phase debriefing model. A key limitation was its tendency toward an overly enthusiastic tone and insufficient critical reflection, consistent with prior observations on chatbot sycophancy [Ma24]. Notably, students appeared unaffected by this bias, maintaining critical self-reflection, which aligns with evidence highlighting learner agency in self-guided debriefing [VMS21].

Despite largely conforming to its assigned role, the chatbot exhibited deviations, particularly when interaction patterns diverged from expectations (e. g., participant JE21). It occasionally breached the "fourth wall" by inserting meta-communicative stage directions (e. g., "waiting for the learner to respond") that potentially disrupted conversational flow [Ka25]. These occurrences may be linked to system parameters—temperature and nucleus sampling fixed at 0.5—and warrant further investigation for optimal configuration to mitigate such behavior.

The probabilistic nature of AI-generated responses complicates consistent conversational quality, as evidenced by limited dynamic engagement and repetitive phrasing [Ba24; Ji24]. Nevertheless, participants did not express dissatisfaction with the chatbot's output and appeared to accommodate its limitations, as reflected by favorable guidance ratings. This suggests that learners' prior awareness of large language model constraints may facilitate adaptive interaction strategies.

From a methodological perspective, employing a relatively small LLM likely contributed to observed limitations. Larger models generally offer enhanced contextual understanding and reduced sycophancy, representing a promising avenue for future research. However, given that many educational institutions have greater access to small- or medium-scale models, our findings demonstrate that complex instructional tasks such as debriefings can be effectively supported by smaller architectures. Future work should systematically evaluate alternative system prompts aimed at curbing exaggerated or overly agreeable responses within structured debriefing protocols.

Critically, this study was conducted with experimenter presence. Subsequent investigations should examine fully autonomous chatbot-based debriefings, which hold promise for resource-efficient and scalable applications, particularly in MOOCs or online training environments with limited human facilitation.

Finally, as participants were university students with foundational knowledge of counseling techniques, generalizability to more diverse or professional populations remains uncertain. Further research should assess differential learner interactions with AI debriefers across varying expertise levels and domains.

In conclusion, notwithstanding occasional role deviations, sycophantic tendencies, and repetitive output, GenAI chatbots demonstrate potential as debriefing facilitators. Learners'

capacity to compensate for chatbot shortcomings underscores their viability as scalable alternatives to human debriefers.

6 Conclusion

The aim of this qualitative study was to examine whether generative chatbots are capable of adhering to theor assigned role as debriefers within more complex conversational settings, and how learners interact with them. The results indicate that chatbots are capable of conducting reflective conversations in the role of a debriefer, thereby supporting learners in reflecting on educational content.

The relevance of this study becomes evident in light of the resource-intensive nature of guided debriefings conducted by instructors, as recommended by the International Nursing Association for Clinical Simulation and Learning [Co16]. In higher education in particular, the personnel effort required to conduct individual debriefings is often unsustainable [Ch17]. Our findings support the notion that chatbots can effectively guide students through reflective processes. As such, AI-supported debriefings may serve as a ressource-efficient complement to moderated debriefing in (higher) education.

Building on these findings, future research should focus especially on the quantitative evaluation of chatbot debriefings and systematically compare their effectiveness with human-led debriefing sessions.

References

[At21] Atthill, S. et al.: Exploring the Impact of a Virtual Asynchronous Debriefing Method after a Virtual Simulation Game to Support Clinical Decision-Making. en, Clinical Simulation in Nursing 50, pp. 10–18, 2021, DOI: 10.1016/j.ecns.2020.06.008.

[Ba24] Bai, Y. et al.: Investigating the Efficacy of ChatGPT-3.5 for Tutoring in Chinese Elementary Education Settings. IEEE Transactions on Learning Technologies 17, pp. 2102–2117, 2024, DOI: 10.1109/TLT.2024.3464560.

[Bo11] Boet, S. et al.: Looking in the mirror: Self-debriefing versus instructor debriefing for simulated crises*: en, Critical Care Medicine 39 (6), pp. 1377–1381, 2011, DOI: 10.1097/CCM.0b013e31820eb8be.

[Bo13] Boet, S. et al.: Within-Team Debriefing Versus Instructor-Led Debriefing for Simulation-Based Education: A Randomized Controlled Trial. en, Annals of Surgery 258 (1), pp. 53–58, 2013, DOI: 10.1097/SLA.0b013e31829659e4.

[Ch17] Cheng, A. et al.: Coaching the debriefer: peer coaching to improve debriefing quality in simulation programs. Simulation in Healthcare 12 (5), pp. 319–325, 2017.

[CH23] Chan, C. K. Y.; Hu, W.: Students' voices on generative AI: perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education 20 (1), p. 43, 2023, DOI: 10.1186/s41239-023-00411-8.

[Co16] Committee, I. S. et al.: INACSL standards of best practice: SimulationSM debriefing. Clinical Simulation in Nursing 12, S21–S25, 2016.

[Co68] Cohen, J.: Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological Bulletin 70 (4), Place: US Publisher: American Psychological Association, pp. 213–220, 1968, DOI: 10.1037/h0026256.

[Cr23] Crookall, D.: Debriefing: A Practical Guide. en, Simulation for Participatory Education, 2023.

[Da23] Dai, W. et al.: Can Large Language Models Provide Feedback to Students? A Case Study on ChatGPT. In: 2023 IEEE International Conference on Advanced Learning Technologies (ICALT). IEEE, Orem, UT, USA, pp. 323–325, 2023, DOI: 10.1109/ICALT58122.2023.00100.

[DY14] Dufrene, C.; Young, A.: Successful debriefing — Best methods to achieve positive learning outcomes: A literature review. en, Nurse Education Today 34 (3), pp. 372–376, 2014, DOI: 10.1016/j.nedt.2013.06.026.

[Ev25] Evangelou, D.: Online Supplement_Chatbots as Debriefers, 2025, https://osf.io/qe5xv.

[Fa24] Favolise, M.: Post-Simulation Debriefing Methods: A Systematic Review. en, Archives of Physical Medicine and Rehabilitation 105 (4), e146, 2024, DOI: 10.1016/j.apmr.2024.02.680.

[Ga15] Garden, A. L. et al.: Debriefing after Simulation-Based Non-Technical Skill Training in Healthcare: A Systematic Review of Effective Practice. en, Anaesthesia and Intensive Care 43 (3), pp. 300–308, 2015, DOI: 10.1177/0310057X1504300303.

[Gr10] Grant, J. S. et al.: Using Video-Facilitated Feedback to Improve Student Performance Following High-Fidelity Simulation. en, Clinical Simulation in Nursing 6 (5), e177–e184, 2010, DOI: 10.1016/j.ecns.2009.09.001.

[He09] Hertel, S.: Beratungskompetenz von Lehrern. Waxmann Verlag, 2009.

[Ho23] Hong, W. C. H.: The impact of ChatGPT on foreign language teaching and learning: opportunities in education and research. en, Journal of Educational Technology and Innovation 5 (1), Number: 1, pp. 37–45, 2023, https://jeti.thewsu.org/index.php/cieti/article/view/103.

[Ji24] Jin, H. et al.: Teach AI How to Code: Using Large Language Models as Teachable Agents for Programming Education. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. CHI '24, Number: 652, Association for Computing Machinery, New York, NY, USA, pp. 1–28, 2024, DOI: 10.1145/3613904.3642349.

[JLC23] Jeon, J.; Lee, S.; Choi, S.: A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models. Interactive Learning Environments 0 (0), pp. 1–19, 2023, DOI: 10.1080/10494820.2023.2204343.

[JNH23] Järvelä, S.; Nguyen, A.; Hadwin, A.: Human and artificial intelligence collaboration for socially shared regulation in learning. en, British Journal of Educational Technology 54 (5), pp. 1057–1076, 2023, DOI: 10.1111/bjet.13325.

[Ka25] Kaate, I. et al.: The 'fourth wall' and other usability issues in AI-generated personas: comparing chat-based and profile personas. en, Behaviour & Information Technology, pp. 1–17, 2025, DOI: 10.1080/0144929X.2025.2469659.

[Kh22] Khosrawi-Rad, B. et al.: Conversational Agents in Education – A Systematic Literature Review. In: ECIS 2022 Research Papers. Timisoara, Romania, p. 18, 2022, https://aisel.aisnet.org/ecis2022_rp/18.

[Kl24] Klar, M.: How should we teach chatbot interaction to students? A pilot study on perceived affordances and chatbot interaction patterns in an authentic K-12 setting. In. Gesellschaft für Informatik e.V., 2024, DOI: 10.18420/delfi2024.

[KR20] Kuckartz, U.; Rädiker, S.: Fokussierte Interviewanalyse mit MAXQDA. Schritt Schritt, pp. 55–74, 2020.

[KS07] Kriz, W. C.; Saam, N. J.: Großgruppenplanspiele als Interventionsmethode. de, Planspiele für die Organisationsentwicklung. Schriftenreihe: Wandel und Kontinuität in Organisationen, 2007.

[Ku19] Kuckartz, U.: Qualitative Content Analysis: From Kracauer's Beginnings to Today's Challenges. en, Forum Qualitative Sozialforschung / Forum: Qualitative Social Research 20 (3), p. 20, 2019, DOI: 10.17169/fqs-20.3.3370.

[Le24] Lee, U. et al.: Few-shot is enough: exploring ChatGPT prompt engineering method for automatic question generation in english education. en, Education and Information Technologies 29 (9), pp. 11483–11515, 2024, DOI: 10.1007/s10639-023-12249-8.

[Lu21] Luctkar-Flude, M. et al.: Effectiveness of Debriefing Methods for Virtual Simulation: A Systematic Review. en, Clinical Simulation in Nursing 57, pp. 18–30, 2021, DOI: 10.1016/j.ecns.2021.04.009.

[Ma24] Malmqvist, L.: Sycophancy in Large Language Models: Causes and Mitigations, 2024, DOI: 10.48550/arXiv.2411.15287.

[MHC07] Metcalfe, S. E.; Hall, V. P.; Carpenter, A.: Promoting Collaboration in Nursing Education: The Development of a Regional Simulation Laboratory. en, Journal of Professional Nursing 23 (3), pp. 180–183, 2007, DOI: 10.1016/j.profnurs.2007.01.017.

[Mo22] Molenaar, I.: The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning. Computers and Education: Artificial Intelligence 3, Number: 100070, 2022, DOI: 10.1016/j.caeai.2022.100070.

[Na23] Naveed, H. et al.: A Comprehensive Overview of Large Language Models, 2023, arXiv: 2307.06435.

[Ng23] Ng, D. T. K. et al.: Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. en, Educational technology research and development 71 (1), pp. 137–161, 2023, DOI: 10.1007/s11423-023-10203-6.

[PFS16] Palaganas, J. C.; Fey, M.; Simon, R.: Structured Debriefing in Simulation-Based Education. en, AACN Advanced Critical Care 27 (1), pp. 78–85, 2016, DOI: 10.4037/aacnacc2016328.

[Re12] Reed, S. J.: Debriefing Experience Scale: Development of a Tool to Evaluate the Student Learning Experience in Debriefing. Clinical Simulation in Nursing 8 (6), e211–e217, 2012, DOI: 10.1016/j.ecns.2011.11.002.

[Ro57] Rogers, C. R.: The necessary and sufficient conditions of therapeutic personality change. Journal of Consulting Psychology 21 (2), Place: US Publisher: American Psychological Association, pp. 95–103, 1957, DOI: 10.1037/h0045357.

[Ro62] Rogers, C. R.: The interpersonal relationship: The core of guidance. Harvard Educational Review 32 (4), Place: US Publisher: Harvard Education Publishing Group, pp. 416–429, 1962.

[SFE16] Sawyer, T.; Fleegler, M. B.; Eppich, W. J.: Essentials of Debriefing and Feedback. In (Grant, V. J.; Cheng, A., eds.): Comprehensive Healthcare Simulation: Pediatrics. Springer International Publishing, Cham, pp. 31–42, 2016, DOI: 10.1007/978-3-319-24187-6_3.

[Sk24] Skulmowski, A.: Placebo or Assistant? Generative AI Between Externalization and Anthropomorphization. en, Educational Psychology Review 36 (2), p. 58, 2024, DOI: 10.1007/s10648-024-09894-x.

[So25] Song, Y. et al.: Interactions with generative AI chatbots: unveiling dialogic dynamics, students' perceptions, and practical competencies in creative problem-solving. en, International Journal of Educational Technology in Higher Education 22 (1), Number: 12, 2025, DOI: 10.1186/s41239-025-00508-2.

[SQ24] Sommer, B.; von Querfurth, S.: "In the end, the story of climate change was one of hope and redemption": ChatGPT's narrative on global warming. en, Ambio 53 (7), pp. 951–959, 2024, DOI: 10.1007/s13280-024-01997-7.

[SW25] Sun, Y.; Wang, T.: Be Friendly, Not Friends: How LLM Sycophancy Shapes User Trust, 2025, DOI: 10.48550/arXiv.2502.10844.

[Th11] The Inascl Board Of Directors: Standard VI: The Debriefing Process. en, Clinical Simulation in Nursing 7 (4), S16–S17, 2011, DOI: 10.1016/j.ecns.2011.05.010.

[Th90] Thatcher, D. C.: Promoting learning through games and simulations. Simulation & Gaming 21 (3), pp. 262–273, 1990.

[Ti13] Tilton, K. J.: Non-acute Care Clinical for BSN Students Chronic Illness Care and Diabetes Self-management Support. en, Doctor of Nursing Practice Systems Change Projects. 2013.

[TTH15] Tilton, K. J.; Tiffany, J.; Hoglund, B. A.: Non-Acute-Care Virtual Simulation: Preparing Students to Provide Chronic Illness Care: en, Nursing Education Perspectives 36 (6), pp. 394–395, 2015, DOI: 10.5480/14-1532.

[Ve18] Verkuyl, M. et al.: Comparison of Debriefing Methods after a Virtual Simulation: An Experiment. en, Clinical Simulation in Nursing 19, pp. 1–7, 2018, DOI: 10.1016/j.ecns.2018.03.002.

[VMS21] Verkuyl, M.; MacKenna, V.; St-Amant, O.: Using self-debrief after a virtual Simulation: the process. Clinical Simulation in Nursing 57, pp. 48–52, 2021.

[Wa24] Wang, K.: From ELIZA to ChatGPT: A brief history of chatbots and their evolution. en, Applied and Computational Engineering 39, pp. 57–62, 2024, DOI: 10.54254/2755-2721/39/20230579.

[Za23] Zamfirescu-Pereira, J. et al.: Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI '23, Number: 437, Association for Computing Machinery, New York, NY, USA, pp. 1–21, 2023, DOI: 10.1145/3544548.3581388.

Integration realer Werkzeuge in eine VR-Trainingssoftware für komplexe manuelle Montageprozesse



Vergleichsstudie mit und ohne Haptik

David Mack [1] und Ludger Schmidt [1]

Abstract: Die hier vorgestellte Studie betrachtet das Training der Montage eines Notgriffs zur Türentriegelung von Schienenfahrzeugen an einem originalgetreuen Arbeitsplatz mit verschiedenen Trainingsanwendungen. Dazu wurde eine VR-Trainingsanwendung mit realen Montagevorrichtungen und Werkzeugen neu entwickelt und mit einer VR-Trainingsanwendung ohne solche haptischen Elemente verglichen. Für beide Anwendungen wurde dieselbe Videodatenbrille verwendet. Das Training ermöglicht eine niederschwellige Einarbeitung neuer Mitarbeiter ohne Sprachbarrieren und ohne einen Montagearbeitsplatz in der Produktion zu blockieren. Die untersuchten Hypothesen sind, dass die Art des Trainings einen Einfluss auf die Trainingszeit, die spätere Montagezeit, die Anzahl der Fehler, die Gebrauchstauglichkeit der Anwendung und die Beanspruchung des Montagepersonals während des Trainings hat. Beide Trainings wurden in einem Between-Subject-Design mit jeweils fünfzehn Probanden fünfmal eingesetzt. Anschließend wurden drei reale Montagedurchgänge durchgeführt. Sowohl bei den Trainingszeiten als auch bei den Montagezeiten konnten signifikante Unterschiede zwischen den Trainingsanwendungen festgestellt werden, wobei sich bei der Trainingsanwendung mit Haptik längere Trainings-, aber kürzere reale Montagezeiten zeigten.[2]

Keywords: Virtual Reality, Haptik, realitätsnahes Montagetraining, Anlernen, Digitalisierung

1 Einleitung

Trotz der fortschreitenden Digitalisierung in Produktionsbetrieben und dem zunehmenden Einsatz von Robotern, insbesondere kollaborierenden Robotern in der automatisierten Fertigung [Ja23], gibt es nach wie vor Montagearbeitsplätze, die eine rein manuelle Montage vorsehen. Gründe hierfür sind häufig eine zu hohe Variantenvielfalt in Verbindung mit geringen Stückzahlen. Dies führt zu steigenden Anforderungen an die Mitarbeiter [Fr98]. Montagearbeitsplätze weisen in der Regel eine hohe Fluktuation auf, und neue Mitarbeiter müssen häufig eingearbeitet werden [HSS19].

Ein Ansatz zur Verbesserung der Verständlichkeit von Arbeitsanweisungen und Montageanleitungen kann der Einsatz innovativer Methoden wie Augmented Reality (AR) und Virtual Reality (VR) sein [Sc05]. Diese sind in der Industrie noch nicht weit verbreitet, es

[1] Universität Kassel, Fachgebiet Mensch-Maschine-Systemtechnik, Mönchebergstraße 7, 34125 Kassel,
D.Mack@uni-kassel.de, https://orcid.org/0009-0005-6557-6224;
L.Schmidt@uni-kassel.de, https://orcid.org/0000-0003-4581-1336
[2] Forschungsdatenrepositorium: https://doi.org/10.48662/daks-261

doi:10.18420/delfi2025_06

wird jedoch erwartet, dass ihre Verbreitung in den nächsten Jahren zunehmen wird [KS22].

Studien wie die von [We18] zeigen, dass visuelles Feedback in AR-Umgebungen die Effizienz steigern und Fehler reduzieren kann. Dennoch wird die Kombination von visuellem und haptischem Feedback oft nicht ausreichend berücksichtigt. Diese Multimodalität könnte jedoch das Potenzial haben, die Nutzerleistung weiter zu verbessern und eine realistischere und effektivere Trainingsumgebung zu schaffen. Darüber hinaus konzentrieren sich viele Arbeiten auf sehr abstrakte oder spezialisierte Anwendungsbereiche, wie das medizinische Training, ohne einen Bezug zu Montage- und anderen industriellen Tätigkeiten herzustellen, obwohl gerade diese Bereiche von der Kombination realer Objekte und virtueller Informationen profitieren könnten [Ga22]. Ein zentrales Limit vieler VR-Anwendungen ist die ausschließlich visuelle Wahrnehmungsmöglichkeit physischer Objekte. Um dies zu überwinden, konzentrieren sich Studien wie die von [Go24] und [Ro22] auf die flexible Generierung von haptischem Feedback durch künstliche mechanische Systeme, während die Verwendung feststehender natürlicher physischer Objekte wie z. B. Werkzeuge, die eine realistischere Interaktion ermöglichen würden, als einfacherer Lösungsansatz meist unberücksichtigt bleibt. Eine weitere Begrenzung vorliegender Studien ist der mangelnde Transfer des Gelernten von virtuellen Umgebungen auf reale Aufgaben. Auch wenn VR-Technologien in Trainingsszenarien vielversprechende Ergebnisse zeigen, fehlt häufig eine systematische Untersuchung des Transfers der erlernten Fähigkeiten auf die reale praktische Ausführung. Dies ist insbesondere in industriellen Anwendungsfeldern von großer Bedeutung, in denen Effizienz und Präzision in der realen Arbeit von zentraler Relevanz sind [Hö22]. Auch wenn in vielen Studien die Rolle von haptischem Feedback im Trainingskontext betont wird, werden auch diese Feedbacksysteme jedoch häufig nur in abstrakten oder künstlichen Szenarien getestet, wodurch der Bezug zu praktischen Anwendungen, insbesondere in der Montage, fehlt [Kn18].

Ziel dieser Veröffentlichung ist es, diese Lücken zu schließen, indem die Rolle von VR ergänzt mit realen haptisch greifbaren Elementen im Training aus einer ganzheitlichen Perspektive betrachtet wird. Insbesondere die Kombination von haptischem und visuellem Feedback sowie der Transfer von virtuellen Trainingsumgebungen in reale Anwendungsbereiche, wie z. B. die industrielle Montage, werden eingehend untersucht. Ziel ist es, neue Erkenntnisse zur Optimierung von Trainingsmethoden und zur Verbesserung des Lerntransfers zu gewinnen, um das praktische Potenzial dieser Technologien weiter auszuschöpfen. Ein Training ist dann erfolgreich, wenn es die für die Aufgabe nötigen Kenntnisse und Fähigkeiten vermittelt.

Zur Bewertung der Durchführung einer Montageaufgabe können in diesem Zusammenhang objektiv und subjektiv messbare Indikatoren herangezogen werden. Aufgrund der widersprüchlichen Aussagen in bisherigen Studien [Ar22, DS24, DSI25, Ga13, Ga22, We18] wurden die folgenden ungerichteten Hypothesen gebildet: Die Art der Instruktion beeinflusst die Trainingszeit (H1), die Montagezeit (H2) sowie die Anzahl der Montagefehler (H3. Die Art der Instruktion beeinflusst die wahrgenommene Gebrauchstauglichkeit

der Trainingsanwendung (H4) und die subjektive Beanspruchung der Probanden während des Trainings (H5).

2 Methode

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.

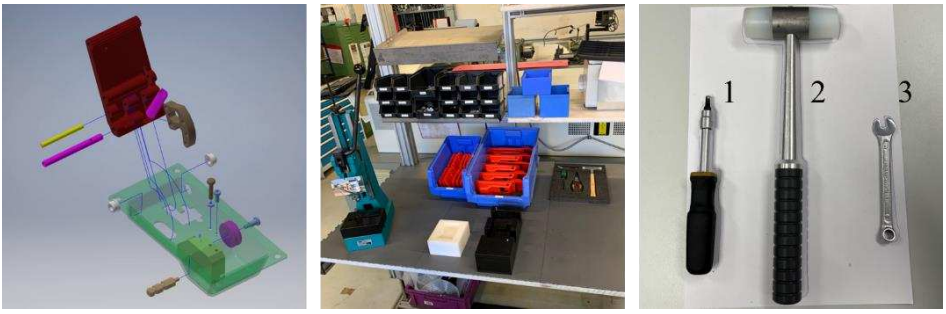


Abb. 1: 3-D-Modell des zu montierenden Notgriffs (links); Montagearbeitsplatz mit Teilelagern, Montagevorrichtungen und einer Handhebelpresse (mittig); für die Montage benötigtes Werkzeug (rechts): Innensechskant-Schraubendreher SW 4 (1), Kunststoffschonhammer (2), Gabelschlüssel SW 8 (3)

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 Handhebelpresse und handgeführtem mechanischem Werkzeug, wie in Abb. 1 dargestellt.

2.1 VR-Montagetraining ohne Haptik

Die Anwendung "VR-Montagetraining ohne Haptik" (siehe Abb. 2) trainiert die Montage des Türnotentriegelungsgriffs ohne haptische Unterstützung. Mit Hilfe des *Meta XR All-In-One SDK (Version 60.0)* und einem von *Meta* entwickelten Plugin für *Unity* wurden die Grundfunktionen der Anwendung umgesetzt. Zunächst wurden das Handtracking und

die Greiflogik implementiert. Die im *Meta SDK* enthaltenen „Building Blocks“ ermöglichen die Integration dieser Funktionen auf Knopfdruck, so dass sie in jede beliebige Anwendung integriert werden können. Der reale Montagearbeitsplatz wurde in der Anwendung „VR-Montagetraining ohne Haptik“ aus CAD-Modellen eins zu eins nachgebildet. Durch die genauen Abmessungen der CAD-Modelle konnte sichergestellt werden, dass die Objekte die richtige Größe und die gleichen Proportionen haben wie der echte Montagearbeitsplatz.

Die für den jeweiligen Schritt benötigten Bauteile werden durch eine grüne Umrandung visuell hervorgehoben und können dann mit einer oder beiden Händen gegriffen und ausgerichtet werden. Die korrekte Positionierung der Bauteile wird durch halbtransparente digitale Zwillinge der Bauteile vorgegeben, die die Zielposition anzeigen. Sobald das Bauteil in die vorgegebene Position gebracht und losgelassen wird, rastet es automatisch in der richtigen Position ein und der nächste Montageschritt wird eingeleitet.

Um den Benutzer durch die einzelnen Montageschritte zu führen, werden außerdem Pfeile und Hinweise eingeblendet, die die erforderliche Bewegungs- oder Drehrichtung anzeigen. Für Schritte, die nicht hinreichend gut direkt in VR dargestellt werden können, wie z. B. die Unterscheidung von Ober- und Unterseite, erscheinen gelbe Ausrufezeichen über dem jeweiligen Bauteil, um diese Besonderheiten hervorzuheben. Diese Hinweise werden auf einem virtuellen Bildschirm, der zentral über dem virtuellen Montagetisch positioniert ist, durch Bilder der realen Montage ergänzt, die mit Emojis und Symbolen versehen sind. Diese Bilder geben nicht nur Anweisungen für die richtige Positionierung, sondern auch zusätzliche Hinweise, die den Montageprozess erleichtern sollen.

Die Anzahl der Durchläufe wird durch einen Zähler in der rechten oberen Ecke des Bildschirms angezeigt. Nach Beendigung eines Durchlaufs erscheint auf der Arbeitsfläche ein roter Knopf, dessen Betätigung die Anwendung zurücksetzt und den nächsten Durchlauf startet. Alle Durchläufe sind identisch.



Abb. 2: VR-Trainingsanwendung ohne Haptik

2.2 VR-Montagetraining mit Haptik

Die Darstellung der Anwendung „VR-Montagetraining mit Haptik“ (siehe Abb. 3 links) erfolgte mit der gleichen Entwicklungsumgebung und der gleichen Videodatenbrille. Die Basis der Anwendung war das zuvor beschriebene VR-Montagetraining ohne Haptik. Dadurch wurden in der Anwendung die gleichen Komponenten verwendet, um eine Vergleichbarkeit der Varianten gewährleisten zu können.

Als reale haptische Elemente werden der Montagetisch, die Handhebelpresse, der Hammer und die Montagevorrichtungen verwendet (siehe Abb. 3 rechts). In der Anwendung wurden die zuvor virtuellen Montagevorrichtungen ausgeblendet, da sie außer ihrer visuellen Darstellung keine weitere Funktion in der Anwendung erfüllen. Auf diese Weise können dann die virtuellen Bauteile in die realen Montagevorrichtungen eingesetzt werden, was die Interaktion verändert, da die Nutzer ein physisches Hindernis spüren. Der reale Tisch, der wie die haptischen Elemente durch die Videodatenbrille sichtbar ist, dient als Auflagefläche für Hand, Arm und reale Werkzeuge. Im Fall der Handhebelpresse und der Werkzeuge wurden deren virtuelle Texturen ausgeschaltet, sodass sie in der Anwendung unsichtbar erscheinen, aber weiterhin auch virtuell gegriffen und für Interaktionen verwendet werden können. Obwohl die Presse und die Werkzeuge in der Realität vorhanden und benutzbar sind, liegt also ein unsichtbares virtuelles Objekt darunter, um die Bedingungen zum Weiterschalten der Anwendung zu erfüllen. Dies sorgt dafür, dass die Nutzer das Gefühl haben, echte Werkzeuge für die Montageschritte zu verwenden, während gleichzeitig die virtuelle Interaktion gewährleistet bleibt. Um Fehler wie das versehentliche Nichtgreifen des virtuellen Teils zu vermeiden, wurde eine grüne Umrandung um die zu greifenden virtuellen Komponenten hinzugefügt, sobald diese verwendet werden müssen.



Abb. 3: VR-Trainingsanwendung mit Haptik (links); Tisch mit haptischen Elementen (rechts)

2.3 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 Trainingszeiten wurden im Nachgang der Studie aus den VR-Aufnahmen, die mittels der Software *OBS-Studie* aufgezeichnet wurden, ausgewertet. Die erhobenen realen Montagezeiten und die Anzahl der Montagefehler (aufgeteilt in „falsches Bauteil“ und „Bauteil falsch montiert“) wurden mittels Videoaufzeichnung der Probanden ermittelt.

Zur Ermittlung der Gebrauchstauglichkeit der Anwendungen 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.

2.4 Verwendete Hardware

Die in der Studie verwendete Videodatenbrille *Meta Quest 3* (siehe Abb. 4 links) ermöglicht die VR-Sicht sowie – mittels zweier außen angebrachter Kameras – die Video-Durchsicht auf die Umgebung durch zwei Pancake-Linsen mit einer Auflösung von 2064 x 2208 Pixeln pro Auge. Das horizontale Sichtfeld beträgt 110 Grad, das vertikale etwa 96 Grad. Zusätzlich sind noch vier IR-Sensoren und ein Tiefensensor eingebaut. Durch das zusätzliche Akkupack ist das Gewicht auf dem Kopf ausbalanciert. Es gab seitens der Versuchsteilnehmer keine Beschwerden bezüglich des Tragekomforts.

2.5 Versuchsablauf

Zu Beginn des Versuchs, der einem Between-Subject-Design folgte, wurden die Probanden über Inhalt und Ablauf der Studie aufgeklärt und unterschrieben eine Einverständniserklärung inklusive Datenschutzerklärung. Die Studie begann mit dem Ausfüllen eines demographischen Fragebogens, der Angaben zu Alter, Geschlecht, Sehhilfe und Vorerfahrungen mit VR-Anwendungen enthielt. Anschließend erhielten die Teilnehmenden eine

Einführung in die Nutzung der Videodatenbrille und in die Aufgabenstellung. Nach Klärung offener Fragen begann die Trainingsaufgabe, bei der fünf Notgriffe mit oder ohne Haptik virtuell montiert werden mussten (siehe Abb. 4 links). Nach dem Training füllten die Probanden die Fragebögen TA-EG, SUS und NASA-TLX aus.

Anschließend montierten die Probanden ohne weitere Anleitung insgesamt drei Notgriffe am realen Arbeitsplatz. (siehe Abb. 4 rechts) Während der Montage wurden die Probanden mit Hilfe von Schildern auf die beiden Fehlerarten „falsches Bauteil“ und „falsche Montage“ hingewiesen. Im Anschluss hatten die Probanden die Möglichkeit, vier offene Fragen zu beantworten und so ein Feedback zu geben.



Abb. 4: Nutzung der Trainingsanwendung mit Haptik (links) und Durchführung der Montageaufgabe am realen Arbeitsplatz (rechts)

2.6 Stichprobe

Die VR-Trainingsanwendung mit Haptik wurde von fünfzehn Probanden genutzt. Es nahmen zwei Frauen und dreizehn Männer mit einem Durchschnittsalter von 23,3 Jahren (SD = 2,3 Jahre) teil. Die VR-Trainingsanwendung ohne Haptik wurde ebenfalls von fünfzehn Probanden genutzt. Davon nahmen sechs Frauen und neun Männer mit einem Durchschnittsalter von 32,0 Jahren (SD = 14,5 Jahre) teil.

Ein Mann-Whitney-U-Test zwischen den beiden Gruppen, der aufgrund einer Verletzung der Normalverteilungseigenschaft laut Shapiro-Wilk-Test durchgeführt wurde, ergab einen signifikanten Unterschied im Alter der Probanden ($U = 53{,}00$; $Z = -2{,}48$; $p = 0{,}012$; $r = -0{,}453$). Es wurden keine signifikanten Gruppenunterschiede in den mittleren TA-EG-Werten zwischen den Probanden der Anwendung mit Haptik ($M = 3{,}70$, $SD = 0{,}34$) und ohne Haptik ($M = 3{,}59$, $SD = 0{,}63$) gefunden ($t(28) = 0{,}617$, $p = 0{,}542$; $r = 0{,}116$). Da diese Daten normalverteilt sind, wurde der ungepaarte t-Test durchgeführt.

Ein χ^2-Test wurde durchgeführt, um die Unterschiede in den Häufigkeiten von Gruppenzugehörigkeit und Vorerfahrung mit VR-Brillen auf Signifikanz zu überprüfen. Alle erwarteten Zellhäufigkeiten waren größer als fünf. Es gab keinen statistisch signifikanten Zusammenhang zwischen der Gruppenzugehörigkeit und Vorerfahrung mit VR-Brillen, $\chi^2(1) = 3{,}33$; $p = 0{,}068$; $\varphi = 0{,}68$.

3 Ergebnisse

Vor der Auswertung wurden alle Daten mit dem Shapiro-Wilk-Test auf Normalverteilung und mit dem Levene-Test auf Varianzhomogenität geprüft. Bei Normalverteilung wurde dementsprechend ein t-Test oder ein Welch-Test für unabhängige Stichproben durchgeführt. Für nicht normalverteilte Werte wurde der Mann-Whitney-U-Test verwendet.

Mit Haptik dauerte das Training insgesamt durchschnittlich 20,84 min (SD 4,63 min), ohne 18,62 min (SD 6,08 min). Die Trainingszeiten der fünf Durchgänge sind in Abb. 5 dargestellt.

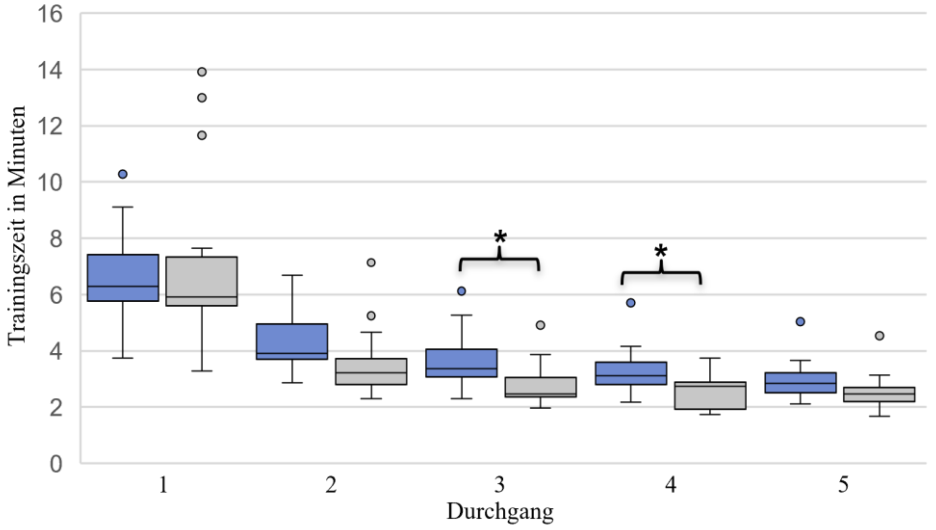


Abb. 5: Trainingszeiten der Gruppe mit (blau) und ohne Haptik (grau)

Bei Hypothese H1 werden die Unterschiede in den Trainingszeiten zwischen den VR-Anwendungen mit und ohne Haptik untersucht. Tab. 1 zeigt die Mittelwerte, Standardabweichungen und die Ergebnisse der Mann-Whitney-U-Tests.

	mit Haptik	ohne Haptik	U-Test
	M [min] (SD [min])		
Dg. 1	6,63 (1,72)	7,17 (3,18)	U = 107; Z = -0,23; p = 0,838[k]; r = -0,042
Dg. 2	4,29 (1,08)	3,55 (1,28)	U = 57,0; Z = -2,30; p = 0,06[k]; r = -0,42
Dg. 3	3,69 (0,98)	2,84 (0,78)	U = 48,5; Z = -2,66; p = 0,035*[k]; r = -0,485
Dg. 4	3,27 (0,86)	2,53 (0,66)	U = 52,0; Z = -2,51; p = 0,044*[k]; r = -0,458
Dg. 5	2,97 (0,73)	2,53 (0,70)	U = 67,0; Z = -1,89; p = 0,12[k]; r = -0,345
Ges.	20,84 (4,63)	18,62 (6,08)	U = 64,0; Z = -2,01; p = 0,045*; r = -0,367

Tab. 1: Mittelwerte und Standardabweichungen der Trainingszeiten (pro Durchgang und insgesamt) der Trainingsgruppe mit Haptik und der Trainingsgruppe ohne Haptik sowie Ergebnisse der statistischen Auswertung des Gruppenvergleichs ([k]Bonferroni-Holm-korrigiert)

Die anschließende Montage dauerte nach vorherigem Training mit Haptik insgesamt durchschnittlich 14,01 min (SD 3,56 min), ohne 20,20 min (SD 6,06 min). Die Montagezeiten der drei Durchgänge sind in Abb. 6 dargestellt.

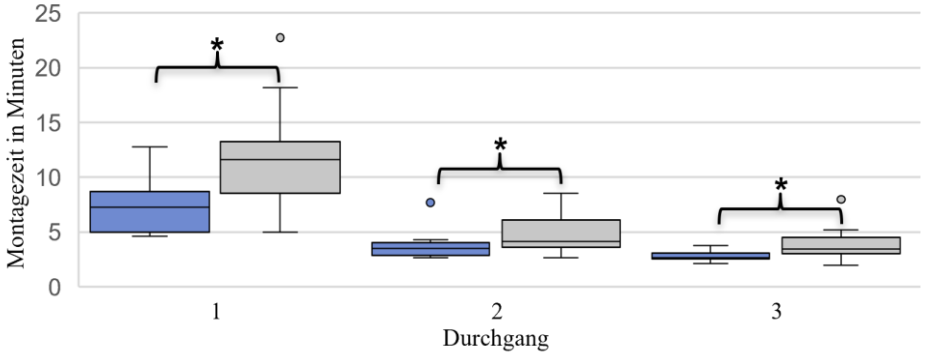


Abb. 6: Montagezeiten der Trainingsgruppe mit (blau) und ohne Haptik (grau)

Bezüglich Hypothese H2 werden die Unterschiede in den Montagezeiten zwischen den VR-Anwendungen mit und ohne Haptik überprüft. Tab. 2 zeigt die Mittelwerte, Standardabweichungen und die Ergebnisse des t- bzw. Mann-Whitney-U-Tests.

	mit Haptik	ohne Haptik	t-test/U-Test
	M [min] (SD [min])		
Dg. 1	7,54 (2,74)	11,47 (4,67)	t(28,00) = -2,83; p = 0,027*[k]; r = 0,471
Dg. 2	3,63(1,26)	4,89 (2,03)	U = 61,00; Z = -2,14; p = 0,033*[k]; r = -0,39
Dg. 3	2,83 (0,49)	3,83 (1,50)	U = 52,00; Z = -2,51; p = 0,027*[k]; r = -0,458
Ges.	14,01 (3,56)	20,20 (6,06)	U = 47,50; Z = -2,70; p = 0,006**; r = -0,492

Tab. 2: Mittelwerte und Standardabweichungen der Montagezeiten (pro Durchgang und insgesamt) der Trainingsgruppe mit Haptik und der Trainingsgruppe ohne Haptik sowie Ergebnisse der statistischen Auswertung des Gruppenvergleichs ([k]Bonferroni-Holm-korrigiert)

Bei Hypothese H3 wird die Anzahl der Fehler nach den beiden Trainingsarten untersucht. Insgesamt machte die Gruppe nach dem VR-Training mit Haptik mit durchschnittlich 4,8 Fehlern (SD = 1,78 Fehler) bei der Montage insgesamt signifikant weniger Fehler als die Gruppe mit dem Training ohne Haptik mit durchschnittlich 13,07 Fehlern (SD = 5,84 Fehler) (t(16.584) = -5,247; p < 0,001***; r = 0,79). Abb. 7 zeigt die durchschnittliche Anzahl der Fehler pro Montagedurchgang.

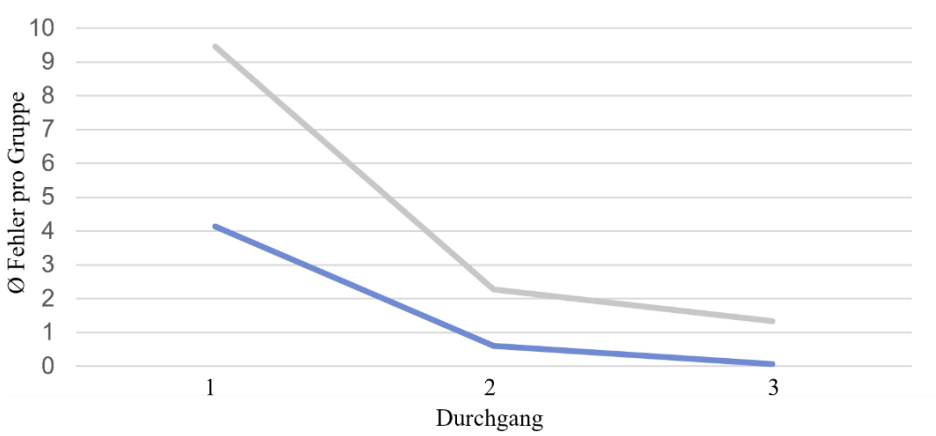


Abb. 7: Durchschnittliche Fehler pro Montagedurchgang der Trainingsgruppe mit (blau) und ohne Haptik (grau)

Bei den Hypothesen H4 (Gebrauchstauglichkeit) und H5 (Beanspruchung) wird geprüft, inwieweit sich die VR-Anwendung mit von der VR-Anwendung ohne Haptik unterscheidet. Die Boxplots der beiden Größen sind in Abb. 8 dargestellt.

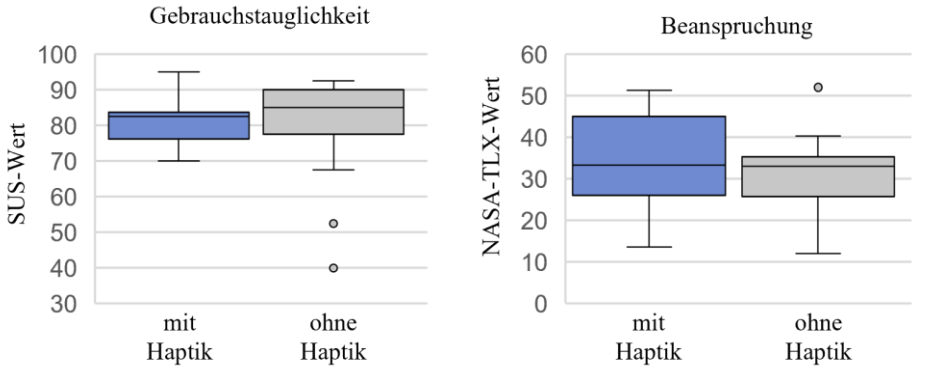


Abb. 8: Boxplots zur Gebrauchstauglichkeit (links) und Beanspruchung (rechts)

Die durchschnittliche Gebrauchstauglichkeit ist bei der VR-Anwendung mit Haptik mit 81,00 (SD = 6,87) minimal größer als ohne Haptik mit 80,00 (SD = 15,50). Die Unterschiede sind jedoch nicht signifikant (U = 88,00; Z = -1,021; p = 0,317, r = -0,186). Die mittlere Beanspruchung ist in der Trainingsgruppe mit Haptik mit 34,56 (SD = 11,61) minimal größer als ohne Haptik mit 31,18 (SD = 9,39). Es liegt auch hier kein signifikanter Unterschied vor (t(28) = 0,876; p = 0,388; r = 0,163).

In der abschließenden Befragung wurde der Trainingsgruppe mit haptischem Feedback die Frage gestellt: *Haben die echten Elemente in der Trainingsanwendung dabei geholfen, die echte Montageaufgabe zu lösen?* Das Ziel bestand darin, den Einfluss realer haptischer Komponenten auf das Training und den Praxistransfer zu bewerten. Beiden Trainingsgruppen – mit und ohne Haptik – wurden zudem folgende Fragen zur qualitativen Evaluation gestellt: *Was ist Ihnen besonders positiv aufgefallen? Was ist Ihnen besonders negativ aufgefallen? Gibt es sonstige Anmerkungen, Infos, ...?* Diese offenen Fragen dienten der Erfassung individueller Eindrücke und Verbesserungsvorschläge. Die Befragungsergebnisse sind hier aus Platzgründen nicht dargestellt, können jedoch in den begleitend veröffentlichten Daten im Detail eingesehen werden [MS25].

4 Diskussion und Fazit

Sowohl für die Trainingszeiten als auch für die Montagezeiten ließ sich ein signifikanter Unterschied zwischen den Varianten mit und ohne haptisches Feedback feststellen. Es zeigte sich eine mittlere Effektstärke [Co88].

Da die Varianten bis auf die haptischen Elemente identisch waren, könnte der signifikante Effekt auf die zusätzlichen haptischen Elemente zurückgeführt werden. Möglicherweise hatten die haptischen Elemente einen bremsenden Effekt und ließen die Probanden bewusster durch die Anwendung gehen. Gerade die Montagevorrichtungen wurden in den Befragungsaussagen mehrfach positiv erwähnt. Auch die reduzierte Immersion durch den Video-Durchsicht-Modus bei der Anwendung mit Haptik könnte zu einem verbesserten Trainingseffekt geführt haben. Darauf deuten Befragungsaussagen der Teilnehmenden hin.

Die Montagezeiten zeigen bei allen drei Durchgängen einen signifikanten Unterschied zu Gunsten der vorherigen VR-Anwendung mit Haptik, ebenfalls mit mittlerer Effektstärke. Vergleicht man die vorliegenden Daten mit einer anderen Publikation mit der gleichen Montage [MFS24], bei der eine VR-Anwendung ohne Haptik auf einer älteren Videodatenbrille mit einem Trainingsvideo verglichen wurde, so zeigt sich, dass gerade beim ersten Montagedurchgang die Zeiten bei der dortigen VR-Anwendungen ähnlich zu den Zeiten bei der jetzigen VR-Anwendung ohne Haptik waren. Dies weist ebenfalls darauf hin, dass die haptischen Elemente einen deutlichen Effekt auf den Trainingserfolg in Bezug auf die spätere Montagezeit haben.

Kritisch muss die Zusammensetzung der verschiedenen Gruppen betrachtet werden, insbesondere im Hinblick auf die Unterschiede beim Alter. Nach [DSI25] hat das Alter der Probanden einen deutlichen Einfluss auf die Leistung bei der Verwendung von VR-Trainingsanwendungen in der Montage. Dies kann aber bei einem detaillierten Blick auf den Datensatz ausgeschlossen werden, da die zwei Probanden, die in der Gruppe ohne Haptik deutlich älter waren als der Mittelwert, nicht deutlich schlechtere oder bessere Montagezeiten hatten als der Rest der Gruppe. Die Stichprobengröße von n = 30 ist ein weiterer Faktor, der die Aussagekraft limitiert.

Die Ergebnisse der Fehler während der Montage zeigen einen signifikanten Unterschied mit hoher Effektstärke [Co88]. Die Probanden in der Variante mit Haptik machten deutlich weniger Fehler als in der Variante ohne Haptik. Vergleicht man die Anzahl der Fehler mit den Daten aus der Publikation von [MFS24], so zeigt sich, dass die Anzahl der Fehler bei der Variante ohne Haptik ähnlich hoch ist wie bei der vorherigen Variante ohne haptische Elemente. Dies weist darauf hin, dass die haptischen Elemente einen deutlichen Effekt auf den Trainingserfolg in Bezug auf die Montagefehler haben.

Die Gebrauchstauglichkeit beider VR-Anwendungen wurde mit „gut“ bewertet [BKM09]. Es gab keine signifikanten Unterschiede zwischen den Varianten, die guten Werte sprechen insgesamt für einen hohen Reifegrad. Zur Einordnung der ähnlichen durchschnittlichen subjektiven Beanspruchungswerte bei beiden Varianten wurden Ergebnisse anderer Studien herangezogen: In einer Metaanalyse 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 können die vorliegenden Werte einer eher geringen Beanspruchung zugeordnet werden.

Insgesamt deuten die signifikanten Unterschiede darauf hin, dass ein VR-Training mit Haptik zu besseren Lernergebnissen bezüglich der sich anschließenden realen Montage führt.

Zum Zeitpunkt der Erstellung der Anwendung gab es aufgrund sicherheitstechnischer Maßnahmen und Datenschutzbedenken keine Möglichkeit, auf die Echtzeitkameradaten der Videodatenbrille zuzugreifen und somit war keine Objekterkennung möglich [Meta24]. Die einzige Methode bestand darin, die Interaktion von realen Objekten mit der virtuellen Welt durch virtuelle Überlagerungen von transparenten Objekten zu ermöglichen. Mit der Veröffentlichung des *Meta XR Core SDK v74* wurde diese Einschränkung aufgehoben und es ist nun möglich, eine Objekterkennung in Echtzeit durchzuführen. Als nächster Schritt ist, geplant, diese zu verwenden, um die haptischen Elemente auf alle verwendeten Werkzeuge auszuweiten.

Förderhinweis

Das Projekt „Zukunftszentrum für menschzentrierte KI in der Produktionsarbeit (ZUKIPRO)“ wird im Rahmen des Programms „Zukunftszentren“ durch das Bundesministerium für Arbeit und Soziales und die Europäische Union über den Europäischen Sozialfonds Plus (ESF Plus) gefördert.

Literaturverzeichnis

[Ar22] Ariansyah, D. et al.: A head mounted augmented reality design practice for maintenance assembly: Toward meeting perceptual and cognitive needs of AR users. Applied ergonomics 98, S. 103597, 2022, DOI: 10.1016/j.apergo.2021.103597.

[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 3/4, S. 114–123, 2009.

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

[Co88] Cohen, J.: Statistical power analysis for the behavioral sciences. Erlbaum, Hillsdale, NJ, 1988.

[DS24] Daling, L. M.; Schlittmeier, S. J.: Effects of Augmented Reality-, Virtual Reality-, and Mixed Reality-Based Training on Objective Performance Measures and Subjective Evaluations in Manual Assembly Tasks: A Scoping Review. Human factors 2/66, S. 589–626, 2024, DOI: 10.1177/00187208221105135.

[DSI25] Daling, L. M.; Schlittmeier, S. J.; Isenhardt, I.: Psychologische und leistungsbezogene Effekte von AR- und VR-basiertem Training in der manuellen Montage. In (Gesellschaft für Arbeitswissenschaft e.V. Hrsg.): Arbeit 5.0: Menschzentrierte Innovationen für die Zukunft der Arbeit. GfA-Press, Sankt Augustin, S. 933–938, 2025.

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

[Ga13] Gavish, N. et al.: Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. Interactive Learning Environments 6/23, S. 778–798, 2013, DOI: 10.1080/10494820.2013.815221.

[Ga22] Gani, A. et al.: Impact of haptic feedback on surgical training outcomes: A Randomised Controlled Trial of haptic versus non-haptic immersive virtual reality training. Annals of medicine and surgery 83, S. 1–6, 2022, DOI: 10.1016/j.amsu.2022.104734.

[Go24] Gomi, R. et al.: InflatableBots: Inflatable Shape-Changing Mobile Robots for Large-Scale Encountered-Type Haptics in VR. CHI ‘24: CHI Conference on Human Factors in Computing Systems, S. 1–14, 2024, DOI: 10.1145/3613904.3642069.

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

[Hö22] Hönig, J. et al.: Mixed-Reality-in-the-Loop Simulation zur Schulung technischer Fachkräfte im Maschinen- und Anlagenbau. In (Härle, C.; Jäkel, J.; Sand, G. Hrsg.): Tagungsband AALE 2022. Wissenstransfer im Spannungsfeld von Autonomisierung und Fachkräftemangel. Hochschule für Technik Wirtschaft und Kultur, Leipzig, 2022, DOI: 10.33968/2022.24.

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

[Ja23] Janson, M.: Mensch und Roboter arbeiten zunehmend Hand in Hand. https://de.statista.com/infografik/24129/absatz-von-konventionellen-und-kollaborativen-industrierobotern-weltweit, Stand: 07.04.2025.

[Ka09] Karrer, K. et al.: Technikaffinität erfassen – der Fragebogen TA-EGDer Mensch im Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme. VDI Verlag GmbH, Düsseldorf, 2009.

[Kn18] Knopp, S. et al.: Using Industrial Robots as Haptic Devices for VR-Training. In (Kiyokawa, K. Hrsg.): 25th IEEE Conference on Virtual Reality and 3D User Interfaces. IEEE, Piscataway, NJ, S. 607–608, 2018, DOI: 10.1109/VR.2018.8446614.

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

[LW12] Lotter, B.; Wiendahl, H.-P.: Montage in der industriellen Produktion. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, DOI: 10.1007/978-3-642-29061-9.

[Meta24] Meta: Implement Passthrough. https://developers.meta.com/horizon/documentation/native/android/mobile-passthrough/, Stand: 08.07.2025.

[MFS24] Mack, D.; Funk, J.; Schmidt, L.: Einsatz von Virtual Reality zum Erlernen eines Montageprozesses. „Arbeitswissenschaft in-the-loop: Mensch-Technologie-Integration und ihre Auswirkung auf Mensch, Arbeit und Arbeitsgestaltung“ H.4.2, 2024.

[MS25] Mack, D.; Schmidt, L.: Erhobene Daten der Veröffentlichung Integration realer Werkzeuge in eine VR-Trainingssoftware für komplexe manuelle Montageprozesse. Universität Kassel, 2025, DOI: 10.48662/daks-261.

[Ro22] Rohith, P. et al.: A Study on the Development of a VR-Haptics Training Module for Cup-Lock Scaffolding. In (Chang, M. Hrsg.): ICALT 2022. International Conference on Advanced Learning Technologies. IEEE, Piscataway, NJ, S. 314–316, 2022, DOI: 10.1109/ICALT55010.2022.00100.

[Sc05] Schmidt, L. et al.: Benutzerzentrierte Gestaltung von Augmented Reality in der Produktion. In (Stary, C. Hrsg.): Mensch & Computer 2005. Kunst und Wissenschaft - Grenzüberschreitung der interaktiven ART. Oldenbourg, München, 2005.

[We18] Werrlich, S. et al.: Assembly Training: Comparing the Effects of Head-Mounted Displays and Face-to-Face Training. In (Chen, J. Y. C.; Fragomeni, G. Hrsg.): Interaction, navigation, visualization, embodiment, and simulation. Springer, Cham, S. 462–476, 2018, DOI: 10.1007/978-3-319-91581-4_35.

Best-Paper-Kandidaten-Kurzbeiträge

Empowering Tutors with Digital Weekly Student Survey Results

Armin Egetenmeier [1] and Sven Strickroth [1]

Abstract: Student feedback on their learning progress or the last lecture is essential for educators to address the actual needs of their students. Approaches such as the One Minute Paper and various survey tools allow educators to collect difficult topics or unresolved questions. However, the information gathered is typically not shared with other persons involved in teaching. This paper presents the results of a group interview with student tutors of an introductory computer science course, who had access to weekly Two-Minute-Paper-like survey results for the lecture. The interview suggests that this information is seen as beneficial for their teaching practice such as preparing their exercise sessions. However, tutors prefer information specifically tailored to the students enrolled to their exercise group. In addition, ideas on how to improve the data gathering and to increase participation are discussed.

Keywords: tutors, technology-enhanced learning, minute paper, feedback, two-minute feedback

1 Introduction

Feedback is an important aspect of the learning and teaching process [HT07]. Not only can students benefit from feedback to improve their learning, but also educators can use insight into student learning to improve their teaching and meet the actual needs of the group. Universities typically use end-of-term surveys to evaluate teaching, but this is too late to adapt teaching during the semester. Therefore, these surveys have limited impact on improving the teaching in practice due to their summative nature and standardized format [Dö05]. To collect formative feedback from students during the semester, some lecturers use technology such as audience response systems (ARS, e. g. [Ló22; MB19]) or simple survey tools to quickly collect specific feedback, e. g. on teaching methods, student comprehension or specific difficulties (e. g., [St05; Vo04]). Such feedback on how teaching is received or perceived is important for educators, especially when using new instructional approaches or technology to enhance their courses. It allows instructors to adapt their teaching to the actual needs of students and to provide targeted support. However, the collected results are typically used only once by the collector, although they have potential value for others.

Introductory and/or large classes are usually accompanied by exercises and tutorial sessions, often conducted by student assistants, to provide students with more individual attention. Such tutorial sessions have been shown to have a high impact on student learning [GS21]. Although, the support for tutors, who run the tutorial sessions, is often limited to generic

[1] LMU Munich, Oettingenstr. 67, 80538 Munich, Germany,
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/delfi2025_07

training programs [Mi19; RG18]. Tutors face similar challenges to those of lecturers, such as student heterogeneity and passivity [Ri21]. To support them throughout the semester, weekly meetings are usually held to discuss the current exercise sheet, to address difficult topics and for reflection [KO23]. Despite this, tutors often have limited or no insight into students' learning. However, some of the insights gained by the lecturer may also be useful for tutors' preparation, such as difficult topics or unanswered questions. Therefore, we analyzed how an existing e-assessment system can be used to easily share results from lecture surveys in order to enhance tutors' insights into "their" exercise group. The following Research Questions (RQ) are investigated: (RQ1) What results from an existing weekly survey are useful for tutors? (RQ2) How can such surveys or systems be improved to better meet the needs of tutors? (RQ3) What are implications for educational technology?

2 Related Research

An established approach to gathering formative feedback is the One Minute Paper (OMP) method in which several short questions are asked, usually at the end of a lecture [CA88]. While primarily intended to address student misunderstandings or misconceptions [ADM21; St05], this method also provides insights for educators. Typically, questions are used to reflect on learning (e. g., "What was the most important thing you learned today?") or to gather open issues (e. g., "What question(s) remained unanswered?") and suggestions for improvement [St05; Vo04]. To streamline this low-tech approach, there are variations of the OMP using technology, such as online approaches [CAL19; Vo04] or the Two-Minute Feedback (2MF, [ES24c]). In addition to the typical questions mentioned above, the 2MF uses closed questions for quick responses. This weekly approach provided valuable insight from the student group to the lecturer [ES24c]. However, as other studies (e. g., [KGF21]) have shown, participation declined significantly over the course of the semester.

Riese et al. [Ri21] discussed challenges of (CS) tutors based on reflective essays. They identified common challenges such as assessment, defining and using best (teaching) practice, and threats to best practice among others things. Although appropriate training and support are needed to overcome existing challenges, technology use is rarely discussed as a possible solution. Egetenmeier and Strickroth [ES24a] investigated teaching practices of CS tutors, their challenges, and their use of technology. Their results show that teaching staff often provides specialized (e-assessment) systems to ease reviewing and correction tasks, however, other tools are not explicitly provided. In general, knowledge about the tutorial group is rather low, unstructured, and gathered during the sessions. General ideas for tools were discussed (e. g., a system for collecting anonymous feedback) to improve tutors' teaching practice. Still, tutors often confine themselves to the tools and information provided. The authors concluded, that tutors need more, structured information. Educational technology can be used to close this gap. For example, various e-assessment systems have been implemented to provide automatic feedback to students and ease the work of tutors (e. g., in programming courses [SOP11; St16]). Features include, for example, automated

(unit) tests, and overviews of passed/failed tests and common errors (as in [St24], but for tutors). There are also ideas to collect learning related data in "offline" environments in a crowd sourcing approach using existing technologies [ES24b], such as using QR codes as a flexible, low effort approach. This approach also targets educators, but is an additional effort for students and the paper does not include a detailed concept or a case study using this idea.

This short overview shows challenges of tutors about insufficient information about the student group. Information about students is typically collected in the lecture using different educational technologies. However, the results are often limited to lecturers and tutors would have to collect data themselves. A concept for sharing (collected) data with tutors is lacking.

3 Study Design and Methodology

Throughout the winter term 2024/2025 of an introductory Computer Science (CS) course at LMU Munich with about 680 enrolled students, a weekly survey was conducted to collect feedback from students for the lecturer. The survey was voluntary and implemented in the e-assessment system GATE [SOP11], which was already used for the lecture. The survey consisted of two parts (cf. 2MF): The first part comprised six closed questions that used a five-point Likert scale (including an option for abstention). These asked about the student's situation ("I feel stressed/overwhelmed by my studies."), preparation and reviewing their notes ("I prepared/reviewed the lecture this week."), understanding ("I could follow the content of the lecture well."), and motivation ("I am motivated to attend the lecture."). The second part included questions on unclear content, suggestions for improvement, and what the students liked most about the last lecture. Survey participation declined over the course of the semester, starting with 82 responses in the first week to one in the last week ($\overline{x} = 17$). The survey results were initially available only to the lecturer. Tutors were granted access to the results in January, but no announcement was sent to them due to a communication issue.

At the end of the term (on February, 6th), a focus group interview was conducted with 12 of the 15 student tutors of the course by one of the authors. The interview was voluntary. It took about one hour, was held in German, and was recorded after consent was given. Afterwards, it was transcribed into a simplified form (without distinguishing between individuals) using noScibe [Dr24] and the researcher's personal notes. Finally, the answers given were analyzed using Mayring's qualitative thematic analysis method [Ma14]. The interview began by asking the tutors to describe what information they use from the e-assessment system GATE in their teaching practices and for their preparation. Next, the results of the weekly feedback were presented and they were asked, how useful the information is for their teaching. Then, they could comment on details (e. g., current representation, timing of the data collection) and suggest new questions that are relevant for them. The interview concluded with an open discussion. In addition, a semi-structured, one hour interview was conducted with the lecturer on March, 5th with a focus on how he perceived the results of the weekly surveys and used the information. The interview concluded with suggestions on how to improve the survey. This part was also recorded, transcribed and analyzed as described above.

4 Results

Tutors reported using the overview features of the e-assessment system GATE, which summarize the most common errors in an assignment based on automated tests, and also to accessing individual submissions in preparation to address them during their tutorial sessions. Particularly, tutors were interested in common and recurring mistakes and using authentic examples to explain difficult concepts or answering individual questions. The group interview revealed that no tutor discovered the results of the weekly lecture survey before the interview. Therefore, all tutors were given the opportunity to inspect the results during the interview through their usual system access. They skimmed individually through past weeks' results for about 5 minutes, followed by a brief explanation of the underlying survey. First reactions were mainly positive (e. g., "*nice feature*", "*great, totally cool*"). The tutors initially regarded the information as helpful, even though the survey focuses on the lecture. They stated that the results would have helped them understand which topics remained unclear and could have been used to stimulate discussion in their tutorial sessions.

Detailed questions were asked about the perceived usefulness for the tutors of each item of the survey (RQ1). In general, the weekly survey is regarded as good and helpful, especially the question about unclear content was seen as useful (assuming the lecturer refers to it in the lecture, as mentioned by the tutors) – also for tutorial sessions to increase engagement and guide the preparation. Other items, such as the stress level, positive highlights of the lecture, student preparation or follow-up of the lecture are perceived as less relevant, because these aspects are beyond the control of the tutors ("*since I am not the person who can cure personal stress*") or just relevant for students themselves. Suggestions for improving the course were only relevant to tutors, if they related to the tutor or the tutorial sessions.

In terms of improvements to the surveys (RQ2), several ideas were provided to increase the relevance for tutors: First, the relevance of the existing weekly survey results could be increased by filtering and highlighting responses from students of the tutors' group ("*Seeing [...] the questions my people asked, that would be pretty cool*"). Nevertheless, all results should be visible, as there is maybe no feedback included from their tutorial group and some difficult topics may be missed. This could help for their preparation. The responses should not be summarized, e. g. using an LLM, unless the number of the responses exceeds a certain level. Second, they added some specific ideas for useful information to get insight into their teaching and how to improve it: These include questions similar to the already asked ones in the weekly survey, but with a focus on the tutorial session (e. g., suggestions for improvement, preparation for the tutorial session, or open questions that arose while working on the assignments). In addition, the question '*'what the students liked about the tutorial session"* could also be helpful to improve the teaching as well as *"was the discussion of the task sufficient"*. The tutors would also like to get insight into the students' (pre-)knowledge, their experiences or the time spent on the assignments to get an idea of what to expect from the group. To prepare the tutorial session, specific questions such as *"What question or topic should be discussed in the tutorial session?"* were suggested to tailor the teaching to the group. Finally, the tutors shared their opinions about the weekly survey to collect feedback

about the tutorial session in general. As a result, responses in tutorial session surveys may come from students who no longer attend a particular session, which limits the usefulness of the data for specific tutors. The tutors reported a performance difference between what they saw while correcting tasks and what they experienced in tutorial sessions. According to the tutors, survey results are most valuable when they reflect the current group of participants and provide actionable insights. Tutors also took the students' perspective. For example, one person explained that participating in the weekly survey (last semester as a participant) was exhausting. Thus, an additional survey (about tutorial sessions) would probably not be answered or be counterproductive. Another person mentioned that scanning a QR code and answering closed questions by clicking buttons is an easy approach to collect data regularly. However, open-ended questions will likely be ignored ("*For the open-ended questions, I would be pretty sure that I wouldn't write anything.*"). An idea came up to merge the surveys, and to add a part especially about tutorial sessions and the tutors. Maybe this could be done during or immediately after the (tutorial) session, since the survey will probably be answered when included in the session time, and this would likely increase participation. Answering later from home will not work, one tutor noted.

The interview with the instructor showed that he taught the course for the first and temporarily last time this semester, so there was less need for him to improve the course design or materials. Nonetheless, he inspected the results regularly and they provided valuable insights (e. g., "*noise level*"), and boosted the lecturers' confidence in his teaching practice (e. g., "*good humor*", "*knowing all content well*"). Reported issues were acknowledged and resolved as far as possible. He usually shared the summative evaluation results with the tutors, however he did not extend this practice to the formative feedback. Thus, he regards granting access to the weekly results to tutors as useful, since they welcome more insights and can ignore irrelevant survey results. In order to increase its relevance for the lecturer, the weekly survey could include questions similar to end-of-term evaluation (e. g., "*The lecturer was easy to understand*") or focus on the tutorial sessions, as limited insight into work of tutors is given.

5 Discussion, Conclusions and Outlook

This article shows the results from a focus group interview with student tutors who had access to the results of a weekly survey from the lecture. The interview focused on exploring the usefulness of existing information and collecting suggestions to improve the relevance of these survey tools for tutors. Although the survey was collected primarily to provide feedback for the lecturer, some aspects could be reused for tutors. In terms of the usefulness of the results (RQ1), feedback such as unanswered questions is considered very helpful for preparation and to initiate discussion in more passive groups. However, aspects that focus on the lecture part and/or goes beyond the control of the tutors are less relevant (e. g., students' stress level). The interview revealed that tutors actively welcome new information as a way to improve their teaching practice. Although the tutors had only skimmed through the previous weeks, they already had a clear sense of what was helpful. While the lecturer

shares evaluation results with tutors, he had not considered sharing the weekly feedback results. Possible reasons for this might include the amount of feedback early in the semester, time constraints, or other academic responsibilities. Additionally, the lecturer may not have been aware that sharing the survey results can be achieved with the existing system.

The question on improvements of the survey (RQ2) revealed ideas to increase the relevance for tutors, e. g. by filtering the answers of students from their group. Another approach is to add questions with a focus on the tutors' teaching practice (e. g., sufficiency of discussion). However, the tutors were ambiguous about getting a weekly feedback. They are curious about how their tutorials help students and get ideas for improving their teaching, but are also afraid of negative feedback. The lecturer also mentioned that negative feedback focusing on a specific tutor or tutorial session can be problematic. In the interview, the tutors adopted the students' perspective, which led to a small debate about participation in weekly surveys (cf. [ES24c]). While the tutors agreed that open-ended questions are presumably not answered by most students, closed questions are seen as a low-effort approach. However, open-ended questions are necessary to gather unresolved issues and suggestions for improvement. Furthermore, merging surveys and reusing the results could be more accepted than answering multiple surveys in the course.

Implications for the design and use of educational technology (RQ3) can be drawn. For example, using systems (e. g., ARS) with a persistent data tracking could support continuous collecting, making student responses available for both lecturers and tutors. However, these systems require a common identifier (e. g., SSO) to track participation, as well as mechanism to associate student responses with specific tutorial sessions. For example, QR codes can be used to track actual attendance (in tutorial sessions) and to conduct surveys. The resulting data can then be filtered and assigned to the responses of individual tutorial sessions. This, however, raises important questions regarding privacy that must be addressed beforehand.

The interviews are limited to one focus group of tutors all coming from the same university and supporting the same course. No tutor had seen the survey results before the interview, thus they only had limited time to interact with the information. This results in only hypothetical assumptions on how the information could be used. However, this should not influence the opinion on how useful each item would be for their teaching. Further research is needed to provide tutors with relevant, additional insights collected with educational tools and analyze their helpfulness over the course of an entire semester.

In conclusion, a seamless access to data collected in the lecture is regarded as helpful and often possible with low effort; however the data needs to be tailored to the tutors needs. This includes specific questions on the tutorial session. Ideas to improve the weekly survey aim to lower the barrier to participate (e. g., scanning a QR code, answer closed questions). Conducting the survey in the tutorial sessions might increase the feeling of being "seen", since the tutors can interact immediately with the results. Also integrating survey results into existing (e-assessment) systems appears promising for improving tutor support and feedback usage. However, such integration is often overlooked by lecturers.

Acknowledgment: This research was supported by the German Federal Ministry of Education and Research (BMBF), grant number 16DHBKI013 (AIM@LMU).

References

[ADM21] Allen, B.; Devlin, M.; McGough, A. S.: Using the one minute paper to gain insight into potential threshold concepts in artificial intelligence courses. In: Proceedings of the 5th Conference on Computing Education Practice. Pp. 21–24, 2021.

[CA88] Cross, K. P.; Angelo, T. A.: Classroom Assessment Techniques. A Handbook for Faculty. 1988.

[CAL19] Campbell, M.; Abel, E. M.; Lucio, R.: The One-minute Paper as a Catalyst for Change in Online Pedagogy. Journal of Teaching in Social Work 39 (4-5), pp. 519–533, 2019, DOI: 10.1080/08841233.2019.1642977.

[Dö05] Döring, N.: Für Evaluation und gegen Evaluitis: Warum kann und wie sollte Lehrevaluation an deutschen Hochschulen verbessert werden. In (Berendt, B.; Voss, H.-P.; Wildt, J., eds.): Neues Handbuch Hochschullehre. Raabe, Berlin, 2005.

[Dr24] Dröge, K.: noScribe. AI-powered Audio Transcription, version 0.6, [Computer software], 2024, https://github.com/kaixxx/noScribe.

[ES24a] Egetenmeier, A.; Strickroth, S.: Challenges of Computer Science Tutors in Their Teaching. In: Proceedings of DELFI 2024 – Die 22. Fachtagung Bildungstechnologien. Gesellschaft für Informatik e.V., 2024, DOI: 10.18420/delfi2024_36.

[ES24b] Egetenmeier, A.; Strickroth, S.: Data Collection in Traditional Learning Environments: New Opportunities for Analytics? In: International Conference on Interactive Collaborative Learning. Springer, pp. 385–393, 2024, DOI: 10.1007/978-3-031-85652-5_39.

[ES24c] Egetenmeier, A.; Strickroth, S.: Scalable Two-Minute Feedback: Digital, lecture-accompanying survey as a continuous feedback instrument. International Journal of Information and Education Technology 14 (4), pp. 573–580, 2024, DOI: 10.18178/ijiet.2024.14.4.2079.

[GS21] Gregg, D.; Shin, S. J.: Why We Will Not Return to Exclusively Face-to-Face Tutoring Post-COVID: Improving Student Engagement Through Technology. Learning Assistance Review v26 (2), pp. 53–79, 2021, DOI: 10.13016/m2vnut-jbcs.

[HT07] Hattie, J.; Timperley, H.: The power of feedback. Review of educational research 77 (1), pp. 81–112, 2007.

[KGF21] Karlsson-Brown, P.; Gibb, A.; Ferri, P.: Using a digital one minute paper in business school classes to increase participation and engagement and improve student experience. 2021.

[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.

[Ló22] López-Jiménez, J. J. et al.: 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.

[Mi19] Mirza, D. et al.: 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.

[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. et al.: Challenges Faced by Teaching Assistants in Computer Science Education Across Europe. In: Proc. ITiCSE. ACM, pp. 547–553, 2021, DOI: 10.1145/3430665.3456304.

[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., eds.): DeLFI 2011. GI-Edition Proceedings, Ges. für Informatik, Bonn, pp. 115–126, 2011.

[St05] Stead, D. R.: A review of the one-minute paper. Active learning in higher education 6 (2), pp. 118–131, 2005.

[St16] Striewe, M.: An architecture for modular grading and feedback generation for complex exercises. Science of Computer Programming 129, pp. 35–47, 2016, DOI: 10.1016/j.scico.2016.02.009.

[St24] Strickroth, S.: Scalable Feedback for Student Live Coding in Large Courses Using Automatic Error Grouping. In: Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1. ITiCSE 2024, ACM, New York, NY, USA, pp. 499–505, 2024, DOI: 10.1145/3649217.3653620.

[Vo04] Vonderwell, S.: Assessing online learning and teaching: Adapting the minute paper. TechTrends 48 (4), pp. 29–31, 2004.

Assessing GPT-4's Reliability for Japanese Kanji Instruction: Accuracy, Prompt Control, and Contextual Sentences

Sigrid L. Klinger [1] and Sven Strickroth [1]



Abstract: Example sentences and native-language translations are often seen as beneficial for vocabulary learning. However, manually creating such resources in multiple languages is labor-intensive. GPT-4 offers a promising solution by generating both example sentences and translations. This study evaluates GPT-4's performance in an adaptive Kanji learning tool, focusing on translation accuracy, semantic adequacy, and the quality of example sentences. A systematic analysis of 50 Kanji—representing about 4 % of the commonly used characters—their most frequent compound words, and 185 example sentences reveals high translation accuracy and mostly coherent contextual sentences, though occasional hallucinations and inappropriate expressions occur. Strict formatting and content guidelines reduce such deviations but increase prompt complexity. The results indicate that GPT-4 Turbo can support learning effectively, provided careful review are ensured.

Keywords: GenAI, Autogenerated Example Sentences, Language Learning

1 Introduction

Large Language Models (LLMs) such as GPT-4 have gained attention for their ability to generate human-like text and support tasks such as translation, summarization, and feedback. GPT-4 shows strong multilingual performance, especially in widely spoken languages like German and Japanese [Ji23; Op23]. This makes it promising for language learning applications, although issues such as hallucinations and inconsistencies still raise concerns [Ji23]. These limitations are particularly relevant in language learning, where precise and level-appropriate content is essential.

This study explores the integration of GPT-4 into an adaptive Kanji learning tool *Kanji-Compass* [KS25a] designed to support self-regulated learning. Kanji are morphographic characters used in written Japanese. Many common words are compounds formed by combining two or more Kanji. Their polysemy and irregular readings make them particularly difficult to master [YM00]. The tool provides mnemonic aids and structural breakdowns for all 2,136 everyday Kanji and their 520 semantic and/or phonetic components.

A persistent challenge in Kanji learning is the lack of learner-friendly, multilingual resources. Most existing Kanji databases rely on English, limiting accessibility for non-English speakers. Similarly, while example sentences are essential for contextual vocabulary

[1] LMU Munich, Oettingenstr. 67, 80538 Munich, Germany,
sigrid.klinger@ifi.lmu.de, https://orcid.org/0009-0006-9277-8355;
sven.strickroth@ifi.lmu.de, https://orcid.org/0000-0002-9647-300X

doi:10.18420/delfi2025_08

acquisition [BAB24], freely available databases like Tatoeba[2] often contain content that is too complex for beginners. To address this, GPT-4 was integrated into the KanjiCompass tool [KS25a] to automatically generate German translations and beginner-friendly example sentences. Unlike traditional dictionaries, which list numerous meanings regardless of context, GPT-4 was explicitly prompted to return only the two most common meanings in German. This approach provides focused, accessible information rather than overwhelming learners with lexical detail. As Japanese proficiency is typically classified by JLPT levels (from N5 [beginner] to N1 [advanced]), GPT-4 was also prompted to produce N5-level example sentences to ensure suitability for novice learners.

This study investigates GPT-4's effectiveness in an adaptive learning tool for Japanese Kanji, addressing the following two research questions: 1) Accuracy: To what extent does GPT-4 provide correct and semantically appropriate German translations of Japanese Kanji and compound words? 2) Quality of Example Sentences: To what extent can GPT-4 generate simple, context-rich example sentences suited to beginner-level proficiency? We systematically evaluate GPT-4's potential as a translation and example-sentence generator for Kanji learning among German speakers. Our contributions are: 1) an empirical evaluation of GPT-4's Kanji and compound word translations into German, benchmarked against established dictionaries. 2) an assessment of GPT-4's ability to generate illustrative, natural-sounding beginner-level example sentences, highlighting both strengths and limitations.

2 Related Work

Jiao et al. [Ji23] compared GPT-4 with GPT-3.5 and commercial tools like DeepL. Their findings show GPT-4 matches commercial tools in quality while also producing fewer errors [Ji23]. It also excels in language fluency and performs best with resource-rich languages [Ji23; Ro23]. A language is considered resource-rich if it has over one million Wikipedia articles [Ro23], which applies to both German (3.0 million) and Japanese (1.5 million).[3] Robinson et al. [Ro23] analyzed 204 languages, comparing GPT-3.5 with Google Translate and NNLB. GPT-3.5 outperformed NNLB but scored lower than Google Translate in spBLEU200 and chrF2++. No major performance differences were found between German and Japanese when comparing 0-shot and 5-shot (five in-context examples) prompt settings.

Beyond translation, GPT can tailor responses based on prompts. Rüdian and Pinkwart [RP23] evaluated GPT-3.5's ability to generate A1-level German sentences using predefined words within a 12-word limit. Of 200 sentences, 80.5 % were grammatically correct and 83.5 % contextually appropriate. A key issue was GPT-3.5's struggle to incorporate all given words coherently. Additionally, responses varied in structure due to the lack of a strict output form.

In summary, GPT-4 can provide high-quality translations, with accuracy depending on the language. It also shows potential for generating example sentences for language learning.

[2] https://tatoeba.org, last accessed 2025-06-03

[3] See Wikipedia statistics: https://w.wiki/7dGH and https://w.wiki/3uxAlast accessed 2025-06-03

Given GPT-4's improvements over GPT-3.5, it is likely to perform better in both translation and sentence generation. Some errors in Rüdian and Pinkwart's study [RP23] may stem from too strict prompt constraints, such as fixed sentence length, line formatting, or vocabulary level restrictions. These constraints often led to unexpected or invalid outputs, suggesting that more flexible or adaptive prompting could improve reliability.

3 Methodology

GPT-4-0125-preview (GPT-4 Turbo), the latest version available during the data collection period (March 2024), was integrated into a MEN stack architecture (MongoDB, Express, Node.js) and accessed via the OpenAI API. The server handled all API communication, using default parameters. GPT-4 was used in three areas of the Kanji learning tool KanjiCompass [KS25a]: (1) providing the German meaning of each Kanji, (2) translating each compound word into German, and (3) generating Japanese example sentences for compound words. For Kanji meanings and compound word translations, prompts were automatically triggered when a Kanji was accessed. The generated response was stored in a MongoDB database and reused for subsequent requests. In contrast, example sentence generation was user-initiated via a button in the interface. In both cases, the results were displayed in the client interface, either retrieved from the database or generated on demand. In the early development phase, GPT-3.5 Turbo was tested for generating Kanji meanings but produced unreliable results even for common characters, which led to its replacement with GPT-4 Turbo.

Since GPT's output is highly dependent on the given prompt, careful formulation was essential, so we based our prompts on Rüdian and Pinkwart's paper [RP23]. During prompt testing, it became evident that overly detailed instructions led to high response variance, as GPT-4 did not consistently follow all specified constraints. Therefore, the final prompt was kept compact and straightforward to ensure consistency. The prompts were originally written and issued in German to match the target language of the output. For readability, we provide English translations in this paper (the original prompts can be found in [KS25b]). For GPT to provide a German meaning of a Kanji, the following prompt was used: "*If you know the meanings of the Kanji {Kanji}, provide the two most common meanings in German in a short form, separated by a comma. If not, respond with 'unknown'.*". The instruction to return "unknown" for unfamiliar Kanji was necessary, as the tool includes components not found in modern Japanese, which could otherwise lead GPT to generate incorrect or speculative meanings. The translation of common Kanji compounds was generated with the following prompt: "*Please translate the following Japanese compound word into German using a maximum of two words, separated by a comma: {compound}.*". The example sentences were created with the prompt: "*Please create a Japanese example sentence at JLPT N5 level in terms of grammar and vocabulary that includes the word {compound}. The sentence should clearly convey the meaning of the word through its context, without requiring additional explanations.*" This prompt aimed to prevent ambiguous or unhelpful sentences. By explicitly requiring sentences that clarify word meanings through context, the study sought to improve their usefulness for learners.

All German Kanji meanings provided by GPT-4 Turbo were evaluated for orthographic accuracy, ensuring that they were correctly spelled German words rather than English terms or invented words. Additionally, adherence to the prompt was assessed, which required exactly two comma-separated German words or "unknown". To identify errors, the meanings were first checked using a commercial spelling checker, which flagged non-German and misspelled words. The flagged entries were then manually reviewed and categorized into hallucinations, spelling errors, non-German, or format deviations. For a detailed accuracy analysis, ten Kanji from each JLPT level (N5–N1), totaling 50 Kanji, were examined. This number was chosen to ensure balanced coverage across proficiency levels while keeping the manual evaluation effort feasible. The selected Kanji represent approximately 4 % of commonly used Kanji. Their meanings were compared against established German Kanji dictionaries [Bi23; Ha02; Ha95; Ha97]. A meaning was rated correct if it matched a dictionary entry, partially correct if it overlapped or derived from a common compound, and incorrect if it deviated significantly, or exceeded the two-word format.

All German translations of compound words generated by GPT-4 Turbo were analyzed using the same criteria, focusing on orthographic accuracy and adherence to the prompt. For a detailed evaluation the compound word translations linked to the above mentioned 50 Kanji were assessed using major Japanese-German dictionaries [St22; Wa], applying the same rating scale.

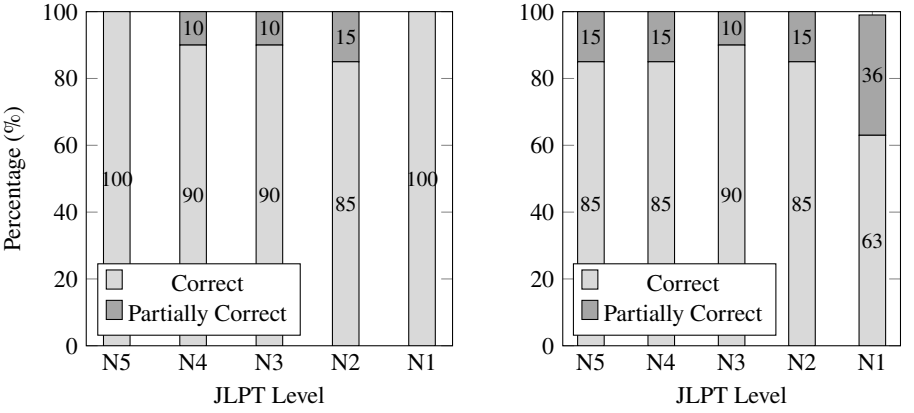
Example sentences were evaluated for clarity, correctness, and naturalness. Sentences lacking clear context (e. g., "I like dogs" for "dog") were marked inadequate, while those providing meaningful context (e. g., "I go for a walk in the park" for "walk") were rated appropriate. Outputs with unnatural phrasing (e. g., "The dog is driving a boat") were deemed incorrect. GPT-4 Turbo was instructed to generate sentences at JLPT N5 level, with difficulty approximated using Kanji lists JLPTSensei [JL] and a general grammar check. For statistical analysis, a Fisher-Yates permutation test ($\alpha = .05$) was used to compare translation and meaning accuracy across JLPT levels. To estimate the strength of association, Cramér's V was computed from the corresponding χ^2 statistic. The dataset used in this study, including the used prompts, Kanji meanings, compound word translations, and example sentences, is available at OSF [KS25b].

4 Evaluation

4.1 Results of GPT-4 Turbo's Responses

GPT-4 Turbo was queried for 1,800 Kanji and 520 components to provide their German meaning (85 % of Kanji and 100 % of components included in KanjiCompass). In 99 % of cases, it followed the instruction to provide exactly two comma-separated correct German words or "unknown". The response "unknown" was returned for 60 % of components and .5 % of Kanji. Errors ranged from English words appearing in the German output (e. g.,

"digit" for 桁) to hallucinations, such as "Raum, Midluft" for 宙 (*Midluft* is not a German word but appears to be a direct translation attempt of "mid-air"). Among the 50 Kanji chosen for a detailed comparison with dictionaries, 92 % were rated correct and 8 % partially correct. None were entirely incorrect. There was no significant difference across JLPT levels in terms of correctness ($p > .05$, $\chi^2(4) = 5.53$, $V = .24$), suggesting a small-to-moderate, non-significant association (see Figure 1 (a)).



(a) Accuracy of German Definitions for Kanji Generated by GPT-4 Turbo

(b) Accuracy of German Translations for Kanji Compound Words Generated by GPT-4 Turbo

Fig. 1: Evaluation of GPT-4's Performance in Generating Translations of Kanji and Compound Word, Compared Across JLPT Levels From N5 (Beginner) to N1 (Advanced).

For compound words, GPT-4 Turbo generated German translations for 4,600 words derived from the 1,800 Kanji queried above. In 99 % of cases, it followed the instruction to provide a maximum of two comma-separated correct German words. Only once— for 荒木—GPT-4 Turbo stated that it could not provide a translation without additional context. In nine cases, English words appeared in the output, such as "regrettable" for 遺憾. In other instances, German umlauts were omitted (e. g., "Hautung" instead of "H*ä*utung" for 脱皮; "skinning"). Six cases contained minor typographical errors, such as letter transpositions. A detailed analysis of the translation of 83 compound words included in KanjiCompass, which correspond to the 50 Kanji examined earlier, showed that 83 % were fully correct, while 17 % were partially correct. Errors in the latter category included the use of unnatural German terms, such as "Seerand" (lit. "lake border") instead of "Seeufer" ("lakefront"), or overly general terms like "Medizin" ("medicine") instead of "medizinische Behandlung" ("medical treatment"). No translation was completely incorrect, and no significant difference in correctness was found across JLPT levels ($p = .43$, $\chi^2(4) = 3.86$, $V = .21$). However, N1 had a slightly higher proportion of partially correct answers ($p = .079$) (see Figure 1 (b)).

A total of 185 example sentences were generated for 80 compound words. The number was determined based on practical constraints, ensuring a representative sample while maintaining feasibility for evaluation. Four sentences contained clear errors (e. g., the target

compound word was missing, or GPT-4 Turbo included an explanation before providing the sentence). Hence, the remaining 181 sentences (98 %) were analyzed in detail. In 4 % of the sentences, the target word's meaning was not clarified by context (e. g., この数字は何ですか (What is this number?) for the word 数字 "number"). Such sentences could easily be substituted with other nouns/verbs. In 3 % of cases, the sentences sounded unnatural from the perspective of someone trying to explain a word (e. g., 私は海洋を守るために毎週ビーチクリーニングに参加しています ("I participate in beach cleanups every week to protect the ocean") for the word 海洋 "ocean"). The rest of the sentences were grammatically correct, provided suitable context, and used an appropriate register. 70 % of the Kanji came from JLPT N5 or N4; the rest included Kanji from higher levels.

5 Discussion, Conclusions and Future Work

Our evaluation shows that GPT-4 Turbo generates highly accurate Kanji meanings and compound word translations in German. Most outputs closely match established dictionary entries. Inaccuracies were limited to issues such as spelling or overly broad translations, and none fully distorted the intended meaning. Since the analysis is based on a small random sample of 50 Kanji, the results should be interpreted with some caution. A few responses showed recognizable hallucinations, which were generally easy to identify for native German speakers. The error rate was slightly higher for Kanji than for components. This may be because GPT-4 Turbo returns "unknown" when uncertain, but produces more elaborate outputs for familiar characters, which can occasionally lead to mistakes.

GPT-4 Turbo also performed well in generating example sentences in Japanese. These were usually natural and clearly conveyed the meaning of the word. However, strict adherence to JLPT N5 level was not always consistent. The model tended to prioritize contextual clarity over grammatical simplicity. This could be addressed through more constrained prompts or post-processing to filter out advanced structures. Personalizing outputs based on a learner's known Kanji or grammar level could further support learning, but also increases prompt complexity and token usage. These factors need to be balanced against the potential benefits of more targeted, context-rich content.

Overall, GPT-4 Turbo is well-suited for generating adaptive learning materials such as Kanji meanings, compound word translations, and beginner-level example sentences. Issues like hallucinations or formatting inconsistencies can often be reduced through prompt engineering or simple user feedback mechanisms. At the time of data collection (March 2024), GPT-4 Turbo was among the most widely used models accessible via API and showed strong performance in Japanese and German [Ji23]. While newer or specialized models may now offer further improvements, this study provides a replicable baseline for evaluating LLM-generated content in Kanji learning. This study focused on the quality of GPT-generated outputs. Future work should explore how such outputs affect actual learning. Planned studies include learner evaluations of comprehensibility, perceived usefulness, and integration into learning workflows.

References

[BAB24] Benedetti, E.; Aizawa, A.; Boudin, F.: Automatically Suggesting Diverse Example Sentences for L2 Japanese Learners Using Pre-Trained Language Models. In: The 62nd Annual Meeting of the Association for Computational Linguistics: Proceedings of the Student Research Workshop. Vol. 4, Association for Computational Linguistics, ACL, Bangkok, Thailand, pp. 114–131, 2024, https://hal.science/hal-04742653.

[Bi23] Bibiko, H.-J.: Japanisch-Deutsches Kanji-Lexikon, Accessed: 2025-02-20, 2023, https://lingweb.eva.mpg.de/kanji/.

[Ha02] Hadamitzky, W.: Japanisch-deutsches Zeichenwörterbuch. Unter Mitarbeit von M. Spahn, O. Putz, H. Arnold-Kanamori, T. Frischkorn, & N. Günther, Helmut Buske Verlag, Hamburg, 2002.

[Ha95] Hadamitzky, W.: Handbuch und Lexikon der japanischen Schrift – Kanji und Kana 1 – Handbuch. Langenscheidt, Berlin and München, 1995.

[Ha97] Hadamitzky, W.: Langenscheidts Großwörterbuch Japanisch-Deutsch: Zeichenwörterbuch. Langenscheidt, Berlin and München, 1997.

[Ji23] Jiao, W. et al.: Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine, 2023, arXiv: 2301.08745 [cs.CL], https://arxiv.org/abs/2301.08745.

[JL] JLPTSensei: Learn Japanese and Pass the JLPT!, Accessed: 2024-05-24, https://jlptsensei.com/.

[KS25a] Klinger, S. L.; Strickroth, S.: KanjiCompass: An Etymology-Driven Adaptive Kanji Learning Tool. In: Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025) - Volume 1. Pp. 472–483, 2025, DOI: 10.5220/0013341000003932.

[KS25b] Klinger, S. L.; Strickroth, S.: Source Material for the Paper: "Assessing GPT-4's Reliability for Japanese Kanji Instruction: Accuracy, Prompt Control, and Contextual Sentences", OSF Project, 2025, DOI: 10.17605/OSF.IO/ZWKMB.

[Op23] OpenAI: Research GPT-4, Accessed: 2025-02-20, 2023, https://openai.com/research/gpt-4.

[Ro23] Robinson, N. R. et al.: ChatGPT MT: Competitive for High- (but not Low-) Resource Languages, 2023, arXiv: 2309.07423 [cs.CL], https://arxiv.org/abs/2309.07423.

[RP23] Rüdian, S.; Pinkwart, N.: Auto-generated language learning online courses using generative AI models like ChatGPT. In: 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., Bonn, pp. 65–76, 2023, DOI: 10.18420/delfi2023-14.

[St22] Stalph, J. et al., eds.: Großes Japanisch-Deutsches Wörterbuch. IUDICIUM Verlag, München, 2022.

[Wa] Wadokijiten: Großes Japanisch-Deutsches Wörterbuch, Accessed: 2024-05-24, https://www.wadokudaijiten.de/.

[YM00] Yamashita, H.; Maru, Y.: Compositional Features of Kanji for Effective Instruction. The Journal of the Association of Teachers of Japanese 34 (2), pp. 159–178, 2000, DOI: 10.2307/489552.

Developing a Graph-based Visualization of Elective Courses to Support Course Selection in Higher Education



Marie Trippel [1] and René Röpke [2]

Abstract: In an open curriculum in higher education, students face the challenge of navigating a broad course catalog independently. Prerequisites and recommended courses can make this even more complex, as they are not always easy to follow. Providing students with clear information is essential to make good choices in study planning. Effective visualization could simplify this process for students by providing structured, understandable, and relevant information. This paper presents a novel tool that helps students choose their elective courses using a graph-based visualization of course relationships, emphasizing prerequisites and recommendations. A first user evaluation indicates a strong interest in a course-planning tool for electives and the use of the graph-based visualization was well-liked. Thereby, this research highlights students' demand for a planning tool and demonstrates the potential of graph visualizations as a valuable resource for future development in this area.

Keywords: Course Planner, Dependency Visualization, Course Prerequisites

1 Motivation and Related Work

An open curriculum in higher education is flexible and allows personalization based on interests, giving students greater autonomy in shaping their academic journeys. However, this increased flexibility also introduces complexities in course selection, often leaving students in need of additional support to navigate their options effectively and consequently have successful studies. Often, students only have limited knowledge about possible choices as well as regulations and restrictions. This can lead to a non-ideal study path if they do not know or miss important information, like course dependencies or semester availability.

To address these challenges, higher education institutions (HEI) provide advisory services and information materials. Traditionally, study advisors help students, offering tailored guidance. While being invaluable, especially in special circumstances, this support is resource-intensive and time-consuming. In recent years, HEI started to develop digital study planning and assistance tools, empowering students to take charge of their study planning and align it with individual requirements and objectives [Na24; Ru24; TP18; Un24a; Un24b]. Within these tools, relevant curriculum information needed for study planning is presented and the planning process can be supported through visual depiction of semesters and course order [HHH21; Ju23a; Ju23b].

[1] RWTH Aachen University, Aachen, Germany,
marie.trippel@rwth-aachen.de, https://orcid.org/0000-0003-4961-986X
[2] TU Wien, Learning Technologies and eDidactics, Vienna, Austria,
rene.roepke@tuwien.ac.at, https://orcid.org/0000-0003-0250-8521

doi:10.18420/delfi2025_09

Visualization is a powerful tool for presenting complex information clearly and understandable [Sh96]. It presents as a promising approach to make complex curricula, including course relations and dependencies, more accessible and easily comprehensible. While there are various types of visualizations for types of data [NJG21], for this work, graph-based visualization is considered—a suitable approach to visualize interrelated entities (i. e. courses and their relations), as seen in social networks [MR20; Ne13].

This work introduces an interactive graph-based visualization tool to assist students in assessing course options and their relations, and selecting suitable courses, thus enabling informed decision-making. As for now, the tool relies solely on data from available module manuals/handbooks, with means to integrate it into the tool landscape of HEI. While it does not yet consider how courses complement each other beyond listed prerequisites, the tool provides a work-in-progress prototype for further investigation and extension.

2 System Design and Implementation

The primary objective of the graph-based tool is to visualize courses, their relations and dependencies, their allocation in different subject areas and additional course information. Courses and subject areas can both be depicted as two types of nodes in graph: Circular nodes for courses and square nodes for subject areas. Relations are depicted as directed edges and two types of course-to-course relations are considered: “recommended”, if A is recommended for B, and “required”, if A is a prerequisite for B, where A and B are courses in the curriculum. Color is used to distinguish the directed edges: “required” edges are red, while “recommended” edges are blue. To visualize the curriculum structure, subject areas are connected to courses via non-directed containment edges (colored in grey). For displaying additional course information, further visual cues are used: node size indicates a course’s credits, while a green ring signifies if a course has been passed.

Requirement Analysis A requirement analysis was conducted using semi-structured interviews[3] to ensure tool meets user needs. Four key stakeholders were interviewed: the head of the Study Center, the Study Program Coordinator, Student Member of the Examination Committee, and an member of the Academic Advisory Service. Their input provided functional requirements for the tool (see Table 1).

Implementation The tool was implemented as an Angular-based web application [Go25] using d3.js [Bo25] for the graph visualization due to its low-level DOM (Document Object Model) control. The component-based design enables the resulting tool to be easily integrated in existing applications, e. g., study planning tools.

Implemented features include filtering by subject areas, credits, semester cycle, course language, and course status, as well as preset filters for quick setup (FR6). Users can hover to highlight linked courses (FR4), click to view course information (FR3, FR5), and

[3] Research data is available under [RT25]

ID	Description
FR1	Courses should be represented as nodes in the graph.
FR2	Prerequisite relations should be represented as a directed edge.
FR3	Nodes should display course name, course area, and credits.
FR4	Users should be able to interact with the graph (e. g., zoom, pan, click on nodes).
FR5	When hovering over or clicking on a node, more course information should appear.
FR6	Users should be able to filter courses based on criteria such as department, level, or semester.
FR7	Users should be able to search for specific courses by name or code.
FR8	Users can set each course to passed, taken, or planned to be taken.
FR9	Users can manage multiple individual pathways.
FR10	When hovering over a course node, all connected nodes should be highlighted.
FR11	The system can offer preset views (e. g., by departments, specializations).
FR12	The system should warn users if a course has requisites that are not in the module catalog.
FR13	Users can create and add new course nodes to the graph.
FR14	Users can rate courses, thereby providing recommendations for other users.
FR15	The system can recommend similar subjects based on course selections.
FR16	The system should warn users if their course selection is too specialized or too broad.

Tab. 1: List of determined functional requirements

toggle graph options such as shortened titles or visualizing prerequisite paths. The tool is standalone and not yet integrated in any study planning tool and stores user data only locally, which is why a export/import functionality is provided for returning users (FR9). Warnings appear when a student selects too few or too many courses in one area (FR16).

3 Evaluation

A user study was conducted to evaluate the work-in-progress prototype and collect first feedback on its features and design. The study involved n=15 participants, all Computer Science (CS) students, enrolled in Bachelor or Master study programs and with prior experience in choosing electives as part of their curriculum. The study was designed as a scenario-based evaluation using the think-aloud method, a short survey on user experience and a brief interview afterwards. The first scenario focused on choosing a course for the winter semester with four credits, to get the students familiar with the interface and functionalities. The second scenario consisted of looking for a specific CS elective course that has prerequisites and seeing whether this course is feasible. Here students had to navigate through the interface and understand a course path with two prerequisites.

3.1 Qualitative Insights from the Evaluation and Interviews

This section presents the qualitative results of the scenario-based evaluation and the interview. All participants reported a positive experience with the tool, with most participants expressing

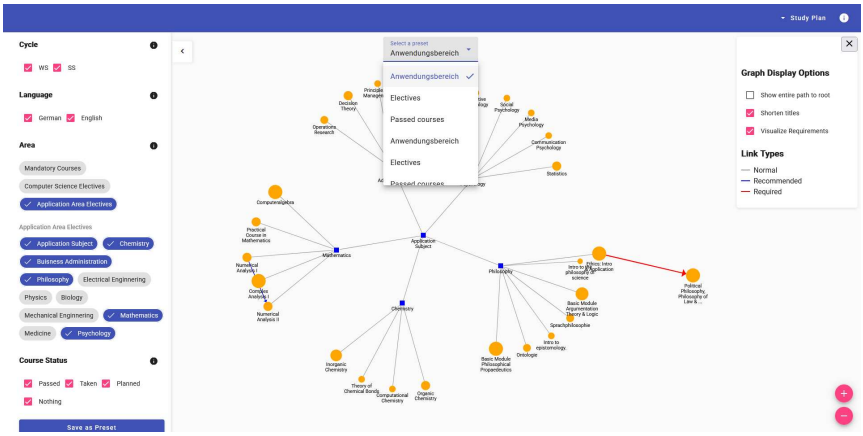


Fig. 1: At the core of the application is the graph structure visualizing a curriculum with courses as nodes (FR1, FR3). A filter menu is on the left side (FR6), integrated in a sidebar. In the top center, the preset drop-down (FR11) is located. On the top right is a box with graph display options and a legend of the different link types (edges). A course recommendation is visualized by blue directed arrows (FR2). Required edges can be recognized by their red color as well as the path forming to further visualize the requirement (FR2).

a desire to use it for their own study planning. The prototype helped to think more strategically about course selection, especially in understanding course relations. In the past, as most participants reported, they had chosen electives rather spontaneously.

The graph visualization of the curriculum was well-received, as it helped participants grasp the range of available options. Many participants appreciated the clear depiction of prerequisites through “recommended” and “required” links (directed, colored edges), and the filter functionality for narrowing down choices. They found it less overwhelming than going through a traditional list, like in their campus management system (CMS) or the module manual. Participants also appreciated tracking course status for long-term planning.

The interface initially appeared overwhelming to some, mainly because of the large graph structure. However, they quickly adapted by using the preset and filter options to get an overview. Interestingly, some participants relied solely on the preset options, while others exclusively used the filter menu. While the preset options sped up the process, some participants found their connection to the filters unclear, calling for clarification in its design.

Participants expressed interest in more detailed course information such as expected workload, information on the instructors, and class schedule. They recommended adding direct links to registration pages, as well as past evaluations or data on average grades. The most requested, yet missing feature was a search function (FR7). Others included side-by-side course comparisons and a course recommendation system providing suitable, personalized course choices.

3.2 Survey Results on User Experience

Next, we present results from the survey portion of our study, evaluating the participants' user experience when testing the tool and completing the scenario-based evaluation. For this, the User Experience (UEQ) was used due to ease of use, wide applicability and reliability [LHS08]. Overall, scores were good across all six dimensions, indicating a generally positive user experience. The dimension "Stimulation" ($M = 1.900$, $SD = 0.667$) and Attractiveness ($M = 1.811$, $SD = 0.842$) scored highest, showing that participants found the tool engaging and pleasant to use. "Efficiency" ($M = 1.617$, $SD = 0.812$), "Dependability" ($M = 1.650$, $SD = 0.766$) and "Perspicuity" ($M = 1.367$, $SD = 0.855$) were rated positively but suggest room for improvement in clarity and task flow. These results are consistent with qualitative feedback on unclear UI (User Interface) elements and the similarity of visual indicators. Results for the "Novelty" dimension ($M = 1.733$, $SD = 0.665$) indicate that participants perceive the tool as innovative and creative. The tool scored overall higher in Hedonic Quality than in Pragmatic Quality ($1.82 > 1.54$), indicating that while participants found it enjoyable and appealing, its usability and functionality can still be improved to better support students' needs.

4 Discussion

Overall, we were able to design and implement a graph-based visualization tool, fulfilling most of the collected requirements. As this prototype presents is work-in-progress, our first goal was to collect feedback by students and assess user experience. The data used in the tool was solely based on module manuals and did not integrate data of a CMS.

Both, the qualitative insights from the scenario-based evaluation and the interview, as well as the quantitative results on the participants' user experience suggest that students are eager for a tool to assist them in choosing elective courses. Especially during the interviews, participants voiced interest in using it for their own study planning in order to compare electives and to make informed choices.

While some participants had initial difficulties grasping the structure of the very large curriculum graph, provided filters and preset options helped navigating and getting oriented quickly. The graph structure itself was easy to understand and the provided information was perceived as helpful, yet, it could be extended in the future. Participants asked for information that could be easily provided from CMS (e. g., instructor information or class schedules), but they also showed interest in data-driven insights (e. g., grade statistics from past semesters). This calls for its integration in an institution's tool landscape as well as the extension with learning analytics. Through learning analytics, data-driven insights on courses could be provided to students when comparing courses.

Limitations This ongoing work has several limitations which need to be addressed. First, the prototype was evaluated exclusively within the CS faculty at RWTH Aachen University,

where the Bachelor study program has rather many restrictions in early semesters and then moves towards almost no restrictions, but only recommendations linking different courses to each other. It was therefore also only tested with CS students, who may have a higher affinity for new digital tools and learn about graph structures as part of their curriculum. This may have biased the results and calls for the evaluation with students from other study programs. Additionally, as already mentioned, this work focused solely on visualizing course relations based on the module manual and did not incorporate any recommendation systems, or look deeper into which courses complement each other based of their actual content. This work is, thus, limited to the quality and availability of data provided in module manuals. Further, informal knowledge of course orders due to course content and difficulty is also not considered. Lastly, as this work does not examine social and educational impact yet, a long-term evaluation would be interesting.

Future Work With this work presenting a prototype, future work calls for user-centered evaluation and systematic integration in HEI. The feedback so far should be considered for implementation, along with missing requirements (e. g., FR7, FR14). The tool should also be tested with students from other study programs to see how results differ when there are more (or less) course prerequisites, and to assess the user experience of students who may be less tech-savvy than CS students. Difficulties with the graph visualization could arise which may require adjustments in the tool design or additional support in learning to use the tool. It should also be considered to explore alternative visualizations, e. g., representing courses within subject areas using bubble charts. Furthermore, the combination of graph-visualization with other visualizations could be explored to provide various ways to interact with a curriculum. Lastly, when developing the prototype further, its integration in study planning tools or CMS should be considered to allow for smooth transitions and ease-of-use for students.

5 Conclusion

This paper addressed the challenge of students when selecting elective courses in multi-semester planning, with a focus on course relations and dependencies. To this, we introduced a graph-based visualization tool to support students in exploring the curriculum and in making informed course choices. The development process included a requirement analysis with stakeholders and the design and implementation as a web application. The resulting prototype was evaluated in a user study with 15 participants to receive feedback and gain insights on its design and the user experience. Results indicate a strong potential and interest for graph-based visualizations for students' course selection. Next steps include improving and further testing the prototype as well as integrating it existing tools for study planning. With it being integrated in tools available to students, further investigations of its benefits to study planning and guidance can be conducted providing valuable insights into students' perspectives.

References

[Bo25] Bostock, Mike and Observable, Inc.: D3 by Observable – The JavaScript library for bespoke data visualization, 2025, https://d3js.org/, accessed: 04/06/2025.

[Go25] Google: Angular, 2025, https://angular.dev/, accessed: 04/06/2025.

[HHH21] Hirmer, T.; Heyne, N.; Henrich, A.: Die kompetenzorientierte Studienplanung - Entwicklung eines Tools zur Unterstützung von (Lehramts-)Studierenden. In: DELFI 2021. Gesellschaft für Informatik e.V., Bonn, pp. 121–126, 2021.

[Ju23a] Judel, S. et al.: KI-gestützte Studienverlaufsplanung und -analyse mit AIStudyBuddy. In: Workshops der 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., Bonn, pp. 193–196, 2023, DOI: 10.18420/WSDELFI2023-55.

[Ju23b] Judel, S. et al.: Supporting Individualized Study Paths Using an Interactive Study Planning Tool. In: 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., Bonn, pp. 225–230, 2023, DOI: 10.18420/DELFI2023-36.

[LHS08] Laugwitz, B.; Held, T.; Schrepp, M.: Construction and Evaluation of a User Experience Questionnaire. In: Symposium of the Austrian HCI and Usability Engineering Group. Springer, pp. 63–76, 2008, DOI: 10.1007/978-3-540-89350-9_6.

[MR20] Majeed, A.; Rauf, I.: Graph Theory: A Comprehensive Survey about Graph Theory Applications in Computer Science and Social Networks. Inventions 5 (1), p. 10, 2020, DOI: 10.3390/inventions5010010.

[Na24] National University of Singapore: NUSMODS, https://nusmods.com/timetable, 2024, https://nusmods.com/timetable, accessed: 04/13/2024.

[Ne13] Nettleton, D. F.: Data mining of social networks represented as graphs. Computer Science Review 7, pp. 1–34, 2013, DOI: 10.1016/j.cosrev.2012.12.001.

[NJG21] Nguyen, V. T.; Jung, K.; Gupta, V.: Examining data visualization pitfalls in scientific publications. Visual Computing for Industry, Biomedicine, and Art 4 (1), p. 27, 2021, DOI: 10.1186/s42492-021-00092-y.

[RT25] Röpke, R.; Trippel, M. E.: Graph-based Visualization of Elective Courses to Support Course Selection in Higher Education, 2025, DOI: 10.17605/OSF.IO/ZTQ2U.

[Ru24] Rutgers University: Rutgers Course Schedule Planner, 2024, https://sims.rutgers.edu/csp/, accessed: 08/30/2024.

[Sh96] Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings 1996 IEEE Symposium on Visual Languages. IEEE Comput. Soc. Press, Boulder, CO, USA, pp. 336–343, 1996, DOI: 10.1109/VL.1996.545307.

[TP18] Tomy, S.; Pardede, E.: Course Map: A Career-Driven Course Planning Tool. In (Gervasi, O. et al., eds.): Computational Science and Its Applications – ICCSA 2018. Vol. 10961, Springer International Publishing, Cham, pp. 185–198, 2018, DOI: 10.1007/978-3-319-95165-2_13.

[Un24a] Universeiteit Utrecht: CoursePlanner Utrecht University, https://cursusplanner.uu.nl, 2024, https://cursusplanner.uu.nl, accessed: 08/30/2024.

[Un24b] University of Helsinki: Sisu Instructions: Study Plan, 2024, https://studies.helsinki.fi/instructions/article/sisu-instructions-study-plan, accessed: 09/06/2024.

Künstliche Intelligenz

Adding the Human Error: Exploring the Impact of GenAI on Student Writing in Information Systems

Philipp Matter [1] and Ninja Leikert-Böhm [1]

Abstract: This study explores the use of Generative Artificial Intelligence (GenAI) in academic writing in higher education. In 2023, students wrote short essays on AI regulation using ChatGPT. We examined AI-assisted writing quality, student experience, and possible influences on student opinions. Analysis of 52 essays showed that while they were of acceptable quality and unique at the text level, they suffered from content repetition and linguistic redundancy. Argumentation tended toward permissionless innovation, even when students perceived themselves as neutral. The study highlights the effort required to verify AI-generated content and the need to raise students' awareness of potential framing effects when using ChatGPT. Overall, our findings suggest that while GenAI tools can support academic writing, they require critical human oversight to ensure quality and originality.

Keywords: GenAI, ChatGPT, Education, Academic Writing

1 Introduction

Research on Generative Artificial Intelligence (GenAI) such as ChatGPT [Op24] in higher education is developing rapidly, with a focus on its impact on writing, critical thinking, and student engagement. There is a substantial body of conceptual work and literature reviews on the benefits and challenges of GenAI in education [Al23; DLL23; LKM25]. Empirical research, including surveys on student use and attitudes [CH23; Ra25], is also increasing. However, comparatively few studies empirically examine how students use GenAI specifically for writing in higher education and what the effects are [Ba23; Fy23; NDS25; NZ23]. Moreover, existing evidence is often conflicting, highlighting the need for further empirical research.

We present a study with master's students in information systems (spring 2023), exploring student experiences, essay quality patterns, and possible influences associated with ChatGPT use in academic writing. Our study is situated within the ongoing debate on AI in education, framed by the dichotomy between the precautionary principle — favoring cautious regulation — and permissionless innovation, which encourages open experimentation [Th16]. This contrast provides fertile ground for examining how AI tools such as ChatGPT may shape academic discourse and student perspectives.

Our research involved 52 students writing essays about the potential regulation of AI, primarily using ChatGPT to generate content, and then reflecting on their experiences.

[1] Zurich University of Applied Sciences, School of Management and Law, Theaterstrasse 17, 8401 Winterthur, Switzerland, philipp.matter@zhaw.ch, https://orcid.org/0009-0002-8086-5719;
ninjairina.leikert-boehm@zhaw.ch, https://orcid.org/0000-0002-8815-0198

doi:10.18420/delfi2025_10

This approach allowed us to assess the average quality of the AI-generated texts (research question RQ1), the overall learning experience and challenges faced by students (RQ2), and how ChatGPT use related to students' opinions and arguments (RQ3).

2 Related Work

We present related work on the benefits and challenges of GenAI in higher education and academic writing, along with survey results on students' usage and attitudes. Our focus is on (quasi-)experimental studies that quantitatively and qualitatively assess GenAI's impact, particularly in writing.

Benefits and challenges: On the positive side, GenAI can address writer's block and anxiety and provide new ideas and perspectives [Al23; LKM25]. It can provide targeted, real-time support for individual learning needs [DLL23; Fa24] and increases efficiency and fluency in writing [De23; Di23; LKM25]. It has also been found to have a positive influence on the development of critical thinking skills [DK25; Me25], as well as students' motivation, persistence, and enjoyment of learning [DK25].

As challenges, the literature has found that GenAI lacks a deep understanding of theoretical concepts, lacks higher-order thinking skills, and has difficulty evaluating the quality of its responses [Fa24]. GenAI can struggle with the relevance and precision requirements of scientific writing and also risks reducing original thought and authenticity [Al23; De23]. Threats to academic integrity, such as lack of accuracy and bias, are also treated very prominently in the literature [DK25; Ek23; Fa24; Pe23]. In addition, and in contrast to its benefits, overdependence and overreliance on GenAI can lead to a decline in higher-order cognitive skills such as creativity, critical thinking, reasoning, and problem solving [DK25; Fa24]. Simplifying the process of finding answers can lead to decreased motivation to conduct independent research [Fa24] and thereby hinder the development of analytical skills. The literature also highlights ethical concerns such as privacy and security, accountability, and fairness [Ra23].

Student usage and attitudes: In a large global survey of higher education students, Ravšelj et al. [Ra25] found that students used ChatGPT primarily for brainstorming and summarizing texts, as well as for searching research articles, with few using it for writing. They found it useful for simplifying information, though not reliable, and believed that ChatGPT could increase their access to knowledge and improve their learning experience and efficiency. Students strongly agreed that universities need to provide ethical guidelines for the use of GenAI. A generally positive attitude towards GenAI was also found in a survey of students by Chan & Hu [CH23], where their willingness to use GenAI was positively correlated with their knowledge and frequency of use. Smolansky et al. [Sm23] reported moderate use of GenAI by students, and that students were concerned about academic integrity and loss of creativity.

Effect of using GenAI on writing: The impact found is mixed. Bašić et al. [Ba23] found no significant improvement in writing when comparing AI-assisted essays to traditional essays, with AI providing neither higher quality content nor greater efficiency. However, AI-assisted essays did not show increased inauthenticity, an important concern in academic writing. The authors suggest that the overall better score for traditional essays may be due to a possible overreliance on ChatGPT in the AI-assisted group, or their unfamiliarity with its use; thus, the challenging task of writing an argumentative essay may have been even more difficult for ChatGPT users. This observation is echoed by Fyfe [Fy23], in whose study the vast majority of AI-assisted students found using GenAI more complicated than writing themselves, especially in terms of integrating the generated text into their own writing. In contrast to both studies, Noy & Zhang [NZ23] found that AI significantly improved writing productivity, reduced writing time, and improved text quality. However, most participants modified the ChatGPT output only superficially or not at all, with grades reflecting ineffective editing.

Participants in a study by Van Niekerk et al. [NDS25] found that ChatGPT improved the structure and flow of their writing, but it lacked insight into the subject matter and depth of analysis, and thus was unable to provide useful evidence to support arguments. Many were also unaware that ChatGPT was 'hallucinating,' and lost confidence in the output once they realized this, emphasizing the need for fact-checking. Oates & Johnson [OJ25] found that while ChatGPT demonstrated proficiency in structure and language, the content generated was also simplistic, with repetitive arguments undermining the quality. Overall, they found that ChatGPT did not improve students' essay grades relative to their overall academic performance. However, it did improve their grades on the critical evaluation of other essays, suggesting that ChatGPT was more effective as an assessment tool.

In addition to studies on writing in higher education, experimental research on GenAI has explored its broader impact. Kanabar [Ka23] found that while students considered ChatGPT helpful for academic assignments, they preferred to use it as a supplemental resource because they were concerned about their actual learning and acquiring critical thinking skills. This sentiment is echoed in Swargiary's study [Sw23] on learning behaviors: more than half of ChatGPT users reported a decline in critical thinking, especially in verifying information, while a third became overly reliant on it. The control group, which used traditional learning methods, outperformed the AI-assisted group.

3 Methods

Our study was conducted in the spring semester of 2023 with 52 students of the master's program in information systems at a Swiss applied university. They were given the task of writing an essay (in German) on the topic of AI regulation ("How should ChatGPT be handled, with bans or in the sense of permissionless innovation? And why?"). Students were instructed to use ChatGPT. At the time of the assignment they could either use the free version (ChatGPT 3.5) or the paid version (ChatGPT 4). Students were also encouraged to use additional tools such as Elicit [El25] for source work. The assignment was an essay of

no more than three pages with the ChatGPT prompts used attached as an appendix, plus one page of reflection. The reflection guiding questions covered the writing process, lessons learned, intended future use of ChatGPT, and perceived influences on opinions. As this was a real-world classroom intervention, all students were required to use ChatGPT, and no control group was included. As a result, the study is exploratory and descriptive, focusing on observed patterns rather than causal relationships.

We addressed our research questions using the following study results. To assess the quality of the essays (RQ1), we analyzed the 52 grading schemes of the instructor (the first author) and his critical feedback. Grading was based on a standardized departmental assessment scheme, which defines qualitative performance descriptors for the following criteria (each rated on a 1–10 scale): understanding of the topic (32 % of the grade), individuality (32 %), structure (8 %), source work (8 %), language (10 %), and formatting (10 %). We gained insight into the students' reflections on the use of ChatGPT (RQ2) from 52 written student self-reflections as well as the critical feedback from the instructor. Finally, to examine how ChatGPT use may have related to student opinion formation (RQ3), we analyzed the 52 essays along with the student reflections and the instructor's feedback.

We analyzed all source material — including essays, student reflections, and instructor feedback — using the approach by Kuckartz and Rädiker [KR22]. Coding was conducted in two cycles, with sentences serving as the smallest coding unit. In the first cycle, main categories were deductively derived from the research questions (e. g., grading criteria, reflection questions, argumentative positions on AI regulation) and inductively expanded where necessary using descriptive coding. This resulted in 42 main codes. In the second cycle, these codes were refined and reduced through axial coding, leading to 20 main categories, which were further detailed into 227 subcodes across different levels. To ensure rigor and reduce coder bias, the first author conducted the initial coding, while the second author critically reviewed the full coding and analysis process, acting as an advocatus diaboli and challenging category definitions and interpretations in line with qualitative research best practices [GCH13, p. 19].

For RQ3, which focused on students' argumentative positions regarding AI regulation, essays were initially coded into broad stance categories: permissionless innovation, precautionary principle, responsible innovation, middle-ground positions, or unclear (no discernible stance). During the second coding cycle, we analyzed the middle-ground essays in more detail to identify underlying argumentative tendencies and also cross-checked these interpretations against the students' reflections.

Results were analyzed categorically by summarizing each code and including illustrative quotes. To illustrate how different views were distributed among students, we include the percentage of students assigned to each code in the results. Direct quotes from student reflections are identified by the participant number (e. g., S5 = Student 5).

4 Results

4.1 Essay Quality (RQ1)

Overall, the average essay grade was quite good, with almost two thirds (33, 63.5 %) of the students receiving a good grade (5 or higher according to the Swiss grading system, which gives passing grades of 4–6, 6 being the best grade). The remainder were graded from satisfactory (4.5–4.75) to sufficient (4.25).

In general, all submissions were distinct and authentic: the majority of the essays (43, 82.7 %) showed individual student achievement in that they not only used ChatGPT output directly, but also incorporated more or less own editing (corrections and adaptations, additions) into the text. In these essays, however, only just over half of the students managed to integrate the ChatGPT output with their own work into a coherent whole (24, 55.8 %). The remaining students were not able to successfully integrate the few adaptations they made into the generated content.

Structuring the essays proved to be a common challenge. Less than a third of the essays were rated as well or very well structured (15, 28.8 %), with the remainder lacking structure and transitions between sections of text, which limited the comprehensibility of the essays. The use and citation of sources was rated very good or good in 40 essays (76.9 %), and as satisfactory to poor in 9 (17.3 %). Only 3 essays (5.8 %) had no sources. This very good average rating is consistent with many students' comments that they had to invest a lot of time in source work due to the poor quality of literature references from ChatGPT (see RQ2), and with some students using Elicit to check and find sources.

The essays were also very well done in terms of language quality and presentation (correct spelling and grammar, clean formatting): 20 essays (38.5 %) were rated as virtually error-free, 26 (50 %) had few, and only 6 (11.5 %) essays had very many errors. Most of the errors were due to the authors' interventions — there were a particularly high number of errors in the students' reflections, as they were mostly formulated without the help of ChatGPT. However, these 'imperfections' often added individuality and increased the overall quality and authenticity of the texts, in contrast to the polished but generic appearance of the generated texts.

In terms of content and argumentation, the instructor found the majority of essays to be rather superficial (45 essays, 86.5 %) and repetitive (30 essays, 57.7 %). In particular, he criticized that the students' own opinions were often difficult to discern, i. e., that the primary question of how technologies like ChatGPT should be regulated (permissionless vs. precautionary) was often not very clearly answered. A first analysis revealed that 37 (71.2 %) essays argued for a middle ground between permissionless innovation and the precautionary principle, 12 (23.1 %) essays argued for permissionless innovation, and only one (2 %) essay explicitly argued for responsible innovation. No opinion was expressed in 2 (3.8 %) essays. However, a more detailed analysis of the essays arguing for a middle ground showed that 23 (44.2 %)

actually argued for regulation compatible with permissionless innovation, advocating a minimum of restrictions while addressing risks with flexible (self-)regulation — a critical comparison with the reflections showed that students were often unaware of this.

4.2 Student Reflection (RQ2)

4.2.1 General Experience

While students could use the free or the paid version of ChatGPT, 7 students indicated that they switched to the latter version because they were not satisfied with the results of the free version. One reason was the better answer and content quality, as noted by S13: "While I often got unusable or unsatisfactory answers with the older version, the experience with 4.0 is completely different. The answers are much more logical, more precise, and more satisfying." Another reason was the improved source handling in ChatGPT 4, where "[. . .] I was pleased to find that all the sources were real." (S9). We found no significant differences between students using ChatGPT 3.5 and ChatGPT 4 in the quality of the essays and the student reflections.

Only 16 students (30.8 %) explicitly mentioned in their reflections that they had used ChatGPT before, especially for programming tasks (5 students) or simple tasks like writing emails or searching for information and learning (4 students). Half of the experienced students stated that they had used ChatGPT for the first time for creative production of larger amounts of text, as noted by S34: "Until this essay, I have never used ChatGPT for such a task, or rather, never to this extent." Only one student declared that he had used ChatGPT for the first time.

Regarding their experiences, students generally found the experimental use of ChatGPT interesting and instructive, and there were no negative comments about the experiment itself. However, the students' experiences varied. Among the positive comments, 11 students found ChatGPT helpful overall for writing texts — e. g., S2: "It made a valuable contribution to the structure and formulation of the text." Another 9 students were impressed with the quality of the answers, e. g., S42: "Surprisingly, ChatGPT was able to handle complex queries quickly and reliably, providing appropriate answers to a large number of requests." 7 students emphasized that ChatGPT was particularly helpful for researching topics or getting an overview, e. g., S16: "[It was] very easy to get an overview of the topic without having to search for information on different websites on the Internet."

The most common negative experience was problems with sources (22 students, 42.3 %), e. g., non-existent sources or incorrect citations: "I [was] often forced to search for suitable sources myself, as ChatGPT gave either incorrect or non-existent sources." (S30). This led 7 students to comment that answers from ChatGPT should generally be viewed with skepticism and not be blindly trusted. In this context, 3 students (5.8 %) commented on the importance of background knowledge, e. g., S45: "It is therefore advisable to always be

informed about the subject and to question the output of ChatGPT with common sense." To improve the quality of references, 24 students (46.2 %) reported using Elicit to check or find additional sources. In addition, 9 students (17.3 %) mentioned that ChatGPT was often repetitive in content and wording: "A major limitation of ChatGPT is the repetition of content." (S3). 8 students (15.4 %) commented negatively on the quality of the language and complaining about style (long sentences, exaggerations).

4.2.2 Approach to ChatGPT usage

Almost two thirds of the students shed some light on their approach to using ChatGPT to write their essays (33 students, 63.5 %). From their total of 42 remarks, we extracted different approaches. Only 6 students reported that they tried to create the whole essay directly with ChatGPT (e. g., S1: "I first took the approach of writing the entire text with the chatbot and instructing it to act like a scientist in order to find appropriate sources."). 23 students proceeded step by step, either following a predefined document structure (introduction, etc.) or by topic (e. g., arguments for and against regulation) and combining or, in the words of S28, 'orchestrating' the individual parts. 3 students reported using ChatGPT primarily to consolidate or summarize text from different sources (e. g., S48: "[I used ChatGPT] to summarize relevant passages from books and articles. This saved time by reducing the effort of reading irrelevant texts that do not add value to the essay [. . .].").

In general, 12 students found it helpful to read or gather information about the topic first, in order to provide ChatGPT with as much contextual information as possible. Two of them noted that this allowed them to form their own opinions before using ChatGPT, in the words of S28: "As a tactic, I wrote down my own opinion at the beginning and used the tool as a basis for formulating various arguments. However, I did not try to use it as the main source for forming an opinion." Another 5 students said that they benefited from (iteratively) generating multiple versions of the text to quickly change and adapt their argument or to select the text they were most satisfied with.

9 students stated that they changed their usage approach during the work because they were not satisfied with the result. 5 of them had started with the approach of creating the essay directly and completely through ChatGPT. They switched to the step-by-step approach. However, S8 was not satisfied with this approach either: "Creating the text and then searching for suitable sources [was] very unfamiliar."; so he decided to research sources first and then have ChatGPT summarize them. After their first attempt, one student each reported that they provided more contextual information to ChatGPT, did additional research on their own sources, or relied primarily on their own sources and had ChatGPT summarize them. Another student noted that he translated answers backwards after switching from German to English prompts, assuming that the English content was better due to more training data.

In terms of their approach to using ChatGPT, 3 students explicitly reflected on their own role as being that of an assistant, contributing less to their work than they actually wanted

to; as S33 put it: “In retrospect, I have to admit that I tried to let ChatGPT do the thinking and writing for me, while I merely acted as a source assistant and ‘fact checker.’ In fact, the roles should be reversed, with ChatGPT being my assistant and me doing the thinking.” Also questioning the value of her own contribution, S28 mused: “I quickly understood the human task: I am responsible for the ‘human error’ to make this essay a little less perfect (less ‘creepy robot’).”

4.2.3 Efficiency and Effectiveness

A total of 32 students (61.5 %) explicitly stated that they had to intervene in the text in some way. However, students did not agree on whether ChatGPT actually made the writing process more efficient. Of the 43 students who explicitly commented on this (82.7 %), 22 (42.3 %) stated that the work was more efficient, e. g., S1: “The tools provided made it possible to process questions more quickly and access relevant information.” S35 found that “[w]orking with ChatGPT and Elicit was much easier than writing the text myself. The time required was greatly reduced.”

19 (36.5 %) students had mixed feelings. S22 stated that the advantages and disadvantages partly canceled each other out: “[T]he writing part actually seemed to be done quickly. Besides the sources, I mainly had to change the wording and enrich it with my own statements. But then it took a few hours to check all the statements and find suitable sources.” S18 found that efficiency depended on the level of quality and detail in the text: “An advantage of this tool is that when you are writing a text, you can get a rough draft very quickly. This can then be used as a basis for the actual text. In my opinion, the efficiency gain promised by using ChatGPT decreases the more detailed and specific a topic is treated.” This sentiment was shared by S21 regarding bibliography: “[The sources] suggested by GPT-4 were all useful. However, I had to read or cross-read all the sources to check them, which took a lot of time.”

Only 3 (5.8 %) students explicitly stated that they found working with ChatGPT less efficient than working without it. S8 noted that “reworking is more time consuming than writing directly” and S21 concluded that he “[...] could have written a better essay in the [same] time.” S41 attributed the lack of efficiency to the handling of sources, finding it “[...] reasonable to assume that finding suitable sources based on an existing text requires even more effort than formulating suitable text based on an existing source.”

In total, 18 students found that the result of working with ChatGPT was better than expected (e. g., S43: “The result was better than expected, because I received a wealth of information and perspectives that complemented my own research and analysis.”), 8 students found it worse than expected (e. g., S6: “[Writing the essay] was more difficult than expected. [...] This task made me realize that ChatGPT is not yet suitable for writing scientific papers.”), and 8 students had mixed feelings (e. g., S14: “Regarding the quality of the AI-generated results, the author is surprised by their quality. On the other hand, the source information and [...] its quality in terms of completeness, accuracy and precision were disappointing.”).

4.2.4 Future Usage

Finally, students were asked how they would like to use ChatGPT in the future. 37 (71.2 %) students explicitly stated that they would like to continue using tools like ChatGPT in the future, none stated that they would not. Regarding tasks, 16 (30.8 %) students stated that they would like to continue using it for text work (e. g., structuring texts, summarizing, shortening, paraphrasing). 11 (21.2 %) students indicated that they would continue to use it for information retrieval (e. g., finding ideas, locating and analyzing relevant literature) or as a study aid (6 students, 11.5 %). Regarding the division of labor between humans and AI, 15 (28.5 %) students stated that they would like to use ChatGPT in the future as an extension and support of their own skills, not as a replacement.

4.3 Student Opinions (RQ3)

42 (80.5 %) students answered the question whether ChatGPT had an influence on their own opinion, and if so, which one. 12 (23.1 %) students thought that ChatGPT had little or no influence on their opinions. 5 (10 %) of these also stated that they agreed with ChatGPT's opinion, e. g., S34: "I often agreed with ChatGPT's output, so I consider the influence on my opinion to be small." Though without a direct connection to the opinion influence question, another 9 (17.3 %) students expressed a similar sentiment, sometimes more subtly, e. g., S22: "I think ChatGPT always reacted in a nuanced way, even if not quite as critically. The (partly justified) fear of AI and where it will lead is not really the focus, but it is addressed."

24 (46.2 %) students felt that ChatGPT had a positive influence on their opinions, some arguing that it provided arguments from different perspectives that the students had not considered before, e. g., S35: "I have to admit that I agreed with most of the opinions and considerations of ChatGPT, which is why there were practically no changes or adjustments to the content. As a rule, you always approach an issue with one opinion, but ChatGPT can show you other points of view and broaden your horizon." or S15: "All in all, my involvement with ChatGPT and the question of how to deal with disruptive technologies made me rethink and reflect on my own position." Student S9 believed that ChatGPT's opinions reflected the arguments of the human experts on which it had been trained: "It is important to emphasize that the opinions and arguments presented by ChatGPT are based on sources written by experts in their respective fields."

Two students felt that ChatGPT had a negative impact on their opinion because they had to trust ChatGPT and take its answers at face value due to their lack of expertise: "As I approached the assignment without much prior knowledge, I had to rely on ChatGPT's answers. In retrospect, this was not ideal, as I was heavily influenced by the answers. Therefore, I did not find any direct contradictions, but took ChatGPT's statements as truth." (S23).

In total, 12 students found that they had to adapt the text to (better) represent their own opinions — for example S36: "Although I generally found the generated text to be very good, I sometimes had to make some adjustments or additions to ensure that my own opinion and perspective was adequately represented." and S44: "I deliberately did not restrict whether the model should argue for or against permissionless innovation. Ultimately, this led to ChatGPT contradicting itself in the main and final sections. I therefore intervened in the text and argued for a balanced combination, since both positions have advantages and disadvantages." S28 found that ChatGPT often overemphasized political correctness over specificity: "ChatGPT's wording is too politically correct, [often adding] that 'there are different opinions and no clear answers'."

5 Discussion

5.1 Essay Quality (RQ1)

While each of the 52 submitted essays was unique at the text level — consistent with prior findings on ChatGPTs linguistic diversity [Ba23] — our analysis revealed notable repetition at the argument level. This was likely due to the similarity of prompts students used with ChatGPT, leading to a lack of depth and variety in content. In addition, linguistic formulations were often repetitive and redundant, especially in longer texts. Similar to the findings of Oates & Johnson [OJ25], this redundancy diluted the argumentative strength of the essays and made them seem generic and unoriginal. While the literature identifies text structuring as one of the benefits of GenAI [NDS25; OJ25], only a third of the essays in our study were well structured, suggesting some moderating or mediating effects.

The essays generally lacked a clear position on the issue of AI regulation. A possible contributing factor to the prevalent middle-ground positions may have been ChatGPT's tendency to avoid explicit opinions. However, other factors such as students' own attitudes or the task framing may also have played a role. The 'balanced' argumentation often coincided with a superficial treatment of the core issue, lacking depth and critical analysis. This observation is consistent with findings from the literature [NZ23], which show that while GenAI tools can increase productivity, they may also lead to superficial engagement with content if not used critically. Such a lack of engagement raises the possibility of unnoticed and unintended influences on students' argumentation or framing, as discussed in RQ3.

Instructor feedback and gradings suggest that essays incorporating personal adjustments or additions to ChatGPT's output tended to be more effective. These modifications appeared to improve argument quality and often made it more difficult to distinguish between AI-generated and human-generated content, implying that the occasional presence of 'human error' was an indicator of successful human-AI collaboration. This observation aligns with findings from Fyfe [Fy23], who highlights the importance of human intervention in AI-assisted writing.

5.2 Student Experience (RQ2)

In terms of student reflections, the general experience with ChatGPT was largely positive, with many students being positively surprised by the quality of the output generated. This finding is consistent with the positive attitudes found in the literature [CH23; Ra25]. The different approaches that students took to using ChatGPT — ranging from using it as a primary writing tool to using it as a supportive assistant — demonstrate a nuanced understanding of the potential role of AI in the writing process, which is critical to fostering effective learning strategies.

Students also pointed out the steep learning curve and the considerable effort required to check the output for accuracy. The reliability of AI-generated content and the need for verification emerged as critical issues, echoing concerns raised by participants in van Niekerk et al.'s study [NDS25]. In our study, source work was particularly demanding: finding appropriate references and correcting erroneous sources or citations proved time-consuming. Therefore, some students assumed that they would have written the essay faster (and better) — and with greater learning success — without ChatGPT. This is also reflected in the literature on the balance between the time saved by AI assistance and the additional time required for checking and editing [Ba23; Fy23]. Overall, less than half of the students who explicitly commented on efficiency found working with ChatGPT more efficient than writing the essay themselves, in contrast to Noy & Zhang's study [NZ23], which reported significant productivity gains when using AI.

5.3 Student Opinions (RQ3)

As noted above, the essays tended to adpot neutral or middle-of-the-road positions, often lacking detailed justification or critical analysis. Upon closer inspection, and despite the students' perception that ChatGPT had not significantly influenced their opinions, our analysis revealed a trend in the essays toward permissionless innovation, even when students believed they were arguing for a middle ground. This pattern may reflect a combination of factors, including framing effects in ChatGPT's output, students' prior beliefs, underlying biases in ChatGPT's training data (e. g., dominant US-based pro-innovation discourses), and fine-tuning processes aimed at neutrality and safety. Given our research design, we cannot determine which of these factors contributed most to the observed argumentative trend.

This possible latent persuasive effect is further illustrated by instances where ChatGPT's output deviated from students' initial opinions, forcing them to revise and realign the text with their own views. This observation aligns with concerns that GenAI tools could negatively affect critical thinking [DK25; Sw23], and that their ease of use may lead to overreliance and reduced motivation for independent verification of information [Fa24]. At the same time, the fact that almost half of the students reported a positive impact on their opinions — such as broadening their perspective — and only two reported a negative impact highlights the need for critical engagement with AI-generated material.

Overall, our findings underscore the potential of GenAI to influence academic discourse and steer argumentation toward certain viewpoints, even in subtle ways. The use of GenAI in academic writing must therefore always be critically evaluated and guarded against such risks [Di23; Ra23]. However, this is balanced by the possible positive effects of GenAI on students' perspectives, suggesting that such tools can also serve as valuable resources for critical engagement and reflection [De23; Fy23].

6 Conclusion

In this paper, we presented empirical research on the use and potential effects of GenAI in academic writing in higher education, contributing to the limited and sometimes conflicting findings in the literature. Our study highlights both the potential benefits and challenges of AI-assisted writing, showing that while AI-generated essays maintained acceptable quality, they often suffered from superficial argumentation, content repetition, and a lack of strong individual perspectives. Students found ChatGPT useful for structuring their essays and generating text efficiently, but they also struggled with fact-checking and integrating AI-generated content with their own ideas.

A key takeaway from our research is the potential for subtle influence by the AI on students' argumentation and perspective-taking. While most students believed that ChatGPT did not shape their perspectives and defended middle-of-the-road arguments in their reflections, their essays nonetheless tended toward more poignant positions, potentially reflecting GenAI influence. This raises concerns about the potential impact of AI on academic discourse and critical thinking, especially in areas unfamiliar to students, and underscores the importance of fostering independent analytical skills in higher education.

Our study has several limitations. As outlined in the Methods section, the lack of a control group limits the ability to draw causal conclusions about the effects of ChatGPT on writing quality or opinion shifts. Furthermore, as the study involved a single cohort from one institution, the generalizability of the findings may be limited.

Going forward, educators and institutions need to establish clear guidelines for AI-assisted writing to ensure that AI is used as a support tool rather than a substitute for human cognition. Universities should incorporate AI literacy training into their curricula to equip students with the skills to critically evaluate AI-generated content, refine AI-assisted writing, and maintain academic integrity.

References

[Al23] Alkhaqani, A.: Can ChatGPT Help Researchers with Scientific Research Writing? Journal of Medical Research and Reviews 1 (1), pp. 9–12, 2023, DOI: 10.5455/JMRR.20230626013424.

[Ba23] Bašić, Ž. et al.: ChatGPT-3.5 as writing assistance in students' essays. Humanities and Social Sciences Communications 10 (1), pp. 1–5, 2023, DOI: 10.1057/s41599-023-02269-7.

[CH23] Chan, C. K. Y.; Hu, W.: Students' voices on generative AI: perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education 20 (1), p. 43, 2023, DOI: 10.1186/s41239-023-00411-8.

[De23] Dergaa, I. et al.: From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. Biology of Sport 40 (2), pp. 615–622, 2023, DOI: 10.5114/biolsport.2023.125623.

[Di23] van Dis, E. A. M. et al.: ChatGPT: five priorities for research. Nature 614 (7947), pp. 224–226, 2023, DOI: 10.1038/d41586-023-00288-7.

[DK25] Dimeli, M.; Kostas, A.: The Role of ChatGPT in Education: Applications, Challenges: Insights From a Systematic Review. Journal of Information Technology Education. Research 24, 2025, DOI: 10.28945/5422.

[DLL23] Dai, Y.; Liu, A.; Lim, C. P.: Reconceptualizing ChatGPT and generative AI as a student-driven innovation in higher education. Procedia CIRP, The 33rd CIRP Design Conference 119, pp. 84–90, 2023, DOI: 10.1016/j.procir.2023.05.002.

[Ek23] Eke, D. O.: ChatGPT and the rise of generative AI: Threat to academic integrity? Journal of Responsible Technology 13, p. 100060, 2023, DOI: 10.1016/j.jrt.2023.100060.

[El25] Elicit: The AI Research Assistant, 2025, https://elicit.com/, accessed: 05/03/2024.

[Fa24] Farrokhnia, M. et al.: A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International 61 (3), 2024, DOI: 10.1080/14703297.2023.2195846.

[Fy23] Fyfe, P.: How to cheat on your final paper: Assigning AI for student writing. AI & SOCIETY 38 (4), pp. 1395–1405, 2023, DOI: 10.1007/s00146-022-01397-z.

[GCH13] Gioia, D. A.; Corley, K. G.; Hamilton, A. L.: Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. Organizational Research Methods 16 (1), pp. 15–31, 2013, DOI: 10.1177/1094428112452151.

[Ka23] Kanabar, V.: An Empirical Study of Student Perceptions When Using ChatGPT in Academic Assignments. In (Zlateva, T.; Tuparov, G., eds.): Computer Science and Education in Computer Science. Springer Nature Switzerland, pp. 385–398, 2023, DOI: 10.1007/978-3-031-44668-9_30.

[KR22] Kuckartz, U.; Rädiker, S.: Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. Beltz Juventa, 2022.

[LKM25] Liu, Y.; Kong, W.; Merve, K.: ChatGPT applications in academic writing: a review of potential, limitations, and ethical challenges. Arquivos Brasileiros de Oftalmologia 88, e2024, 2025, DOI: 10.5935/0004-2749.2024-0269.

[Me25] Melisa, R. et al.: Critical Thinking in the Age of AI: A Systematic Review of AI's Effects on Higher Education. Educational Process International Journal 14 (1), 2025, DOI: 10.22521/edupij.2025.14.31.

[NDS25] van Niekerk, J.; Delport, P. M. J.; Sutherland, I.: Addressing the use of generative AI in academic writing. Computers and Education: Artificial Intelligence 8, p. 100342, 2025, DOI: 10.1016/j.caeai.2024.100342.

[NZ23] Noy, S.; Zhang, W.: Experimental evidence on the productivity effects of generative artificial intelligence. Science 381 (187), 2023.

[OJ25] Oates, A.; Johnson, D.: ChatGPT in the Classroom: Evaluating its Role in Fostering Critical Evaluation Skills. International Journal of Artificial Intelligence in Education, 2025, DOI: 10.1007/s40593-024-00452-8.

[Op24] OpenAI: ChatGPT, 2024, https://chat.openai.com, accessed: 02/28/2024.

[Pe23] Perkins, M.: Academic integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching and Learning Practice 20 (2), 2023, DOI: 10.53761/1.20.02.07.

[Ra23] Ray, P. P.: ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems 3, pp. 121–154, 2023, DOI: 10.1016/j.iotcps.2023.04.003.

[Ra25] Ravšelj, D. et al.: Higher education students' perceptions of ChatGPT: A global study of early reactions. PLOS ONE 20 (2), Publisher: Public Library of Science, e0315011, 2025, DOI: 10.1371/journal.pone.0315011.

[Sm23] Smolansky, A. et al.: Educator and Student Perspectives on the Impact of Generative AI on Assessments in Higher Education. In: Proceedings of the 10th ACM Conference on Learning @ Scale. L@S 2023. Pp. 378–382, 2023, DOI: 10.1145/3573051.3596191.

[Sw23] Swargiary, K.: The Impact of ChatGPT on Student Learning Behavior: An Experimental Research Study, Rochester, NY, 2023, DOI: 10.2139/ssrn.4611213.

[Th16] Thierer, A.: Permissionless Innovation: The Continuing Case for Comprehensive Technological Freedom. Mercatus Center at George Mason University, 2016.

Smart Start: A Critical Analysis of Chatbots as Tools to Support Study Orientation

Lea Cohausz [1], Thilo I. Dieing [1], Belinda Merkle [2], Marc Scheffler [1], and Stefan Janke [2]

Abstract: In higher education, elevated dropout rates pose a significant problem, leading to vast societal and personal costs and disadvantages. In Germany, in particular, dropout is closely connected to a misfit between students' expectations and their choice of study program. As a countermeasure, a chatbot is proposed that targets prospective students and is supposed to help them with their study orientation by providing them with accurate and personalized information. In this study, we first use a survey including the Technology Acceptance Model to test whether the target group would even accept such a chatbot and whether there are differences in the acceptance for different demographic groups. Our results show high acceptance and no differences among demographic groups, although we propose that differences among demographic groups should be scrutinized in future research. Subsequently, we develop a chatbot architecture using Large Language Models and Retrieval Augmented Generation that enables accurate and personalized answers while simultaneously ensuring appropriate responses to inappropriate user prompts. Our first evaluation of the chatbot shows promising results. The chatbot will become part of a study orientation platform in one of Germany's largest federal states.

Keywords: chatbot, LLM, dropout, service-oriented chatbot, study choice advisor, TAM

1 Introduction

Dropout rates in higher education institutions are high in many countries, particularly in systems that require students to choose a major immediately, such as in Germany or Switzerland. In Germany, 28 % of Bachelor students drop out [He22]. These dropouts incur both societal costs (wasted financial resources) and personal costs for the individuals involved (frustration, time loss)[3] [Fa18].

Although multiple factors contribute to dropping out, a German study by Blüthmann et al. found that the main reasons are a mismatch between study program content and personal interests (35.5 %) or misjudging the program's demands (25.3 %) [BLT12]. Other research

[1] University of Mannheim, Chair of Artificial Intelligence, 68159 Mannheim, Germany,
lea.cohausz@uni-mannheim.de, https://orcid.org/0000-0002-6164-3988;
thilo.dieing@uni-mannheim.de, https://orcid.org/0009-0000-2256-0795;
marcscheffler@outlook.com, https://orcid.org/0009-0004-1249-3487

[2] University of Mannheim, Chair of Educational Psychology, 68159 Mannheim, Germany,
belinda.merkle@uni-mannheim.de, https://orcid.org/000-0003-1799-8850;
stefan.janke@uni-mannheim.de, https://orcid.org/0000-0002-1916-0069

[3] The cost of providing one student with higher education in Germany is around 10,500€ annually [Br22]. With 481,962 first-year Bachelor students in 2023 [Bu24] and an average dropout time of 3.8 semesters, this amounts to 2.69 billion euros in taxpayer money and about 256,403 lost study years per cohort [He17].

doi:10.18420/delfi2025_11

also shows that students with a good match between their interests and study content are more satisfied [MMD24]. Therefore, reducing the mismatch by providing relevant information to prospective students could help lower dropout rates. However, the overwhelming amount of information available—ranging from higher education institutions to online self-assessments (OSAs), financing information, and career fairs—can be difficult to navigate. First-generation academics or students from migrant backgrounds may struggle more due to a lack of social and cultural capital, as evidenced by higher dropout rates in these groups [MM22; MS13]. These students may also be reluctant to approach study advisors.

To address this issue, we aim to develop a study advisor tool that will provide answers to key questions (e. g., about programs, admissions, financial aspects, and OSAs), offering additional guidance for school graduates. This tool needs to: (1) be accepted by prospective students, meaning it is perceived as useful, easy to use, and accessible; (2) be equally accepted across demographic groups; (3) answer a wide range of student questions accurately, with links to external materials where needed; and (4) identify and avoid responding to inappropriate questions, such as those that are discriminatory or sensitive.

We propose using Large Language Models (LLMs) to create a chatbot capable of offering personalized, relevant information with a low threshold for use. LLM-based chatbots can answer a variety of questions without needing predefined responses, and with Retrieval Augmented Generation (RAG), they can provide precise answers tailored to specific queries [Fa24]. However, there are challenges when using LLMs in critical information systems. First, it is unclear whether users will accept the guidance of LLM-based chatbots. Second, there is uncertainty about whether acceptance varies across demographic groups, which is important for ensuring the tool's widespread use. Third, LLMs sometimes provide incorrect answers, a phenomenon known as "hallucination" [Fa24]. Fourth, they tend to respond to sensitive or inappropriate questions [Ba21]. In this paper, we aim to address these challenges:

- We investigate prospective students' acceptance of a study advisor chatbot using a survey based on the Technology Acceptance Model (TAM) in Section 3. Our findings suggest high acceptance [BLT12].
- We explore whether acceptance varies across demographic groups in Section 3.2.2. While our data shows no significant differences, we note that further research is needed.
- We justify the need for a reliable study advisor chatbot that provides sensible answers with minimal hallucination in Section 4. Our evaluation shows positive results.
- We describe a filter in our chatbot design that detects inappropriate questions, ensuring it avoids answering them. Section 4.2.2 demonstrates its effectiveness.

This academic investigation is crucial as the resulting study-advisor chatbot will be integrated into a study orientation platform aimed at helping prospective students navigate the complex decision-making process between higher education institutions and the myriad of available study programs.

2 Background

We first discuss existing literature concerning chatbots in general, chatbots in education, and chatbots' acceptance to subsequently derive our research objective.

2.1 Characteristics and Use of Chatbots

Chatbots have become widely used across various fields, from management to IT support [AM20]. They can be categorized based on four key characteristics: openness, service, goal, and functionality [AM20]. Openness refers to whether a chatbot addresses a broad range of topics (open) or a specific one (closed). Service indicates whether the chatbot offers general or personalized interactions. Goal defines whether the chatbot's purpose is to inform, engage in conversation, or perform tasks. Functionality relates to how the chatbot generates responses, either through rule-based methods or generative Machine Learning, with or without additional information (RAG).

With the rise of advanced LLMs in recent years, most modern chatbots are classified as open or semi-open. These chatbots typically offer personalization, can inform, converse, and solve tasks, and rely on generative Machine Learning along with Retrieval Augmented Generation (RAG), which incorporates domain-specific information to enhance the chatbot's output [Fa24].

2.2 Chatbots in Education

A 2023 survey by Kuahil et al. found that 88.88 % of chatbots used in education are rule-based, relying on predefined question-answer pairs [Ku23]. This limits their functionality, as closed chatbots are suited only for narrow contexts. In contrast, a study-advisor chatbot needs to address a wide range of specific and broad questions. While the number of open chatbots is expected to grow quickly, this finding highlights the current limitations of educational chatbots.

Educational chatbots can also be classified by purpose, either teaching- or service-oriented [OA21; PDP20]. Teaching-oriented chatbots support students and educators in learning tasks, such as teaching coding [DTZ20; Wi20], engineering [Me20], or grading [Ba24]. Service-oriented chatbots, on the other hand, assist students and administrative staff with daily tasks, like helping with curriculum or study decisions [OA21]. A study-advisor chatbot is a clear example of a service-oriented chatbot. While service-oriented chatbots do exist in education [Di18; FSW23], they are less common than teaching-oriented ones [OA21]. The adoption of LLM-based solutions has not yet fully reached this area. As a result, most service-oriented chatbots remain rule-based and are limited to closed knowledge domains [Di18; FSW23]. LLMs supported by RAG offer significant benefits, especially for

study-advisor chatbots, as they need to answer a variety of questions on diverse topics and levels of specificity, such as "How much money do I need per month while studying?" and "How much statistics is required in program X at university Y?"

2.3 Chatbots and their Acceptance

Chatbots can positively impact various sectors, including businesses [AMA23; KK20], healthcare [GSK21; Mi21], and education [DSC24; RRS17]. However, for chatbots to realize their full potential, users must engage with them, trust their guidance, and incorporate the information into their decision-making. In the case of study-advisor chatbots, educational effectiveness depends on users' trust in the information and their integration of it into their study decisions. The Technology Acceptance Model (TAM) is useful for evaluating user acceptance [Da89]. According to TAM, acceptance of a system is based on the perceived benefits it offers, divided into two factors: Perceived Usefulness (PU), which measures how much the system improves the current state, and Perceived Ease of Use (PEU), which reflects the effort required to use the system. PU indicates the system's utility, so a study-advisor chatbot would be considered useful if it aids the study decision process. PEU describes the ease of navigating the system, and systems that are easier to use tend to be more effective. High PU and PEU lead to a positive Attitude toward using the system (ATT), which influences Behavioral Intentions (BI) and, eventually, actual system usage. However, actual use can only be measured once the system is developed. As developing such systems is costly, it is crucial to first assess general acceptance before proceeding with development. Figure 1 illustrates the causal relationships in the TAM model [Da89], and the model's components are typically evaluated through user surveys.

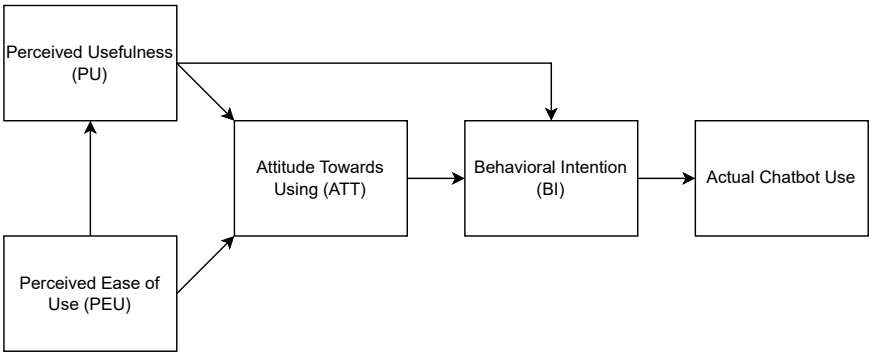


Fig. 1: Technology Acceptance Model based on Davis [Da89].

TAM has been widely used to evaluate chatbot acceptance, including in sectors like banking [AVS21] and tourism [SA24]. In education, studies applying the TAM framework highlight

the significance of PU as a key factor in students' [LCC23] and teachers' [CCM23] willingness to use educational chatbots [BKS21]. Further research has identified potential gender differences in chatbot acceptance [Yi23], with male students generally more willing to use chatbots and expressing fewer concerns [SOM24], while other studies find no significant effects of gender or age [Ra22].

2.4 Research Objectives

Considering existing literature and guided by our requirements formulated in Section 1, we formulate the following research questions: **RQ1:** Do student aspirants in the study orientation phase accept a study-advisor chatbot according to the concepts covered by the TAM? **RQ2:** Are there group differences for any of the concepts covered by the TAM for different demographic groups (gender, first-generation Germans, parental education, age)? **RQ3:** If **RQ1** is answered affirmatively: Can we construct a chatbot using open-source LLMs that can answer a wide range of questions correctly and precisely? **RQ4:** If **RQ1** is answered affirmatively: Is it possible to filter inappropriate questions reliably?

As our later two research questions hinge on a positive answer to **RQ1**, we will first describe the methodology for answering the first two questions and then describe the respective results. Afterward, we will discuss the development and evaluation of the chatbot.

3 Chatbot Acceptance

3.1 Method

To address our first research goal—measuring the user acceptance of a study-advisor chatbot—we conducted an online survey to gather opinions from potential users. Our target group included individuals who are or have been in the process of choosing a study program. This group consisted of young adults still attending school and those who have recently decided on their study program, including those in their first semesters. We focused on the latter group, as they have recently navigated the decision process and can provide insights on what would have been helpful to them.

The survey included various question categories based on the TAM model, incorporating factors that influence system usage, and adapted from Ahn et al. [ARH07] to suit a chatbot and study orientation context. Perceived Usefulness (PU) was assessed with six statements, such as "I anticipate that the chatbot is helpful for my study orientation." Perceived Ease of Use (PEU) was evaluated through four statements, e. g., "I anticipate that the chatbot is easy to use." User Attitude (ATT) was captured with one statement, and Behavioral Intention (BI) with two. All items used a 7-point scale, ranging from strong disagreement to strong agreement. We also collected demographic data, including gender, age, education, parental

education, and immigration background. Additionally, we asked about participants' use of ChatGPT, as we hypothesized that frequent use of chatbots could influence acceptance of the study-advisor chatbot, making users more likely to view chatbots as useful and be open to using others.

We used a convenience snowball sampling method to reach the target group, collecting data from August 18th to September 9th, 2024. The final sample included 80 participants, with 66 fitting our target group and 65 passing an attention test, which ensured that participants answered thoughtfully [MK16]. Of the 65 valid responses, 55 % identified as female, 43 % as male, and the rest as non-binary. The average age was 19.3 years (SD = 2.3), with ages ranging from 16 to 24. Among the participants, 65 % had recently completed their study orientation and started a study program or vocational training, while the others were in or near their study orientation phase. A notable observation from the demographic data is the relatively high level of parental education and low percentage of first-generation Germans, reflecting our focus on young people having achieved or in the processes of achieving a general university entrance qualification (Abitur) [MM22]. At this point in their lives, the social and cultural capital associated with parental education and migration background has already influenced their educational attainment [MM22].

Next, we assessed the reliability of the constructs using Cronbach's alpha and Confirmatory Factor Analysis (CFA) [ACS15; Br15; Cr51]. The results indicated that the items for each construct were reliably measuring their respective constructs. All constructs had Cronbach's alpha values above 0.8, indicating strong reliability. The model fit was also moderate to good: χ^2, $p < 0.000$, RMSEA = 0.014, SRMR = 0.084, CFI = 0.888. We found that some items from PEU and PU also loaded on the other construct, but these cross-loadings were still in an acceptable range, and some covariance between the constructs was expected.

To examine the impact of demographics and ChatGPT usage on the constructs and their interrelationships, we used path analysis. Path analysis extends multiple regression to model direct and indirect relationships between variables in a hypothesized model [Ro12]. It uses simultaneous linear equations to analyze variable dependencies and provides a visual representation through a path diagram. In our analysis, we specified four structural equations, each with one latent construct as the independent variable. The dependent variables (DVs) included demographic factors and other latent constructs from the TAM model. For example, the equation for BI is: $BI = ATT + PU + Age + Gender + FirstGenerationGerman + ParentWithUniversityDegree + ChatGPT$. This approach allowed us to assess how each latent construct was influenced by both demographic variables and other constructs, accounting for interdependencies. We report fully standardized coefficients for easier comparison.

Variable	Mean/ Mode	Minimum / % Smallest Group	Maximum / % Largest Group	N unique values
Age	19.3	16	24	9
Gender	Female	Non-Binary, 2 %	Female, 55 %	3
First Generation Germans	No	Yes, 18%	No, 82 %	2
Any Parent with University Degree	Yes	No, 35 %	Yes, 65 %	2
ChatGPT	Yes	No, 45 %	Yes, 55 %	2

Tab. 1: The table provides key statistics for variables included in the survey.

3.2 Results

3.2.1 General Chatbot Acceptance

Figure 2 shows the average agreement values for each TAM construct, along with their confidence intervals. The light blue areas represent the response distribution. A value of four reflects a neutral stance, with higher values indicating positive and lower values negative stances. All constructs exhibit similar and very positive average values, ranging from 5.60 (ATT) to 5.84 (PEU). Only a small percentage of participants expressed negative views of the chatbot (7.69 % for ATT and BI, 4.63 % for PEU, and 6.15 % for PU).

Regarding **RQ1**, these results demonstrate high acceptance of the study advisor chatbot according to the TAM model, suggesting favorable prospects for its development and use.

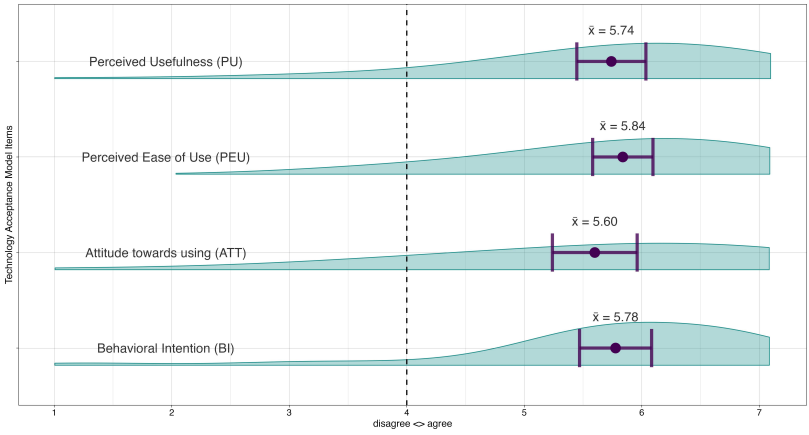


Fig. 2: Acceptance of a study orientation chatbot using the Technology Acceptance Model (TAM) with values ranging from 1 (completely disagree) to 7 (completely agree).

3.2.2 Demographics and Chatbot Acceptance

Table 2 presents the results of the regressions in the path analysis for each TAM construct. Overall, the relationships among constructs align with expectations, though PEU is not significant for ATT, and PU is not significant for BI. This absence of statistical significance may be due to the fact that the other constructs capture much of the variance relevant for the respective dependent variables and that, given the small sample size, smaller remaining variances cannot be utilized. All other relationships in the TAM are significant and strong. For example, a one standard deviation increase in ATT leads to a 0.78 standard deviation increase in BI, on average. Other variables have smaller coefficients in each model.

Age is significant for BI, with a relatively large coefficient for both BI and PEU compared to other demographic variables. The positive coefficient suggests that older individuals may show more agreement, which is surprising as we expected younger individuals to have a higher intention to use the chatbot. This finding may be influenced by the small sample size and limited age range. Other demographic variables are not significant. Although not statistically significant, *First Generation German* shows a relatively large coefficient for PEU. Given the small number of participants in this group, this variable should not be disregarded entirely. Similarly, frequent ChatGPT use has a large coefficient for each construct, although it lacks statistical significance. The coefficient for PEU, PU, and ATT is among the largest, suggesting that a more refined measure of ChatGPT use in a larger sample could reveal interesting patterns.

In conclusion, demographic variables do not show significant associations with the chatbot's acceptance. Only age was significant for BI, while other factors like gender, first-generation status, and parental education were not. However, caution is needed when drawing conclusions. Thus, **RQ2** can be answered by noting that there are few indications of demographic differences in our sample, though further investigation with larger samples is needed. Additionally, we only measured user expectations and behavioral intention, not actual chatbot usage. Future research should assess a larger sample post-launch to connect expectations with real interaction patterns.

4 Chatbot Development and Evaluation

Given the positive results for RQ1, which show that prospective students highly accept study advisor chatbots, we now focus on the development and evaluation of a chatbot designed to answer a broad range of questions accurately (RQ3) and filter inappropriate questions effectively (RQ4).

Variables	PEU	PU	ATT	BI
ATT	-	-	-	**0.78 (0.000)**
PU	-	-	**0.71 (0.000)**	0.05 (0.604)
PEU	-	**0.50 (0.000)**	-0.06 (0.578)	-
Age	-0.19 (0.112)	-0.04 (0.724)	0.02 (0.856)	**0.16 (0.026)**
Gender	0.06 (0.643)	0.05 (0.680)	-0.08 (0.399)	0.03 (0.692)
First Generation German	0.15 (0.209)	-0.08 (0.440)	0.10 (0.236)	0.07 (0.323)
Any Parent with University Degree	-0.03 (0.806)	0.02 (0.840)	0.12 (0.193)	-0.09 (0.170)
ChatGPT	0.19 (0.111)	0.24 (0.389)	0.12 (0.181)	0.07 (0.342)

Tab. 2: The table displays the regression outputs of the path analysis for each of the dependent variables PEU, PU, ATT, and BI (columns). Values significant on the 5 %-level are highlighted in bold. The p-values are added in the brackets. The coefficients are standardized.

4.1 Method

4.1.1 Chatbot Architecture

We built a generative open-domain chatbot using open-source Large Language Models (LLMs) combined with Retrieval Augmented Generation (RAG). The user's query is first embedded and compared to an external data source, also embedded.

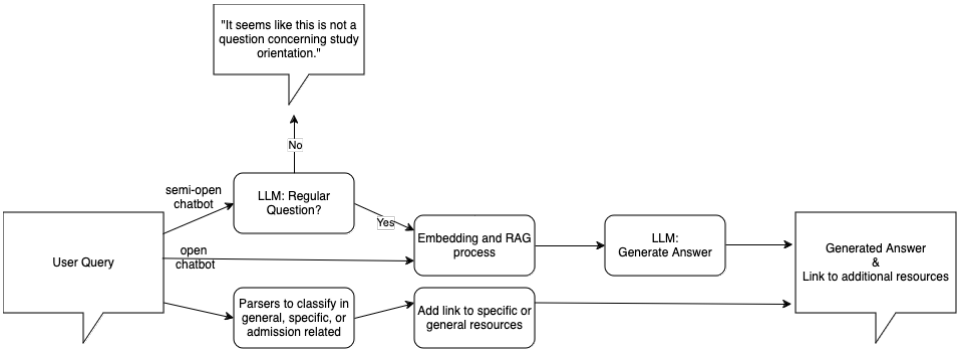


Fig. 3: The architecture of open and semi-open chatbots for study program orientation utilizing RAG and LLM.

This comparison generates a set of best-matching data entries, which are then provided to the LLM to answer the user's question. We tested two RAG match sizes, $N = 30$ and $N = 60$, to handle broader user queries. The more effective size will be used for a semi-open chatbot. This architecture makes the chatbot open, interpersonal, informative/chat-based, and generative. In addition to the open-domain chatbot, we propose a semi-open chatbot with a more limited scope to avoid misuse and prevent the generation of offensive content,

as discussed in Section 1 [Ba21]. The semi-open chatbot focuses solely on assisting student aspirants with their decision process. To prevent inappropriate user inquiries, we filter out rude, discriminatory, or sensitive queries before passing them to the LLM. This is illustrated in Figure 3. Unlike the open-domain chatbot, the semi-open bot first evaluates the user's question with an LLM to determine if it is valid. Inappropriate prompts trigger a predefined response, following the topic-avoidance strategy [Xu21]. Combining RAG and LLM ensures factual accuracy while leveraging natural language capabilities. Using an external database improves control over information, reducing hallucinations and avoiding costly pre-training.

For building the chatbots for a German application, we used the Cross English & German RoBERTa Sentence embedding model [Ma20] and Mistralai's Mixtral 8x7B [Ji24]. These models were selected for their strong performance in German-language applications [DSC24]. Open-source models also enable more reproducible results. We sourced data from the Federal Employment Agency's website[4], which includes program names, descriptions, and university details, all embedded for retrieval. A pipeline was added to detect universities and programs in user queries, providing reliable links to information tools like OSAs, ensuring accurate results without hallucinations.

4.1.2 Chatbot Evaluation Standards

Evaluating LLM-driven chatbots remains an open research challenge with no perfect gold standard. The most reliable method is manual grading by coders [DSC24], though it is time-consuming. We assessed open and semi-open chatbots using 70 study orientation questions. These questions consisted of regular prompts (e. g., "Where can I study program X?"), explorative prompts, unusual prompts, including Reddit questions and critical prompts to test the developed filter layer, and prompts provided by real study advisors. Three independent coders then graded responses on a 3-point scale: 1 (correct), 0 (sufficient), and -1 (wrong).

4.2 Results

4.2.1 Overall Results

The coders presented overall agreement in their grades with a Krippendorff's alpha [HK07; Kr18] of 0.70. To evaluate the three tested chatbots, the individual grades were averaged. Figure 4 displays the performances of the bots grouped by the different prompt types. First, we examine the results for the two sizes of RAG matches in the open-domain chatbots. We can see that there are high levels of uncertainty in the answer quality if only 30 matches are given to the LLM. In contrast, the chatbot with 60 matches is able to generate more stable

[4] https://web.arbeitsagentur.de/berufenet/studienfelder

and correct answers to the various user prompts. Based on this, we chose to implement the RAG with 60 matches in the semi-open chatbot. Comparing the open and semi-open bots revealed no significant difference in their ability to answer regular, explorative, or study advisor prompts correctly.

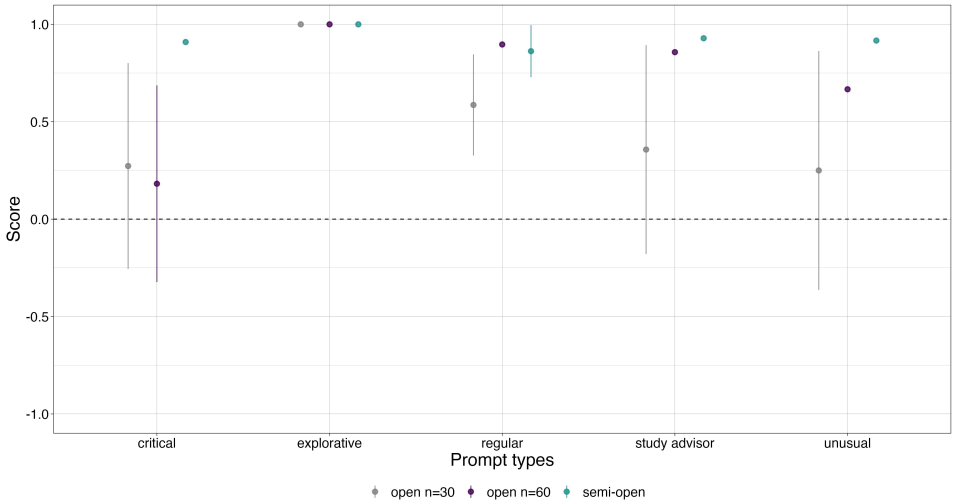


Fig. 4: Performance of both open chatbot versions with 30 and 60 RAG matches and the semi-open chatbot (RAG matches = 60) based on 70 example prompts.

4.2.2 Filtering Inappropriate Questions

Despite strong overall performance, the open-domain chatbot only achieves an average grade of 0.18 for critical prompts, highlighting the need for a filter to manage problematic queries. By applying our filter layer, the semi-open chatbot significantly improves, reaching an average grade of 0.91. These results support the effectiveness of filtering problematic prompts to address potential LLM biases, answering RQ4. Overall, the chatbot evaluation confirms that, as stated in RQ3, using RAG and LLMs is a valid and reliable approach for building a chatbot for study program orientation.

5 Conclusion, Limitations, and Outlook

This paper addresses high dropout rates linked to mismatched student expectations and chosen study programs by proposing a chatbot to support prospective students during their study orientation. The chatbot provides personalized, accurate information to help guide their decisions. Our survey results show high acceptance of the chatbot, with no

significant differences across demographic groups. We also outline the rationale for a reliable study-advisor chatbot that minimizes errors, handles inappropriate questions, and delivers sensible answers on study choice.

While the results are promising, this paper serves as a starting point for further long-term evaluation and implementation. Given the positive feedback, we are confident in providing the chatbot to the target group. However, real-world interactions may reveal unforeseen issues that need addressing. These interactions will also shed light on the chatbot's usefulness and acceptance as an additional resource for prospective students. We plan to expand the survey to include more chatbot users and non-users for more accurate insights into demographic differences in usage and acceptance. A limitation of our current study is the small sample size, which restricts our ability to examine demographic factors in depth, such as the effect of being a first-generation German student. A closer look at the parents' countries of origin would also be valuable, as parents from European countries may have more insight into the study orientation process than those from countries with different higher education systems. Further analysis of parental education and prior chatbot experience is also warranted. Future research should explore whether chatbot users find the chatbot helpful in decision-making and, in the long term, whether these users experience lower dropout rates compared to non-users, which is the ultimate goal of the tool. Finally, our evaluation of the developed chatbot demonstrates the potential of this approach, but there is room for improvement. Using the LLM in the filtering layer may introduce biases, so we suggest developing tailor-made classifiers based on a larger dataset of user queries to improve results and reliability.

References

[ACS15] Armentano, M. G.; Christensen, I.; Schiaffino, S.: Applying the technology acceptance model to evaluation of recommender systems. Polibits (51), pp. 73–79, 2015.

[AM20] Adamopoulou, E.; Moussiades, L.: An overview of chatbot technology. In: IFIP international conference on artificial intelligence applications and innovations. Springer, pp. 373–383, 2020.

[AMA23] Azmi, M.; Mansour, A.; Azmi, C.: A Context-Aware Empowering Business with AI: Case of Chatbots in Business Intelligence Systems. Procedia Computer Science 224, pp. 479–484, 2023.

[ARH07] Ahn, T.; Ryu, S.; Han, I.: The impact of Web quality and playfulness on user acceptance of online retailing. Information & management 44 (3), pp. 263–275, 2007.

[AVS21] Alt, M.-A.; Vizeli, I.; Săplăcan, Z.: Banking with a chatbot–A study on technology acceptance. Studia Universitatis Babes-Bolyai Oeconomica 66 (1), pp. 13–35, 2021.

[Ba21] Bang, Y. et al.: Assessing Political Prudence of Open-domain Chatbots, 2021, arXiv: 2106.06157 [cs.CL], https://arxiv.org/abs/2106.06157.

[Ba24] Baral, S. et al.: Automated Assessment in Math Education: A Comparative Analysis of LLMs for Open-Ended Responses. In: Proceedings of the 17th International Conference on Educational Data Mining. Pp. 732–737, 2024.

[BKS21] Brachten, F.; Kissmer, T.; Stieglitz, S.: The acceptance of chatbots in an enterprise context–A survey study. International Journal of Information Management 60, p. 102375, 2021.

[BLT12] Blüthmann, I.; Lepa, S.; Thiel, F.: Überfordert, enttäuscht, verwählt oder strategisch? Eine Typologie vorzeitig exmatrikulierter Bachelorstudierender. Zeitschrift für Pädagogik 58 (1), pp. 89–108, 2012.

[Br15] Brown, T. A.: Confirmatory factor analysis for applied research. Guilford publications, 2015.

[Br22] Brugger, P. et al.: Bildungsfinanzbericht 2015. 2022.

[Bu24] Bundesamt, S.: Anzahl der Studienanfänger:innen im ersten Hochschulsemester in Deutschland in den Studienjahren 1995/96 bis 2023/24. Statista, 2024, https://de.statista.com/statistik/daten/studie/4907/umfrage/studienanfaenger-in-deutschland-seit-1995/.

[CCM23] Chocarro, R.; Cortiñas, M.; Marcos-Matás, G.: Teachers' attitudes towards chatbots in education: a technology acceptance model approach considering the effect of social language, bot proactiveness, and users' characteristics. Educational Studies 49 (2), pp. 295–313, 2023.

[Cr51] Cronbach, L. J.: Coefficient alpha and the internal structure of tests. psychometrika 16 (3), pp. 297–334, 1951.

[Da89] Davis, F. D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly, pp. 319–340, 1989.

[Di18] Dibitonto, M. et al.: Chatbot in a Campus Environment: Design of LiSA, a Virtual Assistant to Help Students in Their University Life. In (Kurosu, M., ed.): Human-Computer Interaction. Interaction Technologies. Springer International Publishing, Cham, pp. 103–116, 2018.

[DSC24] Dieing, T. I.; Scheffler, M.; Cohausz, L.: Enhancing Chatbot-Assisted Study Program Orientation. In: Proceedings of DELFI Workshops 2024. Gesellschaft für Informatik eV, pp. 10–18420, 2024.

[DTZ20] Daud, S. H. M.; Teo, N. H. I.; Zain, N. H. M.: Ejava chatbot for learning programming language: Apost-pandemic alternative virtual tutor. International Journal 8 (7), pp. 3290–3298, 2020.

[Fa18] Faas, C. et al.: Socioeconomic success and mental health profiles of young adults who drop out of college. Journal of Youth Studies, 21(5), 669–686, 2018.

[Fa24] Fan, W. et al.: A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Pp. 6491–6501, 2024.

[FSW23] Fersch, M.-L.; Schacht, S.; Woldai, B.: Exploring AI in Education: A Quantitative Study of a Service-Oriented University Chatbot. In: The Paris Conference on Education 2023: Official Conference Proceedings. Pp. 2758–0962, 2023.

[GSK21] Gupta, J.; Singh, V.; Kumar, I.: Florence-a health care chatbot. In: 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). Vol. 1, IEEE, pp. 504–508, 2021.

[He17] Heublein, U. et al.: Zwischen Studienerwartungen und Studienwirklichkeit: Ursachen des Studienabbruchs, beruflicher Verbleib der Studienabbrecherinnen und Studienabbrecher und Entwicklung der Studienabbruchquote an deutschen Hochschulen (Forum Hochschule, 1/2017). Deutsches Zentrum für Hochschul- und Wissenschaftsforschung, 2017.

[He22] Heublein, U.: Die Entwicklung der Studienabbruchquoten in Deutschland. 2022.

[HK07] Hayes, A. F.; Krippendorff, K.: Answering the Call for a Standard Reliability Measure for Coding Data. Communication Methods and Measures 1 (1), pp. 77–89, 2007, DOI: 10.1080/19312450709336664.

[Ji24] Jiang, A. Q. et al.: Mixtral of Experts, 2024, arXiv: 2401.04088 [cs.LG].

[KK20] Kushwaha, A. K.; Kar, A. K.: Language model-driven chatbot for business to address marketing and selection of products. In: Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation: IFIP WG 8.6 International Conference on Transfer and Diffusion of IT, TDIT 2020, Tiruchirappalli, India, December 18–19, 2020, Proceedings, Part I. Springer, pp. 16–28, 2020.

[Kr18] Krippendorff, K.: Content analysis: An introduction to its methodology. Sage publications, 2018.

[Ku23] Kuhail, M. A. et al.: Interacting with educational chatbots: A systematic review. Education and Information Technologies 28 (1), pp. 973–1018, 2023.

[LCC23] Lai, C. Y.; Cheung, K. Y.; Chan, C. S.: Exploring the role of intrinsic motivation in ChatGPT adoption to support active learning: An extension of the technology acceptance model. Computers and Education: Artificial Intelligence 5, p. 100178, 2023.

[Ma20] May, P.: Cross English & German RoBERTa for Sentence Embeddings. 2020, https://huggingface.co/T-Systems-onsite/cross-en-de-roberta-sentence-transformer.

[Me20] Mendez, S. et al.: Chatbots: A tool to supplement the future faculty mentoring of doctoral engineering students. International Journal of Doctoral Studies 15, 2020.

[Mi21] Mittal, M. et al.: Web-based chatbot for frequently asked queries (FAQ) in hospitals. Journal of Taibah University Medical Sciences 16 (5), pp. 740–746, 2021.

[MK16] Merkle, B.; Kaczmirek, L.: Du kommst hier nicht rein: Türsteherfragen identifizieren nachlässige Teilnehmer in Online-Umfragen. Planung und Analyse: Zeitschrift für Marktforschung und Marketing, 1, 2–5. 2016.

[MM22] Mishra, S.; Müller, L.: Resources, norms, and dropout intentions of migrant students in Germany: The role of social networks and social capital. Studies in Higher Education 47 (8), pp. 1666–1680, 2022.

[MMD24] Merkle, B.; Messerer, L. A. S.; Dickhäuser, O.: Will I be happy in this major? Predicting intrinsic motivation and subjective well-being with prospective students' well-being forecast and interest-major fit forecast. Social Psychology of Education, 27(1), 237–259, 2024.

[MS13] Müller, S.; Schneider, T.: Educational pathways and dropout from higher education in Germany. Longitudinal and Life Course Studies 4 (3), pp. 218–241, 2013.

[OA21] Okonkwo, C. W.; Ade-Ibijola, A.: Chatbots applications in education: A systematic review. Computers and Education: Artificial Intelligence 2, p. 100033, 2021.

[PDP20] Pérez, J. Q.; Daradoumis, T.; Puig, J. M. M.: Rediscovering the use of chatbots in education: A systematic literature review. Computer Applications in Engineering Education 28 (6), pp. 1549–1565, 2020, DOI: 10.1002/cae.22326.

[Ra22] Ragheb, M. A. et al.: Investigating the acceptance of applying chat-bot (Artificial intelligence) technology among higher education students in Egypt. International Journal of Higher Education Management 8 (2), 2022.

[Ro12] Rosseel, Y.: lavaan: An R package for structural equation modeling. Journal of statistical software 48, pp. 1–36, 2012.

[RRS17] Ranoliya, B. R.; Raghuwanshi, N.; Singh, S.: Chatbot for university related FAQs. In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, pp. 1525–1530, 2017.

[SA24] Solomovich, L.; Abraham, V.: Exploring the influence of ChatGPT on tourism behavior using the technology acceptance model. Tourism Review, 2024.

[SOM24] Stöhr, C.; Ou, A. W.; Malmström, H.: Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. Computers and Education: Artificial Intelligence 7, p. 100259, 2024.

[Wi20] Winkler, R. et al.: Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent. In: Proceedings of the 2020 CHI conference on human factors in computing systems. Pp. 1–14, 2020.

[Xu21] Xu, J. et al.: Recipes for Safety in Open-domain Chatbots, 2021, arXiv: 2010.07079 [cs.CL].

[Yi23] Yilmaz, H. et al.: Student attitudes towards Chat GPT: A technology acceptance model survey. International Educational Review 1 (1), pp. 57–83, 2023.

Virtual Reality

Immersive Virtual Reality Training for Service Personnel in the Power Distribution Grid with Adaptable Training Environments

Pascal Bolz [1], Dominic Fehling [2], and Markus Zdrallek [1]

Abstract: Practice-oriented qualification measures on the power distribution grid present major obstacles such as high dependence on time and location, as well as high costs. Virtual Reality offers a promising approach by providing interactive and adaptable training environments that bridge the gap between theory and practice. This paper introduces the Virtual Energy System (VES), an immersive training solution designed to educate service personnel on the power distribution grid. At the core of the VES is a custom low-code authoring framework that allows power distribution system operators (DSOs) to develop tailored training content without requiring expertise in 3D modeling or software development. To ensure that the VES meets the needs of DSOs, a comprehensive needs analysis was conducted with four DSOs. In a first prototype, the implementation of the initial use case is presented through an exemplary application of the five safety rules at a medium-voltage switchgear, including a detailed showcasing of the structure and use of the authoring tools.

Keywords: Virtual Reality Learning, Low-code Authoring Framework, Power Distribution System Operators, Adaptable Training Environments

1 Introduction

As a result of increasing digitization and changes in the power distribution grid, new switchgear and operating equipment are being introduced, systems becoming more complex and their interrelationships and effects on the power distribution grid more confusing. These changes require better practical training to equip service personnel with the skills required to manage complexity and ensure grid reliability, as incorrect switching operations can have severe consequences. However, practice-oriented qualification measures in energy supply technology present major obstacles such as high time and location dependence as well as high costs. This is especially evident in the switching and disconnecting of secondary substations, which involve logistical effort due to the necessary coordination and can cause operational disruptions [Sc24; SR16].

Current statistics indicate that 86 % of electrical accidents reported in Germany by electricians in the industrial and commercial sectors occur due to violations of the first and third safety rules [BG24]. The Five Safety Rules ensure safe operations in the power

[1] University of Wuppertal, Institute of Power Systems Engineering, Gaußstraße 20, 42097 Wuppertal , Germany, pbolz@uni-wuppertal.de, https://orcid.org/0009-0008-1175-3186;
zdrallek@uni-wuppertal.de, https://orcid.org/0000-0000-0000-0000

[2] University of Wuppertal, Institute for Technologies and Management of Digital Transformation, Gaußstraße 20, 42097 Wuppertal , Germany, fehling@uni-wuppertal.de, https://orcid.org/0000-0002-2607-4409

doi:10.18420/delfi2025_12

distribution grid. They include: (1) Disconnecting the system from the power supply, (2) Securing it against reconnection, (3) Verifying the absence of voltage, (4) Grounding and short-circuiting, and (5) Ensuring nearby live parts are covered or blocked. Given that the first and third safety rules are both simple and quick to implement, their frequent violation is particularly concerning, highlighting the need to continuously train even experienced service personnel.

Virtual reality (VR) represents a promising approach to improving safety training and operational preparedness for power distribution system operators (DSOs) [Ha21; JF18; Ma17]. VR offers immersive, interactive and adaptable environments where service personnel can engage in authentic simulations in a safe setting, bridging the gap between theory and practice while enabling them to internalize safety protocols and routine procedures through experiential learning. By leveraging VR-based training, DSOs can systematically reinforce critical safety behaviors, mitigate workplace risks, and improve overall operational efficiency [HFZ21].

This paper is an extended version of our article published in VDE High Voltage Technology 2024 under the title “Development of an Immersive Educational Offer for the Presentation of Fault and Emergency Situations in the Distribution Grid Infrastructure”.

2 Immersive Learning Experiences in Virtual Reality

Virtual learning environments use three-dimensional and interactive simulations of real or fictional environments. These can be classified as either non-immersive (desktop-based) or immersive. Non-immersive VR primarily enables interaction via traditional input devices such as a keyboard and mouse. In contrast, immersive VR is typically accessed through head-mounted displays, providing a higher degree of sensory and cognitive immersion. Within immersive virtual environments, immersion pertains to the cognitive states users experience while engaging with them, as conceptualized in the Reality-Virtuality Continuum [MK94]. In VR-based learning, immersion plays a central role, defined both as the realism of the virtual environment and the depth of user engagement through interactivity and feedback [NNS16].

When learning complex topics, service personnel often struggle with the large volume of information, which can lead to cognitive overload, especially in immersive media. To address this, structured instructional models are crucial. The Cognitive Affective Model of Immersive Learning (CAMIL) highlights the importance of presence and interactivity, aligning VR with pedagogical frameworks to ensure effective teaching [MP21]. The development of the Virtual Energy System (VES) is grounded in the CAMIL instructional framework. Building on this foundation, the VES authoring tools enable efficient adaptation of learning content, supporting both guided and self-directed learning modes. In the guided learning mode, service personnel receives auditory and visual support, with linear learning paths to help prevent failure of individual subtasks, for example. In contrast, the free learning

mode allows more autonomy, as service personnel can explore and act autonomously. The learning modules are designed so that actions which would pose a real-world danger result in the module being safely aborted and reset, accompanied by targeted feedback to help users understand their error. This allows a clear definition of the learning objective while simultaneously providing service personnel with extensive opportunities to act at their own pace in the virtual learning environment and learn from mistakes. Furthermore, the implementation of selectable learning modes enables DSOs to train service personnel with different levels of experience without having to adapt the learning content.

3 Concept and Structure of the Virtual Energy System

Based on previous research [MFM23] on the adoption and effectiveness of virtual training tools, a user-centered approach was adopted in the development of the VES. The primary focus is on empowering DSOs to create and modify learning content independently. Although standalone software designed for specific purposes can be useful for narrow applications, it often falls short in terms of scalability. For DSOs, the ability to adapt training content to evolving workflows and equipment is essential. Existing VR solutions in the energy sector are typically proprietary and restrict content modification. In contrast, the VES presented in this paper provides an open-source authoring suite that allows DSOs to develop and adapt VR training content independently, without vendor lock-in or the need for programming expertise. Building upon the real-time engine Unity 2021 LTS and the open-source authoring framework Figments.nrw [MF22], VES extends these technologies with domain-specific workflows tailored to the needs of DSOs. These tools allow DSOs to create, customize, and deliver learning content efficiently. The system provides essential components for immersive and collaborative learning experiences while remaining flexible for different use cases. This flexibility is achieved by allowing individual 3D-assets of work and operating equipment to be exchanged within the learning content, enabling the customization and adaptation of learning modules.

3.1 Learning Activities and Low-code Authoring Tools

All learning modules in the VES are built using behavior graphs, which consist of interconnected nodes and edges to define interactions. These graphs add an interactive layer to otherwise static learning content. Fig. 1 illustrates their structure, highlighting two of the seven available node types: input nodes and action nodes.

Nodes can include adjustable parameters, such as initial status, repeatability mode, and debugging options, which can be modified using buttons. Input nodes reference entities within the scene (e. g. objects or players), allowing them to invoke and be influenced by other nodes. Action nodes define system behavior and can serve as triggers, responses, or both. Some are triggered by user interactions or scene changes (actions), some perform

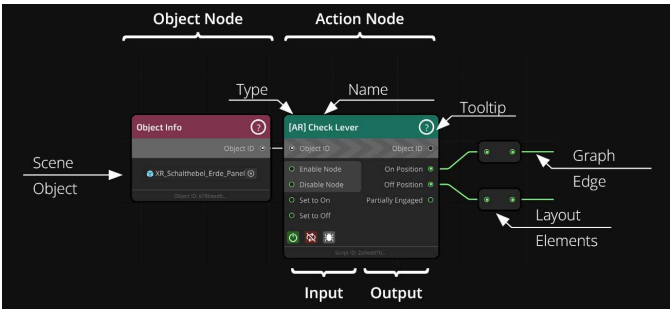


Fig. 1: A simple behavior graph showcasing the node architecture

changes in the scene when triggered (reactions), and others can fulfill both roles depending on context. [Bo24]. In the given example, a behavior graph controls a medium-voltage switchgear, determining and modifying its state (see Fig. 3, right). Nodes are connected by color-coded edges to represent logical relationships: gray edges indicate object references, while green edges show logical flow.

By using behavior graphs, learning content based on 3D models and subject matter can be transformed into structured, interactive tasks. This approach enables DSOs to tailor training content to their needs and create custom use cases independently.

3.2 Identification of Suitable Use Cases and Assets

To evaluate the usability of the VES and showcase the authoring tools, a set of prototypical learning modules is being developed. These modules simulate real-world conditions that are difficult to recreate physically and are often taught only theoretically [MK94]. A needs analysis at four DSOs identified use cases involving inert knowledge or common service errors, along with others essential for training. These were evaluated for VR suitability and grouped into four categories [Bo24]:

- **Five Safety Rules**: Teaching these rules in a repeatable and adaptable manner without exposing service personnel to real risks.
- **Station Inspections**: Repeatable visual inspections of adaptable stations and the regulated recording or fixing of errors
- **Error Handling**: Correction and safe handling of malfunctions, defects or other errors.
- **Critical Situations**: Simulation of critical situations in a safe environment

The feasibility of VR use cases is limited by technical constraints such as the inability to display very small text clearly and the challenges of accurately replicating highly sensitive

work processes. Among the selected use cases, particular emphasis is placed on safety-related training scenarios. A key example is the application of the five safety rules, which are essential for ensuring safe operations in the power distribution grid. Therefore, the following section of this paper will present the first prototype, which contains the learning module “The Five Safety Rules”. The example shows the application and execution of the safety rules that must be carried out when disconnecting a secondary substation.

3.3 Spatial Design of Immersive Learning Environments

When creating content for the VES two design processes intertwine: spatial design of immersive learning environments and low-code interactivity authoring. While the first handles questions around the layout of environments, the latter enriches these environments with interactive and reactive elements. In the module “The Five Safety Rules”, interactions are focused in a range of action around a work vehicle and a mobile switching station. Fig. 2 shows the layout of the environment in a top down view (left) as well as the user’s perspective (right).

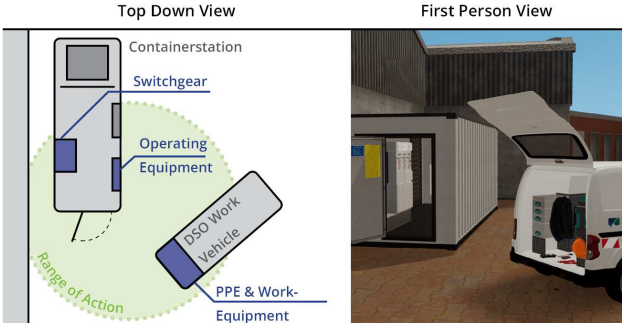


Fig. 2: Different views of the learning environment in the first learning module

The main task service personnel need to perform in this module is to disconnect, earth and short-circuit a medium-voltage switchgear. To introduce the correct procedure, the module begins with a switching conversation that signals the start of the work. After this, the service personnel is tasked with putting on their personal protective equipment and carry out a function test on the mobile voltage test system as shown in Fig. 3 (left).

After a successful self-test, the switch panel can be checked. If voltage is present, the system is first disconnected (Fig. 3, right) and then secured against being switched on again. Once the absence of voltage has been checked, the system can be earthed and short-circuited. Finally, neighboring live parts must be isolated and insulated. The learning module “The Five Safety Rules” ends after the last action.



Fig. 3: Visualization of XR Interactables: Voltage Test System (left), Switchgear Panel (right)

4 Conclusion and Outlook

Practical training in the power distribution grid faces significant challenges, particularly in the switching and disconnecting of secondary substations. These processes require logistical coordination and can cause operational disruptions. At the same time, the increasing complexity of the power distribution grid, driven by the introduction of new switchgear and operating equipment, makes it more difficult to understand system interrelationships and their effects on grid operations. To address these challenges, VR offers a promising solution by providing access to activity-oriented learning content and interactive experiences. This paper introduces the concept and current state of the VES: A modular framework for the creation, teaching and learning of immersive media. Build upon open-source tools (Figments.nrw) and facilitating established realtime application engines (Unity Engine) this can be tailored to the specific needs of DSOs. A key advantage of this solution is its flexibility in adapting learning content to evolving work environments, such as new short-circuit and earth fault indicators or medium-voltage switchgear, ensuring that service personnel receive effective training in their use, handling, and operation. The presented and developed VES is built upon essential components such as interaction metaphors, a set of suitable 3D assets, and behavior graphs, which enable the creation of dynamic and interactive learning content. Several operational use cases have been identified as suitable for development, with a particular focus on the simulation of fault scenarios. These scenarios are often difficult to replicate authentically and are, therefore, usually only considered in theoretical training. The paper presents the first prototype, featuring the learning module "The Five Safety Rules", which demonstrates both the application of these safety procedures during the disconnection of a secondary substation and the structure and the structure and functionality of the authoring tools provided by the VES. Future development efforts will concentrate on expanding the applicability and technical features of the VES while also conducting user studies to assess its effectiveness. Additionally, improvements will be made to the readability of behavior graphs in upcoming prototypes by grouping related nodes, reusing existing node constellations, and providing extensive documentation for individual graph components.

Funding information

This article was written as part of the project “Virtual und Augmented Reality in der elektrischen Energieversorgungstechnik“ (03EI6098A), which is funded by the German Federal Ministry for Economic Affairs and Energy. The authors gratefully acknowledge the financial support of this project. Only the authors are responsible for the content of this contribution.

References

[BG24] Berufsgenossenschaft ETEM, 2024, https://www.bgetem.de/arbeitssicherheit-gesundheitsschutz/institute/institut-zur-erforschung-elektrischer-unfaelle/statistik-der-stromunfaelle, accessed: 03/12/2025.

[Bo24] Bolz, P. and et al.: Development of an Immersive Educational Offer for the Presentation of Fault and Emergency Situations in the Distribution Grid Infrastructure. VDE Hochspannungstechnik 175, 2024.

[Ha21] Hamilton, D.; et al.: Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental desig. Journal of Computers in Education 8.1, pp. 1–32, 2021.

[HFZ21] Hecker, D.; Faßbender, M.; Zdrallek, M.: The implementation of immersive virtual reality trainings for the operation of the distribution grid - development and integration of two operational use cases. The 26th International Conference and Exhibition on Electricity Distribution (CIRED), 2021.

[JF18] Jensen, L.; Flemming, K.: A review of the use of virtual reality head-mounted displays in education and training. Education and Information Technologies 23, pp. 1515–1529, 2018.

[Ma17] Martino, S. and et al.: Enedis field experience of Augmented and Virtual Reality technologies at the service of network operators. The 24th International Conference and Exhibition on Electricity Distribution (CIRED), 2017.

[MF22] Müser, S.; Fehling, D.: AR/VR.nrw – Augmented und Virtual Reality in der Hochschullehre. HMD Praxis der Wirtschaftsinformatik 59(1), 2022.

[MFM23] Maiero, J.; Fehling, C.; Müser, S.: Does Higher Education Need Virtual Reality? A Survey. 2023 11th International Conference on Information and Education Technology (ICIET), 2023.

[MK94] Milgram, P.; Kishino, F.: A Taxonomy of Mixed Reality Visual Displays. Transactions on Information and Systems (IEICE) 77, pp. 1–16, 1994.

[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, pp. 937–958, 2021.

[NNS16] Nilsson, N. C.; Nordahl, R.; Serafin, S.: Immersion revisited: A review of existing definitions of immersion and their relation to different theories of presence. Human technology 12.2, pp. 108–134, 2016.

[Sc24] Scorgie, D. and et al.: Virtual reality for safety training: A systematic literature review and meta-analysis. Safety Science 171, 2024.

[SR16] Steinhöfel, M.; Rosenberg, S.: Herausforderungen und Auswirkungen der Digitalisierung auf die Weiterbildung und das Weiterbildungspersonal. IBBF Institut für Betriebliche Bildungsforschung, pp. 1–15, 2016.

Didaktische Konzepte fürs Onboarding und Tutorials in VR-Lernanwendungen

Birte Heinemann [1], Matthias Ehlenz [1] und Ulrik Schroeder [1]

Abstract: Virtuelle Realität (VR) eröffnet neue Potenziale für immersive Lernumgebungen, stellt jedoch auch besondere didaktische Herausforderungen, insbesondere beim Einstieg und der Steuerung. In vielen Bildungsanwendungen fehlen strukturierte Tutorials und Onboarding-Konzepte, obwohl diese maßgeblich zur Nutzbarkeit und zum Lernerfolg beitragen können. Dieses Paper widmet sich der Frage, welche Konzepte, Gestaltungsprinzipien und Best Practices für das Onboarding und die Einführung in VR-Lernumgebungen existieren. Auf Basis eines Scoping- und Narrative Reviews wurde der aktuelle Forschungsstand analysiert und eine Beispielumsetzung diskutiert. Die Ergebnisse zeigen, dass sich viele relevante Hinweise aus angrenzenden Feldern wie der Spieleentwicklung sowie der Instruktionspsychologie übertragen lassen.

Keywords: Virtual Reality, Tutorial, Bildungstechnologien, Onboarding, Lernapplikationen, Review

1 Einführung

Virtuelle Realität (VR) bietet als Bildungstechnologie zahlreiche Chancen und Potenziale [Pa09; Ze19]. Pantelidis [Pa09] argumentiert, dass VR neue Lernerfahrungen ermöglicht und Lernende aller Altersstufen begeistern kann. Bereits 1993 beschrieb Winn [Wi93] die Vorteile immersiver Lernerlebnisse und betonte, dass VR es ermöglicht, Konzepte durch Interaktion mit einer immersiven Umgebung zu erlernen, ohne ein Symbolsystem des Mediums kennen zu müssen. Dies ist insbesondere für Lernende von Vorteil, die Schwierigkeiten mit symbolbasierten Unterrichtsaktivitäten haben.

Trotz dieser Vorteile stellt VR didaktische und medienpädagogische Herausforderungen dar. Die Steuerung ist nicht immer intuitiv, Lernende sind oft allein in der Umgebung, und Lehrkräfte können nicht direkt unterstützen. Viele Veröffentlichungen fokussieren nicht auf Lernmethoden [Ra20], und Berichte zeigen, dass der Einstieg in VR oft überfordernd ist [FH20]. Trotz ihrer nachgewiesenen funktionalen Wirksamkeit und positiven Effekte auf die Spielleistung sind gut gestaltete VR-Tutorials und Einführungsszenarien bislang selten [Ho17].

Dieses Paper adressiert diese Lücke, indem es Ansätze für didaktisch fundierte Onboarding-Prozesse und Tutorials für VR-Lernanwendungen sowie ein Beispiel für deren Umsetzung vorstellt. Untersucht werden dabei der aktuelle Forschungsstand zu Tutorials und Onboarding

[1] RWTH Aachen University, Templergraben 55, 52062 Aachen, Germany,
heinemann@cs.rwth-aachen.de, https://orcid.org/0000-0002-7568-0704;
ehlenz@lbz.rwth-aachen.de, https://orcid.org/0000-0001-6189-6056;
schroeder@informatik.rwth-aachen.de, https://orcid.org/0000-0002-5178-8497

doi:10.18420/delfi2025_13

in VR sowie eine Sammlung von Best Practices, die sich aus den gefundenen Quellen ableiten lassen. Anhand eines Beispiels wird die Verbindung zwischen den Gestaltungsprinzipien und der konkreten Entwicklung von Tutorials in Bildungsanwendungen diskutiert.

Whittaker [Wh25] definiert Onboarding als "the process followed to guide users to understand technically and conceptually how to participate in a VR experience, facilitating their ability to successfully attend to and participate in the experience". Diese Definition wird als Grundlage für die Betrachtung von Onboarding-Prozessen in diesem Paper verwendet.

2 Methodik

Für diese Arbeit wurde eine Kombination aus Scoping Review und Narrative Review durchgeführt, um den aktuellen Forschungsstand zu Tutorials und Onboarding in VR-Lernanwendungen zu identifizieren.

Dabei wurde initial die übliche Vorgehensweise für ein Scoping-Review angewendet: Es wurde mit einer webbasierten Literaturrecherche gestartet. Das Zeitfenster für die systematische Suche wurde hierbei auf die letzten zehn Jahre (ab 2015) beschränkt, der verwendete Suchterm lautete: (”Onboarding” OR ”Tutorial”) AND ”Virtual Reality” AND ”Edu*”.

Für die systematische Recherche wurden zunächst die Datenbanken IEEE Xplore, ACM Digital Library und Google Scholar herangezogen. In IEEE Xplore und der ACM Digital Library ließen sich keine relevanten Treffer mit Bezug zu VR-Tutorials finden – selbst eine erweiterte Suche in der ACM Digital Library ergab zwar 136 Treffer, jedoch keine spezifischen Publikationen zum Thema. In Google Scholar wurden zunächst rund 16.000 Treffer identifiziert, nach Eingrenzung auf die letzten fünf Jahre verblieben 14.300.

Im Folgenden wurde eine Einschränkung auf spezifische Journals und Verlage (z. B. Computers & Education, Springer, IEEE, ACM) in Google Scholar vorgenommen. Das Ergebnis war eine Reduzierung auf 4.860 Ergebnisse. Hierdurch sollte sichergestellt werden, dass die Publikationen in der Ergebnisliste sich auf solche beschränkten, die einem Peer-Review-Verfahren durchlaufen haben. Es folgten weitere Filter und Verfeinerungen der Suchstrategien mit dem Ziel einer weiteren Reduktion auf Ergebnisse, die im untersuchten Kontext von besonderer Relevanz sind. Insbesondere fand hierbei eine Variation der Suchbegriffe statt (z. B. ”Immersive Environment” anstelle von ”Virtual Reality”, ”User Onboarding” anstelle von ”Onboarding”) und es wurden Alternativen für ”Edu*” wie ”Instructional Design” oder ”Training” exploriert. Trotz der Anpassungen wiesen die Trefferlisten eine hohe Anzahl an False Positives auf. Diese wurden zum Beispiel als solche betrachtet, wenn der Begriff ”Tutorial” zu vielen irrelevanten Treffern führte, da er als Teil der Methodik miterwähnt wurde, aber nicht wesentlicher Untersuchungsgegenstand war.

Da die Effekte dieser Anpassungen der Suchstrategien nicht den nötigen Erfolg mit sich brachten, erfolgte eine Anpassung der Methodik. Aufgrund der hohen Anzahl irrelevanter

Treffer wurde die systematische Suche in einen narrativen Review überführt, ohne jedoch die bisherigen Ergebnisse zu verwerfen. Bereits zu diesem Zeitpunkt konnten viele relevante Studien identifiziert werden, die zentrale Aspekte der Forschungsfragen adressieren und somit als Grundlage des narrativen Vorgehens dienen konnten.

Am Ende des Gesamtprozesses wurden 28 relevante Studien identifiziert, die in die spezifische Analyse zu Onboarding-Prozessen in und für VR aufgenommen wurden, um daraus den aktuellen Forschungsstand zu Tutorials im immersiven Lernen abzuleiten und Best Practices zu identifizieren.

Am Ende des Gesamtprozesses wurden 29 relevante Beiträge identifiziert, die in die spezifische Analyse zu Onboarding-Prozessen in und für VR aufgenommen wurden, um daraus den aktuellen Forschungsstand zu Tutorials im immersiven Lernen abzuleiten und Best Practices zu identifizieren. Hierbei wurden alle relevanten Inhalte in das Review einbezogen – darunter sowohl wissenschaftliche Studien als auch praxisnahe Quellen wie Entwicklerblogs und Leitfäden aus dem Bereich der Spieleentwicklung. Diese methodische Öffnung zielt darauf ab, der interdisziplinären Ausrichtung des Themas gerecht zu werden und auch nicht-akademische, aber etablierte Perspektiven auf Onboarding-Prozesse in immersiven Umgebungen systematisch zu berücksichtigen.

3 Ergebnisse

Die Literaturrecherche hat gezeigt, dass es bislang nicht viel systematische Forschung zu Tutorials in VR-Anwendungen, insbesondere unter Einbeziehung einer didaktischen Perspektive, die für Applikationen zum Einsatz in der Lehre unerlässlich ist, in veröffentlichen, peer-reviewten Publikationen gibt. In Konsequenz wurden drei verwandte Bereiche identifiziert, in denen Erkenntnisse vorliegen, und diese im Bezug auf den hier relevanten Kontext kritisch analysiert, um diese in der Folge in einem Anwendungsbeispiel exemplarisch umzusetzen und daraus eine erste Grundlage für Best Practices auszuformulieren. Hierzu wurden insbesondere zum einen die Erkenntnisse aus dem Bereich der Tutorials im Bereich der Computerspiele-Entwicklung, auch unter Einbeziehung praktischer Umsetzungen aus der Spieleindustrie, zum anderen aber auch der didaktisch geprägte Bereich der (digitalen) Medienpädagogik und deren Erkenntnisse zu Onboarding-Prozessen näher betrachtet. Darüber hinaus wurde auch die Perspektive der Mensch-Maschine-Interaktion einbezogen, der sich ebenfalls mit der Gestaltung intuitiver Nutzerschnittstellen auseinandersetzt.

3.1 Tutorialgestaltung in Spielen

Ein praxisorientierter Einstieg zur Auseinandersetzung mit der Gestaltung von Tutorials in Computerspielen, die ja in der Regel vorrangig einen motivierenden Einstieg ohne Frust ermöglichen sollen, sind zum Beispiel die Materialien, die der Spieleentwickler

Darran Jamieson auf einer Plattform für Lehrmaterialien bereitstellt [Ja15]. Hier werden Computerspiele auf Basis ihrer Tutorials in insgesamt vier Kategorien klassifiziert:

- Nicht-interaktive Tutorials (z. B. Textanleitungen)
- Interaktive Tutorials (z. B. durch Dialoge mit Handlungsaufforderungen)
- Passive Tutorials (fließend in das Erlebnis integriert)
- Kein Tutorial (exploratives Lernen)

Ein passives Tutorial fügt sich besonders fließend in das Spielerlebnis ein und läuft im Hintergrund des Spielgeschehens ab. Die Spielenden entdecken die Tutorialelemente frei in selbstgesteuerter Geschwindigkeit und auf graduelle Weise. Der Übergang zu interaktiven Tutorials kann hierbei fließend sein, da auch diese den Aspekt des Storytellings nutzen, um den Spieler*in aktiv in das Tutorial einzubinden. Dialoge mit Aktionsaufforderung sind ein typisches Element interaktiver Tutorials. Nicht interaktive Tutorials stellen die Inhalte ohne aktive Einbindung der Spielenden dar, z. B. über Anleitungen, die gelesen werden müssen. Die von Jamieson vorgeschlagenen Kategorien können durch nicht spielinterne Tutorials erweitert werden, Ho [Ho17] haben zum Beispiel getestet, in welchem Verhältnis ein Tutorial in VR zu einem Tutorial außerhalb der virtuellen Realität steht. Die Ergebnisse zeigen, wie sich im Folgenden auch zeigen wird, dass Tutorials Auswirkungen auf die Leistung und die Emotionen der Versuchspersonen haben, da die Kontrollgruppe ohne ein Tutorial signifikant unterschiedliche Resultate erzielt hat. Limitierend an den Ergebnissen ist, dass es nur unter spezifischen Rahmenbedingungen überhaupt möglich ist ein durchführbares Tutorial außerhalb der VR-Lernanwendungen zu entwickeln.

White [Wh12] untersucht zwei Strategien: „Flashcard Instructional Strategy“ (Pop-up-Instruktionen) und „Just-In-Time Scaffolded Instructional Strategy“ (kontextabhängige Hilfestellungen). Ergebnisse zeigen, dass unterschiedliche Strategien den Spielfluss unterschiedlich beeinflussen. Quesada Mora et al. [Qu17] argumentieren, dass organische Tutorials, die unauffällig in das Spielerlebnis integriert sind, den Spielspaß erhöhen und Frustration vermeiden.

Über das Zusammenspiel verschiedener Modalitäten, wie Bildern, Texten oder Tonspuren als Repräsentationen für Informationen gibt es zum Teil widersprüchliche Ergebnisse. White [Wh12] berichtet, dass Sound in Instruktionen hilfreich ist. Andere Studien zeigen aber, dass eine gleichzeitige Präsentation von gesprochener Sprache und identischem Text nicht immer hilfreich ist — sie kann sogar hinderlich sein [SAK11]. Mayer [Ma21] und Castro-Alonso; Ayres; Sweller [CAS19], stimmen darüber ein, dass der Effekt verschiedene Modalitäten zu Kombinieren von dem Vorwissen und der Komplexität abhängen, bei lernstarken Personen sind z. B. visuelle Darstellungen oft effizienter, während Noviz*innen eher von auditiver Unterstützung profitieren [CAS19]. Chauvergne; Hachet; Prouzeau [CHP23] stellt bei einer Untersuchung von 21 VR-Tutorials fest, dass unterschiedliche Modalitäten kombiniert werden, allerdings visuelle Hinweise dominieren.

Quesada Mora et al. [Qu17] haben „organische“ Tutorials untersucht, d. h. Tutorials, die spielerisch und indirekt in das Spielerlebnis eingebaut sind. Sie unterscheiden diese Art der Spieleinführungen von direkten Tutorials bei denen Texte und Bilder direkt die zu lernenden Inhalte erklären. Sie untersuchen Tutorials vergleichbar mit den passiven und interaktiven Tutorials von Jamieson [Ja15]. Organische Tutorials unterstützen das unterschwellige Lernen und sollen laut Quesada Mora et al. [Qu17] den Spaß der Spieler und Spielerinnen fördern. Neben einer Analyse von Designtechniken und Videospielen nutzen sie Lerntheorien, um eine Liste von Prinzipien zu erstellen, die für gute organische Tutorials wichtig sind. Die drei wichtigsten Prinzipien, die sie identifizieren sind: Orientierungshilfen anzubieten, indirektes Lernen zu fördern und keine Frustration der Spielenden aufkommen zu lassen. Für diese Prinzipien stellen sie Techniken vor, die diese Prinzipien unterstützen und die sich zum Teil auch auf VR Bildungsapplikationen übertragen lassen, je spielerischer und komplexer die Lernumgebung, desto mehr Techniken sind im Bildungskontext interessant.

Chauvergne; Hachet; Prouzeau [CHP23] schlagen ein konzeptuelles Framework, das zur Gestaltung von VR-Onboarding genutzt werden kann. Dieses umfasst acht Dimensionen. Eine erste Dimension betrifft den Onboarding-Kontext, also ob das Onboarding außerhalb der Anwendung, in einem dedizierten Lernbereich oder kontextsensitiv innerhalb des eigentlichen Nutzungsgeschehens erfolgt. Darüber hinaus wird die Flexibilität des Onboardings erfasst – etwa, ob Nutzer*innen einzelne Schritte überspringen oder Inhalte gezielt auswählen können. Die Abdeckungstiefe beschreibt, ob lediglich grundlegende Funktionen vermittelt oder auch fortgeschrittene Features adressiert werden.

Eine weitere wichtige Dimension betrifft die Vermittlung der Eingabemethoden, also wie Nutzerinnen lernen, mit Controllern oder Handtracking umzugehen – beispielsweise über visuelle Modelle, Gesten oder externe Demonstrationen. Der Interaktivitätsgrad unterscheidet, ob Nutzerinnen aktiv mit der virtuellen Umgebung interagieren können oder lediglich Beobachter*innen von Demonstrationen sind [CHP23].

Ergänzend berücksichtigt das Framework die Art der Unterstützung, etwa durch konkrete Schritt-für-Schritt-Anleitungen, Zielvorgaben oder modellhafte Demonstrationen [CHP23]. Eng damit verknüpft ist der Grad des Feedbacks: Während viele Systeme nur implizit Rückmeldung geben (etwa durch das automatische Fortfahren beim Erfolg), wird explizites oder korrektives Feedback als besonders lernförderlich hervorgehoben. Schließlich spielen auch die verwendeten Instruktionsmodalitäten eine Rolle, also ob Informationen über Text, Audio, Animation oder haptische Signale vermittelt werden – idealerweise in einer redundanten und adaptiven Kombination.

Insgesamt erlaubt das Framework von Chauvergne; Hachet; Prouzeau [CHP23] nicht nur eine differenzierte Beschreibung bestehender Onboarding-Lösungen, sondern dient auch als Inspirationsquelle für die Entwicklung neuer, kontextsensitiver und lernwirksamer Ansätze in VR-Trainingsumgebungen. Dabei ist jedoch zu beachten, dass das Framework nicht speziell für VR-Lernapplikationen entwickelt wurde. In didaktischen Kontexten können zusätzliche Anforderungen und Einschränkungen entstehen – etwa im Hinblick auf Lernziele,

Feedbackqualität, oder die Gestaltung instruktionaler Sequenzen, Zeitrestiktionen -–, die eine Anpassung oder Weiterentwicklung des Frameworks notwendig machen.

3.2 Didaktische Prinzipien für VR-Tutorials

Aus mediendidaktischer Sicht ist der spezifische Bereich der VR-Tutorials vielleicht noch nicht explizit erschlossen, dennoch gibt es hier bereits solide Literatur hinsichtlich allgemeiner Gestaltungsempfehlungen für immersive Lernumgebungen. Eine zentrale Veröffentlichung stellt hier die Forschung von Mulders, Buchner und Kerres dar, die auf der Basis der kognitiven Theorie des Lernens mit Multimedia und dem Instruktionsdesign Hilfestellungen für Lernumgebungen mit immersiver, virtueller Realität hergeleitet haben [MBK20]. Während Tutorials hierbei nicht zentral im Fokus stehen, finden sie in den Gestaltungsempfehlungen durchaus Beachtung, da sie eine Möglichkeit bieten Lernende mit geringem Vorwissen vor einer kognitiven Überlastung zu schützen. Wie in den anderen Quellen zu Tutorials wird darauf hingewiesen, dass Lernende mit Vorwissen durch Tutorials behindert werden könnten („expertise reversal effect“).

Mulders; Buchner; Kerres [MBK20] leiten aus der *Kognitiven Theorie des Lernens mit Multimedia* folgende Empfehlungen ab:

- Das didaktische Ziel sollte zur Immersion passen.
- Unnötige und lernirrelevante Interaktionen sollten vermieden werden.
- Pre-Training zu Konzepten und zur Nutzung von VR-Interaktionen ist essenziell.
- Aufgaben sollten in kleinere Abschnitte unterteilt werden.
- Tutorials sollten kognitive Überlastung verhindern, aber Lernende mit Vorwissen nicht behindern (expertise reversal effect; vgl. [Ar20].

Buchner; Aretz [BA20] stellen ein didaktisches Gesamtkonzept zum exemplarischen Einsatz von VR in der frühen schulischen Bildung vor, bei dem 360°-Medien im Fokus stehen. In den Lessons Learned ergänzt, unabhängig vom Tutorial und der spezifischen Anwendung, eine kritische Reflexion des Mediums als wichtiger Bestandteil das vorgestellte didaktische Design. Diese kritische Auseinandersetzung könnte, falls es zu den Lerninhalten der VR Anwendung passt, Teil der immersiven Didaktik und/oder des Tutorials selbst sein.

Nicht unmittelbar auf VR bezogen, finden sich in der Medienpädagogik darüber hinaus einige (ältere) Publikationen, deren Erkenntnisse sich aber auf den hier besprochenen Kontext durchaus übertragen lassen. So haben sich Fromme; Jörissen; Unger [FJU08] intensiv mit dem Bildungspotential digitaler Spiele auseinandergesetzt und dabei insbesondere den Begriff der immersiven Didaktik mitgeprägt, der in Virtual Reality einen noch stärkeren Stellenwert einnimmt. Die Ergebnisse lassen sich somit aus unserer Sicht gut extrapolieren:

Fromme; Jörissen; Unger [FJU08] erläutern verschiedene Spielelemente, die sie als Lernhilfen klassifizieren, z. B. erläuternde Notizen, die in der Spielwelt gefunden werden oder Dialoge mit Non-Player-Charakteren, die Tipps geben. Sie beschreiben diese pädagogisch konzipierten Hilfestellungen als „immersive Didaktik“, da diese oft gut in den Spielverlauf integriert ist. Sie ordnen die immersive Didaktik in die bildungsrelevante Dimension der „Aneignung und Erweiterung der Weltsichten und –relationen“ ein, welche auf der Interface- und Interaktionsebene wirkt. Zu dieser Dimension zählen sie auch „in-game“-Tutorials. Das mögliche Lernziel, eine Erweiterung der Weltsicht, ermöglichen Tutorials durch die Chance den Mediennutzer in eine kritische Reflexion des medial Präsentierten zu bringen ohne die Immersion zwingend unterbrechen zu müssen [FJU08].

3.3 Mensch-Computer-Interaktion Perspektive

Viele mögliche Hindernisse für die Nutzung von virtuller Realität im Bildungsbereich werden auch von den Forschenden der Mensch-Computer-Interaktion (MCI) aufgegriffen. Ein Tutorial in Virtual-Reality-Umgebungen sollte über die Vermittlung von Steuerungstechniken hinausgehen. Aus Sicht der MCI stellt das Tutorial den kritischen Einstiegspunkt dar, an dem Nutzende nicht nur die Interaktionslogik, sondern auch ein Gefühl von Präsenz in der virtuellen Umgebung entwickeln. Jerald [Je15b] betont, das VR eine Kommunikationsform darstellt, in der das Interface möglichst transparent gestaltet werden sollte. Ziel ist es, das technologische Medium als Vermittler in den Hintergrund treten zu lassen, sodass Nutzende das Gefühl erhalten, direkt mit den Inhalten und Objekten der virtuellen Welt zu interagieren.

Zentral ist dabei die gezielte Gestaltung sensorischer und kognitiver Prozesse: Tutorials sollten interaktive Lernprozesse durch visuelle, auditive und – wenn möglich – haptische Rückmeldungen unterstützen, dabei jedoch die kognitive Belastung gering halten. Dies gelingt etwa durch gestufte Einführungen in Interaktionsmöglichkeiten (progressive disclosure), unmittelbares Feedback sowie eine Einbettung der Lerninhalte in einen narrativen oder funktionalen Kontext. Außerdem müssen Gestaltungsentscheidungen hinsichtlich Immersion, Interaktionsgenauigkeit und Vorerfahrungen immer im Hinblick auf Nutzerkomfort und Präsenzwirkung getroffen werden, um sogenannte Breaks-in-Presence und VR-spezifische Barrieren wie Cybersickness zu vermeiden. [Je15b]

3.4 Zwischenresumee

Die systematische Literaturrecherche zum (didaktischen) Wert von Onboarding-Strategien in Virtual Reality hat sich als komplexes Unterfangen erwiesen. Zwar ist das Medium inzwischen seit einigen Jahrzehnten in der Theorie vorhanden, kommt allerdings gerade erst so richtig in der Breite in der Bildungslandschaft an und erfordert daher auch jetzt erst gute, möglicherweise sogar selbstgesteuerte, Mechanismen um Lernende mit dem

Medium vertraut zu machen. Es hat sich gezeigt, dass hier eine Herangehensweise aus mehreren Perspektiven aus praktischer Sicht die sinnvollste darstellt. Während hier nur in begrenztem Umfang ein Einblick in die ausführliche Recherche sowohl in mediendidaktische Gestaltungsempfehlungen für immersives Lernen als auch in Tutorial-Guidelines für Computerspiele gegeben werden kann, haben wir unsere Erkenntnisse für den Transfer in immersive Lernapplikationen in unserem Forschungskontext auf folgende Sammlung von Best Practices heruntergebrochen:

1. Überfordern Sie die Spielenden nicht. [Ja15]
2. Berücksichtigen Sie bewährte Gestaltungsprinzipien aus der Multimediaforschung – viele davon gelten auch für VR. [Su03]
3. Frustrieren Sie die Spielenden nicht. [Qu17]
   a) Lernen durch Scheitern erfordert Versuch und Irrtum. Um zu verhindern, dass die Spielenden frustriert werden, sollten diese Wiederholungszyklen kurz sein.
   b) der Schwierigkeitsgrad des Tutorials sollte anfangs minimal sein.
   c) Platzieren Sie Upgrade-Objekte, um die Erkundung zu belohnen Belohnung für den aufmerksamen Spielenden.
4. Bauen Sie Orientierungsaufgaben ein, damit die Spielenden den Umgang mit der Technik erlernen können. [BA20]
5. Planen Sie Zeit für das Kennenlernen des Medium ein und lassen Sie die Spielenden selbst entscheiden, wann Sie beginnen wollen. [FH20]
6. Lassen Sie die Lernenden entscheiden, ob und ggf. in welcher Form sie an dem Tutorial teilnehmen möchten, um erfahrene Nutzer nicht zu behindern. [Ar20], [OP05]
7. “Stabilisieren“ Sie die VR-Welt durch räumliche Hinweise (z. B. Schatten, Okklusion) und führen Sie die Nutzer behutsam in die Interaktion ein. [Je15a]
8. Verwenden Sie für Tutorials lieber Teleportation oder geführte Bewegung als sanfte Fortbewegung, es sei denn, die Benutzenden sind trainiert oder Bewegung in VR gewohnt. [Je15a]
9. Betten Sie das Tutorial wenn möglich narrativ ein. [Fr17], [Je15b]
10. Tutorials, sollen so gestaltet werden, dass sie erst abgeschlossen werden können, wenn der Benutzer ein klares Verständnis und effektive Interaktionen gezeigt hat, um mentale Modelle im Gedächtnis der Benutzenden zu verankern. Helfen Sie den Nutzenden ein konsistentes Modell der Funktionsweise der Dinge zu entwickeln. [Je15b]

Damit stellt das Tutorial nicht nur ein funktionales, sondern auch ein gestalterisch-psychologisches Element der Anwendung dar. Es verbindet Interaktionsdesign mit Wahrnehmungspsychologie und Usability Engineering. Tutorials, die auf Grundlage dieser Prinzipien

entwickelt werden, fördern nicht nur die Handhabbarkeit komplexer VR-Umgebungen, sondern wirken auch motivationsfördernd und ermöglichen einen niedrigschwelligen Einstieg in immersive Lernprozesse.

4 Beispiel für die Entwicklung eines VR-Tutorials im Bildungsbereich

Aufbauend auf diesen Ergebnissen soll nun exemplarisch ein Tutorial für VR-Lernumgebungen vorgestellt werden. Im hier beschriebenen Use Case ist die Zielgruppe heterogen, so dass hier insgesamt die größte Bandbreite an Gestaltungsempfehlungen eingeflossen ist und das Tutorial in mehreren Iterationen in kleinem Maßstab, aber dennoch hinsichtlich der Systematik empirisch, evaluiert wurde. Aus diesem Grund wird an diesem Beispiel der Fokus auf die Beschreibung des Tutorials im Kontext der VR-gestützten Lehramtsausbildung gelegt.

Kontext Teach-R Klassenraum-Simulation für Lehrkräftebildung

Die Klassenraum-Simulation Teach-R wird bereits seit einigen Jahren an verschiedenen Hochschulen im Bereich der Lehramtsausbildung eingesetzt. Während zunächst Unterrichtsstörungen und der Umgang damit im Vordergrund standen, sind inzwischen viele Anwendungsszenarien vom Fachunterricht verschiedener Fächer bis hin zu Forschungsansätzen zu Unterrichtseinstiegen und Classroom Management hinzugekommen. Die heterogene Studierendenschaft erschwert dabei die strukturierte Beforschung. Der unmittelbare Einstieg im Klassenraum vor einer Gruppe virtueller Schüler*innen überfordert die Immersanten in sehr unterschiedlichem Maße, so dass ein Tutorial sowohl für die Studierenden als auch für kontrollierte Forschungsbedingungen große Vorteile bringen sollte.

Wie in den Best-Pracices dargestellt, fanden Frommel et al. [Fr17] in einer Nutzerstudie heraus, dass Tutorials am besten in die Handlung des Spiels integriert werden. Daher wurde ein separates Level, z. B. ein leerer Klassenraum, verworfen. Auch ein Korridor als Ort des Tutorials erschien ungeeignet, da nicht alle Interaktionen dort realistisch darstellbar wären, etwa das Schreiben an der Tafel. Diese Funktionen lassen sich jedoch narrativ eingebettet im Lehrerzimmer integrieren, weshalb das Tutorial dort platziert wurde.

Um der Heterogenität der Lernenden gerecht zu werden, bietet die Anwendung zwei Tutorialvarianten an, die unterschiedliche Lerntypen und Vorerfahrungen berücksichtigen. Dies wird auch durch Best Practice 5 gestützt. Oliver; Pelletier [OP05] empfehlen, den Nutzenden eine Auswahl zu ermöglichen. Die erste Variante ist eine agentengeführte Tour, die die Grundlagen von Teach-R vermittelt und besonders für Personen mit wenig oder keiner VR-Erfahrung geeignet ist. Auch für VR-erfahrene Nutzende, die Teach-R noch nicht kennen, kann diese Variante hilfreich sein. Sie lässt sich als „Flashcard Tutorial“ im Sinne von White [Wh12] klassifizieren. Die zweite Variante ist ein sogenanntes Hintergrund-Tutorial, das

nur bei Bedarf Unterstützung bietet. Damit dennoch alle relevanten Funktionen entdeckt werden, erfolgt die Vermittlung zentraler Interaktionen spielerisch, wie später im Detail beschrieben. Diese Form lässt sich als „Just-in-Time“-Tutorial einordnen [Gr21; Wh12]. Dabei ist es möglich, sofort mit der Interaktion zu beginnen, ohne das gesamte Tutorial absolvieren zu müssen.

Ein nutzerzentrierter Ansatz zur Entwicklung von Tutorials wird in de Groot [Gr21] vorgestellt. Die beiden Varianten wurden auf Basis theoretischer Grundlagen konzipiert und implementiert. Das geführte Tutorial, das über Tafeln und gesprochene Instruktionen präsentiert wird, beginnt mit einer Begrüßung, bei der die Lernenden wählen können, ob sie dem Tutorial folgen möchten. Dabei werden die Controller eingeblendet und die relevanten Tasten für „Weiter“ und „Zurück“ hervorgehoben. Anschließend folgt eine Einführung in die Navigation, die drei Elemente umfasst: Zunächst das Gehen im Raum, bei dem beim Erscheinen der Begrenzung (Chaperone) deren Funktion erklärt wird; danach die Teleportation, bei der gegebenenfalls erneut Controllerhilfen eingeblendet werden; und schließlich die freie Bewegung, bei der die Lehrkraft Frau Bode zum Fenster geht und die Lernenden auffordert, ihr zu folgen – der nächste Schritt im Tutorial wird dann in ihrer Nähe aktiviert. Daraufhin lernen die Nutzenden, interaktive Objekte durch visuelle Hervorhebungen zu erkennen und diese auf einem Tisch auszuprobieren. Das Greifen von Objekten wird anhand eines Buchs demonstriert, wobei bei Schwierigkeiten zusätzliche Hinweise zur Handposition gegeben werden. Anschließend wird das Schreiben auf einer Liste am Whiteboard geübt, wobei entweder Kreide oder ein Stift verwendet wird. In einer älteren Version wurde auch eine Spracherkennung zur Kommunikation mit virtuellen Agenten getestet, die in aktuellen Versionen jedoch nicht mehr enthalten ist. Das Tutorial endet mit der Option, entweder direkt in den Klassenraum zu wechseln oder das Lehrerzimmer frei weiter zu erkunden. Eine Übersicht der noch offenen Aufgaben aus dem Hintergrund-Tutorial ist dabei auf einem Poster dargestellt.

Das Hintergrund-Tutorial deckt alle grundlegenden Interaktionen ab, die den Lernenden im Lehrerzimmer in Form einer To-do-Liste präsentiert werden. Zu den Aufgaben gehören das Aufnehmen und Umdrehen eines Buches ("Rückseite betrachten"), sowie das Schreiben auf das Whiteboard. Zudem soll Müll in einen Mülleimer geworfen werden – eine Aufgabe, die in späteren Versionen durch einen Zähler gamifiziert wurde und dazu dient, die Scheu vor Bewegung mit dem Headset zu verringern. Eine weitere Aufgabe kombiniert das Greifen und Navigieren mit Objekten: Eine Tasse soll zu Herrn Krause, einem virtuellen Agenten, gebracht werden, der sich bedankt, sobald sich die Tasse in seiner Nähe befindet. Auch eine kommunikative Interaktion ist vorgesehen: Nähert man sich Frau Bode oder spricht sie an, fragt sie, ob ein geführtes Tutorial gestartet werden soll. Die letzte Aufgabe besteht darin, den Klassenraum zu betreten – sie wird nicht explizit als abgeschlossen markiert, soll jedoch den Raumwechsel nach Absolvierung der übrigen Aufgaben anregen. Einige Elemente des Tutorials sind bedingt aktiv: Beispielsweise wird die freie Bewegung nur erklärt, wenn das Programm im Raummaßstab ausgeführt wird. Bei Nutzung im Stehen entfällt die Aufgabe zur Bewegung im Raum, und nur die Teleportation wird erklärt.

Die zweite Version des Lehrerzimmers enthält im Wesentlichen dieselben Aufgaben wie das ursprüngliche Tutorial, mit einigen Änderungen: Es sind mehr Papierbälle im Raum verteilt, und die Aufgabe gilt erst als abgeschlossen, wenn drei davon im Mülleimer gelandet sind. Die Kaffeetasse muss nun auf einer bestimmten Fläche neben der Lehrkraft abgestellt werden – ein bloßes Werfen reicht nicht mehr aus. Um einzelne Schritte wiederholbar zu machen, gibt es eine Pinnwand mit einer visuellen Übersicht des Tutorials. Die Aufgabe „Eintragen zur Pausenaufsicht“ wurde durch das einfache Testen der Kreidetafel ersetzt.

Die im Rahmen der Überarbeitung der zweiten Version des Lehrerzimmers durchgeführten Evaluationen [Al24] führten zu folgenden Anpassungen: Die Pinnwand wurde so überarbeitet, dass während der geführten Tour keine vollständig freie Auswahl der Aufgaben mehr möglich ist. Ein dauerhaft sichtbares Poster mit der Tastenbelegung der Controller soll die Bedienungssicherheit erhöhen und damit insbesondere beim Teleportieren mehr Vertrauen schaffen. Zudem wurden neue Interaktionen, z. B. mit einem Beamer, ergänzt. Nicht sichtbar, aber wirksam umgesetzt wurde eine Reduktion der Hintergrundgeräusche zur Verbesserung der Barrierefreiheit. Auch die Gesamtdauer des geführten Tutorials wurde leicht reduziert.

4.1 Virtuelle Agenten als Schlüsselbestandteil

Ein zentraler Bestandteil des hier beschriebenen Tutorials ist die (nicht-menschliche) virtuelle Agentin, die sich im Lehrerzimmer befindet: die Lehrerin Frau Bode. Ihre Aufgabe besteht darin, die Lernenden durch das Tutorial zu führen und bei Bedarf zu reagieren, z. B. wenn längere Zeit keine Interaktion erfolgt. Eine gute Umsetzung ihrer Rolle ermöglicht standardisierte Experimente, da das Tutorial einen zentralen Bestandteil geplanter Studien darstellt. „Die Instruktion ist nicht die Einführung in das Experiment, sondern ein wichtiger Teil des Experiments selbst“ [HK76]. Je nach Untersuchungsmethode und Forschungsfrage kann ein sinnvoll kontextualisierter Einstieg in ein Experiment hilfreich sein oder zumindest keine Verzerrung verursachen [ACG17].

Bei der iterativen Entwicklung des Agenten wurde versucht, Immersion und Präsenz – siehe [Sl05; WH24] – möglichst zu fördern. Die Aufgaben, die Frau Bode stellt, sind daher in das schulische Narrativ eingebettet [Go11].[2] Ein Beispiel ist die Aufgabe, einer weiteren Lehrkraft eine Tasse zu bringen – eine Herausforderung, bei der viele Lernende erstmals erkennen, dass sie auch mit einem Objekt in der Hand teleportieren können. In jeder Entwicklungsiteration wurden Frau Bodes Handlungen weiter angepasst. Nutzer*innenstudien zeigten wiederholt, an welchen Stellen Lehramtsstudierende Unterstützung benötigen. In der dritten Entwicklungsrunde wurde z. B. die Anleitung zur Bewegung optimiert: Frau Bode kündigt nun zunächst an, dass die Lernenden geführt werden sollen, und beginnt erst danach mit der Bewegung. In früheren Versionen erfolgten Ankündigung und Handlung gleichzeitig, was einige Personen als überfordernd empfanden. Zusammenfassend ist können

[2] Die Wirkung von Narrativen auf Präsenz wird in der Literatur unterschiedlich bewertet, siehe z. B. Park et al. [Pa10]. Eine positive Wirkung auf die Bewertung durch Nutzende ist jedoch anzunehmen.

virtueller Agenten Tutorials in standardisierter Form begleiten, um konsistentes Onboarding und Studienbedingungen zu gewährleisten.

5 Didaktische Einbettung von Tutorials & Fazit

Damit Tutorials in VR-Lernumgebungen nicht nur funktionale Einstiegshilfen bleiben, sondern didaktisch sinnvoll in Lehr-/Lernszenarien eingebettet sind, ist ihre Einbindung in ein übergeordnetes Veranstaltungskonzept zentral. Hierbei sind mehrere Faktoren zu berücksichtigen: Zunächst stellt sich die Frage nach dem didaktischen Ziel des Tutorials: Dient es lediglich der Bedienungseinweisung oder ist es Teil eines komplexeren didaktischen Designs, etwa zur Förderung von Medienkompetenz, zur Simulation beruflicher Handlungssituationen oder zur Reflexion über Unterrichtshandeln? Je nach Zielsetzung muss auch die immersive Qualität des Tutorials abgestimmt sein – von realitätsnahen Alltagssituationen bis hin zu abstrahierten Modellszenarien. Zudem sollten VR-Tutorials so gestaltet sein, dass sie sich flexibel in unterschiedliche Lehr-/Lernszenarien integrieren lassen: Als vorbereitende Übungseinheit, als wiederholt nutzbare Trainingssequenz oder als einmaliger Einstieg. Auch die Berücksichtigung unterschiedlicher Hardware-Setups, Begleitformate, sowie Unterschiede im Vorwissen der Lernenden beeinflussen die didaktische Konzeption. Hierbei stellt sich häufig das „Henne-Ei"-Problem: Müssen bestimmte Interaktionen bereits im Vorfeld erklärt werden – und wenn ja, wie? Schließlich sollten Tutorials in immersive Settings nicht nur die kognitive Zugänglichkeit, sondern auch die emotionale Sicherheit fördern. Elemente wie eine narrative Einbettung, gezielte Gamifizierung, adaptive Hilfestellungen und klar erkennbare Lernziele auf verschiedenen Taxonomiestufen können dazu beitragen, Lernende individuell zu begleiten und langfristig zu motivieren.

Dieser Beitrag zeigt auf, dass VR-Tutorials weit mehr als technische Einführungshilfen darstellen. Sie sind ein zentrales Gestaltungselement immersiver Lernumgebungen und tragen maßgeblich zur Bedienbarkeit, zur Lernwirksamkeit und zum Lernerleben bei. Dabei ist insbesondere die Verbindung aus mediendidaktischen Prinzipien, Erkenntnissen aus der Spieleentwicklung sowie Aspekten der Mensch-Computer-Interaktion vielversprechend, um praxisnahe und motivierende Tutorialkonzepte zu entwickeln.

Die Integration solcher Tutorials in didaktische Gesamtkonzepte erfordert ein Umdenken bei der Planung von VR-Lernanwendungen: Nicht allein die immersive Darstellung, sondern die gezielte Lernbegleitung entscheidet über den Erfolg des Mediums im Bildungskontext. Wie Kerres et al. [KBM21] und Niegemann [Ni19] betonen, braucht es dazu instruktionspsychologische und didaktische Kompetenz – sonst bleiben auch modernste Technologien wie VR oder KI pädagogisch wirkungslos.

Gut konzipierte Tutorials leisten in diesem Sinne nicht nur einen Beitrag zur besseren Nutzbarkeit von VR-Lernanwendungen, sondern eröffnen neue Perspektiven für reflexives, exploratives und immersives Lernen in der Hochschullehre und darüber hinaus.

Literaturverzeichnis

[ACG17] Alekseev, A.; Charness, G.; Gneezy, U.: Experimental Methods: When and Why Contextual Instructions Are Important. Journal of Economic Behavior & Organization 134, S. 48–59, 2017, DOI: 10.1016/j.jebo.2016.12.005.

[Al24] Alti, F.: Enhancing VR Classroom Onboarding: A User-Centric Redesign for Improved Usability, Magisterarb., Aachen, Germany: RWTH Aachen University, 2024, https://publications.rwth-aachen.de/record/986840.

[Ar20] Armougum, A. et al.: Expertise reversal effect: Cost of generating new schemas. Computers in Human Behavior 111, S. 106406, 2020, DOI: 10.1016/j.chb.2020.106406.

[BA20] Buchner, J.; Aretz, D.: Lernen mit immersiver Virtual Reality: Didaktisches Design und Lessons Learned. MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung, S. 195–216, 2020, DOI: 10.21240/mpaed/jb17/2020.05.01.X.

[CAS19] Castro-Alonso, J. C.; Ayres, P.; Sweller, J.: Instructional Visualizations, Cognitive Load Theory, and Visuospatial Processing. In (Castro-Alonso, J. C., Hrsg.): Visuospatial Processing for Education in Health and Natural Sciences. Springer International Publishing, Cham, S. 111–143, 2019, DOI: 10.1007/978-3-030-20969-8_5.

[CHP23] Chauvergne, E.; Hachet, M.; Prouzeau, A.: User Onboarding in Virtual Reality: An Investigation of Current Practices. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI '23, Association for Computing Machinery, New York, NY, USA, S. 1–15, 2023, DOI: 10.1145/3544548.3581211.

[FH20] Fussan, C.; Hanesova, V.: Pre-Testing der Lerneffektivität von 2D-to-3D-Didactics in immersiven VR-Umgebungen. In (Zender, R. et al., Hrsg.): DELFI 2020 – Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V. DELFI. Gesellschaft für Informatik e.V., S. 62–72, 2020, http://dl.gi.de/handle/20.500.12116/34205, Stand: 24. 11. 2020.

[FJU08] Fromme, J.; Jörissen, B.; Unger, A.: Bildungspotenziale digitaler Spiele und Spielkulturen. MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung 15, S. 1–23, 2008, DOI: 10.21240/mpaed/15_16/2008.12.22.X.

[Fr17] Frommel, J. et al.: The Effects of Context-Sensitive Tutorials in Virtual Reality Games. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play. CHI PLAY '17, Association for Computing Machinery, New York, NY, USA, S. 367–375, 2017, DOI: 10.1145/3116595.3116610.

[Go11] Gorini, A. et al.: The Role of Immersion and Narrative in Mediated Presence: The Virtual Hospital Experience. Cyberpsychology, Behavior, and Social Networking 14 (3), S. 99–105, 2011, DOI: 10.1089/cyber.2010.0100.

[Gr21] de Groot, R.: Nutzerzentrierte Einführung in VR-Umgebungen Am Beispiel Des Virtuellen Klassenzimmers, Magisterarb., Aachen, Germany: RWTH Aachen University, 2021, DOI: 10.18154/RWTH-2021-09016.

[HK76] Heller, O.; Krüger, H.-P.: Experimentelle Psychologie, Wahrnehmung: Ein Programmiertes Lehrsystem Mit Selbstversuchen: Lehrbuch. H. Huber, Bern, 1976.

[Ho17] Ho, J. C. F.: Practice in Reality for Virtual Reality Games: Making Players Familiar and Confident with a Game. In (Bernhaupt, R. et al., Hrsg.): Human-Computer Interaction - INTERACT 2017. Lecture Notes in Computer Science, Springer International Publishing, Cham, S. 147–162, 2017, DOI: 10.1007/978-3-319-67684-5_10.

[Ja15] Jamieson, D.: 4 Ways to Teach Your Players How to Play Your Game, Envato Tuts+ How-To Tutorials, 2015, https://gamedevelopment.tutsplus.com/tutorials/4-ways-to-teach-your-players-how-to-play-your-game--cms-22719, Stand: 10. 09. 2021.

[Je15a] Jerald, J.: The Basics: Design Guidelines. In: The VR Book: Human-Centered Design for Virtual Reality. Association for Computing Machinery and Morgan & Claypool, 2015, DOI: 10.1145/2792790.2792797, Stand: 10. 04. 2020.

[Je15b] Jerald, J.: The VR Book: Human-Centered Design for Virtual Reality. Association for Computing Machinery and Morgan & Claypool, 2015.

[KBM21] Kerres, M.; Buchner, J.; Mulders, M.: Immersives Lernen? Didaktisches Design für Augmented / Virtual Reality und reaktive Objekte / Umwelten. In (Wilbers, K., Hrsg.): Handbuch E-Learning. Deutscher Wirtschaftsdienst, Köln, S. 15, 2021, https://learninglab.uni-due.de/publikationen/13437.

[Ma21] Mayer, R. E.: Multimedia Learning. 3. Ed. Cambridge University Press, Cambridge, 2021.

[MBK20] Mulders, M.; Buchner, J.; Kerres, M.: A Framework for the Use of Immersive Virtual Reality in Learning Environments. International Journal of Emerging Technologies in Learning (iJET) 15 (24), S. 208–224, 2020, DOI: 10.3991/ijet.v15i24.16615.

[Ni19] Niegemann, H.: Instructional Design. In (Niegemann, H.; Weinberger, A., Hrsg.): Handbuch Bildungstechnologie: Praxisorientiertes Handbuch zum intelligenten Umgang mit digitalen Medien. Springer Reference Psychologie, Springer, Berlin, Heidelberg, S. 1–57, 2019, DOI: 10.1007/978-3-662-54373-3_7-1.

[OP05] Oliver, M.; Pelletier, C.: The Things We Learned on Liberty Island: Designing Games to Help People Become Competent Game Players. 2005, http://www.digra.org/wp-content/uploads/digital-library/06276.22378.pdf.

[Pa09] Pantelidis, V. S.: Reasons to Use Virtual Reality in Education and Training Courses and a Model to Determine When to Use Virtual Reality. Themes in Science and Technology Education 2 (1–2), S. 59–70, 2009.

[Pa10] Park, N. et al.: Effects of Pre-Game Stories on Feelings of Presence and Evaluation of Computer Games. International Journal of Human-Computer Studies 68 (11), S. 822–833, 2010, DOI: 10.1016/j.ijhcs.2010.07.002.

[Qu17] Quesada Mora, V. A. et al.: Subliminal Learning. What Do Games Teach Us? In (Zaphiris, P.; Ioannou, A., Hrsg.): Learning and Collaboration Technologies. Novel Learning Ecosystems. Lecture Notes in Computer Science, Springer International Publishing, Cham, S. 487–501, 2017, DOI: 10.1007/978-3-319-58509-3_38.

[Ra20] Radianti, J. et al.: A Systematic Review of Immersive Virtual Reality Applications for Higher Education: Design Elements, Lessons Learned, and Research Agenda. Computers & Education 147, S. 103778, 2020, DOI: 10.1016/j.compedu.2019.103778.

[SAK11] Sweller, J.; Ayres, P.; Kalyuga, S.: The Redundancy Effect. In (Sweller, J.; Ayres, P.; Kalyuga, S., Hrsg.): Cognitive Load Theory. Springer, New York, NY, S. 141–154, 2011, DOI: 10.1007/978-1-4419-8126-4_11.

[Sl05] Slater, M.: Presence 2005: The Eighth Annual International Workshop on Presence, 21-23 September, 2005 University College London (Conference Proceedings). Department of Computer Science, UCL (University College London), 2005.

[Su03] Sutcliffe, A.: Multimedia and Virtual Reality: Designing Multisensory User Interfaces. Psychology Press, New York, 2003.

[Wh12] White, M. M. (M.: New Tutorials for Digital Games: Game Design Meets Instructional Design, doctoral, Memorial University of Newfoundland, 2012, DOI: 10/3/White_MatthewM.pdf.

[WH24] Wiepke, A.; Heinemann, B.: A Systematic Literature Review on User Factors to Support the Sense of Presence in Virtual Reality Learning Environments. Computers & Education: X Reality 4, S. 100064, 2024, DOI: 10.1016/j.cexr.2024.100064.

[Wh25] Whittaker, L.: Onboarding and Offboarding in Virtual Reality: A User-Centred Framework for Audience Experience across Genres and Spaces. Convergence 31 (1), S. 87–106, 2025, DOI: 10.1177/13548565231187329.

[Wi93] Winn, W.: A Conceptual Basis for Educational Applications, Seattle, Washington, USA: Human Interface Technology Laboratory, University of Washington, 1993, http://www.hitl.washington.edu/projects/education/winn/winn-paper.html~, Stand: 29. 06. 2021.

[Ze19] Zender, R. et al.: Potentials of Virtual Reality as an Instrument for Research and Education. i-com 18 (1), S. 3–15, 2019, DOI: 10.1515/icom-2018-0042.

Learning Analytics/Gamification

Entwicklung eines Lehrkräfte-Dashboards für ein sicheres und schulgerechtes soziales Netzwerk

Ein partizipativer, nutzungsorientierter Entwicklungsprozess

Claudine Boyer[1], Tobias Berner[1] und Thomas Schmalfeldt[1]

Abstract: Mit opostum wurde ein sicheres soziales Netzwerk für den Einsatz in der Schule entwickelt. Damit kann das Thema Social Media in einem datenschutzkonformen Rahmen handlungsorientiert vermittelt werden. In opostum können die Jugendlichen Texte, Bilder und Videos posten, Kommentare schreiben, liken und anderen Personen folgen. Aber auch in einer geschlossenen Umgebung kann es zu unerwünschtem Verhalten oder problematischen Dynamiken zwischen den Jugendlichen kommen. Damit die Lehrkräfte den Überblick über die Beiträge der Klasse behalten, problematische Inhalte ausblenden und die Rechte der Benutzenden innerhalb von opostum steuern können, wurde ein Dashboard entwickelt. In einem partizipativen, mehrstufigen Prozess, der sich am Design-Thinking-Ansatz orientiert, wurden die Bedürfnisse und Vorbehalte der Lehrkräfte an ein soziales Netzwerk für den Unterricht und dem dazugehörigen Dashboard in einem Fokusgruppeninterview erhoben. Die daraus hergeleiteten Prototypen wurden in Einzelinterviews evaluiert, woraus ein finaler Entwurf für das Dashboard entstand.

Keywords: Soziale Netzwerke, Dashboard, Datenschutz, Lehrkräfte, Unterricht

1 Soziale Medien

1.1 Soziale Medien im Alltag der Jugendlichen und in der Schule

Soziale Medien gehören zum Alltag der Jugendlichen. Gemäß der aktuellen JIM (Jugend, Information, Medien) [Me24] und JAMES (Jugend, Aktivitäten, Medien – Erhebung Schweiz) [Kü24] Studien 2024 sind soziale Netzwerke weiterhin eines der wichtigsten medialen Elemente der 12- bis 19-Jährigen. Über 80 % der Jugendlichen in der Schweiz und 62 % in Deutschland nutzen Instagram mehrmals pro Woche. Instagram gehört neben TikTok zu den beliebtesten Vertretern. Aktuell geraten soziale Netzwerke zunehmend in die Kritik, unter anderem wegen ihrer Auswirkung auf die psychische Gesundheit junger Menschen. Auf politischer Ebene werden in einigen Ländern Verbote diskutiert. Australien verabschiedete im letzten Jahr ein Social-Media-Gesetz, das Jugendlichen unter 16 Jahren die Nutzung sozialer Netzwerke ab Ende 2025 verbietet [De24]. Ob Verbote solcher Netzwerke zielführend sind, wird kontrovers diskutiert. Kritiker:innen vertreten die Haltung, dass die Schule und auch andere Instanzen (Eltern, Jugendliche

[1] Pädagogische Hochschule Zürich, Lagerstrasse 2, 8090 Zürich, claudine.boyer@phzh.ch

doi:10.18420/delfi2025_14

selbst, Regulierungsbehörden, Medienanbieter) Verantwortung übernehmen und Jugendliche im angemessenen Umgang mit diesen Medien unterstützen müssen [Kü24].

Beim Erlernen eines adäquaten Umgangs in und mit sozialen Medien nehmen somit die Schule und die Jugendarbeit wichtige Rollen ein. Die Lehrkräfte und Betreuenden stehen dabei vor der Herausforderung, dass für die Gestaltung der Lernszenarien nicht auf kommerzielle Plattformen wie Instagram, TikTok oder Ähnliches zurückgegriffen werden kann, da der Datenschutz für den Schuleinsatz nicht ausreichend gewährleistet ist. Dies erschwert oder verhindert den explorativen Zugang zu einem höchst relevanten Thema, in dem die Interaktion untereinander und die eigenen Erfahrungen substanziell sind.

1.2 Ein sicheres soziales Netzwerk für die Schule und die Jugendarbeit

Für einen sicheren Einsatz sozialer Medien wurde daher, basierend auf der für den Aufbau von Kompetenzen zu Datenbanken entwickelten Lernumgebung Instahub [Do19], *opostum* entwickelt. Auf dieser geschützten Plattform können Schüler:innen selbst Erfahrungen in und mit sozialen Medien machen. Die Lehrkräfte und Betreuenden haben dabei die vollständige Kontrolle über die Daten, die durch die Nutzung der Plattform entstehen. Die Schüler:innen können Fotos und Videos posten, Kommentare schreiben, Likes setzen, Freunde finden und vieles mehr. *opostum* ermöglicht es, dass beispielsweise für eine Klasse ein geschlossenes soziales Netzwerk eröffnet werden kann, auf das nur die Schüler:innen der jeweiligen Klasse Zugriff haben. Die Lehrkraft in der Rolle als Administrator:in hat dabei die Kontrolle, wer in das Netzwerk aufgenommen wird. Es sind zudem größere Netzwerke möglich (z. B. ein ganzes Schulhaus). Durch die flexible Nutzendenverwaltung ist ebenso ein Einsatz in der außerschulischen Jugendarbeit möglich, in der es keine festen Zuordnungen der Jugendlichen gibt.

Auch wenn sich das soziale Netzwerk auf eine Schulklasse beschränkt, besteht das Risiko, dass Jugendliche unpassende Inhalte veröffentlichen. Um dieses zu reduzieren, kann eine Lehrkraft Interaktionsmöglichkeiten gezielt freischalten und deaktivieren, z. B. dass Kommentare geschrieben oder Beiträge geliked werden können. Des Weiteren können Beiträge unwiderruflich gelöscht werden (*Recht auf Vergessen*). Neben den Möglichkeiten der Nutzung als soziales Netzwerk lässt sich *opostum* für die Behandlung weiterer relevanter Themen wie personalisierte Werbung, User-Tracking oder Empfehlungsalgorithmen nutzen. Die Jugendlichen können eigene Werbekampagnen auf *opostum* gestalten und diese gezielt ausgewählten Nutzenden anzeigen lassen. Dabei kann das Thema Werbung umfassend behandelt und auf soziale Medien erweitert werden.

In diesem Beitrag wird dargelegt, wie Lehrkräfte mit einem Dashboard dabei unterstützt werden können, den Überblick über das Geschehen auf einer Lernplattform zu behalten. Zudem wird aufgezeigt, wie das Dashboard für *opostum* in einem partizipativen Prozess gemeinsam mit den Lehrkräften entwickelt wurde.

2 Lehrkräfte-Dashboards im Bildungsbereich

2.1 Aufbau und Funktionen von Dashboards

Dashboards sind visuelle Anzeigen, mit denen Daten so dargestellt werden, dass die zugrundeliegenden Informationen von den Nutzenden möglichst effizient erfasst und verarbeitet werden können [Fe06]. Sie dienen als Instrument zur Überwachung eines Zustands, sodass auf Situationen schnell und adäquat reagiert werden kann [Kh24]. Im Bildungsbereich dienen Dashboards der Unterstützung der Lernenden oder indem die Daten den Lehrkräften zugänglich gemacht werden. Dies ermöglicht es Letzteren, die Aktivitäten der Klasse und einzelner Schüler:innen in digitalen Lernumgebungen zu erfassen und daraus, basierend auf verschiedenen zusammengeführten Indikatoren [Sc17], gezielte Unterstützungsangebote abzuleiten. Zudem schafft es die Grundlage, eigene Unterrichtspraktiken zu reflektieren und anzupassen [Va21]. Werden die Dashboards in Echtzeit zur Nutzung der Lernumgebung der Schüler:innen im Unterricht genutzt (synchrone Nutzung), ermöglicht dies, dass eine Lehrkraft direkt in den Unterricht eingreifen kann, um einen Input einzuflechten oder um auf oft auftretende Fehler im Klassenverband oder in Gruppen einzugehen [Al10]. In der asynchronen Nutzung konsultiert es die Lehrkraft in der Vorbereitung einer Lektion. Es zeigte sich, dass Lehrkräfte das Dashboard öfter dazu nutzen, um Interventionen für die gesamte Klasse zu planen als für die individuelle Förderung [XAM17].

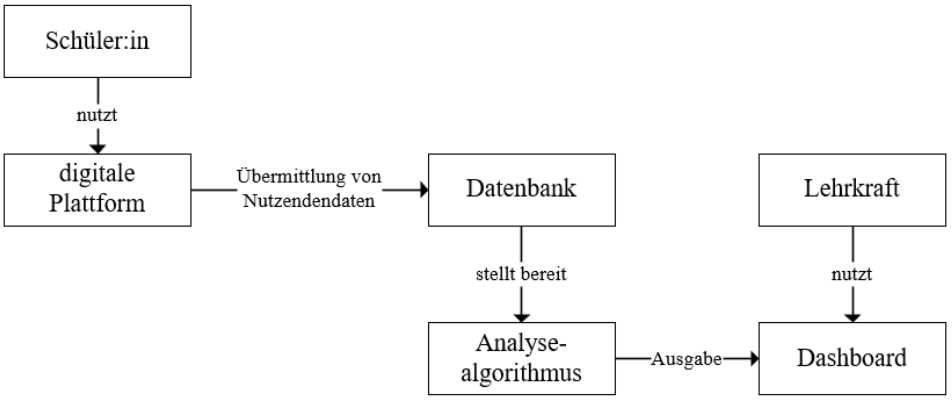


Abb. 1: Funktionsweise eines Dashboards (angepasst nach [Ca19])

Die Funktionsweise eines Dashboards wird in Abbildung 1 dargestellt. Dabei kann die digitale Plattform ein Learning-Management-System (LMS) oder, wie in diesem Artikel ausgeführt, ein soziales Netzwerk sein. Wenn die Daten nur zu Visualisierungszwecken aufgearbeitet werden, wird von *spiegelnden* (mirroring) Systemen gesprochen. Werden die Daten so verarbeitet, dass die Lehrkräfte in konziser Form auf wichtige Ereignisse aufmerksam gemacht werden, werden sie *alarmierende* (alerting) Systeme genannt. Bei

beratenden (advising) Systemen werden die aus den Aktivitätsanalysen gewonnenen Informationen mit Hinweisen zur aktuellen Situation oder zu möglichen Maßnahmen im Unterricht ergänzt [VR19]. Spiegelnde Systeme sind dann besonders hilfreich, wenn durch sie latente Informationen dargestellt werden, die die Lehrkräfte nicht selbst im Unterricht beobachten können [VR22]. Ein Dashboard, das die volle Kontrolle über den Unterricht übernimmt, lehnen Lehrkräfte ab [Al10]. Es zeigte sich, dass sich Lehrkräfte von Dashboards mit beratender Funktion sogar bedroht fühlen können und einen Kontrollverlust über die eigene Klasse und ihr pädagogisches Handeln befürchten. Konnte allerdings ein gewisser Zeitdruck simuliert werden, der schnelle Entscheidungen anhand der Informationen auf einem beratenden Dashboard verlangte, zeigte sich ein anderes Bild: Die Lehrkräfte erkannten die Problemsituationen genauer, wenn ihnen die Hinweise des Dashboards zur Verfügung standen [VR22].

Lehrkräfte berichteten, dass die Alarmhinweise ihre Aufmerksamkeit auf diejenigen Schülerinnen und Schüler lenkten, die normalerweise still für sich arbeiten und bei Schwierigkeiten zurückhaltend darin sind, Hilfe anzufordern. Diesbezüglich konnte zudem festgestellt werden, dass auch die Schüler:innen die Unterstützung der Lehrkraft als hilfreich empfanden, obwohl sie selbst keine Hilfe angefordert hatten [Go23]. Lehrkräfte, die ein Dashboard nutzen können, geben häufiger ein Feedback, als wenn sie kein solches Hilfsmittel zur Verfügung hatten [VR20]. Es ist daher anzunehmen, dass Dashboards mit beratender Funktion gegenüber solchen mit spiegelnder oder alarmierender Funktion bevorzugt werden sollten, da sie der Lehrkraft ein besseres Verständnis der Problematik ermöglichen [Va19].

2.2 Kriterien zur Dashboard-Gestaltung

Für die Darstellungen von Informationen auf Dashboards sollte eine sinnvolle Anzahl an dargestellten Indikatoren und eine für die Zielgruppe hilfreiche Visualisierung gewählt werden, da eine zu umfangreiche Menge an Informationen die Nutzenden überfordern kann. Zu den beliebtesten Diagrammtypen gehören Balken-, Linien- und Kuchendiagramme sowie Tabellen und Netzwerkgrafiken [Sc17]. Diese Typen sind auch im außerschulischen Kontext weit verbreitet und es besteht eine gewisse Routine, die Informationen aus diesen Diagrammen herauslesen zu können. Es hat sich ein Lehrkräfte-Dashboard mit zwei Fenstern bewährt. Ein Übersichtsfenster bildet die Klasse mit Informationen zu den einzelnen Schüler:innen ab und erlaubt es, die Aufmerksamkeit auf einzelne Lernende zu lenken. In einem zweiten Fenster mit Detailinformationen kann sich die Lehrkraft in einzelne personenbezogene Aktivitäten vertiefen, um daraus individuelle Unterstützungsangebote abzuleiten [Al22]. Die individuelle Anpassung des Dashboards ist ein oft von Lehrkräften genanntes Bedürfnis, um den Fokus gezielt auf diejenigen Aspekte richten zu können, die aktuell als relevant erachtet werden [Ma19].

3 Forschungsdesign und Stichprobe

In der hier vorgestellten Erhebung wurde der Frage nachgegangen, welche Ansprüche Lehrkräfte an ein Lehrkräfte-Dashboard für ein sicheres soziales Netzwerk für die Schule identifizieren, sodass sie es in ihrem eigenen Unterricht einsetzen würden. Da es in diesem Bereich noch keine bekannten Publikationen gibt, wurde ein erkenntnisgewinnender, qualitativer Ansatz gewählt. Es wurde auf das Design-Thinking-Framework zurückgegriffen, bei dem eine nutzendenzentrierte Perspektive prägend ist [Ge17]. Diese Herangehensweise wird für die Dashboard-Entwicklung empfohlen (vgl. [Ca19; Ma19; MMM23]).

In unserem Projekt wurden die sechs Design-Thinking-Phasen [GTF20] umgesetzt, basierend auf etablierten empirischen Methoden (siehe Abb. 2). In der Phase des *Verstehens* (1) wurden anhand einer Literaturanalyse die theoretischen Grundlagen geschaffen. Ein Faktor für einen erfolgreichen Entwicklungsprozess für ein Lehrkräfte-Dashboard ist die Beteiligung der Lehrkräfte ab der Bedarfsanalyse über die Konzeptentwicklung bis zur Validierung [Ve20]. Daher wurden in der *Beobachtung*sphase (2) zur Sammlung möglichst vieler Aspekte mit den Lehrkräften leitfadengestützte Fokusgruppeninterviews durchgeführt [SMR12]. Die Lehrkräfte treten als Expert:innen für Unterricht und erfahrungsbasiertes Wissen zu Stärken und Schwächen von Schüler:innen im Klassenverband auf. Dabei rekonstruieren sie den Prozess der Dashboard-Nutzung stellvertretend für die zukünftigen Akteur:innen [GL09] und eventuell latenter Wissensbedarf wird manifest. Der Leitfaden wurde gemäß dem SPSS- (Sammeln, Prüfen, Sortieren, Subsumieren) Prinzip erstellt [He11] und enthielt mehrere Erzählaufforderungen (bspw. „Wie habt ihr bisher das Thema Social Media im Unterricht behandelt?“, „Welche Potenziale seht ihr in der Nutzung von *opostum* für euren Unterricht?“ oder „Welche Informationen sollten auf dem Lehrkräfte-Dashboard angezeigt werden? Welche Funktionen sind für dich wichtig?“), konkrete Unterfragen (bspw. „Welche Herausforderungen und Risiken sehr ihr in Bezug auf die Arbeit mit *opostum* in Unterricht?“, „Wie kann ein Dashboard euch als Lehrkraft unterstützen?“ oder „Welche konkreten Ideen und Vorschläge habt ihr zur Nutzung des Dashboards für euren Unterricht?“) sowie Aufrechterhaltungs- und Steuerungsfragen. Als Stimulus wurde zu Beginn ein Video gezeigt, in dem das sichere Netzwerk kurz vorgestellt wurde. Zur Akzentuierung verschiedener *Sichtweisen* (3) wurden die Interviews transkribiert und mit einem deduktiv-induktiv entwickelten Kategoriensystem kodiert [Ma23].

Die Erkenntnisse wurden in der Phase der *Ideenfindung* (4) mit dem für die Implementierung zuständigen technischen Entwickler bezüglich der Umsetzbarkeit diskutiert. Anschließend wurden in einer Präsentationssoftware Mockups als *Prototypen* (5) erstellt. Dabei handelt es sich um grafische Darstellungen unterschiedlicher Dashboards, die noch nicht funktionsfähig sind. Für die Erstellung wurde die 10-plus-10-Methode verwendet [Gr14], durch die fortlaufend Ideen generiert, reduziert und verfeinert werden. Für das *Testen* (6) und zur Validierung wurden leitfadengestützte Einzelinterviews durchgeführt, bei denen die unterschiedlichen Mockups als Stimuli dienten. So konnten subjektive

Perspektiven und Deutungen zu den Prototypen eingeholt werden. Dies schloss die sechs Phasen des Design-Thinking-Frameworks ab, sodass zur Produkt-umsetzung übergegangen wurde. Die Einzelinterviews wurden analog zu den Fokusgruppeninterviews ausgewertet und ein Mockup des finalen Dashboards wurde zur technischen Umsetzung weitergegeben.

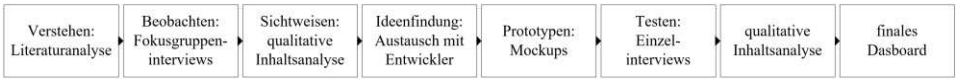


Abb. 2: Projektablauf gemäß den sechs Phasen des Design-Thinking-Prozesses mit abschließender Analyse und Erstellung des finalen Dashboards

An der Fokusgruppendiskussion nahmen sechs Lehrkräfte aus der Schweizer Volksschule teil (eine Primarschullehrkraft, eine Lehrkraft mit Erfahrungen auf der Primar- und Sekundarstufe 1, vier Lehrkräfte der Sekundarstufe 1). Die Teilnehmenden hatten das sichere soziale Netzwerk bereits in einem Workshop kennengelernt, in dem eine Vorabversion vorgestellt wurde. An den Einzelinterviews nahmen neun Personen teil (vier Lehrkräfte, die bereits in der Fokusgruppe waren, drei Personen aus der *opostum*-Projektgruppe und zwei Lehrkräfte, die an einem Barcamp teilgenommen hatten, auf dem eine Zwischenversion vorgestellt und diskutiert wurde).

4 Ergebnisse

Anhand der Aussagen der Fokusgruppe konnten konkrete Dashboard-Elemente herausgearbeitet werden, die zur Strukturierung der nachfolgenden Ergebnisse dienten:

Nutzendenliste mit Zugriff auf zeitliche Nutzungsverläufe zu jedem Profil: Alle Teilnehmenden äußerten den Wunsch, ihre Klasse mit den einzelnen Schüler:innen sowie deren User-Profile auf dem Dashboard einsehen zu können. Mehr als die Hälfte der Teilnehmenden gab an, es zu begrüßen, wenn sie in der Nutzendenliste auf eine Schüler:in klicken könnten, um einen zeitlichen Verlauf und weitere Informationen über das individuelle Nutzungsverhalten der ausgewählten Person zu erhalten.

Suche und Filterung nach problematischen Inhalten anhand von Stichworten und die Möglichkeit, Beiträge zu markieren: Eine teilnehmende Person erwähnte, dass sie die Möglichkeit schätzen würde, problematische Posts zu markieren. Eine weitere Person erwähnte die Möglichkeit, die Posts der Schüler:innen nach Schlüsselwörtern durchsuchen zu können, um zu sehen, wer ein bestimmtes Wort verwendet hat. Es könnte hilfreich sein, wenn das System automatisch darauf hinweist, wenn Nutzende vorher definierte Wörter verwendet haben.

Anzeige von markierten und/oder ausgeblendeten Posts: Der Live-Feed und das Dashboard sollten unabhängig voneinander dargestellt werden können. Dies würde es ermöglichen, dass beispielsweise markierte oder ausgeblendete Posts in der Klasse gezeigt

und mit den Schüler:innen besprochen werden können. Während der Lehrkraft alle Informationen auf dem Dashboard zur Verfügung stehen, müssten die Beiträge für die Präsentation aus Datenschutzgründen im entsprechenden Fenster anonymisiert werden.

Auswahl an statistischen Nutzungsdaten über die Klasse: Eine statistische Auswertung, die Posts mit den meisten Likes und jene mit den meisten Kommentaren aufführt, gaben alle Fokusgruppenteilnehmenden als wünschenswert an. Drei Personen erwähnten, dass ein zeitlicher Filter sinnvoll wäre, um beispielsweise anzuzeigen, welche Beiträge in den letzten sieben Tagen erstellt wurden. Zwei Personen nannten eine Auswertung, die Auskunft darüber gibt, wer wie viel Zeit im Netzwerk verbracht hat und wann jemand zuletzt aktiv war. Es wurde zudem ein erweiterter Modus mehrfach erwähnt, der zusätzliche Auswertungen und Statistiken bietet.

Stopp- und Freeze-Funktion: Alle Teilnehmenden waren sich einig, dass eine Stopp-Möglichkeit benötigt wird, um in einer problematischen Situation alles zu blockieren oder um sicherzustellen, dass alle Schüler:innen aufmerksam zuhören. Allerdings gingen die Meinungen darüber auseinander, was genau durch einen Stopp-Knopf ausgelöst werden soll. Jemand schlug vor, dass der Bildschirm einfach weiß oder leer sein sollte und mehrere Personen hatten die Vorstellung, dass es wie ein Freeze (Einfrieren) der User sein sollte und keine Posts, Kommentare oder Likes mehr abgesetzt werden können.

Hinweise zu den Unterrichtsmaterialien: Die Teilnehmenden erwähnten mehrfach den Wunsch nach vorhandenen Unterrichtssettings inklusive Materialien und Lehrkräftekommentaren. Alle schlugen vor, dass es einen Pool solcher Unterrichtsvorschläge geben sollte. Im Zusammenhang mit dem Dashboard wurde dies so begründet, dass im Dashboard ersichtlich sein sollte, welche Unterrichtssettings im eigenen Netzwerk gerade aktiv genutzt werden. Ob die Unterrichtsvorschläge und Lehrkräftekommentare dazu ebenfalls ins Dashboard gehören oder diese auf einem separaten Wiki platziert und ins Dashboard verlinkt sein könnten, darüber waren die Teilnehmenden unterschiedlicher Meinung.

Gestaltung des Dashboards: Im Zusammenhang mit der Darstellung des Dashboards wurde ein einfacher und ein erweiterter Modus von den Befragten gewünscht. Den erweiterten Modus sahen die Teilnehmenden vor allem für zusätzliche statistische Nutzungsauswertungen. Drei Personen erwähnten, dass sie das Dashboard individuell anpassen möchte.

4.1 Entwicklung der Dashboard-Prototypen

Mit diesen herausgearbeiteten Punkten wurden Dashboard-Prototypen entwickelt. Zuerst wurden verschiedene Skizzen für mögliche Dashboards und deren Elemente auf Papier festgehalten. Dabei erwies es sich als sinnvoll, verschiedene Bereiche des Dashboards zu definieren, die ähnliche oder zusammengehörende Informationen umfassen und dadurch von anderen Bereichen visuell abgegrenzt werden sollten. Mithilfe der Skizzen entstanden zwei Dashboard-Prototypen, die als PowerPoint-Präsentationen umgesetzt wurden.

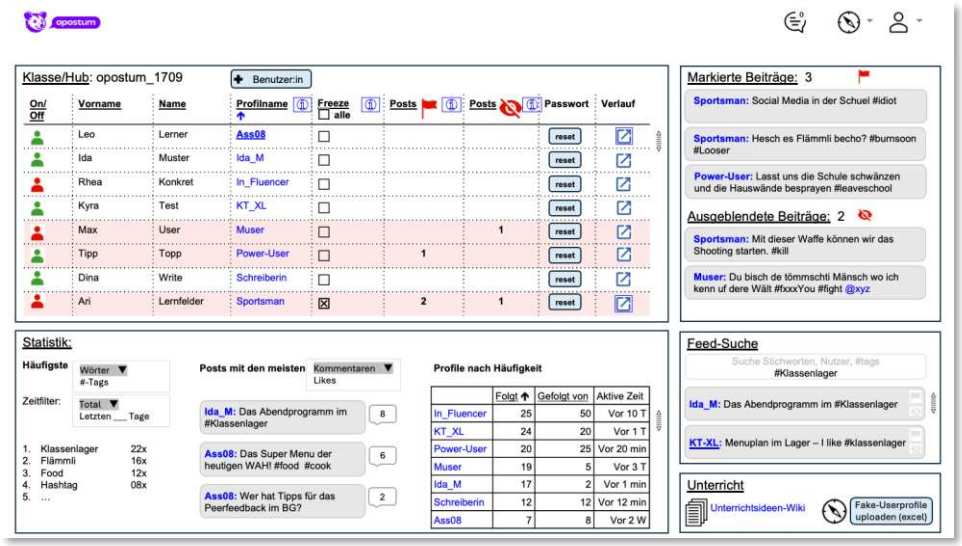


Abb. 3: Prototyp 1, bei dem die meisten Informationen gleichzeitig dargestellt werden.

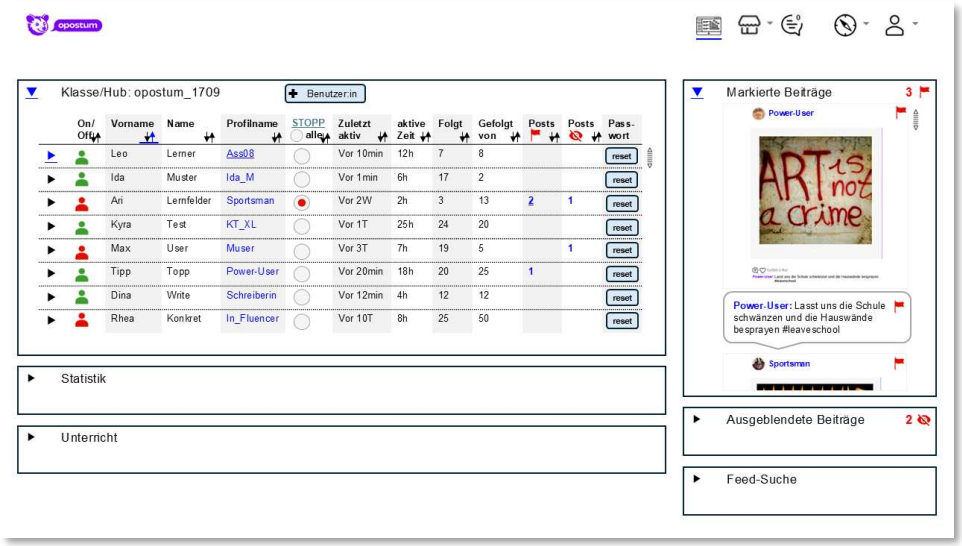


Abb. 4: Prototyp 2 mit offener Klassenübersicht und offenen markierten Beiträgen. Die weiteren Übersichten können bei Bedarf eingeblendet werden.

Die Prototypen unterschieden sich hauptsächlich in der Darstellungsform der Menge an Informationen, die auf den ersten Blick sichtbar waren. Beim Prototyp 1 (siehe Abb. 3) waren die meisten Informationen direkt auf einen Blick sichtbar. Beim Prototyp 2 (siehe Abb. 4) war nur eine konzise Übersicht vorhanden, bei der einzelne Fenster erweitert werden konnten und das Dashboard entsprechend individuell angepasst werden konnte.

4.2 Ergebnisse der Einzelinterviews

Zur Validierung der Prototypen wurden Einzelinterviews durchgeführt. Basierend auf der Analyse der Antworten wurde anschließend ein finaler Prototyp entwickelt, der als Gestaltungsvorlage für *opostum* diente.

Die Mehrheit der Teilnehmenden der Fokusgruppe favorisierte den Prototyp 2 aufgrund der individuelleren Nutzbarkeit. Ihnen gefiel, dass sie die gewünschten Bereiche je nach ihren Bedürfnissen öffnen oder schließen konnten. Das Dashboard erschien den Fokusgruppenteilnehmenden in der Mehrheit übersichtlicher, klarer bzw. aufgeräumter, und eine Person meinte, dass sie den Eindruck habe, damit schneller zu den nötigen Informationen kommen zu können. Ebenfalls merkte eine Person an, dass sie es begrüßte, dass mehrere Bereiche gleichzeitig aufgeklappt sein können. Es wurde zudem von zwei Fokusgruppenteilnehmenden der Wunsch geäußert, dass das Dashboard mit den zuletzt aufgeklappten Bereichen erscheinen solle, falls das Dashboard verlassen und später erneut aufgerufen werde.

Alle Interviewteilnehmenden bevorzugten bei der Darstellung von markierten oder ausgeblendeten Beiträgen neben dem Text auch einen Einblick in zugehörige visuelle Inhalte. Mit Ausnahme von zwei Befragten waren sich die Teilnehmenden einig, dass diese Posts identisch wie der Live-Feed dargestellt werden sollten, um sichtbar zu machen, worum es in diesen geht.

Besonders die Fokusgruppenteilnehmenden äußerten sich mehrheitlich positiv zur Darstellung innerhalb der Statistik, die diejenigen Posts mit den meisten Kommentaren und Likes zeigte. Eine Person fand es wertvoll, dass beide Darstellungen im Prototyp 2 nebeneinander Platz haben und nicht über eine Dropdown-Auswahl das eine oder das andere gewählt werden müsse. Ein *opostum*-Projektmitglied fand die Darstellung der Posts mit den meisten Kommentaren oder Likes zwar hilfreich, konnte sich jedoch nicht vorstellen, wie sie diese Angaben konkret nutzen würde. Für die Darstellung der am häufigsten verwendeten Wörter oder Hashtags im Netzwerk erwähnten speziell die Barcamp-Teilnehmenden und *opostum*-Mitglieder, dass sie eine Darstellung als Wortwolke bevorzugen würden, weil diese eine gute Übersicht liefern würde.

Im Zusammenhang mit der Feed-Suche sprach sich die Mehrheit für eine Darstellung mit Bild wie im Live-Feed aus. Die Möglichkeit, im Prototyp 2 eine eigene Wortliste hochzuladen, um möglichst schnell nach selbst definierten Wörtern in allen Beiträgen der Schüler:innen zu suchen und sich diese Beiträge anzeigen zu lassen, empfand weniger als die Hälfte aller Befragten als notwendig. Eine Person aus der Fokusgruppe erwähnte als

Vorteil der Wortliste, dass harmlose Wörter schnell gesucht werden könnten, die jedoch in der eigenen Klasse als heikel beurteilt werden. Die Mehrheit der Interviewteilnehmenden gab an, dass sie eine solche Wortliste nicht nutzen würden oder keinen Einsatzzweck dafür sahen. Eine Person aus der Fokusgruppe erwähnte, dass sie es als hilfreich empfinden würde, wenn mithilfe einer künstlichen Intelligenz heikle Beiträge automatisch als solche gekennzeichnet und die Lehrkraft über das Dashboard auf diese Posts und Kommentare hingewiesen würde. Dadurch müsste die Lehrkraft diese Beiträge nur noch anschauen und entscheiden, ob sie diese markieren oder ausblenden möchte.

5 Diskussion

Das Lehrkräfte-Dashboard für das schulgerechte soziale Netzwerk *opostum* soll Lehrkräfte beim Einsatz im Unterricht unterstützen, indem es ihnen Informationen sowohl für unmittelbare Reaktionen im Unterricht als auch zur Reflexion zwischen Unterrichtssequenzen anbietet [Al10; Ma19; XAM17]. Die teilnehmenden Lehrkräfte hoben hervor, bei unangebrachten Inhalten schnell reagieren zu wollen. Der Nutzen einer übersichtlichen statistischen Auswertung wird vor allem für die Unterrichtsvorbereitung gesehen, um mehr und besseres Feedback geben zu können [VR20]. Die Ergebnisse der Erhebung bestätigten zudem frühere Erkenntnisse, dass Lehrkräfte soziale Interaktionen über ein Dashboard verfolgen können möchten [Ve20]. Beim vorgestellten Dashboard akzentuierte sich dies besonders bei der Klassenübersicht mit Informationen zu beliebten Posts, Kommentaren und Follower-Angaben. Anders als bei Lernprogrammen wie Mathematik-Tutoren geht es im sozialen Netzwerk weniger um die Priorisierung von Hilfsangeboten [Sc17; Ve20; Va21], sondern eher um die Überwachung von Aktivitäten und um ein schnelles Eingreifen bei problematischen Inhalten. Die Lehrkräfte wünschten sich einen Live-Feed mit spiegelnder (mirroring) Funktion [Va19] und Alarmfunktionen (alerting) bei heiklen Inhalten. Es bestätigte sich, dass es den Lehrkräften ein großes Anliegen ist, selbst die Kontrolle über den Unterricht behalten zu können ([Al10], [VR22]). Die Ergebnisse bestätigen die Wichtigkeit einer benutzendenfreundlichen Gestaltung ohne Informationsüberflutung [Sc17]. Dies zeigte sich in den bevorzugten visuellen Darstellungen wie Wortwolken und bekannten Diagrammtypen. Übereinstimmend mit [Al22] wünschten sich die Befragten mehrere Dashboard-Bereiche: einen Klassenüberblick, Details zu einzelnen Profilen, statistische Auswertungen und einen Unterrichtsbereich. Die Teilnehmenden bevorzugten anpassbare Dashboards mit ein- und ausklappbaren Bereichen [Ma19] und die Möglichkeit, Schülernamen auszublenden [Al22]. Zur Steigerung der Akzeptanz bei den Lehrkräften wurde die Bedeutung einer Dashboard-Einführung [Go23; MK19] hervorgehoben.

Basierend auf der ausgeführten Diskussion wurde ein weiterer, finaler Prototyp für ein Lehrkräfte-Dashboard für *opostum* entwickelt. Abbildung 5 zeigt den Entwurf für das Dashboard, auf dem die relevanten Statistiken zur Klasse dargestellt werden.

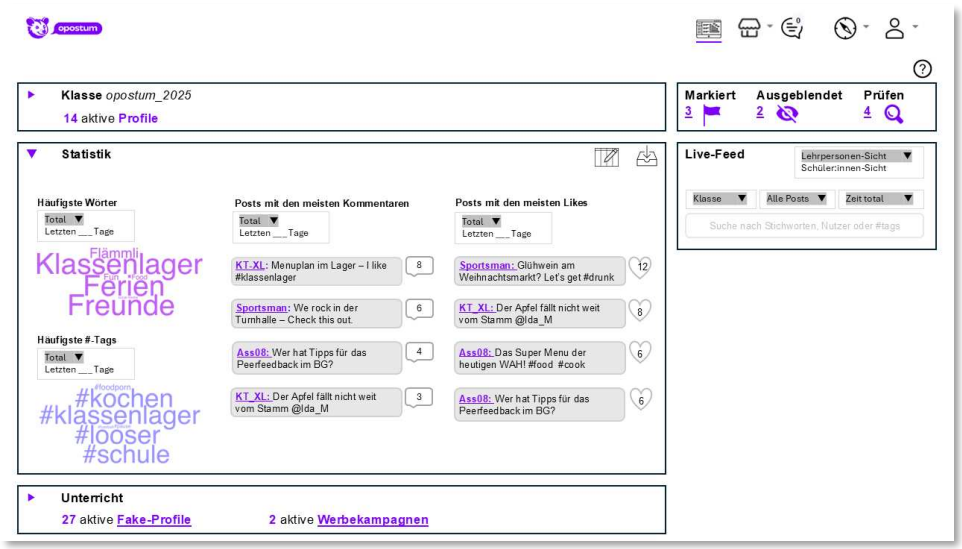


Abb. 5: Finaler Prototyp mit Statistiken für die Lehrkraft

6 Fazit, Limitationen und Ausblick

Die im Rahmen dieser Arbeit entstandenen Ergebnisse ermöglichten es, ein Lehrkräfte-Dashboard für ein schulgerechtes soziales Netzwerk zu entwickeln und eine Empfehlung für *opostum* herauszuarbeiten, die den Erwartungen und Bedürfnissen der Lehrkräfte gerecht werden könnte. Die Stichprobe war mit sechs Lehrkräften in der Fokusgruppendiskussion und mit neun Lehrkräften in den Einzelinterviews zwar ausreichend, aber eher klein. Zusätzlich handelte es sich bei den beteiligten Lehrkräften um digital affine Personen, die bereits aus eigenem Interesse mit *opostum* in Kontakt kamen. Mit weiteren iterativen Testphasen sollten auch Lehrkräfte in die Dashboard-Entwicklung involviert werden, die gegenüber digitalen Lernumgebungen zurückhaltender und weniger vertraut sind. Für die Weiterentwicklung des Dashboards werden Erhebungen zur Nutzung des realen Dashboards eine wichtige Rolle einnehmen. Vor allem betreffend den Umgang und die Nutzung der Daten und Informationen des Dashboards im Unterricht wird weiteres Forschungspotenzial gesehen.

Literaturverzeichnis

[Al10] Aleven, V. et al.: Developing a teacher dashboard for use with intelligent tutoring systems. technology 34/4, 2010.

[Al22] Aleven, V. et al.: A dashboard to support teachers during students' self-paced AI-supported problem-solving practice. In (Hilliger, I. et al., Hrsg.): Educating for a new future: Making sense of technology-enhanced learning adoption, Springer, 13450, S. 16–30, 2022.

[Ca19] Calvo-Morata, A. et al.: Improving teacher game learning analytics dashboards through ad-hoc development. Verlag der Technischen Universität Graz, 2019.

[De24] Department of Infrastructure, Transport, Regional Development, Communication and the Arts: Social media reforms to protect our kids online pass Parliament. Australian Government, 2024.

[Do19] Dorn, J.: InstaHub: Datenbanken und Datenschutz mit einem extra für den Unterricht entwickelten sozialen Netzwerk unterrichten. In: Informatik für alle, 2019.

[Fe06] Few, S.: Information dashboard design: The effective visual communication of data. O'Reilly, 2006.

[Ge17] Gerstbach, I.: Design Thinking im Unternehmen: Ein Workbook für die Einführung von Design Thinking. GABAL, 2017.

[GL09] Glaser, J.; Laudel, G.: Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen. VS Verlag, 2009.

[Go23] Gobert, J. D. et al.: An AI-based teacher dashboard to support students' inquiry. In (Lederman, N. G. et al., Hrsg.): Lederman, Handbook of research on science education. Routledge, S. 1011–1044, 2023.

[GTF20] Graves, M., Tschepe, S., & Fuchs, J.: Coaching Manual: HPI certification program for design thinking coaches. Hasso Plattner Institut, 2020.

[Gr14] Greenberg, S. et al.: Sketching user experiences. Mitp, 2014.

[He11] Helfferich, C.: Die Qualität qualitativer Daten: Manual für die Durchführung qualitativer Interviews. VS Verlag, 2011.

[Kh24] Khan, A.: Visual Analytics for Dashboards: A step-by-step guide to principles and practical techniques. Apress, 2024.

[Kü24] Külling-Knecht, C. et al: JAMES – Jugend, Aktivitäten, Medien – Erhebung Schweiz, 2024.

[Ma19] Mavrikis, M. et al.: Intelligent analysis and data visualisation for teacher assistance tools: The case of exploratory learning. British Journal of Educational Technology, 50(6), S. 2920–2942, 2019.

[Ma23] Mayring, P.: Einführung in die qualitative Sozialforschung: Eine Anleitung zu qualitativem Denken. Beltz, 2023.

[Me24] Medienpädagogischer Forschungsverbund Südwest: JIM 2024 - Jugend, Information, Medien, 2024.

[MMM23] Mohseni, Z.; Masiello, I.; Martins, R. M.: (2023). Co-developing an easy-to-use learning analytics dashboard for teachers in primary/secondary education: A human-bentered design approach. Education Sciences, 13(12), 2023.

[MK19] Molenaar, I.; Knoop-van Campen, C. A. N.: How teachers make dashboard information actionable. IEEE Transactions on Learning Technologies, 12(3), S. 347–355, 2019.

[Sc17] Schwendimann, B. A. et al.: Perceiving learning at a glance: A systematic literature review of learning dashboard research. IEEE Transactions on Learning Technologies, 10(1), S. 30–41, 2017.

[SMR12] Schulz, M.; Mack, B.; Renn, O.: Fokusgruppen in der empirischen Sozialwissenschaft. VS Verlag, 2012.

[Va19] Van Leeuwen, A. et al.: What information should CSCL teacher dashboards provide to help teachers interpret CSCL situations? In International Journal of Computer-Supported Collaborative Learning, 14(3), S. 261–289, 2019.

[VR19] Van Leeuwen, A.; Rummel, N.: Orchestration tools to support the teacher during student collaboration: A review. Unterrichtswissenschaft, 47(2), S. 143–158, 2019.

[VR20] Van Leeuwen, A.; Rummel, N.: Comparing teachers' use of mirroring and advising dashboards. Proceedings of the tenth international conference on learning analytics & knowledge, S. 26–34, 2020.

[Va21] Van Leeuwen, A. et al.: How teacher characteristics relate to how teachers use dashboards: Results from two case studies in K-12. Journal of learning analytics, 8(2), S. 6–21, 2021.

[VR22] Van Leeuwen, A.; Rummel, N. (2022). The function of teacher dashboards depends on the amount of time pressure in the classroom situation: Results from teacher interviews and an experimental study. Unterrichtswissenschaft, 50(4), S. 561–588, 2022.

[Ve20] Verbert, K. et al.: Learning analytics dashboards: The past, the present and the future. Proceedings of the tenth international conference on learning analytics & knowledge, S. 35–40, 2020.

[XAM17] Xhakaj, F.; Aleven, V.; McLaren, B. M.: Effects of a teacher dashboard for an intelligent tutoring system on teacher knowledge, lesson Planning, lessons and student learning. In (Lavoué, E. et al., Hrsg.): Data driven approaches in digital education. Springer, Bd. 10474, S. 315–329, 2017.

Evaluierung einer gamifizierten Lern-App: Usability, User Experience und Lerneffekt in der Hochschulbildung

Josefine Welk [1], Christian Roschke [1], Dominik Breck[1], Max Schlosser [1] und Marc Ritter [1]

Abstract: Gamification-Strategien können motivationssteigernd wirken. Die App StudiSQ nutzt diesen Ansatz, indem sie Lehrinhalte als Quiz vermittelt und Gamification-Elemente, wie Ranglisten und Abzeichen, zur Förderung der Motivation einsetzt. Im Rahmen einer Evaluation wurden ca. 60 Studierende hinsichtlich Usability und User Experience befragt; der Lerneffekt wurde mittels zweier Tests evaluiert. Zur Vorbereitung auf den zweiten Test nutzte eine Gruppe StudiSQ, die andere nicht. Die Usability (65,9/100) und User Experience (0,61/–3 bis 3) wurden durchschnittlich bis leicht positiv bewertet. Im Durchschnitt verschlechterten sich die Testergebnisse aller Teilnehmenden; Studierende, die StudiSQ nutzten, schnitten etwas besser ab (–0,36 vs. –0,63 Punkte). Im Rahmen der Evaluation konnte kein statistisch signifikanter Unterschied zwischen den beiden Testgruppen festgestellt werden. Entsprechende Maßnahmen zur Konzeptionierung und Durchführung von Folgeuntersuchungen wurden diskutiert. Um eine Nutzung durch Studierende und Lehrkräfte zu fördern, sind außerdem Fehlerbehebungen sowie die Verbesserung der Usability notwendig. Dennoch würden ca. 90 % der befragten Studierenden die App wiederverwenden.

Keywords: Evaluation, Usability, User Experience, Lerneffekt, Mobile Learning

1 Einleitung

Gamification wird in der Hochschullehre zunehmend als Mittel zur Steigerung von Motivation und Lernerfolg eingesetzt. Studien zeigen, dass spielerische Elemente wie Fortschrittsanzeigen, Feedback, Abzeichen oder Ranglisten positive Effekte auf die Lernmotivation und das Engagement Studierender haben können [AAS25; GB19; MB23]. Dabei wirken besonders Kombinationen dieser Elemente – z. B. Feedback mit Fortschrittsbalken – motivationsfördernd. Auch mobile Lernanwendungen gelten als vielversprechend, da sie orts- und zeitunabhängiges Lernen unterstützen und sich flexibel in Lernprozesse integrieren lassen [Cr21; Pe24]. Studien zeigen darüber hinaus, dass der positive Einfluss von Gamification sich auch auf den Lernerfolg übertragen lässt [Nu21; SH20], auch wenn die Faktoren, die eine erfolgreiche Gamifizierung ermöglichen, vielfältig und noch ungeklärt sind.

Zugleich wird jedoch auch Kritik am Einsatz von Gamification laut. So betonen Sailer et al. [Sa17], dass Gamification allein nicht automatisch zu besseren Lernergebnissen führt,

[1] Hochschule Mittweida, Institute for Computer Science and Media in Research and Transfer, Technikumplatz 17, 09648 Mittweida, Deutschland, koepke@hs-mittweida.de, https://orcid.org/0000-0002-0160-0505; roschke@hs-mittweida.de, https://orcid.org/0009-0007-2875-5199; breck@hs-mittweida.de; schlosse@hs-mittweida.de, https://orcid.org/0009-0005-2430-9891; ritter@hs-mittweida.de, https://orcid.org/0009-0004-0204-8275

doi:10.18420/delfi2025_15

sondern gezielt und didaktisch durchdacht eingesetzt werden muss. Bai et al. [BHH20] weisen auf mögliche Nebenwirkungen wie Ängste oder Eifersucht hin, denen jedoch positive Effekte gegenüberstehen. So kann Gamification Begeisterung wecken, leistungsbezogenes Feedback darstellen, das Bedürfnis nach Aufmerksamkeit befriedigen und das Setzen eigener Ziele unterstützen.

Wesentlich für die Akzeptanz und Wirksamkeit digitaler Lernanwendungen sind zudem Usability und User Experience (UX). Studien zeigen, dass mangelnde Bedienbarkeit oder negative Nutzungserlebnisse zu seltener Nutzung und geringem Lernerfolg führen können [Is20; Mo23].

Vor diesem Hintergrund wurde die App StudiSQ entwickelt – eine mobile Lernanwendung, die Quizformate mit Gamification-Elementen wie Ranglisten, Abzeichen und Duellen kombiniert. Ziel ist es, eine motivierende und didaktisch wirksame Lernumgebung zu schaffen. Aufbauend auf einer ersten Untersuchung zur Steigerung der Motivation [Sc23] liegt der Fokus dieses Beitrags auf der Evaluation der App im Hinblick auf Usability, User Experience und Lernerfolg. Die Evaluation erfolgte im Rahmen eines Bachelor-Moduls mit Studierenden der Medieninformatik und basiert auf einem mehrstufigen Untersuchungsdesign. Ziel ist es, Potenziale und Optimierungsmöglichkeiten der App aufzuzeigen und evidenzbasierte Handlungsempfehlungen für ihre Weiterentwicklung abzuleiten. Die Methode wird im folgenden Abschnitt beschrieben. In den Abschnitten 3 und 4 werden die Ergebnisse dargestellt und diskutiert. Abschnitt 5 fasst zusammen.

2 Methode

Nachfolgend wird die für die Evaluation der App StudiSQ angewandte Methodik erläutert. Dafür werden zuerst die Rahmenbedingungen dargelegt, anschließend werden die verschiedenen Phasen der Evaluation sowie deren Auswertung beschrieben.

2.1 Rahmenbedingungen

Die Zielgruppe der Evaluation waren Studierende des Bachelor-Studiengangs *Medieninformatik und Interaktives Entertainment*, speziell Teilnehmende des Moduls *Mensch-Maschine-Interaktion*. Im Studienablaufplan ist dieses Modul im ersten Semester verortet, sodass der Fokus der Evaluation auf Studienanfängern lag; aber auch Studierende höherer Semester, die das Modul nach- oder wiederholten, konnten an der Evaluation teilnehmen. Durchgeführt wurde die Evaluation über einen Zeitraum von 1,5 Monaten, die Evaluation umfasste dabei sowohl asynchrone Phasen als auch eine Präsenzphase. Insgesamt haben 58 Studierende die Evaluation vollständig durchlaufen, wobei bei der Evaluation der Usability und User Experience ein Studierender mehr teilgenommen hat als an der Evaluation des Lernerfolgs. Abbildung 2.2 zeigt die Phasen der Evaluation.

2.2 Ablauf der Evaluation

Die Evaluation umfasste vier verschiedene, aufeinander aufbauende Phasen, welche in Abb. 1 dargestellt sind.

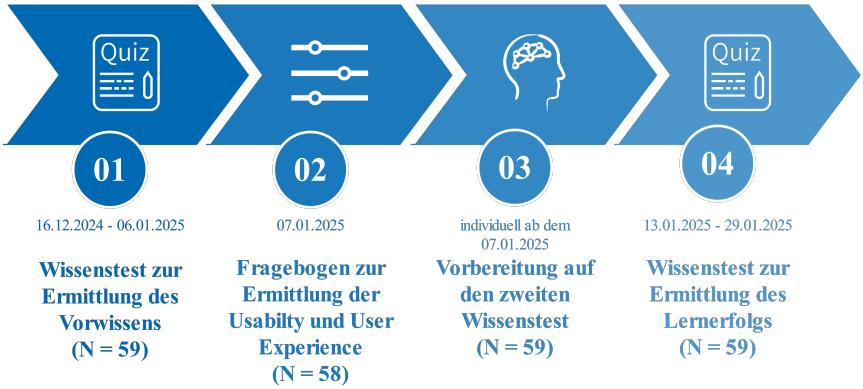


Abb. 1: Ablauf der Evaluation bestehend aus drei Phasen zur Ermittlung des Lernerfolgs (Wissenstest – Vorbereitung – Wissenstest) und der Phase zur Ermittlung der Usability und User Experience (Fragebogen). Für jede Phase sind der Zeitraum der Durchführung sowie die Anzahl an Teilnehmern angegeben.

Nachfolgend werden die vier Phasen der Evaluation im Detail vorgestellt.

Ermitteln des Vorwissens durch den ersten Wissenstest Zur Ermittlung des Vorwissens absolvierten alle Studierende einen Wissenstest mit dem Schwerpunkt der grundlegenden Programmierung mittels der Programmiersprache Java. Diese Programmiersprache kommt im Modul *Einführung in die Informatik* des Studiengangs *Medieninformatik und Interaktives Entertainment* zum Einsatz, weshalb alle Teilnehmenden über Grundkenntnisse der Sprache verfügen. Für die Bearbeitung des Tests standen den Studierenden drei Wochen zur Verfügung, in denen Sie den Test zeitlich flexibel bearbeiten konnten. Der Test umfasste 13 Aufgaben; maximal konnten 25 Punkte erreicht werden. Als Aufgabentypen wurden Auswahlaufgaben, Zuordnungsaufgaben, Reihenfolgeaufgaben sowie Lückentextaufgaben gewählt. Tabelle 1 gibt einen Überblick über die abgefragten Themenkomplexe und listet jeweils die Anzahl der Aufgaben in dieser Kategorie, eine Beispielaufgabe sowie deren Aufgabentyp.

Evaluation der Usability und User Experience Die Usability und User Experience der App StudiSQ wurden im Rahmen einer Präsenzveranstaltung des Moduls *Mensch-Maschine-Interaktion* evaluiert. Dafür wurde den Studierenden ein Fragebogen zur Verfügung gestellt. Dieser umfasste neben allgemeinen Angaben zur Person (Altersgruppe, Geschlecht) und Angaben zur Testumgebung zu bearbeitende Aufgaben, für die die Studierenden bewerten

Kategorie	N	Beispielaufgabe	Aufgabentyp
Zahlen-systeme	1	Ordnen Sie die gegebenen Zahlen den entsprechenden Zahlensystemen zu.	Zuordnung
Variablen	2	Sortieren Sie die nachfolgenden Daten-typen zur Darstellung von Ganzzahlen aufsteigend nach ihrem Wertebereich.	Reihenfolge
Programm-strukturen	5	Welche der nachfolgenden Aussagen über Konstruktoren sind wahr?	Multiple Choice
Speicher-struktur Array	3	Gegeben sei ein Array mit dem Namen array und einer Länge von 10. Geben Sie den Index an, mit dem das letzte Feld des Arrays angesprochen werden kann.	Textaufgabe
Objektorien-tierung	2	Wählen Sie aus den nachfolgenden Optionen die wesentlichen Eigenschaften der Objektorientierung aus.	Multiple Choice

Tab. 1: Inhalte des ersten Wissenstests unterteilt nach Kategorien. Für jede Kategorie ist die Anzahl an Fragen je Kategorie und eine zugehörige Beispielaufgabe inkl. ihres Aufgabentyps aufgelistet.

sollten, wie intuitiv diese lösbar waren (Likert-Skala von 1 bis 5), sowie Aussagen zur Ermittlung der Usability auf Basis der System Usability Scale [Br95]. Für jede Aussage wurden die Studierenden gebeten, anzugeben, in welchem Maß sie der Aussage zustimmen (1 – Stimme überhaupt nicht zu, 5 – Stimme voll und ganz zu). Zur Ermittlung der User Experience erhielten die Studierenden Wortpaare in Anlehnung an Laugwitz et al. [LHS08]. Verwendet wurden dafür die Wortpaare der pragmatischen Qualität sowie Attraktivität. Im letzten Bereich des Fragebogens beantworteten die Studierenden Fragen zur Bewertung des Gesamteindrucks (Bewertung in Sternen, Interesse an Nutzung der App, Allgemeines Feedback). Für die Bearbeitung der Aufgabe und das Ausfüllen des Fragebogens standen den Studierenden im Rahmen der Lehrveranstaltung 180 Minuten zur Verfügung. Zusätzlich wurden die Studierenden gebeten, auftretende Bugs im Rahmen des Fragebogens zu notieren.

Vorbereitung auf den zweiten Wissenstest Vor der Bearbeitung des zweiten Wissenstest erfolgte eine individuelle Vorbereitung durch die Studierenden; die Mindestzeit zur Vorbereitung auf den Test waren dabei 6 Tage, maximal war eine Vorbereitungszeit von 22 Tagen möglich, begrenzt wurde der Zeitraum durch die Prüfung im Modul *Einführung in die Informatik*. Für die Vorbereitung auf den zweiten Wissenstest erfolgte eine Aufteilung der Studierenden in zwei Gruppen. Dafür wurde eine teil-randomisierte Gruppenzuweisung durchgeführt, indem die Studierenden unter der Bedingung, dass möglichst zwei gleichstarke Gruppen entstehen, zufällig auf die beiden Gruppen aufgeteilt wurden. Unter gleichstark ist in diesem Kontext zu verstehen, dass die Gruppen zum einen ähnlich groß sind, zum anderen aber auch jeweils das gesamte Leistungsspektrum der Studierenden abdecken. Durch die teilrandomisierte Einteilung konnten zwei Gruppen erstellt werden, in denen jeweils sowohl

sehr leistungsstarke als auch leistungsschwächere Studierende vertreten waren. Jedem Studierenden wurde individuell mitgeteilt, ob die App StudiSQ für die Vorbereitung auf den zweiten Test verwendet werden durfte. Studierenden, die StudiSQ verwenden durften, stand in der App ein Quiz zum Thema Java-Programmierung zur Verfügung. Dieses umfasste 22 Fragen, wobei die App StudiSQ nur Fragen des Typs Auswahlfragen und Zuordnungsfragen bereitstellt. Die inhaltlichen Schwerpunkte orientierten sich an den in Tabelle 1 aufgeführten Themenkategorien. Dabei umfasste der Test sowohl Aufgaben aus dem ersten Wissenstest als auch weitere Aufgaben, die die Themenkomplexe sinnvoll erweiterten und somit eine breitere Vorbereitung im Hinblick auf den zweiten Wissenstest ermöglichten. Studierende, die nicht auf die App zurückgreifen durften, wurden darum gebeten sich mit herkömmlichen Methoden vorzubereiten, ohne diese spezifisch zu benennen.

Ermitteln des Lernerfolgs durch den zweiten Wissenstest Analog zum in der ersten Phase durchgeführten Wissenstest zur Erhebung des Vorwissens wurde in der vierten Phase ein zweiter Wissenstest durchgeführt, der zur Erhebung des Lernerfolges diente. Auch in dieser Phase konnten die Studierenden den Test innerhalb eines einwöchigen Zeitraums zeitlich flexibel bearbeiten. Der thematische Schwerpunkt der Aufgaben lag ebenfalls auf Grundlagen der Java-Programmierung. Analog zum ersten Test umfasste der zweite Wissenstest ebenfalls 13 Aufgaben mit insgesamt 25 Punkten. Auch die Aufgabentypen und Kategorien entsprachen denen des ersten Wissenstests, die Aufgaben selbst wurden teilweise in ihren Antwortmöglichkeiten modifiziert bzw. vollständig ausgetauscht, um der Wiedergabe auswendig gelernter Antworten entgegenzuwirken.

Kategorie	**N**	**Beispielaufgabe**	**Aufgabentyp**
Zahlen-systeme	1	Ordnen Sie die gegebenen Zahlen den entsprechenden Zahlensystemen zu.	Zuordnung
Variablen	3	Wie viel Byte belegt eine Variable vom Typ int im Speicher?	Textaufgabe
Programm-strukturen	5	Welche Aussagen über die do-While-Schleife sind wahr?	Multiple Choice
Speicher-struktur Array	1	Welche Aussagen über Arrays sind wahr?	Multiple Choice
Objektorien-tierung	3	Wie lautet das Schlüsselwort für die Vererbung?	Textaufgabe

Tab. 2: Inhalte des zweiten Wissenstests unterteilt nach Kategorien sowie die Anzahl an Fragen je Kategorie, zugehörige Beispielaufgaben inkl. Aufgabentyp für die jeweiligen Kategorien

2.3 Auswertung

Zur Ermittlung des Lernerfolgs erfolgte die Auswertung der im ersten und zweiten Wissenstests erzielten Punkte, eine Unterscheidung dieser erfolgte hinsichtlich der Verwendung der

App StudiSQ. Die Auswertung der Usability erfolgte auf Basis der durch Brooke [Br95] beschriebenen Methode zur Auswertung der System Usability Scale. Die für die Ermittlung der User Experience erhobenen Werte wurden in Anlehnung an Laugwitz et al. [LHS08] normiert. Sowohl für die Usability als auch für die User Experience erfolgte die Ermittlung statistischer Maße sowie ein Vergleich dieser mit Hinblick auf die von den Studierenden verwendete Testumgebung. Des Weiteren wurde die Intuitivität bei der Bearbeitung der Aufgaben sowie der Gesamteindruck der App StudiSQ ausgewertet.

3 Ergebnisse

Nachfolgend werden die Ergebnisse der Evaluation vorgestellt. Dabei wird zwischen den Ergebnissen hinsichtlich des Lernerfolgs, der Usability und User Experience sowie der allgemeinen Bewertung der App StudiSQ unterschieden.

3.1 Lernerfolg

In Bezug auf den Lernerfolg verschlechterten sich im Durchschnitt alle Studierenden minimal (20,9 vs. 20,4 Punkte). Die Veränderung der Punkte schwankte dabei zwischen den Extremwerten -10 und 6. Studierende, die StudiSQ zur Vorbereitung auf den zweiten Test verwendeten, verschlechterten sich im Durchschnitt um -0,36 Punkte. Die Gruppe, die sich mit herkömmlichen Lehrunterlagen vorbereitete, verschlechterte sich um -0,63 Punkte. Auch die Extremwerte, zwischen denen die Punkteveränderungen schwankten, unterschieden sich im Hinblick auf die App-Nutzung. Die Punkteveränderungen der Studierenden, die StudiSQ nutzten, schwankten zwischen den Extremwerten -6,5 und 6, die der Studierenden, die die App nicht verwendeten, zwischen den Extremwerten -10 und 4,5. Abb. 2 zeigt die Ergebnisse des ersten und zweiten Wissenstests unterschieden nach Nutzung der App StudiSQ. Erkennbar ist, dass die Streuung der Ergebnisse des zweiten Wissenstests bei Nutzung der App deutlich geringer ausfällt als bei Nichtnutzung der App. Auch verlagerte sich die untere Grenze der Ergebnisse unter Verwendung von StudiSQ im Vergleich zur unteren Grenze der Ergebnisse ohne Nutzung der App stärker nach oben.

Zur weiteren Untersuchung der Ergebnisse wurde Cohen's d zur Ermittlung der Effektstärke berechnet. Darüber hinaus erfolgte eine Untersuchung der statistischen Signifikanz der Ergebnisse mittels eines t-Tests sowie den gegenüber einer hohen Streuung stabileren Mann-Whitney-U-Tests mit anschließender Bestimmung des zweiseitigen p-Werts. Tabelle 3 zeigt die Ergebnisse der statistischen Auswertungen.

3.2 Usability und User Experience

Sowohl die Bewertung der Usability als auch die der User Experience brachte durchschnittliche bis leicht positive Ergebnisse hervor. Die Usability wurde im Durchschnitt mit 65,9 von

100 möglichen Punkten bewertet (σ = 18,3). Hinsichtlich der verwendeten Testumgebungen konnte festgestellt werden, dass die App StudiSQ unter Verwendung von Android-Geräten (70,7 % der Testumgebungen) leicht besser bewertet wurde (67/100, σ = 19,0), während die App bei der Verwendung von iOS-Geräten (29,3 % der Testumgebungen) im Vergleich zum Gesamtergebnis leicht unterdurchschnittlich abschnitt (63,4/100, σ = 16,0). Hinsichtlich der User Experience vergaben die Studierenden im Durchschnitt eine Wertung von 0,61 (σ = 0,27). Auf der zugrundeliegenden Skala von -3 bis 3 stellt dies eine durchschnittliche bis leicht positive Wertung dar. Auch hier konnten wieder Unterschiede zwischen Android- und iOS-Geräten festgestellt werden. Analog zur Usability erzielte die App StudiSQ bei der Verwendung von Android-Geräten eine leicht überdurchschnittliche Wertung von 0,68 (σ = 0,29) während die Wertung bei Verwendung von iOS-Geräten mit einer Wertung von 0,51 (σ = 0,38) unter dem Gesamtdurchschnitt lag. Abb. 3 zeigt die Bewertung der einzelnen Wortpaare; erkennbar ist, dass die Bewertung aller Wortpaare im positiven Wertebereich angesiedelt ist, die Bewertungen einzelner Wortpaare (z. B. verwirrend – übersichtlich) aber nur einen geringen Unterschied zur neutralen Wertung aufweisen.

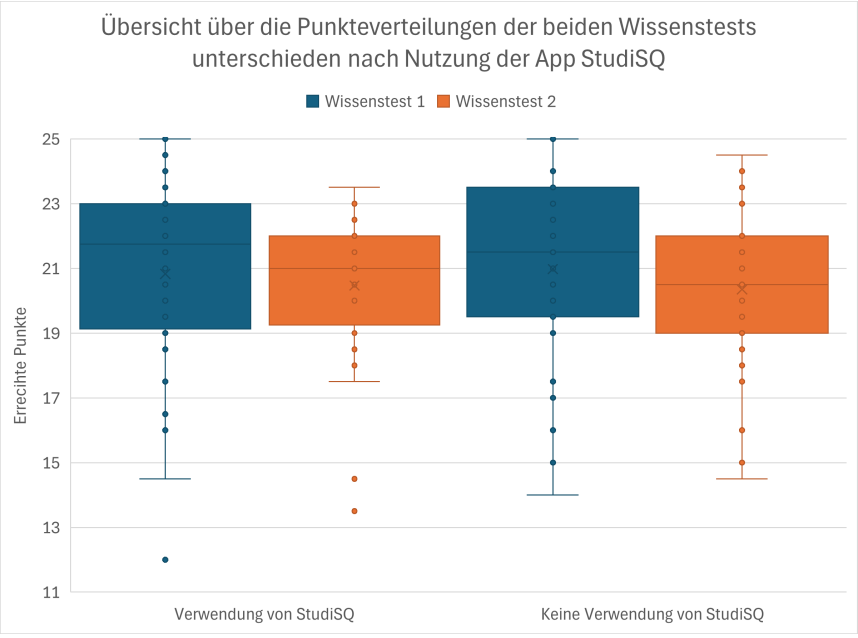


Abb. 2: Ergebnisse der beiden Wissenstests in Abhängigkeit von der Nutzung der App StudiSQ

Auch im Zusammenhang mit der Usability und User Experience erfolgte eine statistische Auswertung, welche Cohen's d, den t-Test sowie den Mann-Whitney-U-Test umfasste. Die Tabellen 4 und 5 zeigen die Ergebnisse der statistischen Berechnungen für die Usability bzw. User Experience.

	N	Durchschnittliche Punktedifferenz	Standard-abweichung	**Cohen's d**	**t-Test**	**U-Test**
Mit StudiSQ	32	-0,36	3,28	0,63	0,76	U = 485 z = 0,81
Ohne StudiSQ	27	-0,63	3,22			p > .05

Tab. 3: Ergebnisse der statistischen Untersuchungen hinsichtlich des Lernerfolgs unterschieden nach Verwendung der App StudiSQ

	N	Durchschnittliche Usability	Standard-abweichung	**Cohen's d**	**t-Test**	**U-Test**
Android	41	66,95	19,04	1,46	0,24	U = 483 z = 2,3
iOS	17	63,38	16,04			p > .05

Tab. 4: Ergebnisse der statistischen Untersuchungen hinsichtlich der Usability unterschieden nach dem verwendeten Betriebssystem

	N	Durchschnittliche User Experience	Standard-abweichung	**Cohen's d**	**t-Test**	**U-Test**
Android	41	0,68	0,95	1,36	0,53	U = 441 z = 1,58
iOS	17	0,51	0,89			p > .05

Tab. 5: Ergebnisse der statistischen Untersuchungen hinsichtlich der User Experience unterschieden nach dem verwendeten Betriebssystem

3.3 Auswertung der Aufgaben

Im Durchschnitt bewerteten die Studierenden die intuitive Lösbarkeit der Aufgaben mit 3,45 von 5 möglichen Punkten. Dabei gabe es Unterschiede zwischen den zu lösenden Aufgaben. Während die Aufgabe, seine eigenen Abzeichen einzusehen, beispielsweise im Durchschnitt mit einer 4,22 bewertet wurde, erhielten die Aufgaben, gegen einen Freund ein Duell zu spielen und seine eigenen Statistiken einzusehen mit 2,36 bzw. 2,86 Punkten die geringsten Wertungen. Tabelle 6 zeigt eine detaillierte Auflistung der Aufgaben sowie ihrer jeweiligen Wertungen inkl. der zugehörigen Standardabweichungen.

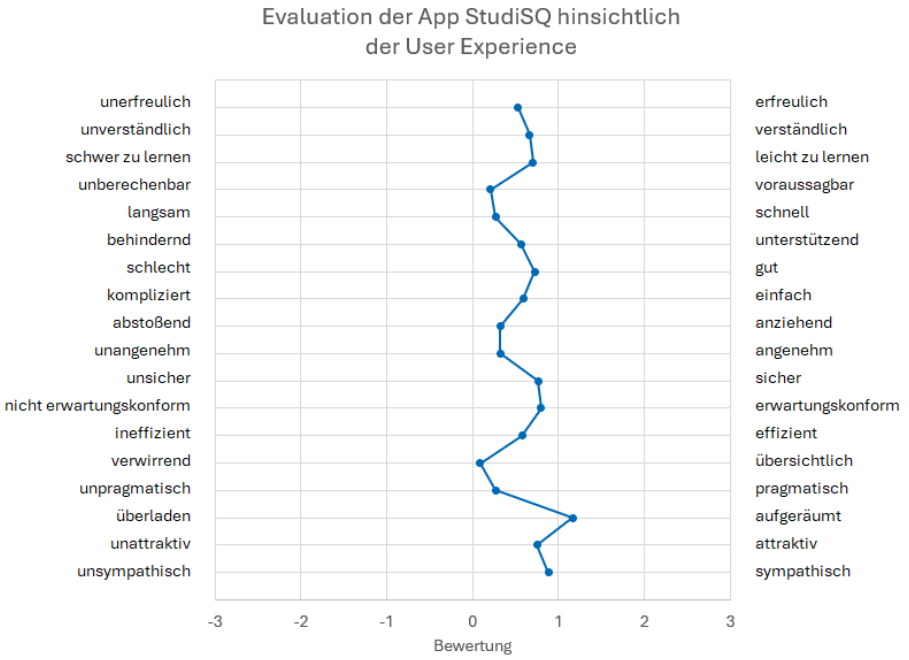


Abb. 3: Evaluation der App StudiSQ hinsichtlich der User Experience in Anlehnung an die von Laugwitz et al. [LHS08] definierten Wortpaare

3.4 Bewertung des Gesamteindrucks

Unter Berücksichtigung des prototypischen Status vergaben die Studierenden im Durchschnitt 3,29 von 5 möglichen Sternen für die App (σ = 0,83). Als Positiveindrücke wurden unter anderem die einfache Möglichkeit der Wissensüberprüfung, die Bandbreite der angebotenen Quizmodi, das grundsätzliche Design sowie die Übersichtlichkeit der App, die Möglichkeit, Erläuterungen zu falsch beantworteten Fragen zu erhalten sowie die Steigerung der Motivation durch das gemeinsame Spielen eines Quiz im Duellmodus genannt. Als Verbesserungsvorschläge führten die Studierenden unter anderem eine Verbesserung der Suchfunktion, das Einbringen eines Tutorials sowie die intuitivere Platzierung einiger Funktionen sowie generell die Behebung bestehender Bugs an. Fünf Studierende wünschten sich einen größeren Aufgabenpool sowie mehr Varianz bei der Auswahl der Fragen. Trotz einiger Kritikpunkte gaben 89,7 % der Befragten an, dass sie die App häufiger verwenden würden, sofern diese relevante Inhalte für ihr Studium bietet. Studierende, die die App nicht verwenden würden, gaben als Gründe zum einen den zu großen Umfang einzelner Quiz, die aktuell bestehenden Bugs, das Zeitlimit bei der Beantwortung der Fragen sowie die mangelnde Passung zwischen der didaktischen Gestaltung und Aufbereitung der Lerninhalte in der App und den persönlichen Lerngewohnheiten an.

Aufgabe	Durchschnittliche Bewertung	Standardabweichung
Quiz suchen und im Übungsmodus starten	3,83	0,81
Quiz über Katalog finden und im Überprüfungsmodus starten	3,83	0,95
Statistiken einsehen	2,86	1,28
Duell gegen einen Freund spielen	2,36	1,11
Quiz über QR-Code starten	3,22	1,25
Quiz über Raumcode starten	3,84	1,01
Abzeichen einsehen	4,22	1,13

Tab. 6: Übersicht über die im Rahmen der Evaluation zu erledigenden Aufgaben, deren zugehörigen durchschnittlichen Bewertungen sowie Standardabweichungen

4 Diskussion

Die in Abschnitt 3 dargestellten Resultate zeigen tendenziell positive Entwicklungen hinsichtlich des Lernerfolgs, der Usability und User Experience, der Bearbeitung der gestellten Aufgaben sowie des Gesamteindrucks. Die berechneten Effektstärken nach Cohen's d deuten auf einen mittleren Effekt im Bereich des Lernerfolgs sowie auf einen starken Effekt im Bereich der Usability und User Experience bei Nutzung unterschiedlicher Betriebssysteme hin. Allerdings erwiesen sich diese Effekte in den durchgeführten Signifikanztests (t-Test und Mann-Whitney-U-Test mit zweiseitigem p-Wert) als statistisch nicht signifikant. Daher sind die Ergebnisse als nicht generalisierbar zu betrachten und mit Vorsicht zu interpretieren. Nachfolgend sollen die Ergebnisse tiefer diskutiert und Ansätze zur Erhöhung der statistischen Signifikanz aufgezeigt werden. Vorab erfolgt eine Diskussion bezüglich möglicher Ursachen für den im Durchschnitt beobachtbaren Leistungsabfall der Studierenden im zweiten Wissenstest.

4.1 Diskussion des Leistungsabfalls

Es kann verschiedene Gründe für den Leistungsrückgang zwischen den beiden Wissenstests geben. Nachfolgend werden ein (subjektiv) abweichender Schwierigkeitsgrad, die Vorbereitung auf den zweiten Test sowie der Einfluss externer Störgrößen diskutiert.

Schwierigkeitsgrad der Aufgaben Eine mögliche Ursache für die durchschnittliche Verschlechterung könnte ein unterschiedlicher Schwierigkeitsgrad der beiden Tests sein. Trotz des identischen Kompetenzniveaus könnten bestimmte Themen im zweiten Wissenstest subjektiv als schwieriger empfunden worden sein, was schlechtere Leistungen erklären könnte. In Hinblick auf diesen Aspekt erfolgte eine genauere Untersuchung der Testergebnisse. Während 40,7 % der Teilnehmer sich um mehr als einen Punkt verschlechterten, konnten 23,7 % der Teilnehmer das Niveau des ersten Wissenstests halten (+/- 1 Punkt) und 35,6 % der Teilnehmer verbesserten sich um mehr als einen Punkt. Die geringe durchschnittliche

Punktabweichung (20,89 gegenüber 20,41) legt nahe, dass der Schwierigkeitsgrad vergleichbar war; dies schließt individuelle Schwierigkeiten mit dem zweiten Test allerdings nicht aus. Neben einem individuell wahrgenommenen Unterschied im Schwierigkeitsgrad kann auch ein ungenaues Lesen der Aufgaben eine Ursache für den Leistungsabfall sein. Einige der Aufgaben wurden beibehalten und variierten nur in ihren möglichen Antworten. Aufgrund eines ungenauen Lesens der Antwortoptionen haben die Teilnehmer möglicherweise falsche Lösungen gewählt, die in modifizierter Form im ersten Wissenstest noch richtig waren.

Vorbereitung Darüber hinaus kann die individuelle Vorbereitung der Studierenden sehr unterschiedlich gewesen sein. Obwohl die Dozentenansicht der StudiSQ-App zeigt, dass der Vorbereitungstest insgesamt 72 Mal absolviert wurde, lassen sich daraus keine Rückschlüsse auf die individuelle Vorbereitung der Studierenden ziehen. Der Umfang der Vorbereitung auf den zweiten Wissenstest ist nicht nachvollziehbar, insbesondere für Studierende, die die App nicht genutzt haben.

Einfluss externer Faktoren Zudem könnten individuelle sowie externe Störgrößen (z. B. Tagesform, Zeitpunkt des Tests, Umgebungsbedingungen) die Leistung beeinflusst haben. Diese wurden nicht erhoben und schränken die Validität ein.

4.2 Diskussion hinsichtlich der statistischen Signifikanz

Weder im Hinblick auf den Lernerfolg noch hinsichtlich der Usability und User Experience konnten statistisch signifikante Unterschiede zwischen den Testgruppen bzw. den eingesetzten Betriebssystemen festgestellt werden. Im Hinblick auf den Lernerfolg wäre ein statistisch signifikanter Unterschied zwischen den Testgruppen aber erforderlich, um die Effektivität der App valide zu untermauern. Nachfolgend werden mögliche Gründe für die ausbleibende statistische Signifikanz sowie Möglichkeiten zum Erreichen dieser diskutiert.

Größe der Stichprobe Eine Möglichkeit, in folgenden Untersuchungen eine statistische Signifikanz zu erreichen, ist die Vergrößerung der Stichprobe. An den Wissenstests nahmen insgesamt 59 Studierende teil, 32 Studierende verwendeten StudiSQ zur Vorbereitung, 27 Studierende nutzten herkömmliche Methoden zur Vorbereitung auf den zweiten Wissenstest. Die berechnete Effektstärke (Cohen's d = 0,63) deutet auf einen mittleren bis starken Effekt der Intervention hin. Eine Power-Analyse zeigt jedoch, dass für die statistische Absicherung eines solchen Effekts bei einem Signifikanzniveau von 5 % und einer Teststärke von 80 % eine Stichprobengröße von mindestens 41 Studierenden pro Testgruppe erforderlich gewesen wäre. Die tatsächlichen Gruppengrößen lagen unterhalb dieses Schwellenwerts, was die Aussagekraft der Signifikanztests einschränkt.

Rahmenbedingungen der Wissenstests Neben der Vergrößerung der Stichprobe sollte für künftige Untersuchungen ebenfalls die Standardisierung der Rahmenbedingungen bei der Durchführung des Tests in Betracht gezogen werden. Durch die Etablierung einer Prüfungssituation können gleichwertige Bedingungen geschaffen werden, die Abweichungen

in der Leistung der Studierenden auf Grund unvorhersehbarer Einflüsse ausschließen oder zumindest deren Dokumentation erlauben.

Rahmenbedingungen der individuellen Vorbereitung Auch die Rahmenbedingungen zur Vorbereitung auf den zweiten Test sollten klarer definiert werden. Im Rahmen der Evaluation erfolgte eine individuelle Vorbereitung der Studierenden, mit dem Ziel, möglichst realistische Lernszenarien abzubilden. Nachteil dieser Vorgehensweise ist hingegen die schlechte Abschätzbarkeit der tatsächlich geleisteten Vorbereitung. Zwar lässt sich aus der statistischen Erhebung der App StudiSQ ableiten, dass der Vorbereitungstest 72 mal durchgeführt wurde; dies lässt sich aber nicht auf die einzelnen Studierenden abbilden, sodass nicht überprüft werden kann, wie oft jeder Studierende das Vorbereitungsquiz durchgeführt hat. Weiterhin war der gewählte Vorbereitungszeitraum auf Grund der Lage der Evaluation zum Semesterende entsprechend kurz gewählt. Um eine Wirkung der App auf den Lernerfolg abschätzen zu können, wäre vermutlich ein längerer Zeitraum notwendig, in dem vermittelte Modulinhalte semesterbegleitend unter Verwendung der App wiederholt und gefestigt werden können und somit der Effekt auf den Lernerfolg der Studierenden untersucht werden kann.

5 Zusammenfassung

In dieser Arbeit wurde eine Evaluation der App StudiSQ mit den Schwerpunkten des Lernerfolgs, der Usability sowie User Experience vorgestellt. Insgesamt nahmen 58 Studierende des Studiengangs *Medieninformatik und Interaktives Entertainment* an der Evaluation teil. Obwohl die Evaluation der Usability und User Experience leicht positiv ausfiel und die Studierenden, die StudiSQ nutzten weniger schlechte Ergebnisse im zweiten Wissenstests zeigten, konnten keine statistisch signifikanten Ergebnisse erzielt werden. Im Rahmen der Diskussion wurden Ursachen diskutiert und Anpassungen für künftige Untersuchungen abgeleitet. Basierend auf den Erkenntnissen sollte eine weitere Evaluation mit dem Fokus auf den Lernerfolg der Studierenden durchgeführt werden, die einheitliche Rahmenbedingungen für den Wissenstest schafft, stärkeren Bezug auf die Nutzungshäufigkeit sowie -intensität der App StudiSQ nimmt, die Stichprobenzahl vergrößert und die Nutzung der App über einen größeren Zeitraum sicherstellt sowie didaktisch in einem Modul verankert werden. Eine solche praktische Anwendung würde außerdem die Evaluation der App aus Sicht der Dozierenden ermöglichen. Darüber hinaus sollte die App hinsichtlich der im Rahmen der Evaluation kritisierten Funktionalitäten überarbeitet und durch die Studierenden dokumentierte Bugs behoben werden. Trotz der im Kontext der Evaluation geäußerten Kritik gaben ca. 90 % der Studierenden an, die App aktiv einzusetzen, wenn für ihr Studium relevante Inhalte bereit gestellt werden, und lobten das zugrundeliegende Konzept.

Literaturverzeichnis

[AAS25] Alonso-Sánchez, J. A.; Alonso, J. L. N.; Santana-Monagas, E.: Gamification in Higher Education: A Case Study in Educational Sciences. en, TechTrends, 2025, DOI: 10.1007/s11528-025-01056-2.

[BHH20] Bai, S.; Hew, K. F.; Huang, B.: Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. Educational Research Review 30, S. 100322, 2020, DOI: 10.1016/j.edurev.2020.100322.

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

[Cr21] Criollo-C, S. et al.: Mobile Learning Technologies for Education: Benefits and Pending Issues. en, Applied Sciences 11 (9), Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, S. 4111, 2021, DOI: 10.3390/app11094111.

[GB19] Groening, C.; Binnewies, C.: "Achievement unlocked!" - The impact of digital achievements as a gamification element on motivation and performance. Computers in Human Behavior 97, S. 151–166, 2019, DOI: 10.1016/j.chb.2019.02.026.

[Is20] Ishaq, K. et al.: Usability and Design Issues of Mobile Assisted Language Learning Application. en, International Journal of Advanced Computer Science and Applications 11 (6), 2020, DOI: 10.14569/IJACSA.2020.0110611.

[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. Springer, Berlin, Heidelberg, S. 63–76, 2008, DOI: 10.1007/978-3-540-89350-9_6.

[MB23] Mazarakis, A.; Bräuer, P.: Gamification is Working, but Which One Exactly? Results from an Experiment with Four Game Design Elements. International Journal of Human–Computer Interaction 39 (3), S. 612–627, 2023, DOI: 10.1080/10447318.2022.2041909.

[Mo23] Mohtar, S. et al.: The usability evaluation on mobile learning apps with gamification for middle-aged women. en, Education and Information Technologies 28 (1), S. 1189–1210, 2023, DOI: 10.1007/s10639-022-11232-z.

[Nu21] Nurtanto, M. et al.: A Review of Gamification Impact on Student Behavioural and Learning Outcomes. en, International Journal of Interactive Mobile Technologies (iJIM) 15 (21), S. 22, 2021, DOI: 10.3991/ijim.v15i21.24381.

[Pe24] Pedraja-Rejas, L. et al.: Mobile Learning and Its Effect on Learning Outcomes and Critical Thinking: A Systematic Review. en, Applied Sciences 14 (19), Number: 19 Publisher: Multidisciplinary Digital Publishing Institute, S. 9105, 2024, DOI: 10.3390/app14199105.

[Sa17] Sailer, M. et al.: How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. Computers in Human Behavior 69, S. 371–380, 2017, DOI: 10.1016/j.chb.2016.12.033.

[Sc23] Schlosser, M. et al.: Evaluation of a Learning Application Designed to Improve Motivation and Engagement of Students in Higher Education. International Journal of Intelligent Computing Research 14, S. 1181–1186, 2023, DOI: 10.20533/ijicr.2042.4655.2023.0145.

[SH20] Sailer, M.; Homner, L.: The Gamification of Learning: a Meta-analysis. en, Educational Psychology Review 32 (1), S. 77–112, 2020, DOI: 10.1007/s10648-019-09498-w.

Fostering Structured Problem Solving in Machine Learning through Collaborative Gameplay

Clemens Witt [1], Nico Reithmeier [1], Thiemo Leonhardt [2], and Mareen Grillenberger [1]

Abstract: To effectively grasp and apply key machine learning concepts, students benefit from learning environments that actively support structured problem-solving strategies. Prior research indicates that learners often rely on exploratory approaches when constructing decision trees, with limited use of systematic problem-solving strategies. This paper examines the design and implementation of collaborative features in a digital learning game focused on developing decision trees as classification models. Building on earlier findings, a multiplayer mode was developed to enhance social interaction through targeted game mechanics, with the aim of fostering the development and persistence of structured problem-solving. Results from an initial pilot study suggest that the multiplayer mode facilitates strategic peer discourse, promoting both collaborative engagement and more systematic problem-solving behavior. These findings highlight the potential of collaborative, game-based learning environments to support effective strategy use in machine learning education.

Keywords: Machine Learning Education, Structured Problem Solving, Collaborative Learning, Digital Learning Game, Multiplayer Learning Game

1 Motivation

Systematic problem-solving is considered a core competence in computer science, especially essential for understanding machine learning (ML) concepts. However, studies consistently show that students – especially novices – often rely on unsystematic trial-and-error approaches rather than following structured and target-oriented strategies [WF22]. This tendency is also evident in ML-related learning settings, where exploratory behavior dominates despite a clear correlation between structured approaches and higher task success [Wi25]. Digital educational games provide engaging, interactive environments to foster cognitive skills. In particular, collaborative learning games designed for mutual problem-solving can promote strategic thinking through social interaction [LSR19; WK16]. Prior research highlights that game-based collaboration can activate beneficial learning processes such as mutual explanation, coordination, and reflection. Yet, there is a lack of systematic investigation into how collaborative educational games specifically support the development and maintenance of structured problem-solving strategies – especially within the context of ML.

[1] TUD Dresden University of Technology,
clemens.witt@tu-dresden.de, https://orcid.org/0009-0005-8160-4029;
nico.reithmeier@gmx.de, https://orcid.org/0009-0001-4211-7624;
mareen.grillenberger@tu-dresden.de, https://orcid.org/0000-0002-8477-1464
[2] RWTH Aachen University, leonhardt@informatik.rwth-aachen.de, https://orcid.org/0000-0003-4725-9776

doi:10.18420/delfi2025_16

This study addresses the following research question: *To what extent can a collaborative educational game support learners in applying and maintaining structured, target-oriented problem-solving strategies in the domain of ML?*

We explore which collaborative mechanisms and didactic design elements in a multiplayer ML game foster such behavior, and how these are reflected in learner interactions. For this purpose, a single-player ML puzzle game was extended for collaborative play and piloted in a workshop setting. The evaluation includes a qualitative analysis of group discourse and problem-solving strategies.

2 Theoretical Background

Effective instructional design requires a balance between cognitive load and task complexity. According to **Cognitive Load Theory (CLT)**, learning materials should be structured to avoid overloading working memory [SMP98]. In collaborative learning, the *collective working memory effect* extends CLT by suggesting that cognitive load can be distributed across group members when solving complex tasks [Ki18]. However, the benefits from load sharing may be offset by coordination demands. When coordination costs exceed cognitive savings, collaboration may hinder performance. Thus, collaborative systems should aim to leverage distributed cognition while minimizing coordination overhead. Game mechanics such as clearly defined rules, explicit role distribution, and visual cues can help reduce this transactional load.

Problem-Based Learning (PBL) advocates active, self-directed engagement with complex, authentic problems [Ba86]. In contrast to instructionist models, PBL fosters learner agency through open-ended assignments. This approach aligns with educational games that embed non-trivial challenges requiring exploration, decision-making, and strategy discussions.

In the context of **Computer-Supported Collaborative Learning (CSCL)**, Johnson [Jo15] identifies five key elements essential to successful collaboration: *positive interdependence*, *individual accountability*, *promotive interaction*, *collaborative skills*, and *group reflection*. These elements served as guiding principles in the design of the game-based environment developed for this study.

From the perspective of educational psychology, cognitive learning strategies are defined as mental procedures used to process and apply information [Po45]. Such strategies often need to be explicitly taught and practiced, particularly when students frequently default to intuitive, exploratory behaviors. In our preliminary study, two distinct patterns were observed: *exploratory strategies* (e. g., random trial of game elements) and *structured strategies* (e. g., hypothesis-driven planning and systematic decision-tree construction). Structured approaches were consistently associated with higher rates of successful game completion [Wi25].



Fig. 1: Multiplayer interface of the *Match the Monkeys* learning game. *Top:* Shared feedback display showing classification status of the collaboratively developed decision tree. *Bottom left:* Playing area of Player 1. *Bottom right:* Playing area of Player 2.

3 Multiplayer Game Design

Building on the theoretical principles outlined above, the digital learning game *Match the Monkeys*[3] (MTM) was developed to foster structured problem-solving through collaborative gameplay [WLB23]. MTM is an adaptation of the learning activity *Good Monkey, Bad Monkey Game*, originally designed as part of the *AI-Unplugged* framework [LSR19]. The primary goal of the game is to distinguish between biting and non-biting monkeys based on their visual characteristics by constructing a decision tree. To achieve this, learners select filter elements represented as puzzle pieces and assemble them into a functional classification model. Throughout the process, they receive continuous feedback via a dedicated display, allowing them to iteratively refine their model and improve its classification accuracy.

The original single-player version of the game was extended into a multiplayer format designed for two participants. Each player interacts with the game through a separate tablet, accessing a distinct subset decision nodes. The shared playing area is synchronized in real time across both devices, creating a unified collaborative workspace. Crucially, players can only manipulate elements from their own inventories. This constraint introduces a form of implicit floor control [Lo07], in which agency is distributed through system design rather than turn-taking mechanics. Since no player possesses all the components required

[3] https://ai.ddi.education/games/match-the-monkeys/

to complete the decision tree independently, successful progression depends on sustained collaboration, coordinated action, and mutual awareness.

This division of resources and interaction privileges inherently fosters *positive interdependence*: each player holds unique information and capabilities, encouraging active coordination and joint decision-making. To further support *promotive interaction*, the game integrates collaborative mechanics such as shared objectives, the absence of intra-team competition, and collectively earned feedback. The decision tree puzzle structure is deliberately challenging, offering multiple solution paths and requiring logical reasoning. Learners must discuss hypotheses, identify information gaps, and coordinate sequencing decisions. These elements reflect the design principles of CLT, PBL, and CSCL, fostering cognitive strategy development through collaborative engagement.

4 Methodology

To examine learners' problem-solving behaviors and interaction dynamics, a qualitative content analysis was conducted based on the definition of specific observation scenarios for the five pillars of collaborative learning and working as defined by Johnson [Jo15], focusing on the application of structured and exploratory problem-solving strategies in the collaborative version of the *Match the Monkeys* learning game.

Data collection. The pilot study was conducted as part of an introductory ML workshop, involving three student pairs (total $n = 6$ learners). Each group played the game using tablet devices. Data sources included video recordings of both the gameplay and the post-game reflection phases (approx. 25 minutes per group). Full transcripts were produced for two groups; for the third group, a keyword-based analysis was performed due to limited audio quality.

Analysis procedure. Coding and analysis of the recorded video material were conducted by the research lead. A structured observational protocol guided the identification of relevant interaction episodes, which were subsequently categorized according to the predefined theoretical dimensions. To ensure consistency, the coding process followed a clearly defined scheme grounded in established theoretical constructs as outlined above. The analysis aimed to identify interaction patterns that either supported or impeded structured problem-solving, with particular attention to how specific game design elements shaped learners' strategic behavior.

Supplementary materials. To ensure transparency and reproducibility, the full coding matrix, anonymized transcript excerpts, and structured observation logs are available on the Open Science Framework (OSF)[4]

[4] https://osf.io/5qxfe/ [LWR25]

5 Results

The qualitative analysis of the recorded problem-solving sessions revealed consistent patterns across two key dimensions: (1) collaborative processes and (2) strategic behavior.

5.1 Collaborative Processes

Across all three groups, learners quickly acknowledged their mutual dependence, demonstrating *positive interdependence*. Statements such as "We have to combine our cards" and "You need my piece" were coded repeatedly. The game mechanics – such as distributed inventories and shared goals – effectively fostered individual accountability, further reinforced by in-game visual cues and self-monitoring remarks (e. g., "I haven't done anything yet", "Now it's your turn").

All groups engaged in strategy discussions (e. g., "What should we test next?", "Let's complete the top branch first"), which were classified as *promotive interaction*. Two groups displayed balanced turn-taking and constructive disagreement resolution (e. g., "Let's try your idea first, then mine"). In one group, a self-assigned facilitator emerged, guiding the process and helping to curb impulsive decisions. During post-game reflections, learners often identified errors and proposed improvements (e. g., "We should've made a plan earlier"), indicating group-level metacognitive development.

5.2 Problem-Solving Strategies

At the beginning of each session, groups primarily engaged in *exploratory behavior*, characterized by rapid, unguided card placement and minimal planning. Between minutes 4 and 6, *structured approaches* began to emerge – frequently triggered by system feedback or peer negotiation (e. g., "Let's do this in order now: top to bottom"). These transitions marked a shift from trial-and-error to planned, target-oriented action.

Metacognitive regulation became increasingly evident as gameplay progressed. Learners verbalized confusion (e. g. "We got mixed up, let's restart") and actively restructured their approaches. The group with the highest performance explicitly developed a top-down plan, followed it consistently, and achieved faster task completion with fewer errors.

Feedback mechanisms (e. g., the traffic light system) were frequently referenced and appeared to prompt strategic reflection. Although some learners critiqued the vagueness of these hints, these moments often led to productive pauses and renewed collaborative discussion.

Dimension	Observed Behaviors
Positive Interdependence	Learners acknowledged mutual dependence (e. g., "I need your card"), solving the puzzle was only possible through collaboration.
Individual Accountability	Players monitored their own participation (e. g., "Now it's your turn") and contributed equally to task progression.
Promotive Interaction	Strategy discussions and negotiation of actions (e. g., "Let's try your idea first") were frequent and constructive.
Metacognitive Regulation	Learners reflected on confusion, restarted approaches, and made explicit strategic adjustments.
Exploratory vs. Structured Behavior	All groups began with exploratory moves; structured strategies emerged after 4–6 minutes, often triggered by discussion or feedback.

Tab. 1: Observed Behaviors by Collaborative Learning Dimension

6 Discussion

The findings indicate that carefully designed collaborative educational games can effectively promote structured problem-solving in ML learning contexts. Game elements such as distributed resources, real-time synchronization, and immediate feedback effectively supported the development of collaborative behaviors and cognitive strategies.

The transition from exploratory to structured problem-solving was typically initiated by in-game feedback or peer negotiation, indicating the combined influence of external scaffolds and social dynamics in fostering *metacognitive regulation*. These observations align with core principles of CLT and PBL: the task was sufficiently complex to necessitate collaboration, yet structured enough to guide learners toward target-oriented behavior. Moreover, the integration of the five CSCL principles – *positive interdependence*, *individual accountability*, *promotive interaction*, *collaborative skills*, and *group reflection* – was evident in learner interactions, affirming the relevance of these design heuristics for game-based environments.

Nonetheless, the study is subject to limitations. The small sample size and brief intervention constrain the generalizability of the results. Additionally, some learners reported ambiguity in the feedback system, indicating a need for more explicit instructional scaffolding. Future research should investigate the long-term effects of collaborative gameplay on strategy retention and explore adaptive feedback mechanisms that respond to group behavior.

Overall, *Match the Monkeys* illustrates the potential of multiplayer game-based learning environments to foster both learner engagement and structured, systematic reasoning – key competencies for developing ML literacy.

References

[Ba86] Barrows, H. S.: A taxonomy of problem-based learning methods. Medical Education 20 (6), pp. 481–486, 1986, DOI: 10.1111/j.1365-2923.1986.tb01386.x.

[Jo15] Johnson, D.: Learning Together and Alone. Better: Evidence Based Education 7, pp. 4–5, 2015.

[Ki18] Kirschner, P. A. et al.: From Cognitive Load Theory to Collaborative Cognitive Load Theory. International Journal of Computer-Supported Collaborative Learning 13 (2), pp. 213–233, 2018, DOI: 10.1007/s11412-018-9277-y.

[Lo07] Lonchamp, J.: Floor Control in Complex Synchronous CSCL Environments. In: International Conference on Web Information Systems and Technologies. Vol. 2, SCITEPRESS, pp. 397–402, 2007.

[LSR19] Lindner, A.; Seegerer, S.; Romeike, R.: Unplugged Activities in the Context of AI. In: Informatics in Schools. New Ideas in School Informatics: 12th International Conference on Informatics in Schools: Situation, Evolution, and Perspectives, ISSEP 2019, Larnaca, Cyprus, November 18–20, 2019, Proceedings. Springer-Verlag, Berlin, Heidelberg, pp. 123–135, 2019, DOI: 10.1007/978-3-030-33759-9_10.

[LWR25] Leonhardt, T.; Witt, C.; Reithmeier, N.: Match the Monkeys: Collaborative Decision Tree Learning, 2025, osf.io/5qxfe.

[Po45] Polya, G.: How to solve it; a new aspect of mathematical method. Princeton University Press, Princeton, NJ, US, 1945.

[SMP98] Sweller, J.; van Merrienboer, J. J. G.; Paas, F. G. W. C.: Cognitive architecture and instructional design. Educational Psychology Review 10 (3), pp. 251–296, 1998, DOI: 10.1023/A:1022193728205.

[WF22] Woo, K.; Falloon, G.: Problem Solved, but How? An Exploratory Study into Students' Problem Solving Processes in Creative Coding Tasks. Thinking Skills and Creativity 46, p. 101193, 2022, DOI: 10.1016/j.tsc.2022.101193.

[Wi25] Witt, C. et al.: From Guesswork to Game Plan: Exploring Problem-Solving-Strategies in a Machine Learning Game. In (Pluhár, Z.; Gaál, B., eds.): Informatics in Schools. Innovative Approaches to Computer Science Teaching and Learning. Springer Nature Switzerland, Cham, pp. 73–84, 2025.

[WK16] Wendel, V.; Konert, J.: Multiplayer Serious Games. In (Dörner, R. et al., eds.): Serious Games: Foundations, Concepts and Practice. Springer International Publishing, Cham, pp. 211–241, 2016, DOI: 10.1007/978-3-319-40612-1_8.

[WLB23] Witt, C.; Leonhardt, T.; Bergner, N.: Über den Bildschirmrand hinaus – Kollaborative Multiscreen-Touch-Lernspiele im Kontext künstlicher Intelligenz. In: INFOS 2023 – Informatikunterricht zwischen Aktualität und Zeitlosigkeit. Gesellschaft für Informatik e.V., pp. 425–426, 2023.

Examiner Perspectives on Assessment Analytics as a Foundation for Developing Quality Assurance Tools in Higher Education

Martin Breuer [1], Annabell Brocker [2], Malte Persike [1], and Ulrik Schroeder [2]

Abstract: Selecting actionable indicators for Assessment Analytics (AA) is challenging due to the diversity of assessment formats, varying quality assurance (QA) practices, and the lack of consensus on which metrics meaningfully support exam improvement. To address this, the present study investigates examiners' QA practices and their perspectives on additional, potentially useful assessment data. A questionnaire was used to gather responses from 79 examiners. The responses were analysed to create a list of ideas, practices and goals as a basis for potential indicators. While current QA practices rely on direct feedback from peers, ideas to facilitate additional assessment data differ in terms of granularity, effort, and potential benefit. The potential benefit of integrating these ideas in an AA component ranges from simply saving time by digitizing manual tasks, to reduce barriers so current QA practices, mainly used by experts, can be applied more widely, and paving the way for novel functionalities not yet well discovered. While initial development steps were undertaken based on these findings, utilizing assessment- and content-related data to inform the decision-making of examiners and institutional stakeholders, the primary focus of this study is to provide an overview of examiners perspectives to inform subsequent development processes.

Keywords: Assessment Analytics, User-Centered Design, E-Assessment.

1 Introduction

Digital assessments offer both organizational and didactical benefits and challenges to academic staff [Ba21]. While e-assessment systems standardize examination procedures, the integration of an Assessment Analytics (AA) component aims to improve teaching and assessment practices [Br23]. However, psychometric analysis as part of AA is usually done with standalone statistical tools [Re24]. To bridge this gap, the e-assessment system Dynexite[3] has laid the groundwork for integrating AA into its platform. The initial development of AA components in Dynexite focused on data transformation and storage using xAPI [Br23], as well as leveraging interoperable components for integration [Sa24]. Building on this foundation, the present study adopts a user-centered approach to inform the further development of AA components, using the assessment system Dynexite as a

[1] RWTH Aachen, Center für Lehr- und Lernservices, Kackertstraße 15, 52072 Aachen, Deutschland, breuer@medien.rwth-aachen.de, https://orcid.org/0009-0008-0749-5110; persike@cls.rwth-aachen.de, https://orcid.org/0000-0002-7825-089X

[2] 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

[3] Dynexite documentation, https://v3.docs.dynexite.rwth-aachen.de last accessed: 07.07.2025

doi:10.18420/delfi2025_17

practical example. This approach acknowledges that simply providing data and indicators does not automatically enhance teaching and learning [HBP17; SS22].

In higher education, AA extends Siemens' Learning Analytics (LA) definition [SB12] to analyse assessment-related data specifically [El13]. Despite significant literature on LA perceptions among various stakeholders in higher education [Ma20], explicit guidance on defining actionable indicators for AA remains scarce. Recent work identifies student assessment data as crucial [HAS24], yet specifying meaningful indicators aligned with learning design, outcomes, and processes continues to be challenging, given the scarcity of well-defined metrics that yield valid data points for AA [IHG23]. Thus, this study adopts a user-centered approach by investigating examiners' existing practices in creating, grading, analyzing, and assuring exam quality to guide AA integration and select indicators that could enhance teaching and assessment. Recognizing that no single study can address all disciplinary nuances, this investigation offers a foundational overview to steer subsequent, more targeted research. Accordingly, we formulate two research questions:

- RQ1: What are current quality assurance processes?
- RQ2: What additional assessment data do examiners perceive as necessary to inform their assessment practices and how could the selection of AA indicators be prioritized?

2 Methods

In this study, we used a questionnaire to gather examiners' QA practices and perceived data needs. Because this study is meant to support the early development of AA components, this approach was more efficient than interviews for obtaining a broad initial overview.

Design of questionnaire The questionnaire, accessible under [Br25], was designed to be completed within five to ten minutes. It begins with an introduction and collects participant details (A1–A3), followed by five parts (B–F). Parts B-E focus on current practices (RQ1), part F addresses additional data needs (RQ2). Each part consists of a closed-ended question, followed by optional open-ended questions to reveal domain-specific needs and contextual factors that inform real-world constraints, thus guiding targeted AA-component design:

Part B: This section queries post-exam calculation practices, including whether metrics such as average grades and passing thresholds are calculated (B1: Yes/No), the specific metrics determined (B2), and how they are used (B3).

Part C: The inquiry into the integration of past correction experiences into current processes begins with a question on whether such experiences are actively incorporated (C1: Yes/No), followed by the methods of integration (C2).

Part D: The questionnaire asks about the reuse of exam questions, including whether existing tasks are used in creating new exam content (D1: Yes, unchanged/Yes, in an adapted form/No; multiple-response), and the methodologies for question reuse (D2).

Part E: QA measures for written exam content are surveyed, starting with a question on their application (E1: Yes/No), followed by the methodologies employed (E2).

Part F: In this part we assess the interest in additional analyses of student behavior during exams, starting with respondents' interest (F1: Yes/No), which specific information would be valuable (F2), and how such analyses would be utilized (F3).

Participants The questionnaire was distributed to users of the Dynexite assessment system at RWTH Aachen University in July 2023 – to this date, there were no AA components available within the assessment system. Of the 123 responses received, 79 were selected for thorough analysis by filtering out incomplete responses. Respondents comprised 31 % lecturers, 63 % scientific employees with joint responsibility, 3 % student assistants, and 3 % others (A3), with participation across all faculties (A1) and four individuals from academic support services. A majority (58 %) were involved in all assessment steps (A2). Individual roles within the assessment process reported were 97 % for creation, 92 % for correction, 84 % for evaluation, and 62 % for QA.

Procedure We analyzed open-ended responses inductively, first reviewing a subset to build a coding scheme, then coding all responses with two coders. To increase confidence in the conclusions drawn from the open responses, we calculated the inter-rater reliability score (κ) [BP81]. For every participant, κ was above 0.87, indicating almost perfect agreement.

3 Results

This section briefly presents each question's results before the detailed discussion in section 4. The results of structured questions (B1, C1, etc.) are depicted as parts of the 79 respondents. The results of the open-ended questions (B2, B3, C2, etc.) are based on the coding mentioned above and presented as proportions of the total valid responses n. The answers were divided into different categories for each part: **i**dea (B_{i1}), **p**ractice (C_{p1}), **g**oal (F_{g1}).

Part B: 79 % of respondents reported conducting calculations after a written exam (B1). Regarding B2 (n=55), 76 % described metrics characterizing the entire exam, such as average grade (B_{i1}), 27 % calculate individual question metrics such as average points per question (B_{i2}), 13 % assess subsets of questions, such as by topic or exam section (B_{i3}), 9 % calculate context-specific metrics, like pre- and post-inspection points, or considering the number of individual exam attempts (B_{i4}). These address the following goals (B3, n=51): exam content evaluation (B_{g1}, 71 %), exam process management (B_{g2}, 26 %, e. g., post-correction tasks, grade scale determination), deriving teaching insights (B_{g3}, 16 %, e. g., constructive alignment [Bi03]), and student feedback (B_{g4}, 16 %). Regarding exam content evaluation, 33 % compare measures with other exam dates (B_{g5}), 24 % primarily reflect on the exam's difficulty, length or question answering difficulties, etc. (B_{g6}). Furthermore, 14 % employ specific QA procedures to identify questions needing further review (B_{g7}).

Part C: 71 % of the respondents indicated active integration of past correction experiences into current correction processes (C1). Responses to C2 (n=51) revealed that 43 % of the participants focused on evaluating and enhancing exam questions and their solutions (C_{p1}). Additionally, 20 % include adjustments to grade scales (C_{p2}), and 8 % about fair grading (C_{p3}). Other responses (13 %) highlighted efforts to streamline the correction process, including the development of correction guidelines aimed at improving efficiency (C_{p4}).

Part D: Among participants, 5 % stated they do not use previously created tasks, 39 % reuse them without alterations, and the majority (87%) adapt tasks in modified forms (D1, multiple response). Responding to D2 (n=50), 40 % explain how they reintegrate old questions (D_{p1}). 18 % consider past experiences in decision-making such as testing new question types in the exercise system (D_{p2}), with 12% focusing on question improvement (D_{p3}). 20 % utilize templating capabilities of the assessment system for minor modifications (D_{p4}).

Part E: 71 % of the participants incorporate QA methods into their assessment processes (E1). In E2 (n=47), 68 % highlight feedback mechanisms like peer reviews and 4-eye principle (E_{p1}). 36 % involve peers in trial calculations of questions or entire exams (E_{p2}). Furthermore, 34 % explicitly mention involving individuals such as student assistants, scientific employees, lecturers/professors, and independent staff (E_{p3}). Lecturers and professors serve as the final QA checkpoint in four responses (E_{p4}). Additionally, four participants use statistical analysis as part of their QA practices (E_{p5}).

Part F: 77 % of respondents expressed interest in additional analyses concerning student task and exam-taking behaviors (F1). Regarding F2 (n=57), 80 % express general interest in time-related data (F_{i1}), with slightly less than half explicitly demanding time-on-task (F_{i2}) and two participants in time-on-task at a subtask level (F_{i3}). 35 % of participants mention detailed answering behavior (F_{i4}). Specifically, 15 % were interested in answering order (F_{i5}), 9 % in changing answer behavior (e. g., iterations, returning to tasks) (F_{i6}), 9 % in reading/thinking time, and time spent reviewing answers (F_{i7}). Result-data, mentioned by 13 %, included correlations between time-on-task and correctness, discrimination index, task difficulty, and error concentration (F_{i8}). Three participants requested more detailed result-data like correctness-per-blank in fill-in-the-blank questions (F_{i9}). Two respondents desired visual representations like histograms (F_{i10}). Regarding usage (F3, n=52), 82 % aim to create or improve assessment content. This includes adjusting point scales based on task weight, e. g. one point per minute (F_{g1}, 39 %), identifying tasks for review due to misunderstandings, difficulty of questions, etc. (F_{g2}, 26 %), and exam creation based on evidence of difficulty and completion times (F_{g3}, 14 %). 12 % aim to validate expected effort with time-on-task (F_{g4}), and 22 % to enhance teaching practices (e. g. by improving constructive alignment, F_{g5}). Other use cases (27 %) include: determining if empty answers result from linear answering or topic difficulty (F_{g6}), assessing question distribution and type (F_{g7}), relationship between scores and time-on-task (F_{g8}), assessing question reusability (also preventing excessive question recurrence) and need for revision (F_{g9}), tracking cheating through performance variance (F_{g10}), aligning task difficulty with competency levels (F_{g11}), and marking accuracy in exam development (F_{g12}).

4 Discussion

The findings from part B to E, as outlined in section 3, directly address RQ1, showing that 70 % of examiners apply QA measures. While examiners calculate exam-wide statistics (part B), QA practices mainly rely on direct feedback from peers before handing out the exams (part E). Moreover, examiners both reuse and adapt examination content (Part C) and leverage past experiences with various techniques in the correction process (Part D).

The first part of RQ2 is addressed by listing the findings consisting of ideas (F_{i1} to F_{i10}) and goals (F_{g1} to F_{g12}) that summarize valuable additional assessment data for examiners. A current discussion in AA research states that AA should go beyond analysis and focus on which actions examiners could derive from indicators [SS22]. Despite efforts to categorize AA indicators [IHG23], a gap remains in utilizing a user-centered, practical perspective. We bridge this gap by highlighting varying efforts and benefits of including ideas and goals in the AA component, answering the second part of RQ2. The potential benefits, categorized as Cat. 1 to Cat. 3, range from saving time by digitizing manual tasks (Cat. 1) to reduce barriers so current QA practices, mainly used by experts, can be applied more widely (Cat. 2) and including novel functionalities enabled through technology enhancements (Cat. 3).

The findings in parts B-E align with Cat. 1 or Cat. 2, as these parts explore current practices. They differ by object to improve, which include overall course, entire exams, individual tasks, subtask, input element, etc., affecting data granularity. For instance, average exam grades (B_{i1}) requires less detail than task-level metrics (B_{i2}). In terms of effort for data collection and analysis, it is considered low since both rely on the final assessment result-data, which is already available to examiners. The frequent calculation of exam-wide statistics (B_{i1}) by examiners suggests that integrating these metrics into an AA component would primarily save time, indicating Cat. 1. Less common practices, such as statistical analysis for QA (B_{g7}, E_{p5}), can be categorized as Cat. 2. Calculating statistical measurements for individual tasks, such as discrimination index [SS22], is straightforward because the available result-data contains all required data points (earned/max score per task). However, performing analysis on more specific objects, such as correctness-per-blank in fill-in-the-blank questions (F_{i9}), requires more detailed result-data (F_{i9}) and therefore increases the effort to analyze the data. However, the process of choosing representations (F_{i10}) from which examiners can deduce actionable insights requires more (medium) effort. Integrating contextual data, such as (B_{i4}), requires additional effort. While contributing to a shared repository could automate parts of the analysis [Br23; St22], merging data from diverse sources remains a high-effort task.

Approaches leveraging temporal or behavioral data, such as adjusting scores with the one-point-per-minute rule (F_{g1}), are classified as Cat. 3. While result-data encompasses the final result of an assessment, finer-grained data includes timing information describing the timespan during the assessment. It can span from aggregated time data such as time-on-task (F_{i2}), time-on-subtask (F_{i3}), to data including timestamps describing detailed answering behavior (F_{i4}). Selecting these data points is medium to high effort, since they require additional data points and are often not well-defined [Go14; IHG23]. The goals identified

in part F, such as F_{g3}, F_{g6}, F_{g8}, indicate that examiners aim to combine insights derived from result-data and from finer-grained data. Since the in-depth analysis (Cat. 3) relies on information from the result-based analysis (Cat. 1, Cat. 2), Cat. 1 and Cat. 2 should precede Cat. 3. Figure 1 visualizes the effort for integrating each potential benefit (Cat. 1 to Cat. 3) based on the granularity of data points and object level. However, finer-grained data points may reveal sensitive information, necessitating careful data privacy considerations including pseudonymisation and access control, e. g. through a central LA system [JS22].

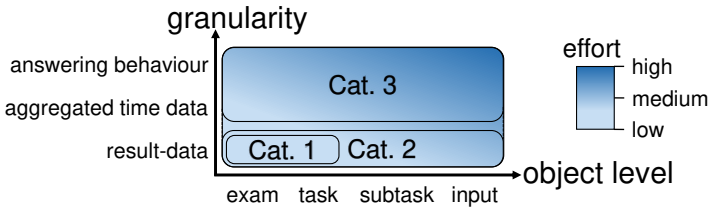


Fig. 1: Relationship of data granularity, object level and effort for each potential benefit Cat. 1 to Cat. 3.

This study offers a novel contribution to the AA literature by systematically capturing examiner-driven practices across multiple phases of the assessment cycle and mapping these to potential AA indicators. Although prior work has acknowledged the relevance of assessment data [HAS24; Ma20], it has lacked concrete guidance on aligning such data with actual workflows and decision points in exam practice. Our findings bridge this gap by introducing a categorization of AA ideas and goals, structured by object granularity, effort, and practical benefit. This framework offers a stepwise integration strategy for AA indicators, distinguishing between easily implementable practices (e. g., digitized exam-wide statistics) and more advanced, effort-intensive features (e. g., behavior-based item diagnostics). As such, it forms a basis for both immediate and long-term development of AA tools.

5 Conclusion

In this study, as part of the NOVA:ea project funded by the Foundation for Innovation in Higher Education, we presented insights into examiners' perceptions of AA for enhancing QA in higher education. We identified a diverse set of ideas and goals targeting potential indicators for AA. Identified categories (Cat. 1 to Cat 3) differ in object to improve, data usage and their effort and serve as a basis for the selection of ideas and goals to include in an AA component for QA in higher education. While the study offers a broad overview of examiners' practices, some phases, such as creation and correction, were less elaborated due to their limited relevance to the research questions, and responses may have been biased by the examples provided. Future work should employ case-based approaches that capture nuanced, domain-specific practices and pilot the AA component in authentic teaching contexts. This strategy will help align the component with existing assessment processes and ensure a seamless integration of AA components fostering iterative improvement.

References

[Ba21] Bandtel, M. et al.: Digitale Prüfungen in der Hochschule Whitepaper einer Community Working Group aus Deutschland, Österreich und der Schweiz. Stifterverband für die Deutsche Wissenschaft, 2021.

[Bi03] Biggs, J.: Aligning teaching and assessing to course objectives. en, International Conference on Teaching and Learning in Higher Education: New trend and innovations 2, 2003.

[BP81] Brennan, R. L.; Prediger, D. J.: Coefficient Kappa: Some Uses, Misuses, and Alternatives. Educational and Psychological Measurement 41 (3), pp. 687–699, 1981, DOI: 10.1177/001316448104100307.

[Br23] Breuer, M. et al.: 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.

[Br25] Breuer, M.: Examiner Perspectives on Assessment Analytics as a Foundation for Developing Quality Assurance Tools in Higher Education, 2025, DOI: 10.17605/OSF.IO/A89ZR.

[El13] Ellis, C.: Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics: Colloquium. en, British Journal of Educational Technology 44 (4), pp. 662–664, 2013, DOI: 10.1111/bjet.12028.

[Go14] Goldhammer, F. et al.: The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. en, Journal of Educational Psychology 106 (3), pp. 608–626, 2014, DOI: 10.1037/a0034716.

[HAS24] Hershkovitz, A.; Ambrose, G. A.; Soffer, T.: Instructors' Perceptions of the Use of Learning Analytics for Data-Driven Decision Making. en, Education Sciences 14 (11), p. 1180, 2024, DOI: 10.3390/educsci14111180.

[HBP17] Hora, M. T.; Bouwma-Gearhart, J.; Park, H. J.: Data driven decision-making in the era of accountability: Fostering faculty data cultures for learning. en, The Review of Higher Education 40 (3), pp. 391–426, 2017, DOI: 10.1353/rhe.2017.0013.

[IHG23] Ifenthaler, D.; Heil, J.; Greiff, S.: Toward a categorisation of indicators for assessment analytics. Learning Letters 1, 2023, DOI: 10.59453/CCTB2003.

[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). ISSN: 2161-377X, pp. 155–157, 2022, DOI: 10.1109/ICALT55010.2022.00053.

[Ma20] Mahmoud, M. et al.: Learning analytics stakeholders' expectations in higher education institutions: a literature review. en, The International Journal of Information and Learning Technology 38 (1), pp. 33–48, 2020, DOI: 10.1108/IJILT-05-2020-0081.

[Re24] Retnawati, H. et al.: A Systematic Review of the Use of Technology in Educational Assessment Practices: Lesson Learned and Direction for Future Studies. International Journal of Robotics and Control Systems 4, pp. 1656–1693, 2024, DOI: 10.31763/ijrcs.v4i4.1572.

[Sa24] Salmen, F. et al.: FAIR Learning Technologies with Web Components and Packages. en, Publisher: Gesellschaft für Informatik e.V., 2024, DOI: 10.18420/DELFI2024_23.

[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, New York, NY, USA, pp. 252–254, 2012, DOI: 10.1145/2330601.2330661.

[SS22] Scheidig, F.; Schweinberger, K.: Assessment Analytics – Daten digitaler Prüfungen auswerten. de, 2022.

[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 - Volume 1: CSEDU. SciTePress, pp. 157–164, 2022, DOI: 10.5220/0010879700003182.

Open Science/Software

A Publication Framework for Research Data in the EdTec Community

Sandra Schulz [1], Daniel Schiffner [2], Natalie Kiesler [3], and Katarzyna Biernacka [4]

Abstract: The interdisciplinary nature of the Educational Technology (EdTec) field is resembled in its research data, which comprises quantitative and qualitative data, but also software. Basic research data management principles can help ensure that our research is understandable, reproducible, and transferable to peripheral research and practice communities. However, publishing research data has not yet gained traction within the DELFI community. Therefore, the goal of this position paper is to propose a baseline for reporting and publishing typical EdTec research data (i. e., software, qualitative, quantitative) at the DELFI conference. We further aim to initiate a discussion within the community and ease data sharing and publication efforts.

Keywords: Research Data Publication Framework, Open Data, Open Science, FAIR, Software, EdTec, DELFI, conference

1 Introduction

The acronym FAIR refers to research data that is findable, accessible, interoperable, and reusable [Wi16]. There is no doubt that FAIRness and open data are crucial components in the epistemological process. Despite their value, publishing research data is not yet a common practice in the German-speaking Educational Technology (EdTec) community. Year after year, publishing research data is neither incentivized nor recognized, or worse, it causes punishments, misconceptions, and even rejections, e. g., at the DELFI conference [KS22; KS23b; SLK25]. A study on behalf of the European Commission calculated the cost of not having FAIR research data in 2018. It resulted in 10.2 billion Euros of opportunity costs [Eu18]. A recent search for research data in the field confirms the lack of FAIR, or even open, available research data [Ki24b].

In the past few years, Open Data and Open Science have received increasing attention within the EdTec community. Workshops at the DELFI conference, for example, discussed authors' needs and concerns regarding data publications, as well as their expectations toward the review of open data [KS23a; SK24a]. In addition, the DELFI conference recently dedicated

[1] 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
[2] 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
[3] Nuremberg Tech, Faculty of Computer Science, Keßlerplatz 12, 90489 Nuremberg, Germany, natalie.kiesler@th-nuernberg.de, https://orcid.org/0000-0002-6843-2729
[4] Technische Universität Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany, katarzyna.biernacka@tu-berlin.de, https://orcid.org/0000-0002-6363-0064

doi:10.18420/delfi2025_18

a keynote [Sc24] and supported the presentation of various submissions on the challenges and current practices w.r.t. open research data [Ki24a; MS23; St24]. However, at the 2023 DELFI conference, only 5 % of the submitted articles and posters had published their data; only 24 % intended to publish their data in the future [Ki24a]. Furthermore, Schulz and Kielser anaylzed the DELFI 2024 submissions and proceedings [SK24b] and pointed out that many research data were not recognized as such by the authors [SK25].

The **goal** of this position paper is to present a baseline of a data publication framework for the DELFI conference as mandated by good scientific practice (GSP). Thus, we summarize a minimum of requirements for research data (i. e., software, qualitative, and quantitative) that should be published, and introduce basic relevant principles and conditions. The **contribution** is thus to raise awareness and provide a short, concise guideline for members of our community based on existing guidelines from other communities. This is to support them in the research data publication process. We are **not** presenting a comprehensive research data management approach supporting archiving. Neither do we want to define the level of detail of single metadata or other documents.

2 Basic Principles of Research Data Management

Research Data Management (RDM) is becoming increasingly important in today's scientific landscape, and it is a central component of Good Research Practices (GRP) [De25]. RDM encompasses all measures related to handling research data throughout a project's lifecycle: from planning and gaining ethical approval, collection, and analysis to archiving, publication, and possibly deletion [Bi24]. RDM aims to organize data systematically, ensure quality, and guarantee reusability.

The GRP explicitly requires researchers to handle data responsibly. Researchers are obliged to document the nature and scope of the data they generate and to manage them "*in accordance with the requirements of the relevant subject area*" [De25, Guideline 7, p. 14]. The GRP further states: "*As a rule, researchers make all results available as part of scientific/academic discourse*" and "*Where possible and reasonable, this includes making the research data, materials, and information on which the results are based, as well as the methods and software used, available and fully explaining the work processes.*" [De25, Guideline 13, p. 17]. Additionally, transparency and reusability are a core goal of the GRP and that is why "*whenever possible researchers make the research data and principal materials on which a publication is based available in recognized archives and repositories in accordance with the FAIR principles (Findable, Accessible, Interoperable, Reusable).*" [De25, Guideline 13, p. 18]. A well-structured RDM and the publication of research data and outcomes is therefore not an administrative formality, but a mandatory element of quality assurance in research.

To meet these GRP requirements and to conduct transparent and sustainable research, researchers must engage with the basic principles of RDM and integrate them systematically

into their workflows. A wide range of generic guidelines, checklists, and tools are available to support researchers in implementing effective and FAIR RDM strategies [Au15; Bi24; En22; JG17; Li20]. Many universities and research institutions also publish their own guidelines, often tailored to the specific services of the institutions [Od20; Se23]. Schulz and Jacob encourage and evaluate the implementation of RDM into curricula [SJ25] to foster RDM knowledge according to the RDM learning objective matrix [Pe23].

In addition to the general and institutional recommendations, numerous discipline-specific guidelines have been developed in recent years to address their particular requirements, data types, and standards. In the life sciences, the ELIXIR [EL25a] initiative provides concrete guidance on metadata standards, data formats, and repository use, ensuring responsible RDM [EL25b]. In the social sciences, the Consortium of European Social Science Data Archives (CESSDA) [Co25] offers best practices for survey data, including anonymization [CE20]. In the humanities, domain-specific recommendations have been formulated by DARIAH [Tó19] or the Text Encoding Initiative Consortium [Te25], which offer frameworks for managing and publishing digital editions, corpora, and other complex textual data. Similarly, in the geosciences, the GEO Data Working Group [Gr25] supports researchers with discipline-tailored guidelines [GE24]. In chemistry, NFDI4Chem [NF25b] offers a guide to publishing research data [NF25a]. Similarly, this trend is supported in EdTec-related fields such as Learning Analytics [BP21], or computing education research [Ki24b; KS23b].

3 Basic Requirements for the EdTec Community

Unlike other related conferences that have some guidelines for reporting (e. g., ICER [As; FR19; No23], or SIGSOFT [Ra20]), the DELFI does not require authors to report on empirical research in a specific way, i. e., post data, code, and relevant material. The Transparency and Openness Promotion (TOP) guidelines [No23], for example, emphasize that data and code must be posted to a trustworthy repository to ensure the independent reproducibility of the publication. In alignment with these guidelines, we propose minimal requirements for future DELFI and other EdTec conferences.

Table 1 presents the baseline recommendations for data publications with selected examples of software, qualitative, and quantitative data. In the following subsections, we briefly introduce and discuss the various artifacts, focussing on (1) what to share, (2) what to anonymize, (3) sustainable data formats, (4) suitable repositories, and (5) access levels. In general, pseudonymization may be applied instead of anonymization if it ensures that no individual or the like can be identified.

3.1 Software

(1) *What to share* – Source code is very flexible and has different levels of readiness. The names of the authors, a ReadMe and a license must be included. A citation, e. g.,

	Software	**Qualitative Data**	**Quantitative Data**
(1) What to share	– Code (e. g., of a prototype, for data analysis, etc.) – ReadMe – Licence	– Instruments, data processing, and resulting data (e. g. interview guide, transcripts, think-aloud-protocol, coding scheme, etc.) – ReadMe – Licence	– Instruments, data analysis, and resulting data (e. g. survey, csv files) – ReadMe – Licence
(2) What to anonymize	Code (e. g., tools for identifying personal information)	Data that can identify participants (e. g., institution, names, specific tools in interview transcripts)	Data that can identify participants (e. g., demographic data of small samples)
(3) Sustainable data formats	– txt files – Schemas for creating binary or encoded data	– txt files	– csv files – xlsx files
(4) Suitable repositories	– GitHub – GitLab	– OSF – Zenodo	– OSF – Zenodo
(5) Access levels	– Open access, – Individual access or custom repository	– Open or restricted access depending on sensitivity of data, and sample (e. g., vulnerable individuals, minors)	– Open access depending on level of anonymization (e. g., open, restricted or closed)

Tab. 1: Proposed Summary of Recommendations for Data Publications with Selected Examples

a .cff-file, should be added. The ReadMe is a textual summary of the purpose and role of the software. The target audience and the level of detail is left to the authors. The ReadMe should contain a statement, whether additional software or a specific infrastructure is required. It is not necessary that the software is able to run, but it should be explained what is required to run the code. To make the source code better understandable, comments can be added. We acknowledge that this heavily depends on the context and the capabilities of the reader.

(2) *What to anonymize* – All information in data (see sections 3.2 and 3.3) and source code should be free of references to any participants or test subjects. If present, this data must be anonymized accordingly. Existing tools [Du25] should be used to support this.

(3) *Data formats* – Common text-formats used for the programming language of choice. An additional textual description must be given for binary data, for why it is included and what it contains in form of a schema, so that it can be decoded using other tools.

(4) *Repositories* – Source code management repositories, such as GitHub or GitLab are mandatory. A tag or branch of the submitted software state should be created to allow for verifications and traceability.
(5) *Access* – Open access should be the default for the source code. If limitations or restrictions apply, custom repositories should be created, which are not limited. Otherwise, individual access should be granted to interested researchers.

3.2 Qualitative Data

(1) *What to share* – Qualitative data can come in many forms. Yet we can generalize that instruments for data gathering (e. g., interview guide, thinking-aloud protocol, experimental setup), the data that is analyzed (e. g., transcripts, documents, observations), and resulting abstractions (e. g., coding scheme, themes, theory) are relevant. In addition, the context of the applied methodology must be shared (e. g., recruiting, sampling, transcription rules, etc.), along with licensing information.
(2) *What to anonymize* – By design, samples are usually small [Bo16]. So, it is likely that many aspects need to be anonymized if they can easily identify an individual, their institution, etc. To reduce the loss of information, it is advisable to anonymize e. g. names of institutions or study subjects in the following way: @@*institution name*## or @@*study subject*##.
(3) *Data formats* – According to the KOST guidelines [Ko21], the following data formats are considered as sustainable. For *text files* txt, pdf, pdf/a, odf, ooxml and for *pictures* tiff, jpeg, jp2 (JPEG2000), png, svg. *video data* can be stored and published as mkv, mp4, avi, mj2 (MJPEG2000).
(4) *Repositories* – Platforms like OSF [Ce25] and Zenodo [CE25] to publish qualitative and quantitative data and metadata easily. Both platforms provide DOIs.
(5) *Access* – Qualitative data is naturally rich in information. Hence, caution is required. Only those data should be collected that are required to answer the research question. In addition, the context and individuals need to be considered. Are they minors, or public figures? As we cannot predict, for example, how good text analysis tools will be in the future, it may be advisable to only allow restricted access to sensitive data. This would also impact the choice of licensing.

3.3 Quantitative Data

(1) *What to share* – In the EdTec community, quantitative data may be used in, e. g., benchmarking, data mining, longitudinal or survey studies. Similar to the qualitative paradigm, authors must share all data used to gather (e. g., questionnaire, benchmarking dataset), pre-process (e. g., csv files, aggregation methods), and analyze (e. g., csv files, scripts) their findings. In alignment with section 3.2, the context of the applied methodology must be shared, along with the licensing

information. For quantitative data, however, caution is advised, as many of these types of data – e. g. geo data, measurement data, sensor data, observational data, and metadata – are, in most cases, inherently in the public domain and therefore not subject to licensing.

(2) *What to anonymize* – Quantitative data may also contain direct and indirect identifiers, and sensitive information (e. g., demographics). Such data should be erased, aggregated, or changed so it becomes unrecognizable [BKL21]. Another option is pseudonymization via a codebook. Choosing an adequate method depends on the sample size, the topic, and ethics board approval, among other factors.

(3) *Data formats* – According to the KOST guidelines [Ko21], sustainable data formats for *spreadsheets* comprise, e. g. xlsx/xls, odt, pdf, pdf/a. For *structured data* csv, json, and xml are recommended. Formats for *databases* comprise, e. g., siard, and sql.

(4) *Repositories* – As mentioned in section 3.2, OSF [Ce25] and Zenodo [CE25] are platforms where all types of data and metadata can be published. Both platforms provide DOIs.

(5) *Access* – Open access may be an option, if, for example, no personal data has been gathered, the sample is sufficiently large, and possibly identifying data (e. g., demographics) are sufficiently aggregated.

4 Discussion and Conclusions

In this paper, we present a baseline of requirements for the publication of research data in the context of the DELFI, and other EdTec conferences. It should be noted that the presented framework is a lower bound regarding existing RDM standards. Hence, we only included the minimal requirements we consider essential for data to be reused and research to become more transparent and comprehensive. This is crucial, as general RDM guidelines need to be transferred to the needs of specific communities. Yet, extensive guidelines for research data management are available and recommended for those who want to do more [DV24]. Similarly, research data centers (e. g., DZHW [Ge25], FDZ Bildung [DI25]) can offer an external quality check and feedback before the publication of research data. Biernacka and Schulz [BS22] created a starting point for transferring RDM into computer science research. However, a lot of work is ahead of us. For example, our community has to discuss and determine how to live by such guidelines, realize and stand by them in peer reviews – which is currently not the case for all submissions.

In the future, we also need to identify domain-specific aspects for the application of RDM in EdTec contexts. Moreover, we as a community need to identify the breadth and depth of the details needed to describe our research data and their context. Thus, it is important to determine whether and which discipline-specific aspects we may need for managing research data. Finally, we encourage the DELFI community to provide the tools, infrastructure, and support to researchers so they can share their data wherever possible. This includes publication venues and conferences, and respective incentives for researchers.

References

[As] Association for Computing Machinery: ACM ICER Author Guidelines (2024), accessed on April 3, 2025, https://docs.google.com/document/d/15F7T0J4_YkVplAvgF5zF4FtaA35s9GY--ZFNJhyybHk/edit?tab=t.0.

[Au15] Austin, C. C. et al.: Key Components Of Data Publishing: Using Current Best Practices To Develop A Reference Model For Data Publishing. 2015, DOI: 10.5281/ZENODO.34542.

[Bi24] Biernacka, K. et al.: Train-the-Trainer Concept on Research Data Management, 2024, DOI: 10.5281/ZENODO.13927614.

[BKL21] Baumann, P.; Krahn, P.; Lauber-Rönsberg, A.: Forschungsdatenmanagement und Recht: Datenschutz-, Urheber-und Vertragsrecht. Wolfgang Neugebauer Verlag, 2021.

[Bo16] Boddy, C. R.: Sample size for qualitative research. Qualitative market research: An international journal 19 (4), pp. 426–432, 2016.

[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, DOI: 10.4018/978-1-7998-7103-3.ch002.

[BS22] Biernacka, K.; Schulz, S.: Forschungsdatenmanagement in der Informatik. 2022, DOI: 10.30819/5490.

[CE20] CESSDA Training Team: CESSDA Data Management Expert Guide. 2020, DOI: 10.5281/ZENODO.3820473.

[CE25] CERN: Zenodo, accessed on April 8, 2025, 2025, https://zenodo.org/.

[Ce25] Center for open Science: Open Sciene Framework, accessed on April 8, 2025, 2025, https://osf.io/.

[Co25] Consortium of European Social Science Data Archives: CESSDA, accessed on April 8, 2025, 2025, https://www.cessda.eu/.

[De25] Deutsche Forschungsgemeinschaft: Guidelines for Safeguarding Good Research Practice. Code of Conduct. de, 2025, DOI: 10.5281/ZENODO.14281892.

[DI25] DIPF | Leibniz-Institute for Research and Information in Education: Research Data Centre for Education, accessed on April 8, 2025, 2025, https://www.fdz-bildung.de/home?la=en.

[Du25] Durieux, T.: Anonymous GitHub, accessed on April 8, 2025, 2025, https://anonymous.4open.science.

[DV24] DDP-Bildung; VerbundFDB: Stamp – Standardisierter Datenmanagementplan. Version 1.0, accessed on April 8, 2025, 2024, https://www.forschungsdaten-bildung.de/datenmanagement/planen/datenmanagement-planen/.

[EL25a] ELIXIR: ELIXIR: Welcome to ELIXIR, accessed on April 8, 2025, 2025, https://elixir-europe.org/.

[EL25b] ELIXIR: The Research Data Management toolkit for Life Sciences, accessed on April 8, 2025, 2025, https://rdmkit.elixir-europe.org/.

[En22] Engelhardt, C. et al.: How to be FAIR with your data, A teaching and training handbook for higher education institutions. Universitätsverlag Göttingen, Göttingen, 2022.

[Eu18] European Commission and Directorate-General for Research and Innovation: Cost-benefit analysis for FAIR research data – Cost of not having FAIR research data. Publications Office, 2018.

[FR19] Fincher, S. A.; Robins, A. V., eds.: The Cambridge Handbook of Computing Education Research. Cambridge University Press, 2019.

[GE24] GEO Data Working Group: GEO Data Management Principles Implementation Guidelines. en, 2024, DOI: 10.60566/V56MD-KQG97.

[Ge25] German Centre for Higher Education Research and Science Studies (DZHW): Find Higher Education and Science Research Data Packages, accessed on April 8, 2025, 2025, https://metadata.fdz.dzhw.eu/en/start.

[Gr25] Group on Earth Observations: Data and Knowledge Working Group, accessed on April 8, 2025, 2025, https://earthobservations.org/organization/working-groups/geo-data-working-group.

[JG17] Jones, S.; Grootveld, M.: How Fair Are Your Data?, en, 2017, DOI: 10.5281/ZENODO.1065991.

[Ki24a] Kiesler, N. et al.: Towards Open Science at the DELFI Conference. In (Schulz, S.; Kiesler, N., eds.): 22. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., Bonn, pp. 251–265, 2024, DOI: 10.18420/delfi2024_22.

[Ki24b] Kiesler, N. et al.: Where's the Data? Finding and Reusing Datasets in Computing Education. In: Working Group Reports on 2023 ACM Conference on Global Computing Education. CompEd 2023, ACM, Hyderabad, India, pp. 31–60, 2024, DOI: 10.1145/3598579.3689378.

[Ko21] Koordinatiosstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST): Empfehlungen, accessed on April 3, 2025, 2021, https://perma.cc/AQP7-F2LD.

[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 V. 1. ITiCSE 2023, ACM, Turku, Finland, pp. 348–353, 2023, DOI: 10.1145/3587102.3588860.

[Li20] Liffers, M. et al.: 23 (research data) Things, en, 2020, DOI: 10.5281/ZENODO.3955523.

[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.

[NF25a] NFDI4Chem: A Simple Guide to Publishing Your Research Data, accessed on April 8, 2025, 2025, https://knowledgebase.nfdi4chem.de/knowledge_base/docs/publishing_standards_authors/?userLocale=true.

[NF25b] NFDI4Chem: NFDI4Chem, accessed on April 8, 2025, 2025, https://www.nfdi4chem.de/.

[No23] Nosek, B. A. et al.: Transparency and Openness Promotion (TOP) Guidelines, 2023, osf.io/9f6gx.

[Od20] Odebrecht, C.: Leitfaden zum Forschungsdatenmanagement für Abschlussarbeiten, 2020, DOI: 10.5281/ZENODO.3931920.

[Pe23] Petersen, B. et al.: Lernzielmatrix zum Themenbereich Forschungsdatenmanagement (FDM) für die Zielgruppen Studierende, PhDs und Data Stewards, version 2, 2023, DOI: 10.5281/zenodo.8010617.

[Ra20] Ralph, P. et al.: Empirical Standards for Software Engineering Research, 2020, DOI: 10.48550/arXiv.2010.03525, arXiv: 2010.03525 [cs.SE].

[Sc24] Schiffner, D.: (Re)-Using the Data: Address Challenges in Research Data via FAIR and APIs, ed. by Schulz, S.; Kiesler, N., Proceedings of DELFI 2024, 2024, DOI: 10.18420/delfi2024_02.

[Se23] Servicezentrum Forschungsdatenmanagement (SZF): 10 Steps to Successfull Research Data Management, Online, accessed on April 3, 2025, 2023, https://www.static.tu.berlin/fileadmin/www/40000027/Dokumente/Checklist_10_Steps_RDM_2023_EN.pdf.

[SJ25] Schulz, S.; Jacob, J.: Lecturers' perspectives on the integration of research data management into teacher training programmes, 2025, DOI: 10.48550/arXiv.2505.21704, eprint: arXiv.

[SK24a] Schiffner, D.; Kiesler, N.: Open Science in den Bildungstechnologien: Zur Unterstützung bei der Veröffentlichung von Forschungsdaten. In (Kiesler, N.; Schulz, S., eds.): Proceedings of DELFI Workshops 2024. Gesellschaft für Informatik eV, pp. 10–18420, 2024, DOI: 10.18420/delfi2024-ws-11.

[SK24b] Schulz, S.; Kiesler, N., eds.: Proceedings of DELFI 2024. Gesellschaft für Informatik e.V., 2024.

[SK25] Schulz, S.; Kiesler, N.: The Data Dilemma: Authors' Intentions and Recognition of Research Data in Educational Technology Research. arXiv preprint arXiv:2506.04954, 2025, DOI: 10.48550/arXiv.2506.04954.

[SLK25] Schneider, J.; Limbu, B.; Kiesler, N.: Of house of cards and air castles, a deep dive into the fertile fields of educational technologies and technology enhanced learning. Journal of Computing in Higher Education, pp. 1–53, 2025, DOI: 10.1007/s12528-025-09450-8.

[St24] Striewe, M.: Forschungsdaten in der Bildungstechnologie: Worüber sprechen wir und was ist wann relevant? In (Schulz, S.; Kiesler, N., eds.). Gesellschaft für Informatik e.V., 2024, DOI: 10.18420/delfi2024_44.

[Te25] Text Encoding Initiative Consortium: Guidelines for Electronic Text Encoding and Interchange, version 4.9.0, 2025, https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf.

[Tó19] Tóth-Czifra, E.: DARIAH Pathfinder to Data Management Best Practices in the Humanities, 2019, https://hdl.handle.net/21.11159/019595b2-cca1-70ad-b1e3-d088b4409de5.

[Wi16] Wilkinson, M. D. et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific data 3 (1), pp. 1–9, 2016.

Nutzungserfahrungen und partizipative Entwicklung eines Scaffolding-Frameworks zur Unterstützung von Lernprozessen

Sarah Schneeweiß [1] und Andreas Harrer [1]

Abstract: Im Rahmen eines Forschungsprojektes haben wir ein Scaffolding-Framework konzeptionell und softwaretechnisch entwickelt und in verschiedenen Pilotstudien praktisch in der Hochschullehre eingesetzt. Die Umsetzung der Pilotstudien gemeinsam mit den jeweiligen Lehrenden hat im Rahmen der partizipativen Entwicklung sowohl die Features des technischen Systems als auch die Einsatzbedingungen der Anwendung beeinflusst. Abschließend berichten wir über die entstandenen und nachwirkenden Erzeugnisse des Projektes in Form von Software und methodischen Werkzeugen.

Keywords: Scaffolding, Scaffolding-Framework, Lernprozessunterstützung, Praxiseinsatz

1 Einleitung

In der Förderinitiative „Künstliche Intelligenz in der Hochschulbildung“[2] des Bundesministeriums für Bildung und Forschung wurden zwischen 2021 und 2025 lehr-/lernunterstützende Projekte gefördert. Eines davon ist **IPPOLIS**[3], das in Teilprojekten das Schreiben wissenschaftlicher Arbeiten, die interaktive Bearbeitung statistischer Fallstudien, die chatbotbasierte Informationssuche und die Unterstützung formalisierter Lernprozesse behandelt.

Über die praktischen Erfahrungen, erarbeiteten Konzepte und Artefakte und deren mögliche nachwirkende Nutzung im Teilprojekt **Scaffolds** zur Unterstützung von Lernprozessen wollen wir in diesem Artikel berichten. Ausgangspunkt der Überlegungen ist die Existenz von systematischen Praktiken und Verfahrensweisen in zahlreichen wissenschaftlichen Disziplinen, die zum Teil als theoretische Modelle begründet werden und sich zum Teil aus Praxiserfahrung ergeben. Beispielsweise gibt es im Design interaktiver Anwendungen Vorgehensmodelle, wie den *nutzerzentrierten Entwicklungsprozess*, der typische Aktivitäten mitsamt deren Reihenfolge und iterativer Wiederholung beinhaltet. Aus der praktischen Erfahrung gewonnen sind solche Vorgehensweisen teils als ISO/DIN-Norm [In19] kanonisiert und gehören somit wesentlich zu den Methodenkompetenzen der jeweiligen Disziplin.

Um die Relevanz dieser Verfahrensweisen und deren praktische Anwendung verstehen zu können, ist Erfahrung nötig, über die Studierende anfangs kaum verfügen. In dieser Situation

[1] Fachhochschule Dortmund, Fachbereich Informatik, Emil-Figge-Strasse 42, 44227 Dortmund, Deutschland, sarah.schneeweiss@fh-dortmund.de, https://orcid.org/0009-0002-4281-8356; andreas.harrer@fh-dortmund.de, https://orcid.org/0009-0001-1076-7964

[2] https://www.bmbf.de/SharedDocs/Bekanntmachungen/DE/2021/02/3409_bekanntmachung.html

[3] (Intelligente Unterstützung projekt- und problemorientierter Lehre und Integration in Studienabläufe) https://fh.do/ippolis – Gefördert vom BMBF und MKW NRW unter dem Förderkennzeichen 16DHBKI050

doi:10.18420/delfi2025_19

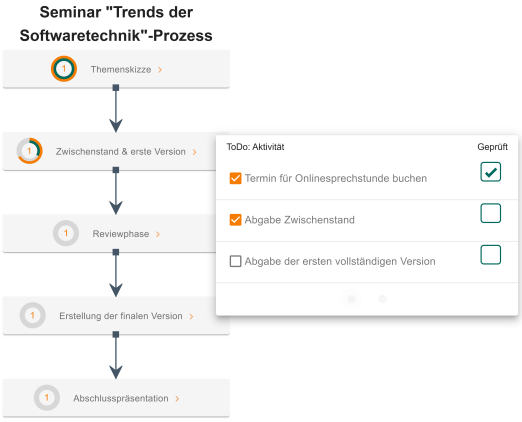
beziehen wir uns in der didaktischen Konzeption des Teilprojektes auf das „Scaffolding“ [WBR76], bei dem im Wortsinne eines Stützgerüsts den Lernenden solange Unterstützung gegeben wird, bis die Studierenden diese Hilfen nicht mehr benötigen, sodass diese nach und nach entfernt werden können (sogenanntes „Fading“). Unsere Konzentration auf *Scaffolding von Vorgehensweisen* ist vergleichbar mit der Prozessorientierung in den Arbeiten zur Unterstützung von forschend-entdeckendem Lernen (Science Inquiry) von Quintana et al. [Qu04; Qu21], soll aber auf eine breitere Spanne von Domänen und einen flexibleren Ansatz von Scaffolds ausgerichtet sein. Ziel unseres Teilprojektes war es, ein konzeptionelles Rahmenwerk mitsamt softwaretechnischen Lösungen zum Scaffolding von Lernprozessen zu schaffen, welches wir im folgenden Kapitel beschreiben.

2 Scaffolding-Framework

Wie bereits erwähnt, besteht die Grundidee des Teilprojekts Scaffolds darin, bestimmte formalisierbare Vorgehensweisen und damit verbundene Lernprozesse in der Lehre zu unterstützen. Das **Scaffolding-Framework** bleibt dabei soweit wie möglich entkoppelt von fachspezifischen Inhalten, nicht nur, um vielseitig einsetzbar zu sein, sondern insbesondere, damit die Expertise hier durch die entsprechenden Lehrenden eingebracht wird.

Der Kern des konzeptionellen Scaffolding-Frameworks mit Prinzipien der Strukturierung, unterstützenden Aktivitäten und Feedback zur Selbstreflexion wird durch das **Scaffolding-System** in Form einer Webanwendung, die interaktive Prozesse abbildet, umgesetzt. Damit lassen sich Prozesse für unterschiedliche Anwendungsfälle darstellen, wie z. B. im Bereich Engineering in [RSH24] genauer beschrieben. Die Grundstruktur jedes Prozesses (s. Abb. 1a) besteht dabei aus mehreren *Prozessschritten*, die jeder für sich wiederum *Aktivitäten* für die Absolvierung dieses Prozessschritts enthalten. Aktivitäten können von den Nutzenden zweistufig als abgeschlossen markiert werden: die Erledigung der Aktivität und die Überprüfung durch eine Lehrperson. In Form von zirkulären Fortschrittsanzeigen je Prozessschritt wird der aktuelle Bearbeitungsstand dargestellt. Aktivitäten können über *Unterseiten* verfügen, auf denen Informationen und Hilfestellungen zur Bewältigung der Aktivitäten bereitgestellt werden. Im Engineering-Anwendungsfall bieten die Unterseiten eine Überprüfung von erarbeiteten Ergebnissen und Feedback dazu. Prozessschritte können beliebig aufeinander verweisen. Auch eine Verzweigung im Prozess, bei der Lernende eine Entscheidung z. B. im Rahmen einer Methodenauswahl treffen, ist implementiert.

Beispielsweise ist in Abb. 1a ein Prozess mit insgesamt fünf Prozessschritten für eine Lehrveranstaltung in Seminarform zu sehen (vgl. Abschnitt 3). Die Aktivitäten des ersten Prozessschritts „Themenskizze“ und damit der Schritt selbst sind bereits als abgeschlossen markiert. Zurzeit sind die Informationen und die Übersicht der Aktivitäten des zweiten Prozessschritts „Zwischenstand & erste Version“ geöffnet. Dort sind die ersten beiden Aktivitäten bereits „erledigt“, bei ersterer auch die Überprüfung. Der Fortschritt spiegelt sich in den Fortschrittsanzeigen in der Gesamtübersicht der Prozesschritte wider. Unterseiten sind in diesem Beispiel nicht angelegt und damit nicht in der Abbildung vorhanden.



(a) Darstellung im Scaffolding-System



(b) Ausschnitt der JSON-Spezifikation

Abb. 1: Prozess- und Aktivitätsfortschritte im Anwendungfall „Seminar Trends der Softwaretechnik“

Ein Prozess wie in Abb. 1a wird über eine Datei in der JavaScript Object Notation (JSON)[4] spezifiziert, wie in Abb. 1b dargestellt. Durch den Aufbau der Spezifikation in JSON lässt sich ein Prozess mit geringem Aufwand um weitere Schritte und Aktivitäten erweitern oder verändern. Jeder Prozessschritt verfügt über einen eindeutigen Identifikator (*id*), einen Namen (*title*), eine Liste der Aktivitäten des Prozessschritts (*checklist*), die aktuellen Werte der beiden Fortschrittsanzeigen (*processValue* bzw. *checkTutorValue*) sowie die direkt folgenden Prozessschritte des aktuellen Schritts (*next*). In Abb. 1b enthält der Prozessschritt überschrieben mit „Zwischenstand & erste Version“ ein Array mit verschiedenen Aktivitäten. Eine Aktivität wird dabei über die Tätigkeit (*description*), optionale Zusatzinformationen (*information*), den Status der Aktivität – erledigt (*check*) und durch eine Lehrperson geprüft (*tutorCheck*) – sowie optionale Unterseiten (*content*) beschrieben.

Das Scaffolding-System ist als Webanwendung mit dem JavaScript-Framework vue.js[5] realisiert, welches sich u.a. für die graduelle komponentenweise Erweiterung um Funktionalitäten anbietet. Dabei setzen wir auf eine Fat-Client-Architektur, um die Datensouveranität der nutzenden Personen zu gewährleisten. Das bedeutet, dass jegliche Daten zum Fortschritt eines Prozesses nur lokal im „Local Storage“ des Browsers gespeichert werden und serverseitig keinerlei Zugriff darauf besteht. Zusätzlich kann der persönliche Prozessfortschritt in Form einer JSON-Datei ex- und importiert werden. Der Fortschritt ist dadurch zwischen verschiedenen Geräten portabel, was beispielsweise für den Praktikumsbetrieb in der Lehre relevant ist. Das Logging des Verhaltens in der Webanwendung, z. B. das Abhaken einer

[4] https://www.json.org
[5] https://vuejs.org/

bestimmten Aktivität, kann wahlweise an- oder ausgeschaltet werden und wird ebenfalls nur lokal gespeichert. Hier setzen wir auf freiwillige Datenspenden der Loggingdaten seitens der Nutzenden, welche diese aus der Webanwendung heraus initialisieren können.

Zu dem Rahmen des Scaffolding-Frameworks gehört ein **Erstellungswerkzeug** (Authoring Tool) für Prozesse, durch das Lehrende auch mit weniger formalen Kenntnissen mithilfe von Eingabemasken Prozessschritte definieren, beschreiben und in Reihenfolge bringen können. Ebenfalls ist es über das Werkzeug möglich, Unterseiten zu erstellen und mit Inhalten anzureichern. Die so geschaffene Prozess-Spezifikation kann anschließend exportiert und in das Scaffolding-System als Prozess für einen (neuen) Anwendungsfall eingepflegt werden.

3 Erprobung in der Praxis: Pilotstudien

Ausgehend von unserem Scaffolding-Framework stellte sich im Projektverlauf wiederholt die Frage, ob der vorgeschlagene Ansatz flexibel und über Anwendungsdomänen hinweg nutzbar ist, also auch einen verallgemeinerbaren Beitrag leisten kann. Zu diesem Zweck und zu der praktischen Erprobung des Frameworks haben wir als Minimalrahmen drei Pilotstudien in den Domänen Engineering, Science Inquiry sowie Data Science vorgesehen und erfolgreich in Lehrveranstaltungen in den Präsenz-Bachelorstudiengängen des Fachbereichs Informatik an der Fachhochschule Dortmund durchgeführt. Bei der Auswahl der Lehrveranstaltungen war für uns ein wesentlicher Punkt, unterschiedliche Fachkulturen und unterschiedliche Lehrende miteinzubeziehen, um eine Selbstinspektion und Verengung der Perspektiven bewusst auszuschließen. Dazu wurde für jede der geplanten Pilotstudien mindestens eine Lehrperson ausgesucht, die außerhalb des direkten Projektkontexts stand und somit 'stakeholder' für die jeweilige Disziplin ist.

In Tabelle 1 findet sich eine Übersicht der durchgeführten Pilotstudien. Im Folgenden heben wir die Besonderheiten und Merkmale der Studien hervor. Für die Pilotstudien wurde ein Fragebogen als Evaluationsinstrument entwickelt (s. Abschnitt 4). Dieser wurde in den Pilotstudien mit höheren Studierendenzahlen eingesetzt und weiterentwickelt, während in der Pilotstudie mit einem kleineren Kurs qualitativ Feedback eingeholt wurde. Erkenntnisse aus den Evaluationen wurden mit den jeweiligen Lehrenden betrachtet und diskutiert.

Lehrveranstaltung	Anzahl Studierende	Einsatz des Scaffolds
Softwaretechnik 1, Vorlesung mit Übung und Praktikum	ca. 120	Erarbeitung von UML-Diagrammen in einer Gruppenprojektarbeit
Seminar Trends der Softwaretechnik, Seminar	ca. 50	Unterstützung der Vorgehensweise und Einzelprojektarbeit über den ganzen Seminarzeitraum
Data Science Projekt, Projekt	ca. 10	Darstellung von CRISP-DM [WH00] zur Orientierung im Projektablauf

Tab. 1: Übersicht über die Anwendungsfälle in den Pilotstudien

Softwaretechnik 1: In dieser Pilotstudie standen die Aktivitäten der Erstellung eigener UML-Diagramme im Vordergrund, was eine zusätzliche Unterstützung auf den Unterseiten in Form einer Analyse von UML-Diagrammen mit zugehörigem Feedback gefordert hat. Von der ersten Iteration [RSH24] hin zur zweiten wurde das Feedback um verschiedene Darstellungsformen und -stufen erweitert. Basierend auf den Evaluationsergebnissen der ersten Iteration wurde die Anwendung stärker in die Lehrveranstaltung eingebunden und durch eine Anleitung zur Handhabung der Webanwendung ergänzt, was zu einer höheren Verwendung des Scaffolds seitens der Studierenden geführt hat.

Seminar Trends der Softwaretechnik: Der Prozess wurde mithilfe des Erstellungswerkzeugs auf Basis der Struktur der Lehrveranstaltung selbstständig und in kurzer Zeit durch die Lehrperson erzeugt. Das Scaffolding-System wurde von Teilnehmenden vereinzelt verwendet, da die Veranstaltung in der Vorgehensweise bereits umfassend unterstützt war, und eine detailliertere Beschreibung der Aktivitäten durch Unterseiten gewünscht.

Projekt Data Science: In intensiver Rücksprache mit den Lehrpersonen wurde das CRISP-DM-Modell spezifiziert, um den Studierenden, die im Rahmen der Lehrveranstaltung in kurzen Iterationen arbeiten sollten, eine Orientierung für das Gesamtvorgehen zu bieten. Im Nachgang der Veranstaltung identifizierten die Lehrpersonen mit uns gemeinsam weitere Anforderungen zur Visualisierung der Bearbeitungsreihenfolge von Prozessschritten.

Neben zusätzlichen bereits entwickelten Szenarien sind weitere Iterationen der Pilotstudien geplant und werden auch über die Projektlaufzeit hinaus fortgesetzt.

4 Ergebnisse und Vermächtnis des Projektes

Aufgrund des hohen Nutzungspotenzials der Resultate liegt sowohl für das Gesamtprojekt IPPOLIS als auch das Teilprojekt Scaffolds nahe, die entwickelten Ansätze, Verfahren und Werkzeuge nachhaltiger Nutzung in der Hochschulpraxis zuzuführen. Das „Vermächtnis“ des Teilprojektes, das momentan geplant und realisiert wird, erstreckt sich auf:

Software: Die entwickelten Webanwendungen für das Scaffolding-System und das Erstellungswerkzeug sollen frei zugänglich sowohl als *Code* als auch als *ausführbare Software* für die jeweiligen Nutzungsgruppen zur Verfügung stehen. Der Programmcode der Anwendungen soll im FLOSS-Ansatz (Free/Libre Open Source Software) interessierten Entwickelnden zum Projektende zur Weiterentwicklung bereitgestellt werden. Die entsprechende Plattform wird momentan einheitlich für alle zu veröffentlichen Code-Teile des gesamten Projektes bestimmt. Das in der Softwaretechnik-Veranstaltung etablierte Scaffolding-System soll dort auch zukünftig im Praktikum eingesetzt werden. Dazu wird aktuell ein Hosting durch eine von der Hochschule bereitgestellte Infrastruktur angestrebt, die bereits seit einigen Jahren eine verstetigte und weiterhin nachgefragte Nutzung eines anderen - bereits seit 2022 beendeten – Forschungsprojektes FALEDIA[6] erfolgreich anbietet. Somit soll den

[6] https://faledia.de

Lehrenden eine stabile lauffähige Version unter erfahrener Wartung zur Verfügung stehen. **Evaluationsinstrument:** Das im Rahmen unserer Pilotstudien konzipierte und im Verlauf weiterentwickelte Evaluationsinstrument in Form eines Fragebogens ist nach einer weiteren Evaluation im Sommersemester 2025 und finaler Revision zum Zeitpunkt der Konferenz öffentlich auf unserer Webseite und in einem OER-Archiv unter Creative Commons BY-Lizenz bereitgestellt. Der im Praxiseinsatz genutzte Fragebogen besteht neben Fragen zur allgemeinen Verwendung des Scaffolding-Systems aus Fragen zur Zielerreichung [Zi00], Selbstwirksamkeit [Ba97] und – außerhalb des CC-BY-Teiles – User Experience [LHS08]. **Partizipatives Vorgehen zur Entwicklung von Scaffolds:** Bei der ausgiebigen Diskussion mit den Lehrenden der verschiedenen Lehrveranstaltungen sowohl vor, während als auch nach den Pilotstudien, ergaben sich wertvolle Ideen aus der Praxis für neue Funktionalitäten und Anforderungen bezüglich der Darstellung und Visualisierung von Prozessinformationen, die in die (Weiter-)Entwicklung und Revision unserer Webanwendung einflossen. **Erstellung von fachspezifischen Inhalten:** Ein begrenzender Faktor bei der Erstellung von fachspezifischen Scaffolds ist der hohe Aufwand, den die Erarbeitung von detailliertem Feedback und domänenspezifischen Inhalten mit sich bringt. Die Erstellung von Unterseiten ergab in der Softwaretechnik-Studie hohen Implementierungsaufwand für die UML-Analysen und die Spezifikation von Dashboard-Indikatoren. Diese Abwägung von domänenspezifischem Aufwand gegenüber des Nutzens für die Wirksamkeit des Prozess-Scaffolds sollte bei der Planung vorgenommen und mit den Lehrenden diskutiert werden.

Aufgrund der sehr praktischen Natur unseres Gesamtprojektes mit diversen Unterstützungsmaßnahmen für die Hochschulpraxis innerhalb und außerhalb von Lehrveranstaltungen ist uns die potenzielle Weiternutzung der entwickelten Ansätze sehr wichtig und nimmt eine wesentliche Rolle in unserem aktuell laufenden finalen Projektjahr ein.

5 Fazit und Ausblick

Im IPPOLIS-Teilprojekt Scaffolds haben wir das erschaffene Scaffolding-Framework als flexibles Instrument zur Prozessunterstützung eingesetzt, was durch die Spannbreite der Pilotstudien in sehr unterschiedlichen Lehrveranstaltungsformen gezeigt werden konnte. Durch den Einsatz in authentischer Lehre konnten wir sowohl das technische System als auch das Evaluationsinstrument so validieren und weiterentwickeln, dass sie weitergenutzt werden können und sollen, was wir durch eine gezielte Bereitstellung unserer Hinterlassenschaft sowohl in unserer Hochschule als auch für die Hochschul-Allgemeinheit ermutigen wollen.

Insgesamt stellen wir fest, dass die Kooperation mit (projekt-)externen Lehrenden sehr positive Wirkung auf die Entwicklung des Scaffolding-Frameworks hatte und die Einsatzmöglichkeiten in anderen Domänen deutlich erhöht hat, wobei die bisherigen Pilotstudien in technischen Fächern stattfanden. Die maßgeschneiderte Unterstützung innerhalb der Fachdomäne ist allerdings entweder begrenzt oder mit enormem Erstellungsaufwand verbunden. Lose Kopplung von Domänenwerkzeugen, wie z. B. Datenanalyse-Frameworks oder Computer-Algebrasystemen, und Nutzung ihrer Kennzahlen könnten hilfreich sein.

Literaturverzeichnis

[Ba97] Bandura, A.: Self-Efficacy: The Exercise of Control. Worth Publishers, 1997.

[In19] International Organization for Standardization: ISO 9241-210:2019 Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems, https://www.iso.org/standard/77520.html, Accessed: 2025-05-22, 2019.

[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, DOI: 10.1007/978-3-540-89350-9_6.

[Qu04] Quintana, C. et al.: 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, DOI: 10.4324/9781315685779-12.

[RSH24] Reinartz, L.; Schneeweiß, S.; Harrer, A.: Design, Realisierung und Evaluation von Scaffolds zur Unterstützung von Lernprozessen im Hochschulkontext. In (Schulz, S.; Kiesler, N., Hrsg.): DELFI 2024 - Die 22. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V., DELFI 2024, Fulda, Germany, September 9-11, 2024. Bd. P-356. LNI, Gesellschaft für Informatik e.V., 2024, DOI: 10.18420/DELFI2024_35.

[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, DOI: 10.1111/j.1469-7610.1976.tb00381.x.

[WH00] Wirth, R.; Hipp, J.: CRISP-DM: Towards a standard process model for data mining. In: Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining. Bd. 1, Manchester, S. 29–39, 2000.

[Zi00] Zimmerman, B. J.: Chapter 2 - Attaining Self-Regulation: A Social Cognitive Perspective. In (Boekaerts, M.; Pintrich, P. R.; Zeidner, M., Hrsg.): Handbook of Self-Regulation. Academic Press, San Diego, S. 13–39, 2000, DOI: 10.1016/B978-012109890-2/50031-7.

Auf dem Weg zur Umsetzung der Open Educational Resources Strategie (OER-Strategie)

von OER zu einem OEP-förderlichen digitalen Ökosystem

Nadja Dietze [1] und Caroline Surmann[2]

Abstract: Der Beitrag thematisiert gegenwärtige Bemühungen auf Bundesebene im Kontext offener Bildungsressourcen (Open Educational Resources, OER) und offener Bildungspraktiken (Open Educational Practices, OEP). Im Fokus stehen drei zentrale Herausforderungen, die wesentliche Potenziale von OER und OEP zur Förderung von Kooperation und Kollaboration zwischen Bildungseinrichtungen, Lehrenden und Lernenden adressieren. Erstens reflektiert der Beitrag das Verhältnis der Konzepte OER und OEP und betont dabei die vorherrschende ressourcenorientierte Perspektive auf OER, während die Prozessperspektive von OEP bisher nur begrenzt berücksichtigt wird. Zweitens beleuchtet er die Komplexität und diskutiert die pädagogische Wirksamkeit kooperativer und kollaborativer Lehr-Lern-Ansätze aus sozial-konstruktivistischer Sicht. Drittens adressiert der Beitrag bestehende infrastrukturelle Lücken, die eine nachhaltige Implementierung offener Bildungspraktiken erschweren. Der Beitrag plädiert für eine Erweiterung der Perspektive: (1) von OER zu OEP, (2) von Instruktion zur Ko-Kreation und (3) von Insellösungen zu einem vernetzten Ökosystem, um neue Formen pädagogischer Praxis im Bildungssektor zu etablieren.

Keywords: Open Educational Resources, Open Educational Practices, OEP-förderliches Ökosystem, OER-Strategie

1 Einleitung

Open Educational Resources (OER) wird das Potenzial zugeschrieben, Kollaboration und Kooperation in der Bildungspraxis zu fördern, Kompetenzen zu stärken und neue pädagogische Ansätze zu unterstützen – insbesondere im Kontext einer zunehmend digitalen Lebens- und Arbeitswelt [BM22; Ga24a]. In Anlehnung an die meist zitierte Definition von der UNESCO werden unter OER in der Regel Bildungsmaterialien verstanden, die entweder gemeinfrei oder unter einer offenen Lizenz veröffentlicht werden [UN19]. Dieser Definition liegen die fünf Aktivitäten von David Wiley zugrunde, die im deutschsprachigen Raum als sogenannte „5V-Freiheiten“ bezeichnet werden [Be19]. Diese Freiheiten – das Verwahren, Verwenden, Bearbeiten, Vermischen und Verbreiten von offenen

[1] Deutsches Zentrum für Luft- und Raumfahrt e. V. (DLR), DLR Projektträger, Digitalisierung in der Bildung, Heinrich-Konen-Straße 1, 53227 Bonn, Nadja.Dietze@dlr.de, https://orcid.org/0009-0008-9418-9508
[2] Deutsches Zentrum für Luft- und Raumfahrt e. V. (DLR), DLR Projektträger, Digitalisierung in der Bildung, Heinrich-Konen-Straße 1, 53227 Bonn, Caroline.Surmann@dlr.de

doi:10.18420/delfi2025_20

Bildungsressourcen – sollen nicht nur flexiblere individuelle Einsatzmöglichkeiten eröffnen, sondern auch zugleich zur Umsetzung offener und kollaborativer Lehr- und Lernformate beitragen.

Mit der ersten bildungsbereichübergreifenden Förderrichtlinie „Open Educational Resources – OERinfo“ (2016–2020) setzte das BMBF einige wichtige Impulse im Bildungsbereich und identifizierte weitere Handlungsbedarfe [Su18]. Mit der Veröffentlichung der OER-Strategie im Jahr 2022 knüpft das BMBF an diese Entwicklungen an und schafft einen Rahmen, um OER und OEP als zentrale Elemente einer modernen Bildung langfristig zu verankern [BM22].

Im vorliegenden Artikel werden die aktuellen Entwicklungen im Bereich OER und OEP anhand der drei Kernherausforderungen skizziert: (1) das Verhältnis der Konzepte OER und OEP, (2) die Komplexität kooperativer und kollaborativer Lehr-/Lernstrategien aus sozial-konstruktivistischer Sicht sowie (3) bestehende Insellösungen im Bereich offener Bildungsinfrastrukturen.

2 Von Open Educational Resources (OER) zu Open Educational Practices (OEP)

Trotz der zunehmenden Verbreitung von Open Educational Resources (OER) bleibt der Bereich der Open Educational Practices (OEP) in der wissenschaftlichen Diskussion nach wie vor relativ unterrepräsentiert [Be19]. Bellinger weist in diesem Zusammenhang auf das sogenannte „Henne-Ei-Problem“ hin: Es bleibt unklar, inwieweit die Etablierung von OEP von der Verfügbarkeit von OER abhängt oder ob OER erst durch die Umsetzung von OEP tatsächlich ihre volle Wirkung entfalten können [ebd.].

Mayrberger (2019) hebt hervor, dass der stärkere Fokus auf OER nicht überraschend ist, da die Material- bzw. Produktperspektive – also die Betrachtung von OER als konkretes, zugängliches Produkt – deutlich einfacher zu fassen ist. Im Gegensatz dazu erfordert die Prozessperspektive, die auf die Entwicklung und Umsetzung von OEP abzielt, tiefere Reflexionen und komplexe Handlungsansätze. Während die Diskussion zu OER als materielles Produkt zunehmend an Bedeutung gewonnen hat, ist die Realisierung von OEP eine größere Herausforderung [Ma19].

Die Definition von OEP variiert stark, da je nach Perspektive unterschiedliche Schwerpunkte gesetzt werden. Bellinger und Mayrberger (2019) identifizieren vier Dimensionen von OEP, die deren Vielschichtigkeit verdeutlichen [Be19]:

- OEP im weitesten Sinne: Vollständige Übernahme der Prinzipien einer offenen Pädagogik inkl. institutionellen Rahmenbedingungen und Governance.
- OEP im weiteren Sinne: Offene, partizipative Lernszenarien und Nutzung von OER auf Mikroebene.

- OEP im engeren Sinne: Nutzung und Remix von OER gemäß den „5R"-Prinzipien (Retain, Reuse, Revise, Remix, Redistribute).
- OEP im engsten Sinne: Einsatz von OER in spezifischen Lernsettings (z. B. MOOCs) zur Förderung kollaborativer und partizipativer Lernmethoden.

In diesem Beitrag liegt der Schwerpunkt auf den bildungspolitischen und wissenschaftlichen Diskussionen, die einen Übergang von der rein produktorientierten Perspektive von OER hin zu offenen Bildungspraktiken und einer Kultur des Teilens fördern [Be19; Bo19; BM22]. Der Begriff „Open Educational Practices" umfasst dabei sowohl die Nutzung von OER in Lehr- und Lernszenarien als auch kollaborative und partizipative Lehr- und Lernansätze mit sozialen Medien.

Gantenbrink und Tibbe (2024) weisen jedoch darauf hin, dass die Diskussionen um OER-Praktiken häufig auf deren Nutzung fokussiert sind, während die Auswirkungen des Einsatzes von OER und OEP im Hinblick auf kollaborative Zusammenarbeit bislang weitgehend unerforscht bleiben. Einige Studien stellen in Nebenbemerkungen fest, dass kollaborative Lehr- und Lernszenarien zeitintensiv sind und häufig zunächst erlernt werden müssen [Ne18]. Diese Beobachtung spiegelt zentrale Annahmen des sozial-konstruktivistischen Lernverständnisses sowie Erkenntnisse der empirischen Lernstrategieforschung wider: Die bloße Verfügbarkeit von OER fördert weder partizipative Wissensaneignung noch kollaborative Interaktionen. Vielmehr bedarf es gezielter Impulse und inspirierende Lehr-/Lernarrangements, um Lerngelegenheiten zu ermöglichen und Lernende bei der aktiven Auseinandersetzung mit offenen Bildungsmaterialien zu begleiten. Die Erweiterung der Produktperspektive um die Prozessperspektive, die über die reine Nutzung von OER hinausgeht und kollaborative Lehr- und Lernansätze mit sozialen Medien initiiert, ist ein wesentlicher Schritt, um OER in pädagogisch tragfähige, kollaborative Bildungssettings zu integrieren.

3 Von Instruktion zur Ko-Konstruktion und Ko-Kreation

Der Paradigmenwechsel der letzten Jahrzehnte beschreibt den Weg vom Instruktionismus zum Konstruktivismus, der auf kognitivistischen und konstruktivistischen Ansätzen in ihren unterschiedlichen Varianten beruht. Konrad hebt hervor, dass es in der vergleichsweise kurzen Entwicklung der Lerntheorien bislang kaum eine Zeit gab, in der so viele theoretische Ansätze in ihren grundlegenden Annahmen übereinstimmten [Ko04]. Zu diesen zählt die Auffassung, dass Lernen ein aktiver, konstruktiver, kumulativer, selbstgesteuerter und sozial eingebetteter Prozess ist. Zentrale Bedeutung kommt dabei dem sozialen Kontext sowie dem Konzept der Ko-Konstruktion von Wissen zu, das Konrad als eine Form der kollaborativen Aktivität versteht, die Partizipation fördert und kognitive Veränderungen anstößt, wenn Lernende gemeinsam Probleme bearbeiten [ebd.].

Das Konzept der Ko-Konstruktion wird häufig synonym mit dem Begriff der Ko-Kreation verwendet [Sa25]. Ko-Kreation, ursprünglich in den Wirtschaftswissenschaften verortet,

beschreibt einen kollaborativen und kooperativen Ansatz, der gleichberechtigte Zusammenarbeit, Perspektivenvielfalt und die aktive Einbindung weiterer Zielgruppen umfasst [ebd.]. Im Zentrum steht dabei die Annahme, dass durch ergebnisoffene und kreative Zusammenarbeit qualitativ bessere Lösungen bzw. Ergebnisse entstehen. Im Kontext von OER und OEP bezieht sich dies auf die ko-kreative Entwicklung von Bildungsmaterialien durch Lehrende und Lernende, bei der Inhalte gemeinsam erstellt und zur freien Weiterverwendung bereitgestellt werden [Rö23]. Besonders die ko-kreative Beteiligung von Lernenden an der Erstellung offener Bildungsmaterialien gilt aus sozial-konstruktivistischer Perspektive als zentraler Ansatzpunkt zur Förderung partizipativer Bildungsprozesse. Im Fokus steht hierbei weniger das entstandene OER als Produkt, sondern vielmehr der soziale, partizipative Prozess der Ko-Kreation sowie der damit verbundene individuelle Kompetenzerwerb der Lernenden.

Beide Konzepte werden als kollaborative und kooperative Ansätze verstanden, begründen aus sozial-konstruktivistischer Sicht jedoch nicht automatisch die Wirksamkeit sozialer Interaktionen. Unumstritten bleibt, dass kollaborative und kooperative Lernformen in allen Bildungsbereichen das Potenzial besitzen, den individuellen Lernerfolg positiv zu beeinflussen [Dä24]. Ob Kooperations- und Kollaborationsprozesse tatsächlich lernförderlich sind, hängt von einer Vielzahl von Faktoren ab. Für den erfolgreichen Einsatz solcher Lernarrangements sind bei den Lernenden spezifische Fähigkeiten für die kollaborative Zusammenarbeit erforderlich, deren Vorhandensein nicht selbstverständlich ist [ebd.].

Offene Bildungspraktiken bieten zwar besondere Potenziale für Kollaboration und Kooperation, doch kollaborative und kooperative Lernstrategien sind nicht automatisch vorhanden, sondern müssen gezielt erlernt werden. Damit kollaborative Lernformen tatsächlich zu verbesserten Lernergebnissen führen, sind ergänzende Förderansätze erforderlich, die Lehrende sowie Lernende beim Erwerb ko-konstruktiver und ko-kreativer Lern- und Arbeitsstrategien unterstützen. Die Konzepte der Ko-Konstruktion und Ko-Kreation erweisen sich dabei als aufwändige und zeitintensive Lehr-/Lernformate, deren potenzielle Effekte den damit verbundenen Aufwand rechtfertigen.

4 Von Insellösungen zu einem vernetzten Ökosystem

Die ko-konstruktive und ko-kreative Ansätze in offenen Bildungspraktiken erfordern nicht nur spezifische Kompetenzen seitens der Lehrenden und Lernenden, sondern stellen auch hohe Anforderungen an digitale Bildungsinfrastrukturen. Einige Studien zeigen, dass die Nutzung von OER sowie deren Integration in OEP, insbesondere im Hinblick auf die 5V-Freiheiten, noch nicht in allen Bereichen ausreichend technisch unterstützt wird [Ga24b; Rö21]. Diese Lücken in den technischen Infrastrukturen können die kollaborative Zusammenarbeit erheblich einschränken und die Potenziale offener Bildungsressourcen nicht vollständig ausschöpfen. Röwert und Kostrzewa [Rö21] heben die besonderen Anforderungen an Plattformen hervor, die es den Lernenden und Lehrenden ermöglichen

sollen, OER-Inhalte eigenständig zu erstellen und zu bearbeiten. Als wichtiger Aspekt werden hier Schnittstellen genannt, die als kritische Erfolgsfaktoren für die skalierbare Nutzung von OER gelten [Rö21].

Vor diesem Hintergrund engagieren sich zahlreiche Initiativen auf Bundes- und Landesebene sowie informelle Zusammenschlüsse für die Weiterentwicklung offener Bildungsinfrastrukturen. Der vorliegende Beitrag konzentriert sich auf aktuelle Maßnahmen des Bundes, insbesondere auf die OER-Strategie, die als lernender Ansatz in Zusammenarbeit mit Akteuren verschiedener Bildungsbereichen weiterentwickelt wird [BM24a; BM24b]. Dabei werden laufend neue thematische Schwerpunkte identifiziert und in interdisziplinären Fachgesprächen vertieft. Die Weiterentwicklung offener Bildungsinfrastrukturen wurde in diesen Gremien wiederholt betont. Aus diesem Grund fand am 28. November 2023 ein Fachgespräch zum Thema „Offene Bildungsinfrastrukturen" statt, bei dem, auf Grundlage vorbereitender Interviews, infrastrukturelle Lösungen aus unterschiedlichen Perspektiven beleuchtet wurden [BM24c].

Die Experten und Expertinnen des Fachgesprächs waren sich einig, dass die bestehende Vielfalt an OER-Repositorien und -Referatorien, Informations- und Weiterbildungsportale sowie Installationen von Lernmanagement-Systemen (LMS) mit OER sowohl mit Chancen als auch mit Herausforderungen verbunden ist. Einerseits existieren vielfältige Lösungsansätze für unterschiedliche Bedarfe und Präferenzen der Zielgruppen, d. h. es gibt eine hohe Passgenauigkeit. Andererseits erschwert eine solche Vielfalt oftmals die Auffindbarkeit, die Bearbeitungsmöglichkeiten und die partizipative Gestaltung von OER in OEP.

Um den genannten Herausforderungen zu begegnen, verfolgt das BMBF einen dezentralen Ansatz, Vernetzung bestehender Lösungen im Bereich offener Bildungsinfrastrukturen abzielt. Interoperable Bildungsinfrastrukturen stellen dabei eine grundlegende Voraussetzung dar, um die Auffindbarkeit, Nutzung und Weiterentwicklung von OER zu erleichtern und ein dynamisches Ökosystem zur Unterstützung von OEP zu schaffen. Im Rahmen dieses Ansatzes plant das BMBF, mit Förderinitiativen nicht nur die technische Interoperabilität zu stärken, sondern auch didaktische, rechtliche und organisatorische Dimensionen zu berücksichtigen. Ein Beispiel hierfür ist der Ideenwettbewerb „OE_Space", der im Mai 2025 testweise für Teilnehmende der Konferenz „OER im Blick"[3] gestartet wurde. Die thematischen Schwerpunkte bzw. die adressierten Herausforderungen (Challenges) wurden im Vorfeld in einem Expertengespräch identifiziert und durch die OER-Community auf der Konferenz „OER im Blick" im Jahr 2024 ergänzt [BM24c]. Dabei stehen folgende fünf Challenges im Fokus:

1. **Didaktik & Usability:** Wie können offenen Bildungsinfrastrukturen und digitale Tools sinnvoll in den Lehr- und Lernprozess integriert werden, um den Bildungsalltag bestmöglich zu unterstützen?

[3] Weitere Informationen zur Konferenz „OER im Blick" und zum Ideenwettbewerb „OE_Space" unter: https://www.oer-strategie.de/konferenz/dokumentation-der-statuskonferenz-2025. Stand: 04.03.2025.

2. **Automatisierung & Integration:** Wie können automatisierte Prozesse und verbundene Systeme den Bildungsalltag erleichtern? Wie können Lösungen zur reibungslosen Integration unterschiedlicher Systeme aussehen?
3. **Aktualität & Qualität:** Wie bleibt das Lehr-/Lernmaterial immer auf dem aktuellsten Stand und datenschutzkonform? Welche Mechanismen zur Sicherung der Qualität und Relevanz von OER/OEP bieten sich an?
4. **Tradition & Innovation:** Was hat sich bewährt und sollte in andere Bereiche übertragen werden, und wo ist Veränderung nötig, um bestehende Bildungsinfrastrukturen OEP- und OER-förderlich zu gestalten?
5. **Change Management & Kollaboration:** (Wie) Können innovative Bildungsinfrastrukturen die Kultur des Teilens und die OEP unterstützen?

Im Rahmen der Testphase sollen bis zu zehn kreative Projekte ausgewählt werden, die im Zeitraum vom 1. September 2025 bis zum 30. November 2025 finanzielle Unterstützung für die Umsetzung ihrer Ideen erhalten. Nach Abschluss der Testphase sind in den kommenden Jahren drei weitere Förderaufrufe (Calls) mit jeweils sechsmonatiger Laufzeit für Projekte geplant. Der Ideenwettbewerb OE_Space stellt eine von mehreren Fördermaßnahmen im Kontext der OER-Strategie dar und trägt dazu bei, infrastrukturelle Lücken zu schließen sowie die Entwicklung eines digitalen Ökosystems zu fördern, das offene Bildungspraktiken unterstützt.

5 Zusammenfassung und Ausblick

Die Etablierung eines OEP-förderlichen digitalen Ökosystems erfordert die enge Zusammenarbeit verschiedenster Akteure im Bildungsbereich. Die OER-Strategie des Bundes stellt dabei einen wichtigen Rahmen dar, um die Potenziale von OER und OEP systematisch zu integrieren. Der vorliegende Beitrag ist ein Plädoyer für:

- die Erweiterung der Produktperspektive um die Prozessperspektive im Kontext von OER/OEP, um Lernanlässe zu initiieren und Lernende bei der Auseinandersetzung mit offenen Bildungsmaterialien zu unterstützen.
- die Weiterentwicklung sozialer Formate der Ko-Konstruktion und Ko-Kreation im Kontext offener Bildungspraktiken, um durch partizipative Handlungsansätze eine tiefere Wissenskonstruktion zu ermöglichen und die Kultur des Teilens zu fördern.
- die Vernetzung bestehender Insellösungen sowie deren Ausbau um OEP-förderliche Komponenten, um ko-konstruktive und ko-kreative Prozesse im Sinne der offenen Bildungspraktiken zu unterstützen und Medienbrüche zu vermeiden.

Ein digitales Ökosystem für offene Bildungspraktiken weiterzuentwickeln heißt, technische, didaktische, rechtliche und organisatorische Perspektiven zu verbinden. In dieser Vielfalt liegt die Chance, nachhaltige Strukturen für partizipative Bildungsprozesse zu schaffen – ein aufwändiger Prozess, dessen Potenzial den Einsatz rechtfertigt.

Literaturverzeichnis

[Be19] Bellinger, F.; Mayrberger, K.: Systematic Literature Review zu Open Educational Practices (OEP) in der Hochschule im europäischen Forschungskontext. In: MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung 34 (Research and Open Educational Resources – A snapshot on Europe), S. 19-46, 2019. DOI: 10.21240/mpaed/34/2019.02.18.X

[BM22] BMBF. Bundesministerium für Bildung und Forschung: OER-Strategie – Freie Bildungsmaterialien für die Entwicklung digitaler Bildung. 2022. Verfügbar unter: https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/3/691288_OER-Strategie.pdf?__blob=publicationFile&v=6. Stand: 04.03.2025.

[BM24a] BMBF. Bundesministerium für Bildung und Forschung: Impulsgremium – Futurium für Zukunftsperspektiven. 2024. Verfügbar unter: https://www.oer-strategie.de/aktuelles/impulsgremium-futurium-fuer-zukunftsperspektiven/. Stand: 04.03.2025.

[BM24b] BMBF. Bundesministerium für Bildung und Forschung: „Wo der Schuh drückt" – Aktuelle Aufgabensetzung des OER-Beirats. 2024. Verfügbar unter: https://www.oer-strategie.de/aktuelles/wo-der-schuh-drueckt-aktuelle-aufgabensetzung-des-oer-beirats/. Stand: 04.03.2025.

[BM24c] BMBF. Bundesministerium für Bildung und Forschung: Gemeinsam zukunftsfähige OER-förderliche Bildungsinfrastrukturen gestalten. 2024. Verfügbar unter: https://www.oer-strategie.de/aktuelles/gemeinsam-zukunftsfaehige-oer-foerderliche-bildungsinfrastrukturen-gestalten/. Stand: 04.03.2025.

[Bo19] Bock, A.; Tribukait, M.: Kultur des Teilens. Ein kritischer Blick auf ein zentrales Konzept der OER-Bewegung. In: MedienPädagogik 34, S. 47-66, 2019. Verfügbar unter: https://www.medienpaed.com/article/view/642/652. Stand: 04.03.2025.

[Dä24] Dählin, C.; Standop, J.: Kollaboratives Lernen lernen. Seminarsitzung zum Pre-Training von Kompetenzen des kollaborativen Lernens. In: Die Materialwerkstatt, 6(1), S. 1–7, 2024. DOI: 10.11576/dimawe-7002

[Ga24a] Gantenbrink, S.; Tibbe, T.: Kollaborative Nutzung und Weiterentwicklung von Open Educational Resources. Eine Studie zu Gelingensbedingungen am Beispiel der inklusionsbezogenen Lehrer:innenbildung in NRW im Projekt InDigO. MedienPädagogik 62 (InDigO), S. 163-186, 2024. DOI: 10.21240/mpaed/62/2024.07.10.X

[Ga24b] Gantenbrink, S.; Knurr, M.: Wo muss ich klicken? Gestaltungsprinzipien für ein OEP-förderliches digitales Ökosystem. MedienPädagogik 62 (InDigO), S. 41-67, 2024. DOI: 10.21240/mpaed/62/2024.07.03.X

[Ko04] Konrad, K.: Förderung und Analyse von selbstgesteuertem Lernen in kooperativen Lernumgebungen: Bedingungen, Prozesse und Bedeutung kognitiver sowie metakognitiver Strategien für den Erwerb und Transfer konzeptuellen Wissens. Habilitationsschrift, Pädagogische Hochschule Weingarten, 2004.

[Ne18] Neuschäfer, M.; Kolbe, C.; Voigt, M.: edulabs. In: Projekte der BMBF Förderung OERinfo 2017/2018. Sonderband zum Fachmagazin Synergie. Universität Hamburg, 2018, S. 222–231.

[Rö21] Röwert, R.; Kostrzewa, M.: Phasenübergreifende Lehrkräftebildung mit und durch OER – Thesen und Impulse für eine vernetzte Lehrkräftebildung. In: C. Gabellini, S. Gallner, F. Imboden, M. Kuurstra & P. Tremp (Hrsg.): Lehrentwicklung by Openness – Open Educational Resources im Hochschulkontext. Dokumentation der Tagung vom 06. März 2021.

[Rö23] Röwert, R.: Offenheit im Bildungsbereich: Keine OER ohne Open Pedagogy? 2023. Verfügbar unter: https://open-educational-resources.de/keine-oer-ohne-open-pedagogy. Stand: 10.06.2025.

[Sa25] Sars, P.L.M.; Jentges, S.: „Wir haben alles selbst machen müssen“ – Ko-Kreation in internationalen Lehr-Kooperationen zur Materialentwicklung im schulischen und universitären Kontext. In (Jentges, S.; Siebold, K.; Adamczak-Krysztofowicz, S., Hrsg.): Internationale Bildungskooperation im Bereich Deutsch als Fremdsprache. State of the Art & Perspektiven. Narr Francke Attempto, Tübingen, S. 273–292, 2025.

[Su18] Surmann, C.; Echterhoff, C.: OERinfo – Die Richtlinie zur Förderung von offenen Bildungsmaterialien. In: Projekte der BMBF-Förderung OERinfo 2017/2018. Sonderband zum Fachmagazin Synergie, S. 10–17, 2018.

[UN19] UNESCO: Guidelines on the Development of Open Educational Resources Policies. UNESCO & Burnaby, BC, Canada: Commonwealth of Learning, 2019.

Ein Data Warehouse für die Zusammenführung von Hochschuldaten



Sven Judel [1], Erik Schnell[2] und Ulrik Schroeder [1]

Abstract: Dieser Beitrag präsentiert die Konzeption und Umsetzung eines Data Warehouses zur gemeinsamen Speicherung von Studiengangsmodellen und Studienverlaufsdaten verschiedener Hochschulen. Die aufzunehmenden Daten wurden zuvor in einem Datenreferenzmodell aufbereitet, welches dem System die Struktur der Datenverwaltung diktiert. Zur Erstellung und Bearbeitung der Studiengangsmodelle werden API Endpunkte bereitgestellt, welche u.a. von einem dedizierten Editor genutzt werden. Studienverlaufsdaten, die Hochschulen aus ihren lokalen Managementsystemen extrahieren und selbst in das Datenreferenzmodell überführen, werden als Datendump an eine Import-Schnittstelle geschickt. Zur zusätzlichen Sicherung der Datenübertragung an das Data Warehouse werden Zertifikate bereitgestellt, die Hochschulen für GnuPG Verschlüsselungen nutzen können. Das System wurde mit einem Zwischenstand und der finalen Version des erstellten Datenreferenzmodells erfolgreich auf Korrektheit und gute Performance getestet und mit Projektabschluss Open Source veröffentlicht.

Keywords: Data Warehouse, Studienverlaufsdaten, Curriculum Analytics, Learning Analytics

1 Einleitung und Konzept

Hochschulsysteme, wie Campus- oder Prüfungsmanagementsysteme halten Studierendendaten in verschiedenen Formaten und Strukturen fest, welche nicht immer schnell zusammenführbar sind. Eine vereinheitlichte Repräsentation dieser Daten birgt jedoch den Vorteil, die Daten mit den gleichen Analysen, teilweise sogar gemeinsam, verarbeiten zu können und damit den Aufwand der Analyseerstellung zu reduzieren und Ergebnisse zwischen Hochschulen, in ihrem jeweiligen Kontext, vergleichbar zu machen. Dieses Ziel wurde im Projekt AIStudyBuddy verfolgt, in dem Studiengangsmodelle und Studienverlaufsdaten von drei Universitäten in einem einheitlichen Modell an einer zentralen Stelle zusammengeführt und anschließend verarbeitet werden. Dafür wurde ein zentrales Data Warehouse (DWH) geschaffen, in dem die Daten gesammelt und abgerufen werden können und dessen Anforderungen, Konzeption und Umsetzung im Folgenden vorgestellt werden.

Da alle Universitäten unterschiedliche Managementsysteme (MS) mit unterschiedlichen Datenmodellierungen nutzen, war die Übernahme eines dieser Modelle, welche zudem stets mehr Daten, als für das Projekt notwendig, enthalten, keine Option. Stattdessen wurde ein Data Reference Model (DRM) geschaffen, in welches die lokalen Daten zu überführen

[1] RWTH Aachen, Lehr- und Forschungsgebiet Lerntechnologien, Ahornstraße 55, 52074 Aachen, judel@cs.rwth-aachen.de, https://orcid.org/0000-0002-0424-0402; schroeder@cs.rwth-aachen.de, https://orcid.org/0000-0002-5178-8497
[2] RWTH Aachen, Templergraben 55, 52062 Aachen, erik.schnell@rwth-aachen.de

doi:10.18420/delfi2025_21

sind, ehe sie an das DWH geschickt werden. In Anlehnung an das Extract-Transform-Load Verfahren beziehen die Universitäten die notwendigen Daten aus ihren MS (Extract), überführen sie in das DRM (Transform) und schicken anschließend die überführten Daten auf sicherem Wege an das DWH, welches die Daten prüft und im Anschluss speichert (Load). Im Projekt gemachte Erfahrungen zeigen, dass die Überführung der lokalen Daten in das DRM teils aufwendige Interpretationen von Daten erfordert, welche aufgrund von benötigtem Kontextwissen nicht an einer zentralen Stelle erfolgen können. Entsprecht wurden alternative Ansätze wie Extract-Load-Transform nicht umgesetzt. Auch war ein Data Lake keine Alternative zum DWH, da die zu speichernden Daten durch das DRM strukturiert sind und wenig Flexibilität benötigt wird [El22].

Die Datenquellen für das DWH umfassen die lokalen Managementsysteme, auf welche direkter Zugriff, Zugriff über eine API oder lediglich durch die Bereitstellung eines Datendumps benötigt ist, sodass die Daten transformiert und übertragen werden können. Zusätzlich wurde für die Studiengangsmodellierung im Projekt ein dedizierter Editor entwickelt, mit dem neue Modelle eingepflegt oder bestehende editiert werden können. Verarbeitet werden diese Daten durch verschiedene Komponenten, z. B. deskriptive Analysen, welche die Einschreibungen pro Semester bestimmen, sowie Techniken des Process Minings und regelbasierter KI. Im interaktiven Studienplaner *StudyBuddy* [Ju23] können Studienverlaufsdaten einzelner Studierender dargestellt werden. Abbildung 1 stellt diese Datenflüsse dar.

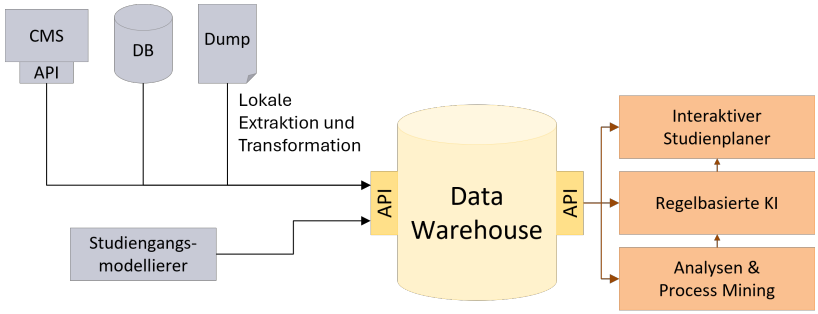


Abb. 1: Datenfluss um das Data Warehouse

Die Umsetzung des DWH umfasst entsprechend nicht nur das Bereitstellen eines Datenbankmanagementsystems, sondern auch Schnittstellen für den Datenfluss, für welche mehrere Anforderungen berücksichtigt werden mussten.

2 Anforderungsanalyse

Viele Anforderungen an das Data Warehouse leiten sich aus dem Data Reference Model (DRM) ab, welches im Projekt entworfen wurde [JJ25]. In diesem werden die projektrelevanten Daten zusammengebracht, sodass alle Arbeitspakete ihre Aufgaben auf der gemeinsamen Datenbasis ausführen können. Das inkrementell entwickelte DRM war in

seiner ersten Version auf Studienverlaufsdaten fokussiert, welche den Managementsystemen der Partner entnommen werden sollten. Alle betrachteten Systeme verwendeten relationale Datenbanken, sodass auch dieser Teil des DRM relational modelliert wurde. Mit der Erweiterung um die Studiengangsmodellierungen mussten Modelle integriert werden, welche dokumentenorientiert aufgebaut wurden. Eine Übertragung dieser Teile in ein relationales Modell wären möglich gewesen, der Aufwand einer Überführung dieser Daten zurück in eine dokumentenorientierte Struktur bei Datenabfrage aber groß. Da die zunächst relational modellierten Teile jedoch ohne Anpassungen des Modells auch dokumentenorientiert abgebildet werden konnten, wurde das gesamte DRM auf Dokumentenorientierung umgestellt. Die Auswirkungen auf die technische Umsetzung beschränken sich auf eine explizite Prüfung von Datentypen und Fremdschlüsselbeziehungen, da dies nicht mehr implizit durch eine relationales Datenbankmanagementsystem erfolgt.

Neben der Angabe von Datentypen wurden für einige Attribute Datenformate oder eine feste Wertemenge festgelegt. So sind z. B. Datumsangaben als String im Format `YYYY-MM-DD` anzugeben und der Hörerstatus als einer von fünf Werten, z. B. *Ersthörer*in* oder *Promovend*in* [JJ25]. Weitere Anforderungen wurden dem Projektplan selbst entnommen. Dieser sieht vor alle Projektergebnisse Open Source zu veröffentlichen. Entsprechen müssen die verwendeten Technologien damit vereinbar sein. Zudem muss das DWH selbst gehostet werden können, damit die zusammengeführten Daten weiterhin auf Servern in Deutschland liegen und verarbeitet werden. Abschließend muss das DWH sicherstellen, dass die empfangenen Daten der verschiedenen Universitäten keine ID-Konflikte erzeugen, wenn z. B. an zwei Standorten die gleiche ID für ein Modul verwendet wird.

3 Related Work und eigene Umsetzung

Während das Konzept von Data Warehouses technologieunabhängig ist [BO17], werden oft relationale Datenbankmanagementsysteme (DBMS) genutzt (vgl. [Or25; SA25]). Aufgrund der Dokumentenorientierung sind solche Implementierungen, wie z. B. Apache Hive [Ap25b] nicht für den Projektkontext geeignet. Die Nutzung von NoSQL Datenbanken ist noch vergleichbar jung und bringt einige Herausforderungen mit sich [PHV21], jedoch finden sich in der Literatur bereits Berichte über erfolgreiche Umsetzungen [Be23; BO17]. Die eigene Umsetzung orientiert sich an diesen Beispielen und den Technologieevaluationen.

Technologiewahl Als zentrale, dokumentenorientierte Datenbank wurde MongoDB gewählt. Bei dieser Entscheidung wurde berücksichtigt, dass dieses DBMS bereits in zwei anderen Komponenten der Projekt-Gesamtarchitektur genutzt wird. Geringe Latenzen sind bei Datenabfrage zu vernachlässigen, Konsistenz jedoch relevant, sodass ansonsten gleichwertige Alternativen wie Couchbase [Co25; Gu24] und CouchDB [Ap25a; Si24] keinen Mehrwert bieten, der eine Erweiterung des Projekt-Technologiestacks gut begründet hätte. Da das DWH nicht als *Database as a Service* angeboten wird, sondern als Aufnahme für Daten zur Analyse, erlegt die Server Side Public License [Mo18] der Veröffentlichung der DWH-Software keine Restriktionen auf [Sc21; Za22].

Data Warehouse und Import Pipeline Für die Datenaufnahme wird eine REST API nach gängigen Best Practices (vgl. [Mi25; Sm25]) erstellt und die Endpunkte für alle Entitäten der Studiengangsmodellierung bereitstellt, sodass diese abgerufen, erstellt und aktualisiert werden können. Diese Endpunkte sind die Schnittstellen des Studiengangsmodellierer und der Analysekomponente der regelbasierten KI. Der Modellierer betreibt Fehlerprävention und -handling im Nutzerinterface, die API prüft zusätzlich Fremdschlüssel und Datentypen mit geeigneten Bibliotheken (siehe unten). Zur Aufnahme der Studienverlaufsdaten stellt die API einen Endpunkt bereit, welcher eine ZIP Datei erwartet, in der die Daten jeder Entität der Studienverlaufsdaten als CSV Datei enthalten sind. Diese *Dumps* werden nach dem Einlesen entpackt und weiterverarbeitet. Der Prozess, dargestellt in Abbildung 2, beginnt mit einer HTTP Post Anfrage und dem *Preprocessing*. In diesem Schritt wird die Anfrage authentifiziert, die anfragende Universität anhand des Authentifizierungstoken identifiziert und, sofern keine Pipeline für diese Universität besteht, die erhaltene ZIP Datei entpackt.

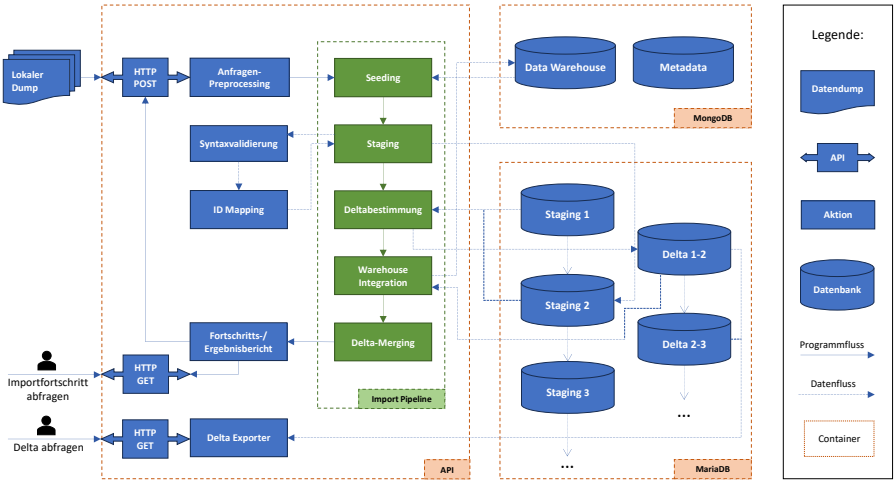


Abb. 2: Daten- und Programmfluss der Import Pipeline

Ist der Import der erhaltenen Daten erlaubt, wird im *Seeding* Schritt eine Staging Datenbank aufgesetzt. Diese Datenbank dient der Aufnahme der Daten in den extrahierten CSV Dateien und deren Überprüfung. Da die Studienverlaufsdaten relational modelliert sind, wird hierfür ein relationales DBMS genutzt. Mit dieser werden Fremdschlüssel sowie Datentypen impliziert beim Befüllen der Datenbank geprüft und bei Verletzungen des Schemas Fehlermeldungen ausgegeben, die aufbereitet und an den Client geschickt werden können. Diese Staging Datenbank wird für jeden Import einer Universität neu erstellt, um die parallele Bearbeitung mehrerer Imports zu ermöglichen und nicht für jede angeschlossene Universität eine permanente Datenbank aufrecht erhalten zu müssen. Als Technologie wurde hier zunächst MariaDB aufgrund der einfachen Handhabung und der vollständigen Open Source Bereitstellung im Vergleich zu ihrem Ursprung, MySQL, genutzt. Auch wenn

PostgreSQL als schneller beim Schreiben großer Datenmengen beschrieben wird [Ki24; Re24], erwies sich MariaDB in eigenen Evaluationen [JSS25] als performanter.

Im *Staging* Schritt werden die einzelnen CSV Dateien nacheinander in die zugehörigen Tabellen der Staging Datenbank geladen. Sollte an einer Stelle das CSV Format nicht mit dem der Tabelle übereinstimmen oder eine Fremdschlüsselbeziehung nicht bestehen, wird die Import Pipeline abgebrochen und eine Fehlermeldung an den Client ausgegeben. Teil des Stagings ist neben der Validierung von Datenstruktur und -typen das Mappen der Primärschlüssel jeder Entität. IDs können von den Universitäten selbst gesetzt werden, was das Potenzial von Dopplungen birgt. Um diesen vorzubeugen, werden die erhaltenen IDs mit der ID der Universität zusammengebracht. IDs vom Typ String wird die Universitäts ID unter Verwendung eines Trennzeichens vorangestellt, während IDs vom Typ Integer nach einer bitweisen Verschiebung mit der Universität ID bitweise verodert werden. Da Trennzeichen und die Anzahl der Bits für die Verschiebung beim Setup des Data Warehouses festgelegt werden, ist auch eine Wiederherstellung der ursprünglichen ID möglich. Für die Verschiebung der Integer-IDs muss darauf geachtet werden, die Anzahl der Bits so groß zu wählen, dass die vorgesehene Zahl an Universitäten gedeckt werden kann. Von einer internen Überführung aller Integers in Strings wurde abgesehen, da das Data Warehouse dem Data Reference Model folgen und dieses nicht anpassen soll.

Nach erfolgreichem Staging wird das *Delta* des erhaltenen Dumps im Vergleich zu den bereits im DWH vorhandenen Daten der Universität bestimmt. Neue, aktualisierte und gelöschte Datenreihen werden identifiziert und in jeweils eigene Tabellen einer neuen Datenbank übertragen, ehe die Staging Datenbank zerstört wird, um später einen neuen Import für die gleiche Universität zu ermöglichen. Dieses Delta wird im nächsten Schritt für die *Befüllung des Data Warehouses* genutzt. Neue Daten werden integriert und Änderungen angewandt. Ob gelöschte Daten aus dem Data Warehouse entfernt werden, ist von der Konfiguration des DWHs abhängig. Nach Rücksprache mit den technischen Ansprechpartnern an allen am Projekt beteiligten Universitäten wurde festgestellt, dass Daten nur dann aus den Systemen gelöscht werden, wenn die Speicherfrist abgelaufen ist oder eine Löschanfrage gestellt wird. Für die Aufnahme solcher Löschanfragen wird ein weiterer Endpunkt bereitgestellt. Im Import fehlende oder gelöschte Daten nicht aus dem DWH zu entfernen ermöglicht den importierenden Universitäten datensparsam nur ein Datendelta zu verschicken und nicht immer den gesamten Datendump der eigenen Systeme.

Der letzte Schritt, *Delta-Merging*, prüft, ob eine frühere Delta-Datenbank für die Universität besteht und wenn ja, merged diese mit der für den aktuellen Import erstellten. Somit existiert pro Universität nur eine Delta-Datenbank. Im Rahmen der Projektarchitektur ist vorgesehen in jedem Analysezyklus das Delta seit der letzten Analyse abzurufen und den Analysekomponenten zu übergeben. So können Analysen, wenn möglich, inkrementell erfolgen und verarbeiten nur neue Daten, mit welchen bestehende Ergebnisse aktualisiert oder erweitert werden. Die Komponenten, welche auf dem holistischen Datensatz arbeiten müssen, können diesen separat abrufen. Nach dem Durchlauf aller Analysen wird das Delta bereinigt und der Speicherplatz wieder freigegeben.

Datenschutz und Sicherheit Aufgrund der geringen Menge an Clients, die mit dem DWH interagieren, wurde keine Nutzerverwaltung umgesetzt, sondern Authentifizierungstoken auf sicherem Wege ausgegeben. Die beiden in Abbildung 1 dargestellten APIs des DWH sind eigenständige, individuell konfigurierbare Komponenten. So kann der Zugriff auf die API zum Lesen der Daten durch die Analysekomponenten durch eine Firewall auf ein Subnetz beschränkt werden, während das Schreiben weiterhin allen authentifizierten Clients möglich ist. Neben den klassischen technischen und organisatorischen Maßnahmen zum Datenschutz und der Gewährleistung einer sicheren Datenübertragung (vgl. [IH25; St25]), unterstützt das DWH die Verschlüsselung der zu übertragenden Daten mit GnuPG [Gn25]. Für jede Universität wird ein regelmäßig aktualisiertes Schlüsselpaar gepflegt und der öffentliche Schlüssel über einen authentifizierten Endpunkt bereitgestellt. Erhält die Importschnittstelle statt einer ZIP Datei eine GPG Datei, wird der private Schlüssel geladen und die Datei damit entschlüsselt. Ist diese erfolgreich, wird die entschlüsselte ZIP Datei wie oben beschrieben weiterverarbeitet, und andernfalls eine entsprechende Fehlermeldung ausgegeben. IDs oder demografische Datenpunkte werden nicht zusätzlich anonymisiert, diese Aufgabe obliegt den datenbereitstellenden Hochschulen. Falls benötigt kann ein solches Feature durch eine neue Komponente in der Import Pipeline nach dem Schritt *ID Mapping* umgesetzt werden.

4 Zusammenfassung und weitere Arbeiten

Das Data Warehouse wurde erfolgreich auf einer Zwischen- und der finalen Version des im Projekt erstellten Data Reference Model getestet. Erhaltene Daten werden korrekt validiert, Schlüsselkonflikte verhindert und bestehende Daten inkrementell aktualisiert. Die im Projektkontext betrachteten Studiengänge umfassen die Daten von ca. 10.000 bis 51.000 Studierender pro Universität, welche wenigen hundert Megabyte entsprechen. Ein erfolgreicher Durchlauf der Import Pipeline erfolgt bei dieser Datenmenge in unter 10 Minuten. Auch längere Verarbeitungszeiten sind akzeptabel, sodass dem DWH eine gute Performance im Projektkontext zugesprochen wird.

Die Performance sollte bei fortlaufendem Einsatz mit größeren Datenmengen weiter überwacht werden. Import Pipelines sind als Ganze entsprechend der Skalierung des Servers parallelisierbar, die einzelnen Schritte wurden diesbezüglich aber noch nicht betrachtet. Für auftretende Fehler während der Pilotierung wurden bei Bedarf erweiterte Fehlermeldungen bereitgestellt. Solche Bedarfe sollten weiterhin identifiziert werden, um Clients bestmögliche Reaktionen zu ermöglichen und eventuelle Hilfestellungen für die Fehlerbehebung zu geben.

Eine Nutzung des Data Warehouses mit anderen Datenmodellen ist dort möglich, wo relational abbildbare Daten importiert werden. Die Validierung von nicht relationalen Daten erfordert die Erweiterung des DWHs um entsprechende Komponenten. Ermöglicht ist dies, da sowohl das DWH als auch der Code zur Evaluation der Import Pipeline Open Source veröffentlicht sind [JSS25].

Literaturverzeichnis

[Ap25a] Apache Software Foundation: Apache CouchDB, https://couchdb.apache.org/, Zuletzt aufgerufen am 01.04.2025, 2025.

[Ap25b] Apache Software Foundation: GitHub - apache/hive: Apache Hive, https://github.com/apache/hive, Zuletzt aufgerufen 31.03.2025, 2025.

[Be23] Belaroussi, A. et al.: From spatial data warehouse into NoSQL document-oriented database: case study Forest fire risk management. In: Proceedings of the 2023 7th International Conference on Computer Science and Artificial Intelligence. ACM, Beijing China, S. 146–151, 2023, DOI: 10.1145/3638584.3638595.

[BO17] Bicevska, Z.; Oditis, I.: Towards NoSQL-based Data Warehouse Solutions. en, Procedia Computer Science 104, S. 104–111, 2017, DOI: 10.1016/j.procs.2017.01.080.

[Co25] Couchbase, Inc.: Couchbase: Best Free NoSQL Cloud Database Platform, https://www.couchbase.com/, Zuletzt aufgerufen am 01.04.2025, 2025.

[El22] El Aissi, M. E. M. et al.: Data Lake Versus Data Warehouse Architecture: A Comparative Study. In (Bennani, S. et al., Hrsg.): WITS 2020. Bd. 745, Series Title: Lecture Notes in Electrical Engineering, Springer Singapore, Singapore, S. 201–210, 2022, DOI: 10.1007/978-981-33-6893-4_19.

[Gn25] GnuPG Project: The GNU Privacy Guard, https://gnupg.org/, Zuletzt aufgerufen am 01.04.2025, 2025.

[Gu24] Gulia, R.: Couchbase Vs MongoDB Key Differences, Use Cases, and Similarities- Scaler Topics, https://www.scaler.com/topics/couchbase-vs-mongodb/, Zuletzt aufgerufen am 01.04.2025, 2024.

[IH25] IHK Regensburg für Oberpfalz / Kelheim: Anforderungen an Websites nach der DSGVO - IHK Regensburg für Oberpfalz / Kelheim, https://www.ihk.de/regensburg/fachthemen/recht/online-recht-und-datenschutz/eu-datenschutzgrundverordnung/anforderungen-an-websites-nach-der-ds-gvo-4158848, Zuletzt aufgerufen am 04.04.2025, 2025.

[JJ25] Judel, S.; Johnen, T.: Data Reference Model 3.6.0 of the joint centure project AIStudyBuddy, Techn. Ber., Projektverbund AIStudyBuddy, 2025.

[JSS25] Judel, S.; Schnell, E.; Schroeder, U.: Quellcode und Evaluationsskript des Data Warehouses, 2025, DOI: 10.17605/OSF.IO/RXH73.

[Ju23] Judel, S. et al.: KI-gestützte Studienverlaufsplanung und -analyse mit AIStudyBuddy. In (Thiemo Leonhardt et. al., Hrsg.): Workshops der 21. Fachtagung Bildungstechnologien (DELFI). 2023.

[Ki24] Kinsta Inc.: MariaDB vs PostgreSQL: 14 Critical Differences, https://kinsta.com/blog/mariadb-vs-postgresql/, Zuletzt aufgerufen 31.03.2025, 2024.

[Mi25] Microsoft: Bewährte Methoden beim Web-API-Design - Azure Architecture Center | Microsoft Learn, https://learn.microsoft.com/de-de/azure/architecture/best-practices/api-design, Zuletzt aufgerufen 17.03.2025, 2025.

[Mo18] MongoDB, Inc.: Server Side Public License, https://github.com/mongodb/mongo/blob/master/LICENSE-Community.txt, Zuletzt aufgerufen 30.12.2024, 2018.

[Or25] Oracle: Was versteht man unter einem Data Warehouse? | Oracle Deutschland, https://www.oracle.com/de/database/what-is-a-data-warehouse/, Zuletzt aufgerufen 31.03.2025, 2025.

[PHV21] Petricioli, L.; Humski, L.; Vrdoljak, B.: The Challenges of NoSQL Data Warehousing. In: E-business Technologies Conference. Belgrad, Serbien, S. 44–48, 2021.

[Re24] Regu, A.: MariaDB vs PostgreSQL: Detailed Comparison for Developers, https://blog.tooljet.ai/mariadb-vs-postgresql-a-detailed-comparison-for-developers/, Zuletzt aufgerufen 31.03.2025, 2024.

[SA25] SAP SE: Was ist ein Data Warehouse? | Definition, Komponenten, Architektur | SAP Insights, https://www.sap.com/germany/products/data-cloud/datasphere/what-is-a-data-warehouse.html#:~:text=Data%20Warehouse%20und%20Datenbank&text=Data%20Warehouses%20verwenden%20einen%20Datenbankserver,Integration%20von%20Datenquellen%20und%20mehr, Zuletzt aufgerufen 31.03.2025, 2025.

[Sc21] Schlauri, S.: Die «Server Side Public License», https://www.digitale-nachhaltigkeit.unibe.ch/unibe/portal/fak_naturwis/a_dept_math/c_iinfamath/abt_digital/content/e90973/e1045995/e1046107/e1092040/e1092047/07SimonSchlauriRonzaniSchlauri-SSPL_ger.pdf, Zuletzt aufgerufen am 31.03.2025, 2021.

[Si24] Singh, S.: Understanding Document Stores: A Look At MongoDB Vs. CouchDB | by Shivani Singh | Medium, https://medium.com/@shivanisingh16012004/understanding-document-stores-a-look-at-mongodb-vs-couchdb-9dbb3e469805, Zuletzt aufgerufen am 01.04.2025, 2024.

[Sm25] SmartBear Software: Best Practices in API Design, https://swagger.io/resources/articles/best-practices-in-api-design/, Zuletzt aufgerufen 17.03.2025, 2025.

[St25] Stiftung Datenschutz: Technische und organisatorische Maßnahmen - Praxisratgeber, https://stiftungdatenschutz.org/ehrenamt/praxisratgeber/praxisratgeber-detailseite/technische-und-organisatorische-massnahmen-318, Zuletzt aufgerufen am 04.04.2025, 2025.

[Za22] Zaitsev, P.: The Case Against the Server Side Public License (SSPL) - The New Stack, https://thenewstack.io/the-case-against-the-server-side-public-license-sspl/, Zuletzt aufgerufen am 31.03.2025, 2022.

Poster-Beiträge

Synergising Human Expertise and Generative Artificial Intelligence for Curriculum Design

Aakash Ahmad[1] and Thomas Schmid[2,1]

Abstract: With this currently running study, we are exploring the role of Generative Artificial Intelligence (GenAI) as an intelligent recommender system to design and deliver a sustainable and inclusive learning environment aligned with goals of curriculum transformation. In particular, we investigate the extent to which GenAI can be safely and reliably incorporated in curriculum design processes by synergising human intellect (lecturers' knowledge) and machine intelligence (data-driven recommendations) to design curriculum tailored to the needs of diverse learners. Overall, this project aims to develop a proof-of-concept and conduct controlled experimentation for empirically-grounded guidelines about safe and reliable integration of GenAI in the curriculum design process.

Keywords: Curriculum design, Outcome-based education, Generative AI, Large language models

1 Introduction

Since the introduction of the first computer science programmes in the 1960s [Gu07], the design and redesign of the curricula of computer science programmes has become a constant exercise at universities around the world. Curriculum changes are often driven by changes in recommendations and guidelines by professional organisations such as ACM [AC23] and GI [Ge05], which many universities follow for their degree programmes. In the past, such changes have been motivated by integrating new emerginging topics, such as security [Si15], human-computer interaction [Do02], and ethics and social responsibility [MW99]. On a larger scale, also national and government-driven curriculum changes represent a reoccurring phenomenon. The British government, for example, has in 2014 initiated the establishment of a radical new computing curriculum [MC18]. On university level, too, continuous evolvement of curricula can be observed, for example, where quality assurance and assessment of study programmes lead to revisions of existing curricula [MM99].

Outcome-based education (OBE) aims to outline several measurable outcomes (e. g., gained knowledge and/or acquired skills) that should be attained via pedagogically-aligned and objective-driven curriculum [SAJ25]. In practice, designing and revising curricula therefore includes extensive writing tasks such as designing and revising learning outcomes [CC17], competency statements [Cl20], and assessment requirements [LB14]. Due to the repetitive character and quantity of such paperwork, curriculum design tasks present a strong potential for automation through the application of computer-assisted text processing techniques.

[1] Lancaster University Leipzig, School of Computing and Communications, Nikolaistrasse 10, 04109 Leipzig, Germany, a.ahmad13@lancaster.ac.uk; thomas.schmid@medizin.uni-halle.de
[2] Martin-Luther-Universität Halle-Wittenberg, Medizinische Fakultät, Magdeburger Strasse 8, Halle, Germany, thomas.schmid@medizin.uni-halle.de

doi:10.18420/delfi2025_22

2 State of the Art

Mimicking human alike communications via natural language processing (NLP), recent developments have demonstrated the potential of Generative Artificial Intelligence (GenAI) to act as a decision support agent in various domains such as content generation, academic research, and learning support. The research reported by [Pe24; Ul24b], for example, aims to understand how GenAI can assist in producing course outlines along with assessment strategies that can be mapped to learning outcomes in the context of OBE [SAJ25]. These studies indicate that human knowledge (educators' experience) and GenAI capabilities (intelligent recommendations) can be unified to ensure the quality and relevance of the designed curriculum. AI-based tools support brainstorming and structuring tasks effectively, but their output consistently requires expert revision and input to finalise the contents [Ul24a].

Conrad and colleagues, for example, explore the role of ChatGPT in instructional design for public health education [CH24]. ChatGPT acts as a virtual assistant to generate learning contents for a diverse cohort of learners where formal training such as traditional lectures, classroom based are limited. In a similar study [WC23], the authors aim to leverage GenAI-designed curriculum to develop students' skills in understanding analytical thinking, identifying biases, and applying appropriate tools for problem-solving. The findings of this study support active learning and engagement with AI as an instructional design agent to enhance digital pedagogy by aligning educational outcomes with curriculum design.

With a more fundamental claim, Dickey and colleagues have proposed comprehensive guidelines termed *GenAI in the Development of Educational Content* (GAIDE) [DB24]. Based on human supervision, the GAIDE framework supports a semi-automated generation of course outline and curriculum by facilitating educators to design course contents, assessment material, and teaching strategies. In the same spirit but with a focus on Massive Online Open Courses (MOOCs), Ruiz-Rojas and colleagues developed scalable strategies and workflows for personalising learning experiences and supporting instructional coherence using GenAI in line with the 4PADAFE instructional design matrix [Ru23].

In practice, several GenAI tools are available that support support in their efforts for AI-assisted curriculum design. Developed by Khan Academy, *Khanmigo* aids in adaptive lesson planning and formative assessment generation aligned with learning outcomes. Similar tools such as *Coursemojo* and *MagicSchool AI* offer a platform for designing modular and standard-based curricula and courseware by following pedagogical approaches such as OBE and constructive alignement. These tools enable educators to design curriculum content, course planning, and lesson generators. The tool *AI Diffit* enables the transformation of curriculum topics into concrete learning material and assessment design.

3 Project Goals

This project aims to support Lancaster University (LU) in its goal to unify diverse curricula around its global network of campuses (e. g., in Leipzig and Ghana via its so-called

curriculum transformation programme (CTP), which focuses on integrating innovation and sustainability in curriculum development. Given a global and diverse LU campus network, however, this ambition faces a multitude of challenges for the quality and consistency of the curriculum, such as limited teaching resources, ambitious but inexperienced faculty, lack of training, diverse learners, and the need for continuously updated pedagogical approaches.

To address these challenges, educators require support mechanisms, such as pedagogically customised recommender systems that can help them to design and deliver a consistent, inclusive and high-quality curriculum. While GenAI could function as such, its integration in curriculum design process raises ethical, data privacy, and accuracy concerns. To this end, we aim to find an empirically grounded answer to the question: 'How can GenAI be effectively integrated into the curriculum design process – synergising human knowledge with machine intelligence – to design an inclusive and pedagogically aligned curriculum?'

Compared to ongoing initiatives on AI in higher education, we aim to document empirically grounded guidelines, develop a proof-of-concept, and conduct controlled experimentation about safe and reliable integration of GenAI in curriculum design. Our main objectives are:

- To put a GenAI system to action that will provide automated curriculum templates, assists in learning outcome alignment, and streamlines course redesign to ensure consistency across disciplines and campuses;
- To investigate the potential and limitations of GenAI in curriculum design by conducting an empirical study on its role in enhancing student experience, flexible academic structures, and inclusive curriculum development;
- To validate the human-AI collaborative approach for curriculum design, leveraging adaptation of existing tools to improve accuracy, inclusivity, and sustainability in course development, while addressing ethical, privacy, and pedagogical concerns.

4 Approach and Methodology

After exploration of the state-of-the-art to identify potential and limitations, the project will employ a two-step research method (Fig. 1) to (i) design and develop a prototype, and (ii) provide preliminary validations of the human-AI collaborative curriculum design process:

(i) **Prototype Development.** A core outcome of the project will be AI-driven models tailored for curriculum development. Key technologies used will be large language models (LLM) and Retrieval-Augmented Generation (RAG). As LLM model, we will built upon the Llama 3 family, which is open-source and has proven state-of-the-art capabilities. The complementary use of RAG will help minimise bias and hallucination in curriculum generation by grounding the model's output in diverse and factual data sources, such as existing curricula, related manuals and guidelines, etc. RAG reduces AI biases by retrieving factually accurate sources before generating content and will ensure that AI-assisted curriculum design is credible and inclusive.

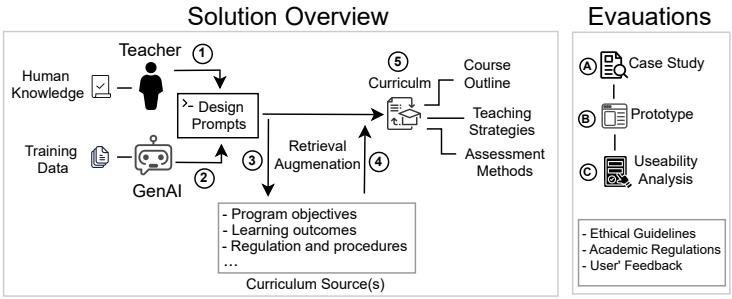


Fig. 1: Overview of the Proposed Solution and Evaluation Methods

Overall, the prototype will showcase how generative AI can assist in creating adaptive, personalized, and dynamic curricula that cater to diverse learning needs.

(ii) **Empirical Evaluation.** A multitude of socio-technical issues – including but not limited to: ethical considerations, availability and privacy and training data, and scepticism around accuracy of GenAI produced contents – can hamper the integration of pedagogical approaches and GenAI tools for curriculum design. To mitigate these issues, we will conduct an empirical evaluation of human-AI collaborative curriculum design. By applying human inspection and discussion of AI-generated curricula with other experts, we carry out a rigorous evaluation and provide valuable insights into the effectiveness, limitations, and potential improvements of using AI in education, helping educators make informed decisions about its integration.

5 Conclusions & Outlook

Building on existing frameworks and tools that support personalised content creation and improve efficiency, this project will strengthen and extend the exchange between AI research and real-world applications, ensuring that findings can be directly applied in educational settings. With the development of a prototype for GenAI-based curriculum design in particular, educators will gain access to practical tools and resources that will demonstrate the capabilities of GenAI to streamline curriculum development, promote personalisation, and align content with pedagogical frameworks. Existing research, while acknowledging the benefits such as acceleration of developing and updating curricula, also points to ethical risks such as data privacy, algorithmic bias, copyright concerns, and environmental impact, emphasizing the need for responsible use of GenAI in teaching and learning support activities. By sharing insights and outcomes across the Global network of Lancaster University as well as educational communities, we will foster collaboration and inspire further exploration of AI's role in transforming education.

Acknowledgements

We would like to thank the Data Science Institute (DSI) of Lancaster University for financial support of this AI Innovation Study.

References

[AC23] ACM, IEEE Computer Society, and AAAI: Computer Science Curricula 2023: Curriculum Guidelines for Undergraduate Programs in Computer Science. ACM, 2023.

[CC17] Chatterjee, D.; Corral, J.: How to write well-defined learning objectives. The journal of education in perioperative medicine: JEPM 19 (4), E610, 2017.

[CH24] Conrad, E. J.; Hall, K. C.: Leveraging generative AI to elevate curriculum design and pedagogy in public health and health promotion. Pedagogy in Health Promotion 10 (3), pp. 178–186, 2024.

[Cl20] Clear, A. et al.: Designing computer science competency statements: A process and curriculum model for the 21st century. In: Proceedings of the working group reports on innovation and technology in computer science education. Pp. 211–246, 2020.

[DB24] Dickey, E.; Bejarano, A.: GAIDE: A Framework for Using Generative AI to Assist in Course Content Development. In: 2024 IEEE Frontiers in Education Conference (FIE). IEEE, pp. 1–9, 2024.

[Do02] Douglas, S. et al.: Incorporating human-computer interaction into the undergraduate computer science curriculum. ACM SIGCSE Bulletin 34 (1), pp. 211–212, 2002.

[Ge05] Gesellschaft für Informatik: Empfehlungen für die Ausgestaltung von akkreditierten Informatik-Studiengängen, tech. rep., Stand: 2005, fortlaufend aktualisiert, 2005.

[Gu07] Gupta, G. K.: Computer science curriculum developments in the 1960s. IEEE Annals of the History of Computing 29 (2), pp. 40–54, 2007.

[LB14] Lile, R.; Bran, C.: The assessment of learning outcomes. Procedia-Social and Behavioral Sciences 163, pp. 125–131, 2014.

[MC18] Moller, F.; Crick, T.: A university-based model for supporting computer science curriculum reform. Journal of Computers in Education 5 (4), pp. 415–434, 2018.

[MM99] McDonald, M.; McDonald, G.: Computer science curriculum assessment. Acm Sigcse Bulletin 31 (1), pp. 194–197, 1999.

[MW99] Martin, C. D.; Weltz, E. Y.: From awareness to action: Integrating ethics and social responsibility into the computer science curriculum. ACM Sigcas Computers and Society 29 (2), pp. 6–14, 1999.

[Pe24] Pesovski, I. et al.: Generative AI for Customizable Learning Experiences. Sustainability, 16 (7), 3034, 2024.

[Ru23] Ruiz-Rojas, L. I. et al.: Empowering education with generative artificial intelligence tools: Approach with an instructional design matrix. Sustainability 15 (15), p. 11524, 2023.

[SAJ25] Shanto, S. S.; Ahmed, Z.; Jony, A. I.: A proposed framework for achieving higher levels of outcome-based learning using generative AI in education. Educational Technology Quarterly 2025 (1), pp. 1–15, 2025.

[Si15] Siraj, A. et al.: Integrating security in the computer science curriculum. ACM Inroads 6 (2), pp. 77–81, 2015.

[Ul24a] Ullmann, T. et al.: Towards generative AI for course content production: Expert reflections. European Journal of Open, Distance and E-Learning, In–press, 2024.

[Ul24b] Ullmann, T. D. et al.: Teaching with Generative AI: moving forward with content creation. Ubiquity Proceedings 4 (1), 2024.

[WC23] Wu, T.; Chang, M.: Application of generative artificial intelligence to assessment and curriculum design for project-based learning. In: 2023 International Conference on Engineering and Emerging Technologies (ICEET). IEEE, pp. 1–6, 2023.

FAIR4EDTECH: Towards an Artifact Evaluation Track

Jan Bernoth [1], Birte Heinemann [2], Matthias Ehlenz [2], Martin Armbruster [3], and Ulrike Lucke [1]

Abstract: This poster proposes a community-driven Artifact Evaluation Track for the DELFI conference to promote FAIR principles and recognize research artifacts such as software and data. Building on recent initiatives within the AK Open Science of the section Educational Technology of the German Informatics Society, we outline requirements and processes for submitting, evaluating, and awarding badges to artifacts. Emphasis is placed on domain-specific standards in Educational Technology, including suitable metadata ontologies and data exchange formats like xAPI. Review procedures should be time-bound, criteria-based, and ideally include interactive rebuttal phases. The goal is to foster transparency, reproducibility, and reusability in educational technology research by aligning with practices from established computer science venues. The poster invites discussion on how to implement and sustain such a track within DELFI.

Keywords: FAIR Principles, Open Science, Software Evaluation, Metadata Standards, Educational Technology, Artifact Evaluation

1 Motivation

The topic of Open Science, as well as FAIR data and software, has become increasingly important within the DELFI community. Building on a recent position paper, several key challenges were identified: recognizing and crediting software and infrastructure as valuable scientific contributions; improving and diversifying publication practices for software; establishing publication standards that reflect development processes; and fostering reuse and collaboration through open science principles aligned with the FAIR paradigm [KS22]. The position paper led to the formation of a new Working Group (AK Open Science)[4] as part of the section Educational Technologies (FG Bildungstechnologien) within the German Informatics Society (GI). Both workshops [KS23; KS24] of the group were centered on discussions about requirements of publication of research artifacts, conception of peer-review process, and shaping new publication formats. In summary, wishes were that (1) artifact evaluations are divided between research data and research software evaluation, (2) having a more interactive rebuttal phase between reviewers and authors, (3) align reviewing criteria

[1] Universität Potsdam, Institut für Informatik, An der Bahn 2, 14476 Potsdam, Germany,
jan.bernoth@uni-potsdam.de, https://orcid.org/0000-0002-4127-0053;
ulrike.lucke@uni-potsdam.de, https://orcid.org/0000-0003-4049-8088
[2] RWTH Aachen University, Templergraben 55, 52062 Aachen, Germany,
heinemann@cs.rwth-aachen.de, https://orcid.org/0000-0002-7568-0704;
ehlenz@lbz.rwth-aachen.de, https://orcid.org/0000-0001-6189-6056
[3] Karlsruhe Institute of Technology, Am Fasanengarten 5, 76131 Karlsruhe,
martin.armbruster@kit.edu, https://orcid.org/0000-0002-2554-4501
[4] https://fg-bildungstechnologien.gi.de/fachgruppe/arbeitskreise/ak-open-science

doi:10.18420/delfi2025_23

to other software oriented conferences, and (4) present the best artifacts in a conference track.

It is a clear mission that artifact evaluation differs from paper reviewing. Designing research software and evaluating its impact in teaching and learning scenarios is an essential part of educational technology. The research artifacts could differ based on the research question [St24], but mostly the focus is on the educational technology and its impact on the teaching and learning scenario. Due to this, the authors focus in this work on the artifact evaluation track on evaluating software, and expect that most of the measurements could support most of the research data evaluation.

Artifact evaluation tracks are already established in computer science conferences (e. g., the SuperComputing[5] or International Conference on Software Engineering[6]). In these tracks, authors of accepted papers submit the accompanying artifacts for a separate peer review of the artifacts [Ar25]. In addition, authors apply for specific badges, for instance, of the ACM badges [As20]. During the review, reviewers and authors can communicate to clarify upcoming issues or questions. Based on the final reviews, the badges can be awarded for the artifacts.

The goal of this poster is to spark a discussion around community-driven solutions for implementing the FAIR principles within the DELFI community. The focus lies on the idea to establish a dedicated Artifact Evaluation Track as part of the DELFI conference, aiming to make research outputs such as software and data more visible, reusable, and systematically evaluated. Based on the already named preliminary works, the poster presents concrete requirements, formats, and evaluation criteria to enable transparent and fair review processes for research artifacts. Key elements for discussion include the separation of software and data artifacts, the introduction of interactive rebuttal phases, and the alignment with established practices from the software engineering community.

2 Community Solutions for Artifact Creators

Submission to the evaluation track requires the work to be at least FAIR. The recommendations in FAIR4RS [Ch22] are addressed not only to creators of artifacts but also give the research community the task of defining guidelines to achieve FAIR compliance. In the following list, principles from these recommendations, which need to be supported by community solutions, will be cited and possible solutions will be proposed:

1. *F2 – Software us described with rich metadata.* Rich metadata are important for creating findable research artifacts, and using a controlled vocabulary supports this purpose by ensuring consistency and clarity. To provide a broader description of the field to which the artifact belongs, the Computer Science Ontology[7] [Sa18] and

[5] https://sc25.supercomputing.org/program/papers/reproducibility-initiative/
[6] https://conf.researchr.org/track/icse-2025/icse-2025-artifact-evaluation
[7] https://cso.kmi.open.ac.uk/home

the Software Ontology[8] [Ma14] can be used, but these will not deeply describe the purpose. More fitting metadata can be produced by using an ontology that is tied to educational technologies. This compendium about didactic metadata [OR24] provides a good overview of current work on metadata in this field.

2. *I1 – Software reads, writes and exchanges data in a way that meets domain-relevant community standards.* In Education Technology, xAPI is commonly used to log data about learners and teachers in educational systems, although the flexibility of this standard hinders compatibility between systems [Eh20]. Using a centralized registry makes it easier [He22]. For this purpose, we propose to use and, if necessary, to extend the xAPI Definitions Registry[9]

3. *R3 – Software meets domain-relevant community standards.* This principle is also connected with *R1 – Software is described with a plurality of accurate and relevant attributes.* There are three areas that have domain-relevant standards: software or package registries, licenses, and coding practices. For registries, there is a clear recommendation to use language-relevant registries, for example, for Python use PyPI, for JavaScript use npm, or for Unity use Unity's Package Manager. For licenses, we propose to use open licenses, for example, MIT or GNU GPLv3. For coding standards, it is very difficult for such a diverse community to create a standard. A minimal baseline could be created from the guidelines from Lee et al. [Le21], by e. g. using version control, documenting your software and how to set up, and testing mostly at the unit level.

Once these standards are set, the authors can prepare their artifacts for the review phase based on them. For preparing the artifacts themselves, general recommendations like Five Recommendations for FAIR Software [eS] can be used. To put different artifacts into one submission and prepare an overview paper about the artifact submission, the creation workflow of the Research Data Management Container can be used [Be25], a demo can be seen here[10]. The submitters need to clarify for which badge they want to achieve, if the community agrees on ACM Badges [As20] these can be used to choose from.

3 Proposed Process for an Artifact Evaluation Track

For reviewing an artifact, there should be an explicitly mentioned expected time frame, for example, two hours, so that reviewers know the required time investment ahead and authors can prepare their artifact accordingly. As a consequence, reviewers need to efficiently and effectively evaluate and check the quality criteria corresponding to the badges for which the authors apply. They can encompass, for instance, the metadata, data formats for data exchange, or, inspired by the ACM badges [As20], if the artifacts are documented, consistent,

[8] https://github.com/allysonlister/swo
[9] https://xapi.elearn.rwth-aachen.de/
[10] https://rdmc.create.nfdixcs.org/

complete, and exercisable. Regarding “exercisable”, reviewers should execute the artifact to check the results with the associated paper or claims. To simplify reviewing, all checks should be automated as much as possible.

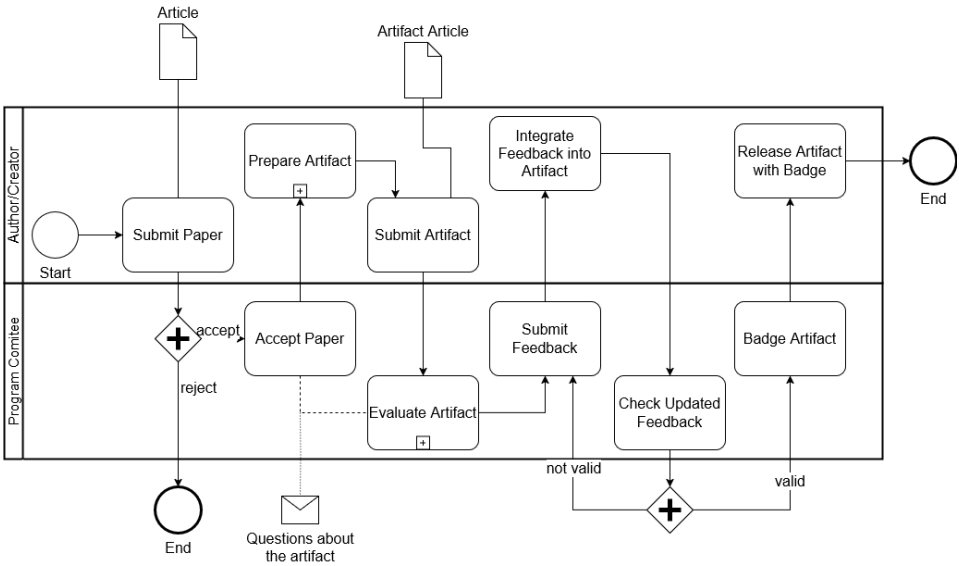


Fig. 1: Proposed Workflow for the Artifact Evaluation in BPMN 2.0

In Figure 1, a possible process of the Artifact Evaluation Track is proposed. The authors should indicate their intent to submit an artifact for evaluation, to be provided upon paper acceptance. After paper acceptance, the authors should get time to prepare their artifact. Some conferences also have a communication phase in the evaluation period in which reviewers can communicate with the authors to quickly resolve issues and have a running system in their environment. After the artifacts are evaluated, the committee of reviewers may award the badge. The authors need to create a new release of their artifacts with eventual updates from the review phase and the awarded badge.

In summary, this poster proposes a community-driven Artifact Evaluation Track for the DELFI conference to promote FAIR principles and recognize research artifacts like software and data. Building on prior Open Science initiatives and related efforts, it outlines requirements, like metadata standards and review processes, including interactive rebuttals and badge systems to foster transparency, reproducibility, and reusability.

Acknowledgments

This work was mainly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – NFDI 52/1 – 501930651.

References

[Ar25] Armbruster, M.: Artifact Evaluation in the Context of NFDIxCS, 2025, https://nfdixcs.org/meldung/artifact-evaluation-in-the-context-of-nfdixcs.

[As20] Association for Computing Machinery: Artifact Review and Badging Version 1.1, 2020, https://www.acm.org/publications/policies/artifact-review-and-badging-current, accessed: 03/30/2023.

[Be25] Bernoth, J. et al.: Workflow for Creating and Sealing a Research Data Management Container (RDMC). In (Feichtinger, K.; Sonnleithner. Lisa; Hajiabadi, H., eds.): Software Engineering 2025 – Companion Proceedings. Gesellschaft für Informatik e.V, Bonn, 2025.

[Ch22] Chue Hong, N. P. et al.: FAIR Principles for Research Software (FAIR4RS Principles), 2022.

[Eh20] Ehlenz, M. et al.: Eine forschungspraktische Perspektive auf xAPI-Registries. In: DELFI 2020 – Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V. Gesellschaft für Informatik e.V., Bonn, 2020, http://dl.gi.de/handle/20.500.12116/34179.

[eS] eScience Center Netherlands: Five Recommendations for FAIR Software, https://www.fair-software.eu/, accessed: 06/06/2025.

[He22] Heinemann, B. et al.: xAPI Made Easy: A Learning Analytics Infrastructure for Interdisciplinary Projects. International Journal of Online and Biomedical Engineering (iJOE) 18 (14), pp. 99–113, 2022.

[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, Bonn, pp. 201–206, 2022.

[KS23] 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.

[KS24] Kiesler, N.; Schiffner, D.: Open Science in den Bildungstechnologien: Zur Unterstützung bei der Veröffentlichung von Forschungsdaten. In (Kiesler, N.; Schulz, S., eds.): Proceedings of DELFI Workshops 2024. Gesellschaft für Informatik e. V. (GI), Bonn, pp. 77–86, 2024.

[Le21] Lee, G. et al.: Barely sufficient practices in scientific computing. Patterns (New York, N.Y.) 2 (2), p. 100206, 2021.

[Ma14] Malone, J. et al.: The Software Ontology (SWO): a resource for reproducibility in biomedical data analysis, curation and digital preservation. Journal of biomedical semantics 5, p. 25, 2014.

[OR24] Oellers, M.; Rörtgen, S.: Kompendium: Didaktische Metadaten, 2024.

[Sa18] Salatino, A. A. et al.: The Computer Science Ontology: A Large-Scale Taxonomy of Research Areas. In (Vrandečić, D. et al., eds.): The Semantic Web – ISWC 2018. Vol. 11137, Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 187–205, 2018.

[St24] Striewe, M.: Forschungsdaten in der Bildungstechnologie: Worüber sprechen wir und was ist wann relevant?, 2024.

Prompt Engineering or Creating Custom GPTs, or both?

Exploring interventions to support students in developing AI literacy skills

Andrea de Santiago [1], Carina Ziegler [2], and Isa Jahnke [3]

Abstract: Student use of GenAI tools, like ChatGPT, remains a prominent topic in higher education. Techniques like prompt engineering and creating Custom GPTs are promising tools to support student learning. However, there is more data needed on how students create their own Custom GPTs to support their learning. Thus, we propose a study to explore how prompt engineering training and Custom GPT development training influence student learning. The study is work-in-progress and data is collected throughout the summer of 2025. The first results of this pilot will be presented as a poster at DELFI 2025.

Keywords: Generative AI, Educational Technology, Higher Education, Prompt Engineering, Custom GPTs

1 Introduction

While the release of OpenAI's ChatGPT might have been groundbreaking to some, the application of AI in education (AIEd) has been a prominent research domain for years [MVM24; CB23]. Since the release of ChatGPT, however, publications about AIEd have significantly increased [Og24; Za19]. Within these recent publications, academic integrity is one of the reoccurring themes, with educators voicing their fears that students might be misusing Generative AI (GenAI) tools. While several studies show that students are using GenAI for their studies [Ba24; CTS24; GM23], a consensus as to how it should or should not be integrated has yet to be found. There is a need to better understand how AI skills can be included in curriculum, study programs, and courses.

[1] University of Technology Nuremberg, Learning Technologies Group, Dr.-Luise-Herzberg-Straße 4, 90461 Nürnberg, andrea.de.santiago@utn.de https://orcid.org/0009-0009-2267-5336
[2] University of Technology Nuremberg, Learning Technologies Group, Dr.-Luise-Herzberg-Straße 4, 90461 Nürnberg, carina.ziegler@utn.de https://orcid.org/0009-0002-5341-9342
[3] University of Technology Nuremberg, Learning Technologies Group, Dr.-Luise-Herzberg-Straße 4, 90461 Nürnberg, isa.jahnke@utn.de https://orcid.org/0000-0001-6255-6416

doi:10.18420/delfi2025_24

2 Literature Review – What is already known from research

Artificial intelligence (AI) is omnipresent in our world. A variety of specialized AI tools for different purposes like translation, searching, planning, and more is available. Similarly, AI is a prominent topic in education and very visible in current education research. One recurring theme of AIEd is AI literacy. We use the definition from Long and Magerko [LM20, p. 598], "AI literacy (is) as a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace". This definition is still prominent in the literature today [AJL24; Be23; SH24; So23; CTS24; Ya25], and indicates the importance of acquiring these skills. Today, many higher education institutions include AI literacy in their program (e. g., Univ. of Florida).

AI literacy includes teaching students prompt engineering [LP25]. Prompt engineering is a process of creating a prompt or prompt function to effectively perform a task [Li23]. Therefore, prompt engineering is more than searching for quick solutions. Research shows that knowing how to prompt yields more accurate GenAI outputs [STS24; LP25]. There is a variety of different techniques for prompting, from "Zero shot" to complex "Tree of Thought Prompting". Independently of the used approach, effective prompt engineering requires content knowledge, critical thinking skills, and iterative design [Ca24]. This makes using GenAI a deliberate, reflective learning task. Only if the students have enough content knowledge, they are able to evaluate the outcomes critically and spot hallucinations. They need to carefully consider what to ask for and how to frame their prompts to get the intended outcome [Ca24].

However, with Custom GPTs, users have the possibility to tailor a model that is fine-tuned for a specific task. This is one difference to regular prompting in ChatGPT or other tools. Furthermore, users add specific data (e. g., upload material), define the 'behavior' of the Custom GPT, and configure the Custom GPT to apply a certain or selected conversational style. In addition, Custom GPTs can be connected to other outside applications and retrieve real time information, like the weather. Custom GPTs act as tailored assistants. In the end, the user created a shareable artifact, a Custom GPT with a logo. The customized GPT becomes a personal tutor that can be adjusted or tailored to specific needs, and may serve as a substitute human partner.

Research focusing on students designing their own GPTs for educational purposes is sparse. In a general study on AI in education, Lyu et al. [Ly25] mentioned one educator, who asked students to create a custom GPT: "The results forced the students to think about what information the GPT needed to be helpful. It forced the students to revise their queries until they got better results". But the authors did not elaborate the study further. Other papers focused on educators creating bots to support administrative planning [Ch24] and feedback [ALJ24; Ki25; Li24]. Some studies indicate that humanizing GenAI tools can be useful to improve reflective practice (e. g. [YMD25, Zh24]). It seems users believe AI outputs more than other humans. Humans are viewed more critically, so humanizing custom GPTs might be effective in supporting reflective thinking.

This study proposes to explore the training of higher education students in prompt engineering and Custom GPT creation. The training includes methods to "humanize" the machine. The central research questions are:

- RQ1: How can prompt engineering training support students in the effective and reflective use of such tools to support student learning?
- RQ2: How can Custom GPT training support students in the effective and reflective use of such tools to support student learning?

3 Methods

There will be two studies, the pilot and the actual study. The pilot study with four participants investigated how students use ChatGPT as a study tool, and how brief instructional interventions influence their prompting strategies. Participants first completed the Metacognitive Awareness Inventory (MAI) [SD94] in order to reflect on their learning, then engaged in a task using ChatGPT, with their screen interactions being recorded using the Think Aloud Method [ES93]. In a second session, they watched a short video on prompt engineering and completed a second task. Finally, after viewing a video on creating Custom GPTs, each student built their own personalized "Study Buddy" and completed the MAI again. The pilot study was conducted in the beginning of July 2025 and evaluated in August. The recordings are being analyzed for prompting structure and patterns. The students' reflections are also being analyzed according to Braun and Clarke's thematic analysis [BC06]. Thirdly, the results of the MAI pre- and post-interventions will be compared to see if there has been a change in their metacognitive awareness

This pilot study helps to inform the actual study, i. e., to test the methods of training students how to prompt and how to customize the GPT for their learning. In the actual study with approx. 30 participants will be randomly selected, into three groups (students of a master level course in an AI and Robotics master program). Besides the training received, they will be treated identically, completing the same assignments in the course, and attending the same classes. The three groups will use OpenAI's ChatGPT Team license, which all students at the university have access to. They have unlimited, pseudonymized chats as well as the license to create Custom GPTs. We plan the following two groups plus a third control group:

Group 1-Prompt Training (group P): 45 minutes in-person kick-off session including a warm-up activity (e. g., asking participants: How do you currently use GenAI in your learning? What has worked for you, what hasn't?), followed by prompt engineering training, and hands-on practice of different techniques for summarizing, explaining new concepts, generate example problems etc. At the end, they create a prompt for a given problem.

Group 2-Group Custom GPT (group C): 90 minutes in-person kick-off session, that includes a warm-up where students reflect on how they usually learn and come up with a list of methods, strategies, and activities they use. Then, an introduction to Custom GPTs and examples follow, and concludes with the design challenge “Build Your AI Learning Partner” (using results from previous reflection). To guide students in creating their own Custom GPT "Study Buddy," a structured template will be provided. It includes: (a) defining the GPT’s purpose, (b) specifying desired ‘behavior’ (how the machine interacts with the student, prompt the tone and interaction style, e .g., behave like a tutor, like an academic advisor, etc.), (c) writing key instructional prompts, (d) uploading relevant base materials, and (e) testing and refining the bot through iterative use.

The interventions will take place during the winter semester 2025/26. During the study period, group P will use the regular ChatGPT, while group C will use their own Custom GPT. A third group will use the regular ChatGPT but without any training. All participants will write down what worked well, what did not work well; and students from group C adjust their Custom GPT as needed.

Data from the courses pre- and post-test to determine the knowledge gain and learning growth will be analyzed in accordance to La Barge’s qualifier method [La07]. The students will turn in a reflective journal, documenting their uses of AI during the course of a unit. They will document the prompts they used, the answers they got and comment on these answers. We will use Braun and Clarke’s thematic analysis approach [BC06]. At the end of the course, the students will be briefly interviewed.

4 Next Steps

From the pilot study, we will gain insights into students using different prompting techniques. We want to explore how and to what extent the creation of Custom GPTs influence the students’ learning growth. Further, we will use this pilot to reflect on the methods and interventions for fostering AI literacy skills in higher education courses, and how to design these. Then, the next step is to repeat the study with different groups, including a control group for more reliable research results.

The study may inform future research in two key areas: the potential of student-created Custom GPT as Study Buddies to support learning, and the critical role of balancing effective prompting with content knowledge—both guided by reflective thinking when integrating GenAI in learning.

References

[AJL24] Almatrafi, O.; Johri, A.; Lee, H.: A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). Computers and Education Open, p. 100173, 2024.

[Ba24] Balabdaoui, F.; Dittmann-Domenichini, N.; Grosse, H.; Schlienger, C.; Kortemeyer, G.: A survey on students' use of AI at a technical university. Discover Education 3(1), p. 51, 2024.

[BC06] Braun, V.; Clarke, V.: Using thematic analysis in psychology. Qualitative Research in Psychology (3), pp. 77-101, 2006.

[Be23] Benvenuti, M.; Cangelosi, A.; Weinberger, A.; Mazzoni, E.; Benassi, M.; Barbaresi, M.; Orsoni, M.: Artificial intelligence and human behavioral development: A perspective on new skills and competences acquisition for the educational context. Computers in Human Behavior 148, p. 107903, 2023.

[Ca24] Cain, W.: Prompting Change: Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education. TechTrends 68(1), pp. 47-57, 2024.

[CB23] Crompton, H.; Burke, D.: Artificial intelligence in higher education: the state of the field. International Journal of Educational Technology in Higher Education 20(22), 2023.

[Ch24] Chukhlomin, V.: Exploring the Use of Custom GPTs in Higher Education Strategic Planning: A Preliminary Field Report. SSRN Electronic Journal, 2024.

[CTS24] Chen, K.; Tallant, A. C.; Selig, I.: Exploring generative AI literacy in higher education: student adoption, interaction, evaluation and ethical perceptions. Information and Learning Sciences 126(1/2), pp. 132-148, 2024.

[ES93] Ericsson, K.; Simon, H.; Protocol Anaysis: verbal reports vs data. MIT Press, 1993.

[GM23] Von Garrel, J.; Mayer, J.: Artificial Intelligence in studies—use of ChatGPT and AI-based tools among students in Germany. Humanities and Social Sciences Communications 10(1), pp. 1-9, 2023.

[Ki25] Kim, N.; Vickers, J.; Alyehaibi, S.; Ozuzu, C., Ren, X., Kotcherlakota, S.: Leveraging Custom GPTs for Formative Assessment and Feedback. Proceedings of Society for Information Technology & Teacher Education International Conference, pp. 1124-1130, 2025.

[La07] La Barge, G.: Pre- and Post-Testing with More Impact. The Journal of Extension, 45(6), p. 17, 2007.

[Li23] Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G.: Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM computing surveys, 55(9), pp. 1-35, 2023.

[Li24] Lievens, J.: An exploratory assessment of the usability and potential of generative pretrained transformers (GPTS) as feedback assistants for long-format academic writing tasks. In (INTED proceedings): pp. 1280-1286, 2024.

[LM20] Long, D.; Magerko, B.: What is AI Literacy? Competencies and Design Considerations. Proceedings of the 2020 CHI conference on human factors in computing systems, pp. 1-16, 2020.

[LP25] Lee, D.; Palmer, E.: Prompt engineering in higher education: a systematic review to help inform curricula. International Journal of Educational Technology in Higher Education, 22(1), p. 7, 2025.

[Ly25] Lyu, W., Zhang, S., Chung, T., Sun, Y., & Zhang, Y.: Understanding the practices, perceptions, and (dis) trust of generative AI among instructors: A mixed-methods study in the US higher education. Computers and Education: Artificial Intelligence, 8, p. 100383, 2025.

[Ma25] Ma, T.: Systematically visualizing ChatGPT used in higher education: Publication trend, disciplinary domains, research themes, adoption and acceptance. Computers and Education: Artificial Intelligence, p. 100336, 2025.

[MVM24] Mohammed, A.T.; Valander, J.; Milrad, M.: A Retrospective Analysis of Artificial Intelligence in Education (AIEd) Studies: Perspectives, Learning Theories, Challenges, and Emerging Opportunities. In: Burgos, D. et al.: Radical Solutions for Artificial Intelligence and Digital Transformation in Education, p.127-141, 2024.

[Ng21] Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., Qiao, M. S.: Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence 2, p. 100041, 2021.

[Og24] Ogunleye, B.; Zakariyyah, K.;Ajao, O.; Olayinka, O.; Sharma, H.: A Systematic Review of Generative AI for Teaching and Learning Practice. Education Science, 14(6), p. 636 2024.

[SD94] Schraw, G.; Dennisn, R.: Assessing Metacognitive Awareness. Contemporary Educational Psychology 19, pp. 460-47, 1994.

[SH24] Stolpe, K.; Hallström, J.: Artificial intelligence literacy for technology education. Computers and Education Open 6, p. 100159, 2024.

[So23] Southworth, J., Migliaccio, K., Glover, J., Glover, J. N., Reed, D., McCarty, C., Brendemuhl, J., Thomas, A.: Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. Computers and Education: Artificial Intelligence 4, p. 100127, 2023.

[STS24] Sawalha, G.; Taj, I.; Shoufan, A.: Analyzing student prompts and their effect on ChatGPT's performance. Cogent Education, 11(1), p. 2397200, 2024.

[Ya25] Yang, Y., Zhang, Y., Sun, D., He, W., Wei, Y.: Navigating the landscape of AI literacy education: insights from a decade of research (2014–2024). Humanities and Social Sciences Communications 12(1), pp. 1-12, 2025.

[YMD25] Yusuf, H.; Money, A.; Daylamani-Zad, D.: Pedagogical AI conversational agents in higher education: a conceptual framework and survey of the state of the art. Education Tech Research Dev **73**, pp. 815–874, 2025.

[Za19] Zawacki-Richter, O.; Marin, V.; 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.

[Zh24] Zhang, S.; Zhao, X.; Nan, D.; Kim, J.: Beyond learning with cold machine: interpersonal communication skills as anthropomorphic cue of AI instructor. International Journal of Educational Technology in Higher Education 21, 2024.

Server or Cloud? Who needs it! Peer-to-Peer Remote Labs with edrys-Lite.

Modular & Reconfigurable & Reusable & Browser-Based Real-Time Experiments.

André Dietrich[1], Jihad Hyadi[1] and Sebastian Zug[1]



Abstract: Remote labs often depend on centralized servers for authentication, storage, and streaming, introducing complexity and bottlenecks. edrys-Lite offers a browser-based P2P framework using WebRTC for secure, direct connections and Yjs for CRDT-driven state synchronization. Sessions launch via a shared URL, enabling instructors, learners, and hardware nodes to join without extra software or elaborate backends. When WebRTC is blocked, a lightweight self-hosted WebSocket relay maintains connectivity. We outline the core architecture, modular extensibility, and showcase an Arduino sensor lab. By eliminating server dependencies, edrys-Lite simplifies deployment and sharing while preserving real-time collaboration and remote hardware control in STEM education.

Keywords: Peer-to-Peer, Remote Laboratories, Serverless Architecture, WebRTC, WebSocket, CRDT, Browser-Based Education, Decentralized Synchronization

1 Introduction

Remote laboratory environments are increasingly vital for engineering and science education, especially when access to physical labs is limited by geography, cost, or logistics. Traditional remote labs are complex to build, requiring authentication systems, scheduling backends, data persistence, and low-latency connections—plus the challenge of safely exposing real hardware. Centralized systems introduce single points of failure, high maintenance needs, and latency issues when many users connect simultaneously.

edrys-Lite[2] avoids this complexity with a modular peer-to-peer design. It uses WebRTC[3] Data Channels and CRDT[4]-based synchronization (via Yjs[5]) to enable direct browser-to-browser collaboration. When direct peer connections are not possible—such as in restricted network environments—WebSockets[6] can be used as a fallback, relayed through

[1] TU Bergakademie Freiberg, Institute of Computer Sciene, Bernhard-von-Cotta-Straße 2, 09599 Freiberg, {Andre.Dietrich | Jihad.Hyadi | Sebastian.Zug}@informatik.tu-freiberg.de
[2] Project website: https://edrys-labs.github.io
[3] Web Real-Time Communication: https://en.wikipedia.org/wiki/WebRTC
[4] Conflict-Free Replicated Data Types: https://en.wikipedia.org/wiki/Conflict-free_replicated_data_type
[5] CRDT-based real-time collaborative data structures for JavaScript: https://docs.yjs.dev
[6] Persistent bi-directional web communication: https://en.wikipedia.org/wiki/WebSocket

doi:10.18420/delfi2025_25

lightweight, self-hosted servers. This maintains real-time interactivity without requiring a full backend. There's no user management, no scheduling logic, and no persistent server storage. Sessions are launched via shared URLs, with no installations or setup required.

To the best of our knowledge, edrys-Lite is the only fully browser-based, peer-to-peer remote laboratory implementation available. The only loosely related efforts [AG23; THB23] also leverage WebRTC for remote experiment control, but they both depend on fixed hardware setups and a locally hosted server to deliver the user interface. In contrast, edrys-Lite employs a completely modular architecture, including its communication schema, enabling dynamic composition of experiment stations and true peer-to-peer interaction without any central server mediation.

We present edrys-Lite's architecture, covering communication, synchronization, and modular components. An Arduino sensor station is used to illustrate how shared terminals and collaborative hardware control are achieved in practice.

2 Technical Background

Edrys-Lite utilizes a fully modular, peer-to-peer architecture that eliminates the need for any centralized backend. Every user—whether instructor, student, or hardware node—simply opens the edrys-Lite client in their browser. To interface with physical devices, modern browsers offer native APIs such as WebSerial, WebUSB, and WebBluetooth (cf. [Bf24]); if a particular integration requires it, a lightweight localhost proxy can be spun up as a fallback.

2.1 Peer-to-Peer Connectivity

Rather than channeling sensor data, control commands, or shared documents through a central server, edrys-Lite establishes direct connections between all clients via WebRTC Data Channels. WebRTC (Web Real-Time Communication) provides secure, low-latency, browser-to-browser transport for arbitrary data streams, enabling real-time updates without an intermediary. If direct peer-to-peer links cannot be formed—due to strict NATs, firewalls, or isolated networks—an instructor can fall back to a lightweight WebSocket server. In this relayed mode, all session messages (CRDT updates, control packets, and sensor values) are tunneled through a simple self-hosted WebSocket endpoint, preserving functionality without requiring a full backend infrastructure.

2.2 CRDT-Based State Synchronization

Application state is maintained using Yjs, a CRDT (Conflict-Free Replicated Data Type) library that ensures eventual consistency even if participants momentarily disconnect. In edrys-Lite, each module—whether it's a collaborative code editor, a shared terminal, or a

real-time plotting canvas—binds to its own Yjs document. Every peer keeps a local replica of each document or state object. When a user types code, issues a command, or modifies module settings, those changes are incrementally broadcast via the peer-to-peer network. Yjs deterministically merges concurrent updates, so that despite message reordering or brief network interruptions, all peers ultimately synchronize to the same state without relying on a central merge server.

2.3 Modules

Modules in edrys-Lite are the fundamental building blocks of any classroom or lab. Each module is simply an HTML page loaded in an `<iframe>` that must include the `edrys.js` script. Via the Edrys API, a module:

- Registers lifecycle hooks (`Edrys.onReady`, `Edrys.onUpdate`) to initialize itself and react to real-time state changes.
- Sends and receives scoped messages (`Edrys.sendMessage`, `Edrys.onMessage`) over the underlying P2P channels.
- Declares metadata (via standard HTML `<meta>` tags) and accepts runtime parameters through `Edrys.module.config`, `Edrys.module.studentConfig`, `Edrys.module.teacherConfig`, and `Edrys.module.stationConfig` to adapt its behavior based on role or deployment context.
- Next to this, methods for sharing state and media streaming are provided.

2.4 Laboratories

A complete edrys-Lite lab is nothing more than a text-based configuration (JSON or YAML) that lists the URLs of each module, assigns each a unique document name, and specifies any module-specific settings. When you load that file, edrys-Lite instantiates every referenced module, binds each to its own Yjs-backed CRDT document, and the communication is set up automatically WebRTC (or WebSocket). Sharing or reloading a lab is then as simple as distributing that single configuration file—with no additional backend or service configuration required.

3 Applications

A list of open laboratories can be found on GitHub at https://github.com/topics/edrys-lab. Whereby we will focus on one laboratory only in our demonstration.

3.1 Micro-controller[7] – Lab (Arduino C++)

Fig. 1. shows one of three laboratory setups used in embedded systems lectures. Each station integrates five core modules: a collaborative code editor for real-time co-authoring, a camera module for visual feedback, a terminal client that compiles and flashes the code to the Arduino board, a Markdown module for displaying lab instructions, and a station-local code logger that stores each flashed version anonymously in the browser's IndexedDB for later analysis. This setup was evaluated during the winter semesters 2023/24 and 2024/25 with students from the computer science programs across three different stations. In addition, edrys-Lite serves as a foundation for implementing student-led labs in the Master's program during the summer semesters, thus building a bridge to higher-level teaching and project-based learning.

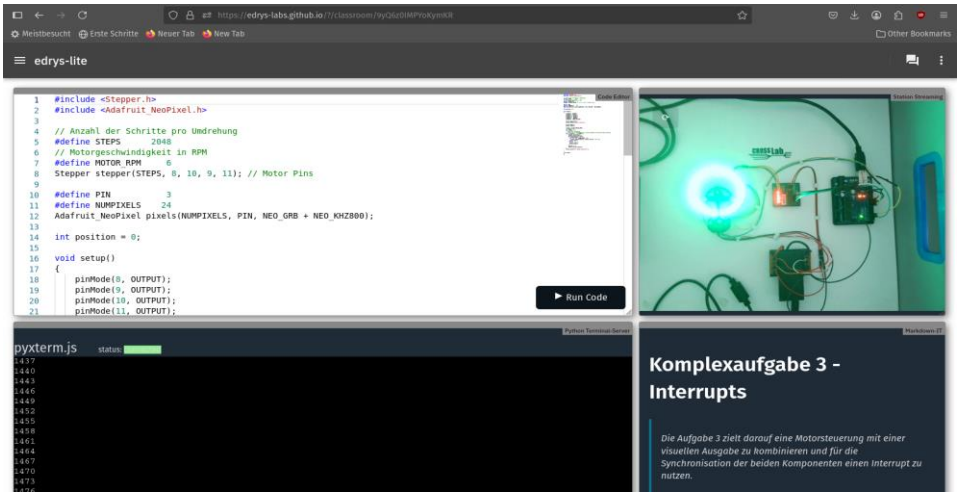


Fig. 1: Screenshot of the Arduino-Lab with 4 distinct modules.

Bibliography

[AG23] Ariza, J.Á.; Galvis, C. N.: RaspyControl Lab: A fully open-source and real-time remote laboratory for education in automatic control systems using Raspberry Pi and Python. HardwareX 13/2023.

[Bf24] Beaufort, F.: Accessing hardware devices on the web. Available at: https://web.dev/articles/devices-introduction (Accessed: 2025-07-09).

[7] See the lab-configurations at: https://github.com/edrys-labs/lab-tubaf-embedded-systems

[THB23] Tiberkak A.; Hentout A.; Belkhir A.: WebRTC-based MOSR remote control of mobile manipulators. International Journal of Intelligent Robotics and Applications 7/2023, 304-320, 2023.

Virtual Reality in Speech Therapy Education

A new approach to teach the Flexible Endoscopic Evaluation of Swallowing (FEES)

Flavia Gentile[1,2], Mascha Wanke[2], Evi Hochuli[2] and Wolfgang Mueller[1]

Abstract: Recent studies have demonstrated the efficacy of virtual reality as a powerful tool for clinical training in healthcare. It offers a safe and controlled environment where students can develop essential hands-on skills without exposing patients to risk. In the domain of speech therapy, the utilisation of this technology for educational purposes remains limited. Virtual reality holds considerable potential in the training of students in the performance of complex procedures, such as the Flexible Endoscopic Evaluation of Swallowing (FEES), which require a high level of precision, familiarity with instrumentation, and an efficient clinical decision-making process. Our research project aims at developing a virtual reality training module able to empower students' knowledge, foster their decision-making process and enhance their clinical skills while gaining hands-on experience.

Keywords: Speech Therapy, Speech Language Pathology, Virtual Reality, VR, Clinical Training, Flexible Endoscopic Evaluation of Swallowing, FEES

1 Background and Motivation

There are several reasons why the way knowledge and skills are transferred to speech therapy students needs to be innovated. The expected shortage of healthcare workers, the demand for highly qualified professionals and the ongoing shift towards university-level education focusing on scientific knowledge and evidence-based practice all call for a revolution in teaching methods. As well as theoretical knowledge, practical skills training are essential in speech therapy to train qualified professionals. Speech therapists must apply theoretical knowledge through direct patient interaction, often with complex cases. Initial training should occur in safe environments that allow mistakes without risking patient safety, enable trial, reflection, and repetition to build (confidence and) skills.

Flexible Endoscopic Evaluation of Swallowing (FEES) is a key diagnostic procedure for swallowing disorders (e. g., dysphagia). It is complex and invasive, requiring strong anatomical knowledge and advanced skills. For beginners, mastering endoscope handling and interpreting findings is challenging. Ethical and safety concerns demand structured training, making FEES integration difficult for novices. Despite the relevance of this

[1] University of Education Weingarten, Media Education and Visualisation Group (MEVIS), Kirchplatz 2, 88250, Weingarten, {flavia.gentile, mueller}@ph-weingarten.de

[2] University of Education Weingarten, German with Speech Training, Kirchplatz 2, 88250, Weingarten, {flavia.gentile, mascha.wanke, evi.hochuli}@ph-weingarten.de

doi:10.18420/delfi2025_26

In the era of digital transformation, Virtual reality (VR) has emerged as an increasingly relevant technology in the educational context, particularly in the healthcare fields [KPS16; Qu25]. The potential of VR to create immersive and interactive environments has been identified as a valuable tool in educational settings, offering novel approaches to learning that enhance traditional teaching methods [Mc14]. This technology has been demonstrated to be particularly effective in the acquisition of theoretical knowledge and practical skills [Ky19; Su24], with the ability to simulate complex situations in a controlled and safe environment. In addition, the World Health Organization [HW13] has begun to encourage and promote the development of new simulation systems with the aim of integrating them into academic healthcare curricula to foster learning and improve clinical and practical skills.

Although speech therapy is still an under-explored field in the application of e-learning, it offers interesting opportunities for the development and implementation of virtual reality. This is primarily due to the fact that it involves complex procedures that necessitate the acquisition of advanced clinical skills, which can be effectively developed through specially designed, high-level interactive experiences.

This work targets to address the specific demand for novel simulation-based training methods in ST, and to provide most realistic and scalable alternatives to state-of-the-art training methods in this field based on VR. Specifically, this research will design novel VR-enhanced training scenarios and corresponding VR training environments that allow for an individualised and repeated training of practical clinical skills in speech therapy students. We expect to develop a more comprehensive understanding of opportunities and barriers of VR in the health sector in general. In addition to design guidelines and best practices for the health sector, findings are also expected on the possibility of implementing training sequences with high requirements with regard to feedback on practical, manual procedures based on state-of-the-art interaction techniques and devices.

2 Project Objectives

This research project aims to develop a virtual reality system that incorporates speech therapy-specific clinical scenarios. This system will utilise *ad hoc* interaction metaphors to facilitate the acquisition of specific competencies, with the purpose of assisting students in the practice of clinical procedures and the development of practical skills. One scenario focuses on the FEES diagnostic procedure, supporting the translation of theoretical knowledge into hands-on experience and enhancing technical competence.

In order to implement a training module able to promote and strengthen the decision-making process in students while exploiting all the potential of the simulation-based learning, we are designing our scenario guided by the following research questions: (1) How realistically can VR-based systems replicate a speech therapy clinical behaviour? (2) To what extent and in what kind of form might haptic feedback be required to simulate

these processes faithfully? (3) In the absence of full-fledged haptic feedback, what learning objectives can be realised?

Moreover, given the proven effectiveness of simulation-based systems in combination with traditional learning methods [Mc14], the extent to which VR technology can be integrated into ST curricula to provide students with comparable skill levels to those in traditional learning settings is also being investigated.

3 Methodology and Innovation

While the project is designed and implemented using methodologies that are already known from technical, medical and educational perspectives, the interdisciplinary nature of the project brings with it innovative elements. Although immersive VR is increasingly used in healthcare training, its adoption in speech therapy remains limited. In speech and language pathology, VR is mainly applied for rehabilitation and therapy, both immersive and non-immersive, targeting communication disorders such as cognitive impairments [Br23], developmental language disorders [Ca23], neurological conditions [KPB23; St21], autism [Ha17], and stuttering [AAA21]. VR is also used alongside therapies like transcranial Direct Current Stimulation (tDCS) [Co24; Gi20]. For training speech therapy students, non-immersive simulations—including role-plays, virtual platforms, and manikins—are common [RCS17], whereas immersive VR for practical skill training is still limited, mostly in prototype form, or only evaluated from students' perspectives [KWU23].

The project uses a design-based research (DBR) methodology [AR08; MR18; Th03; Va06], enabling a three-iteration process to develop and refine the FEES scenario. Its efficacy will be evaluated after each cycle, with adjustments implemented in subsequent stages. The scenario is designed following instructional design principles [Ga05], based on learning objectives and skills identified by speech therapy experts. Personas and scenario-based design methods [PG03; RC02] guide the design and development. It features realistic clinical cases reconstructed by experts, enhanced with artificial intelligence for detailed learning. Users progress through interactive, scenario-based challenges simulating clinical reasoning, supported by an integrated feedback and error system that offers immediate, constructive responses to foster practical skills and critical judgment. Fig. 1 provides a depiction of a flow-chart for the overall VR-enhanced training scenario.

The VR training module development is based on the VR learning scenario design and follows two main phases: prototyping in Spatial and full development in Unity. Spatial is used to quickly create an initial prototype and enable collaborative spaces for onboarding and gathering student feedback. This allows assessment of student interest, acceptance, and engagement with immersive VR, informing further development. Unity is being used to enhance system interactivity and user feedback, streamlining the development process.

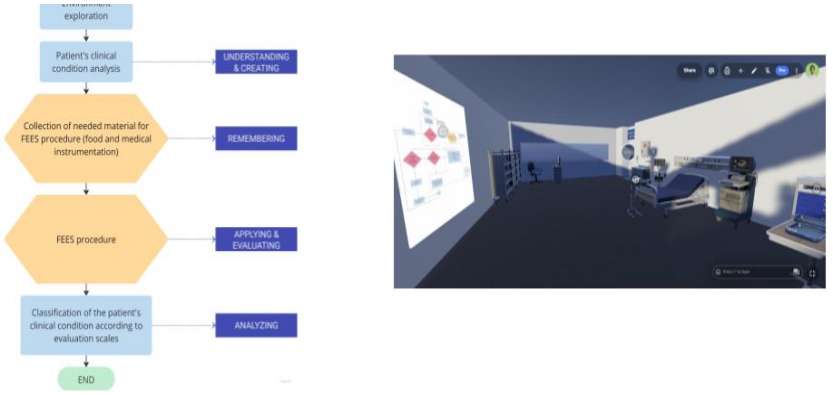


Fig. 1: FEES flowchart and Spatial environment

The project evaluation is going to be performed under two main aspects: the VR-based technology assessment and the speech therapy appraisal. On the one hand, the VR-technology will be evaluated according to three key perspectives, i.e. acceptance, user experience and usability. Data will be collected through already recognised methods, the Technology Acceptance Model (TAM) [Da89], User Experience Questionnaire (UEQ) [SHT14; SHT17] and System Usability Scale (SUS) [Br96]. Additionally, VR-specific measures such as Presence (Slater-Usoh-Steed Questionnaire, SUS) [SUS94] will be assessed. On the other hand, speech therapy knowledge and practical skills' acquisition will be evaluated. Since standardized instruments are lacking in the area [Ge25; KE23], we target to evaluate the effectiveness of VR-based training within a controlled study. Existing instruments aimed at evaluating patients' severity in real-life scenarios, such as the Interaction of Assigned Safety and Efficiency Grades (DIGEST Score), might be integrated in the evaluation process to assess students' progress. Although formal evaluation tests have not been performed yet, a first qualitative assessment of one of the scenarios of the system was conducted. A small sample of users was asked to try the system out and to answer a short questionnaire addressing key themes such as usability, learning effect and physical reactions to VR. Overall, the VR-based was positively accepted by users since it was rated as intuitive, user-friendly and fostered motivation. Aspects such as dizziness or eye fatigue were reported as well but did not affect the overall experience.

4 Conclusions

The project addresses a current demand for training modules in the field of Speech Therapy, a specific field of the health sector. Specific VR-enhanced training scenarios have been designed in collaboration with experts from the field, and first prototypes for VR training modules for the FEES process have been designed. First formative evaluations with Speech Therapy students provided positive results and feedback.

Bibliography

[AAA21] Al-Nafjan, A.; Alghamdi, N.; Almudhi, A.: Virtual Reality Technology and Speech Analysis for People Who Stutter. EMITTER International Journal of Engineering Technology, 9(2), 326–338, 2021.

[AR08] Amiel, T.; Reeves, T. C.: Design-based research and educational technology: Rethinking technology and the research agenda. Journal of Educational Technology Society, vol. 4, no. 11, 2008.

[Br23] Brassel, S.; Brunner, M.; Power, E.; Campbell, A.; Togher, L.: Speech-language pathologists' views of using virtual reality for managing cognitive-communication disorders following traumatic brain injury. American Journal of Speech-Language Pathology, vol. 32, no. 2S, pp. 907–923, 2023.

[Br96] Brooke, J.: SUS: A "quick and dirty" usability scale. In: Jordan, P.; Thomas, B.; Weerdmeester, B.; McClelland, I. L. (Eds.): *Usability Evaluation in Industry*, pp. 189–194, Taylor & Francis, 1996.

[Bro23] Brown, K. M.; Swoboda, S. M.; Gilbert, G. E.; Horvath, C.; Sullivan, N.: Curricular Integration of Virtual Reality in Nursing Education. Journal of Nursing Education, 62(6), 364–373, 2023.

[Ca23] Cappadona, I.; Ielo, A.; La Fauci, M.; Tresoldi, M.; Settimo, C.; De Cola, M. C.; Muratore, R.; De Domenico, C.; Di Cara, M.; Corallo, F.; Tripodi, E.; Impallomeni, C.; Quartarone, A.; Cucinotta, F. Feasibility and Effectiveness of Speech Intervention Implemented with a Virtual Reality System in Children with Developmental Language Disorders: A Pilot Randomized Control Trial. Children 2023, 10 (8), 1336.

[Co24] Contrada, M.; Scarfone, F.; Iozzi, A.; Carozzo, S.; Vatrano, M.; Nicoletta, M. G.; Nudo, G.; Quintieri, M.; Tonin, P.; Cerasa, A.: Combining computer-based rehabilitative approach with tDCS for recovering of aphasia: Implications from a single case study. Clinical Case Reports, 12(6), e8928, 2024.

[Da89] Davis, F. D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319–340, 1989.

[FH24] Frieling, T.; Heinrich, H. S.: Dysphagie – Ursachenspektrum, Differenzialdiagnostik und Therapie. Gastroenterologie, no. 19, pp. 474–493, 2024.

[Ga05] Gagné, R. M.; Wager, W. W.; Golas, K. C.; Keller, J. M.; Russell, J. D.: Principles of Instructional Design, 5th ed. Wadsworth Cengage Learning, 2005.

[Gi20] Giachero, A.; Calati, M.; Pia, L.; La Vista, L.; Molo, M.; Rugiero, C.; Fornaro, C.; Marangolo, P.: Conversational Therapy through Semi-Immersive Virtual Reality Environments for Language Recovery and Psychological Well-Being in Post Stroke Aphasia. Behavioural Neurology, 1–15, 2020.

[Ge25] Gentile, F.; Wanke, M.; Mueller, W.; Hochuli, E.: Virtual Reality in Speech Therapy Students' Training: a Scoping Review. *Preprints*, 2025.

[Ha17] Halabi, O.; Abou El-Seoud, S.; Alja'am, J.; Alpona, H.; Al-Hemadi, M.; Al-Hassan, D.: Design of Immersive Virtual Reality System to Improve Communication Skills in

[HW13] Health Workforce (HWF); WHO Headquarters (HQ): Transforming and scaling up health professionals' education and training. World Health Organization, 2013.

[KE23] Kalunga, R.; Elshobosky, F.: Using Immersive Technologies to Enhance Student Learning Outcomes in Clinical Sciences Education and Training. *International Conference on Education and New Developments*, 2023.

[KPB23] Kuznetsov, A. A.; Petrenko, T. S.; Borisov, V. I.: Cognitive Rehabilitation in Virtual Reality (Modern Review). 2023 IEEE Ural-Siberian Conference on Computational Technologies in Cognitive Science, Genomics and Biomedicine (CSGB), 94–99, 2023.

[KPS16] Kim, J.; Park, J. H.; Shin, S.: Effectiveness of simulation-based nursing education depending on fidelity: a meta-analysis. BMC Medical Education, vol. 16, 152, 2016.

[KWU23] Kelly, B.; Walters, J.; Unicomb, R.: Speech Pathology Student Perspectives on Virtual Reality to Learn a Clinical Skill. *Teaching and Learning in Communication Sciences and Disorders*, vol. 7, no. 1, 2023.

[Ky19] Kyaw, B.; Saxena, N.; Posadzki, P.; Vseteckova, J.;, Nikolaou, C.; George, P.; Divakar, U.; Masiello, I.; Kononowicz, A.; Zary, N.; Tudor Car, L.: Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res, 21(1):e12959, 2019.

[Le24] Ledl, C.; Frank, U.; Dziewas, R.; Arnold, B.; Bähre, N.; Betz, C. S.; Braune, S.; Deitmer, T.; Diesener, P.; Fischer, A. S.; Hamzic, S.; Iberl, G.; Konradi, J.; Löhler, J.; Platz, T.; Rohlfes, C.; Westhoff, M.; Winkler, S.; Wirth, R.; Graf, S.: *Curriculum „Trachealkanülenmanagement in der Dysphagietherapie"*. Der Nervenarzt, 95(4), 342–352, 2024.

[Mc14] McGaghie, W. C.; Issenberg, S. B.; Barsuk, J. H.; Wayne, D. B.: A critical review of simulation-based mastery learning with translational outcomes. Medical Education, 48(4), 375–385, 2014.

[MR18] McKenney, S.; Reeves, T.: Conducting Educational Design Research, 2nd ed. London, UK: Routledge, 2018.

[PG03] Pruitt, J. S.; Grudin, J.: Personas: Practice and Theory. Designing for User eXperiences, 1–15, 2003.

[Qu25] Quah, T. C. S.; Lau, Y.; Ang, W. W.; Lau, S. T.: Experiences of immersive virtual reality in healthcare clinical training for nursing and allied health students: A mixed studies systematic review. Nurse Education Today, 148, 106625, 2025.

[RC02] Rosson, M.; Carroll, J.: Usability Engineering: Scenario-Based Development of Human-Computer Interaction (Interactive Technologies). Elsevier Science, 2002.

[RCS17] Rose, T. A.; Copley, A.; Scarinci, N. A.: Benefits of providing an acute simulated learning environment to speech pathology students: An exploratory study. Focus on Health Professional Education: A Multi-Professional Journal, 18(3), 44–59, 2017.

[St21] Stasolla, F.; Matamala-Gomez, M.; Bernini, S.; Caffò, A. O.; Bottiroli, S.: Virtual reality as a technological-aided solution to support communication in persons with neurodegenerative diseases and acquired brain injury during COVID-19 pandemic. Frontiers in Public Health, vol. 8, 2021.

[Su24] Sung, H.; Kim, M.; Park, J.; Shin, N.; Han, Y.: Effectiveness of Virtual Reality in Healthcare Education: Systematic Review and Meta-Analysis. Sustainability, 16(19), 8520, 2024.

[SHT14] Schrepp, M.; Hinderks, A.; Thomaschewski, J.: Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In: Marcus, A. (Ed.): *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience*, pp. 383–392, Springer, Cham, 2014.

[SHT17] Schrepp, M.; Hinderks, A.; Thomaschewski, J.: Construction of a benchmark for the user experience questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, pp. 40–44, 2017.

[SUS94] Slater, M.; Usoh, M.; Steed, A.: Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 3(2), 130–144, 1994.

[Th03] The Design-Based Research Collective: Design-based research: An emerging paradigm for educational inquiry. Educational Researcher, vol. 32, no. 1, pp. 5–8, 2003.

[Va06] Van den Akker, J.; Gravemeijer, K.; McKenney, S.; Nieveen, N.: Educational Design Research, 1st ed. London, UK: Routledge, 2006.

Enhancing Vocational Education through AI-driven Learning and Competence Assessment: The DiAgnostiK Project

Volker Göhler [1], Sebastian Zug [1], and Hilke Domsch[2]

Abstract: The DiAgnostiK project develops digital tools for the systematic assessment and support of competencies in inter-company apprenticeship training. Up to now, digital or asynchronous learning methods have not been implemented in this setting. Taking into account the different skill levels of apprentices, DiAgnostiK introduces digital tools for the continuous assessment of theoretical competencies, a framework for the documentation of practical skills, and an integrated platform connecting all involved parties. Using Large Language Models, DiAgnostiK generates adaptive, interactive learning materials and questions specifically tailored to trades such as carpentry, interior design, painting, and dental technology. This significantly reduces the workload for instructors. It is hypothesized that learning methods tailored to the trades increase accessibility for students, while digital and asynchronous access to learning materials allows students to learn at their convenience. Consequently, DiAgnostiK significantly improves the quality of vocational education and reduces dropout rates. To our knowledge, this is the first systematic application of LLM-generated content tailored for German inter-company vocational training.

Keywords: Vocational Education, Adaptive Learning, Large Language Models, Competence Assessment

1 Introduction

Inter-company apprenticeship training (german: Überbetriebliche Lehrlingsunterweisung (IAT)) represents a fundamental pillar of the German dual vocational education and training (VET) system, complementing learning in companies and vocational schools. Conducted in inter-company training centers, IAT aims to impart specialized practical and theoretical competencies that individual companies may not fully provide. Courses typically last one to three weeks and involve 6–12 apprentices. They are usually required to attend between one and five of these courses per training year. The curricula for IAT are developed by the Heinz-Piest-Institut für Handwerkstechnik [3]. Despite clearly defined guidelines outlined in vocational training regulations and curricula, IAT encounters substantial challenges due to the wide variation in prior education, skill levels, and learning needs among apprentices. This heterogeneity complicates efficient and targeted competency transfer, particularly within the constrained timeframes allocated to IAT courses. Consequently, apprenticeship dropout

[1] TU Bergakademie Freiberg,
volker.goehler@informatik.tu-freiberg.de, https://orcid.org/0000-0003-1901-1335;
sebastian.zug@informatik.tu-freiberg.de, https://orcid.org/0000-0001-9949-6963
[2] Geokompetenzzentrum Freiberg e.V., hilke.domsch@gkz-ev.de
[3] See https://hpi-hannover.de, accessed on 2025-04-08.

doi:10.18420/delfi2025_27

rates remain persistently high, specifically at 10 % for carpenters, 20 % for painters, and 28 % for interior designers between 2018 and 2022 [Bu23]. Additionally, examination retake rates are relatively high, averaging around 15 % according to internal surveys conducted by the Handwerkskammer Dresden.

A preliminary study conducted by us involving 11 different crafts revealed significant disparities in course execution. Achievements in only 45 of 133 IAT courses are systematically collected and evaluated. Assessments predominantly focus on vocational-practical competencies. Only 2 courses used IT-supported assessment methods, with paper-pencil tests and oral questioning prevailing. Structured competency assessments typically occur at course completion (in 41 courses), with only one course conducting an initial diagnostic assessment. Thus, a structured, continuous measurement and evaluation of performance and competency development throughout IAT training remains unimplemented.

To address these deficits, we pursue two central goals: (G1) implementing digital instruments for the structured assessment of theoretical competencies and learning progress, and (G2) developing instructor-facing dashboards that visualize apprentices' competencies to enable timely, adaptive support. These goals aim to introduce formative diagnostics, support individualization in heterogeneous groups, and increase transparency in skill acquisition.

2 Background

Multiple drivers are fueling a shift toward digital and AI-supported methods in vocational education. Germany faces a worsening skilled labor shortage and must train large numbers of new workers efficiently. At the same time, many vocational skills can only be learned through practice, which can be costly or risky (e. g. handling expensive equipment or live animals) [Pa25; Re22]. Gessler and Howe [GH15] emphasize the alignment of vocational training with real-world work experiences, often mediated through digital simulations, virtual labs, or workplace-oriented e-learning environments. Ding [Di23] stresses the importance of integrating digital resources to maintain relevance and adaptability in evolving industries.

To address these demands, digital learning platforms are increasingly used to complement traditional hands-on training. Nationwide initiatives such as eCampus Handwerk [Dr23] and widely adopted platforms like OPAL or Moodle[4] support the delivery of course content and learning activities. In addition, sector-specific tools such as Cornelsen eCademy and Tischler-Schreiner-Campus[5] provide tailored learning resources for specific trades.

Recent advances in Large Language Models (LLMs) enable the automated generation of learning tasks that align with pedagogical goals such as competency orientation and formative assessment, while also accounting for domain-specific requirements in vocational contexts.

[4] https://www.opal-schule.de; https://bildungsportal.sachsen.de/opal/; https://moodle.org, accessed 2025-04-09
[5] https://app.ecademy-learning.com/ and https://www.tischler-schreiner-campus.de, accessed 2025-04-09

To leverage this potential, we are developing a toolchain that assists instructors in creating context-aware tasks and quizzes. These quizzes can be rapidly formulated in natural language that reflects the terminology and scenarios of the respective trade, thereby fostering acceptance and recognizability among apprentices.

The feasibility of such approaches is supported by recent research. Gopi et al. [GSD24] present a pipeline that combines retrieval augmented generation (RAG) and LLMs to automatically generate adaptive quizzes in engineering education. Maity et al. [MDS25] show how structured prompting strategies can elicit pedagogically sound and context-sensitive learning tasks across a broad range of subjects. Building on these foundations, the following section outlines our solution concept.

3 AI-Based Solution Concept

Our approach introduces a LLM based toolchain that facilitates domain-specific content creation and competence diagnostics in vocational education – an area largely unexplored in existing research. To this end, we analysed foundational IAT courses in five trades to identify common learning challenges and thematic overlaps. We chose professions with around 150 students each year [6] as well as low-enrolment professions [Ha23]. Our investigation showed that all courses need competencies in safety, workplace conduct, and basic mathematics.

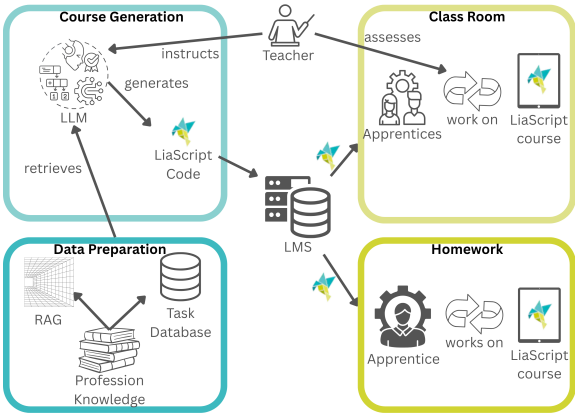


Fig. 1: Overview of the AI-assisted content generation and deployment within the DiAgnostiK project.

In alignment with G1, we developed a library of generic task templates targeting these core areas. LLMs are used to adapt these templates to the context of specific professions – for example, by framing a mathematical area calculation as a paint requirement estimation task.

[6] Plant mechanic for sanitary, heating and air conditioning systems (german: Anlagenmechaniker SHK), carpenter (german: Tischler), and painter and varnisher (german: Maler und Lackierer) had between 100 and 150 apprentices between 2021 and 2023. Dental technician (german: Zahntechniker) had between 35 and 50 apprentices, and interior decorator (german: Raumausstatter) had around 10.

This contextualization increases relevance and comprehension among learners. We use LiaScript [DZ25] to deliver these exercises in an interactive, easy to adapt, markdown-based format. LiaScript enables serverless assessments and real-time feedback through peer-to-peer functionality, and integrates into existing learning management systems (LMS), and as such supports G2.

Fig. 1 illustrates the end-to-end workflow of content generation, classroom usage, and feedback within the DiAgnostiK project[7]. This workflow enables a teacher-driven yet scalable content pipeline that adapts to trade-specific needs. The process begins with a teacher prompting a LLMs to generate LiaScript-based markdown, grounded on structured data from a curated knowledge base and task repository. These resources are provided through RAG and enable the creation of instructional material tailored to specific trades, including appropriate question types from a predefined catalog. The generated content is deployed via a LMS and supports both supervised classroom instruction and self-directed learning. LiaScript's interactive features and its serverless, peer-to-peer assessment approach enable learners to exchange real-time feedback without relying on centralized infrastructure.

Together, these components implement structured digital assessment (G1) and support real-time pedagogical decision-making (G2), contributing to more transparent and individualized vocational training.

4 Conclusion

We propose a novel AI-supported toolchain for vocational education, specifically tailored to the requirements of IAT courses. By contextualizing abstract learning content with domain-specific tasks generated by large language models and delivering them via LiaScript, we aim to create scalable, adaptable, and inclusive digital learning environments. To ensure relevance and usability, a co-creation team comprising instructors, apprentices, and educational researchers will guide the iterative refinement of task templates, feedback mechanisms, and the instructor dashboard. Throughout the development process, we adhere to principles of ethical AI use: all generated content is subject to human review, and data collection complies with GDPR regulations.

To empirically validate our approach, a longitudinal study involving all participating apprentices is planned over a period of three years. We will collect both performance data and qualitative feedback to systematically evaluate the effectiveness, acceptance, and areas for improvement.

Our approach demonstrates how AI can sustainably enhance the quality of vocational education while supporting more personalized and transparent learning processes.

[7] See https://uelu-digital.de for recent information.

Acknowledgements

This project is co-financed by the European Union. The measure is additionally co-financed with tax funds on the basis of the budget adopted by the Saxon State Parliament.

References

[Bu23] Bundesinstitut für Berufsbildung (Hrsg.): Datenreport zum Berufsbildungsbericht 2023. Informationen und Analysen zur Entwicklung der beruflichen Bildung. Bundesinstitut für Berufsbildung, 2023.

[Di23] Ding, T.: Social Dialogue and Its Role in Germany Vocational Education and Training. Lecture Notes in Education Psychology and Public Media 2 (1), pp. 715–722, 2023, DOI: 10.54254/2753-7048/2/2022422.

[Dr23] Dresen, B.: Einer für alle(s): Der eCampus Handwerk bringt digitales Lernen nach vorn - BMBF InnoVET, 2023, https://www.inno-vet.de/innovet/de/aktuelles/ProNet_Handwerk_eCampus.html.

[DZ25] Dietrich, A.; Zug, S.: LiaScript, 2025, https://liascript.github.io, accessed: 04/05/2025.

[GH15] Gessler, M.; Howe, F.: From the reality of work to grounded work-based learning in German vocational education and training: Background, concept and tools, 2015, DOI: 10.13152/IJRVET.2.3.6.

[GSD24] Gopi, S.; Sreekanth, D.; Dehbozorgi, N.: Enhancing Engineering Education Through LLM-Driven Adaptive Quiz Generation: A RAG-Based Approach. In: Proceedings - Frontiers in Education Conference, FIE. Institute of Electrical and Electronics Engineers Inc., 2024, DOI: 10.1109/FIE61694.2024.10893146.

[Ha23] Handwerkskammer Dresden: Entwicklung der Ausbildungsverhältnisse nach Berufen (2023 bis 2021), tech. rep., Handwerkskammer Dresden, 2023, https://www.hwk-dresden.de/service/analysen-berichte-und-statistiken/detail/lehrlingsstatistik.html.

[MDS25] Maity, S.; Deroy, A.; Sarkar, S.: Leveraging In-Context Learning and Retrieval-Augmented Generation for Automatic Question Generation in Educational Domains. In: 16th Meeting of the Forum for Information Retrieval Evaluation. ACM, 2025, http://arxiv.org/abs/2501.17397.

[Pa25] Paaßen, B.: Die Rolle Künstlicher Intelligenz in der beruflichen Bildung. Berliner Bildungszeitschrift 78, pp. 20–21, 2025, https://www.gew-berlin.de/aktuelles/detailseite/die-rolle-kuenstlicher-intelligenz-in-der-beruflichen-bildung.

[Re22] ReferNet Germany: Germany: virtual reality, AI and other digital technologies in VET, tech. rep., Cedefop, 2022, https://www.cedefop.europa.eu/en/news/germany-virtual-reality-ai-and-other-digital-technologies-vet.

Lernendenzentriertes Onboarding in RSE: Ein Forschungsdesign zur Persona-Entwicklung



Jasmin Hartanto [1], Birte Heinemann [1] und Ulrik Schroeder [1]

Abstract: Dieser Posterbeitrag stellt ein geplantes Forschungsdesign vor, das Lernende mit heterogenem Vorwissen in den Mittelpunkt stellt. Ziel ist es, mittels eines strukturierten Fragebogens Informationen zu Motivation, Berufszielen, Herausforderungen und Weiterbildungsbedarfen von Personen im Research Software Engineering (RSE)-Kontext zu erheben. Diese Erkenntnisse sollen zur Entwicklung von Personas beitragen, die als Grundlage für zielgruppenspezifisches Onboarding-Material und didaktische Konzepte dienen können. Der Beitrag regt eine Diskussion über Lernende im RSE-Bereich an – eine Gruppe, die häufig interdisziplinär geprägt ist und vor besonderen Herausforderungen steht. Anders als klassische Softwareentwickler*innen verfügen RSE-Lernende nicht immer über formale Ausbildung in Software Engineering, müssen aber dennoch robuste, nachhaltige und vertrauenswürdige Forschungssoftware entwickeln. Dies erfordert Weiterbildungsangebote, die flexibel, niedrigschwellig und auf Teilhabe ausgerichtet sind. Da viele dieser Lernenden zeitlich eingebunden sind, kommt Konzepten des selbstgesteuerten eLearnings mit hoher Verfügbarkeit und modularem Aufbau besondere Bedeutung zu. Mit dem vorgestellten Fragebogen sollen auch Einflussfaktoren auf ein gelungenes Onboarding identifiziert werden – im Sinne einer praxisnahen, nutzendenzentrierten Weiterbildung, die der Vielfalt in der RSE-Community gerecht wird.

Keywords: RSE, Research Software Engineering, Research Software Development, Education, Training, Forschungsdesign

1 Einleitung und Motivation

Forschungssoftware ist ein unentbehrliches Werkzeug, um Forschungsdaten zu erzeugen, analysieren und visualisieren und bildet somit ein omnipräsentes Thema in der heutigen Forschungslandschaft. Hierbei fällt die Entwicklung dieser Software (Research Software Engineering, RSE) häufig auf die Forscher*innen selbst zurück und zwingt sie, die Rolle von Softwareentwickler*innen anzunehmen – häufig ohne das nötige Vorwissen und Training. Das breite Spektrum an Forschungsdisziplinen, die vielfältigen Anforderungen der Forschungssoftware selbst sowie das heterogene Vorwissen unter den Betroffenen erschwert das Design von einheitlichen Hilfestellungen anzubieten oder konsistente Onboardingstrategien und Trainingsmaterialien für RSE-Lernende zu entwickeln.

Eine vorher durchgeführte Literaturrecherche [HHS25] hat ergeben, dass fehlendes Wissen über Praktiken der Softwareentwicklung eine der primär identifizierten Herausforderungen

[1] RWTH Aachen University, Templergraben 55, 52062 Aachen, Germany,
hartanto@cs.rwth-aachen.de, https://orcid.org/0009-0003-3916-917X;
heinemann@cs.rwth-aachen.de, https://orcid.org/0000-0002-7568-0704;
schroeder@informatik.rwth-aachen.de, https://orcid.org/0000-0002-5178-8497

doi:10.18420/delfi2025_28

ist, der sich RSE-Lernende stellen müssen. Die Konsequenz ist oftmals eine ad-hoc durchgeführte Entwicklung, resultierend in nicht-nachhaltiger Software, die nach Ende der Förderperiode des Forschungsprojekts keine weitere Verwendung findet. Um sich diesem Problem zu stellen, engagiereen sich Organisationen wie de-RSE e.V., Software Carpentry und das Software Sustainability Institute aktiv für eine Verbesserung der Situation und bieten unter anderem Lernprogramme sowie Sammlungen von Lehrmaterialien an. Neben Trainingsmaterialien gibt es jede Menge Guidelines und Best Practices, die jedoch ihrerseits häufig aufgrund der Heterogenität der Zielgruppe sehr allgemein gehalten sind [PP12; Wi16].

Die Zielsetzung dieses Beitrags ist, die Zielgruppe in ihrer Diversität besser einzufangen. Hierzu wird ein Forschungsdesign zur Erstellung von Personas entwickelt, welche die Entwicklung für Onboarding- und Weiterbildungskonzepte für RSE-Lernende ermöglichen soll. Darüber hinaus sollen in der Theorie gefundene Probleme [HHS25] in der Praxis bestätigt werden.

2 Hintergrund und Problemstellung

Die Eingruppierung typischer RSE-Lernender gestaltet sich als komplex, da eine Vielzahl fachlicher Hintergründe vertreten sein kann – von Informatik über Ingenieurswissenschaften bis hin zu Naturwissenschaften. Zudem bringen die Beteiligten oft ein breites Spektrum an Programmierkenntnissen mit: In vielen Studiengängen wird gutes Software Engineering entweder gar nicht oder nur oberflächlich behandelt. Auch die Motivationen, sich intensiver mit guten Praktiken des Software Engineerings auseinanderzusetzen, können äußerst vielfältig sein und müssen deswegen untersucht werden. Dies verdeutlicht die Notwendigkeit, differenzierte Lösungsansätze zu entwickeln, um den individuellen Bedürfnissen dieser Lernenden gerecht zu werden, wobei einigen möglicherweise nicht einmal bewusst sein mag, dass sie Unterstützung benötigen.

Während die Probleme im Kontext von RSE im wissenschaftlichen Bereich durchaus bekannt sind [An20], sind die bestehenden Trainingsmöglichkeiten, die in Kapitel 1 angesprochen wurden, bisher nicht ausreichend, um den Herausforderungen gerecht zu werden. Die Situationen der RSE-Lernenden sind äußerst individuell, und sowohl synchrone als auch nicht-personalisierte Lernangebote reichen nicht aus, um einen erfolgreichen Einstieg in das Gebiet der RSE zu gewährleisten. Daher besteht die Notwendigkeit, ein an die jeweilige Situation angepasstes, teilhabeorientiertes Onboarding zu entwickeln, um RSE-Lernenden die erste Hürde im Prozess des Erlernens und Anwendens guter Software-Engineering-Praktiken im Kontext von RSE zu erleichtern.

Die Heterogenität der Zielgruppe im Bezug auf Vorkenntnisse, Motivation und mehr stellt eine große Herausforderung in der Entwicklung von Trainingsmaterialien dar. Ein vielversprechendes Werkzeug zur Bewältigung dieser Herausforderungen sind Personas, fiktive Charaktere, die auf realen Daten basieren und helfen, verschiedene Lernendengruppen

besser zu verstehen und deren Bedürfnisse abzubilden. Durch den Einsatz von Personas lässt sich der Designprozess der Lehrmaterialien gezielt auf die Bedürfnisse der verschiedenen Lernendengruppen abstimmen und bildet somit einen wichtigen Grundstein für deren Konzeption. Ein Beispiel dafür findet man im Bereich Forschungsdatenmanagement (FDM). FDM weist potenziell ähnliche Zielgruppen auf und sieht sich vergleichbaren Schwierigkeiten gegenüber. Ein mögliche Eingruppierung der Lernenden wurde hier anhand der Variablen Gruppengröße. Disziplinarkultur, Gruppen- und -leitungsqualität, Teamkultur und Heterogenität innerhalb des Teams getätigt, die die Annahmetendenz neuer Techniken und Herangehensweise beeinflussen [LR21]. Eine weitere mögliche Herangehensweise ist die Erstellung der Personas auf Basis von Meinungen anhand eines Clusterings möglicher Statements im Bezug auf relevante Themen im Bezug auf FDM und einer Zuordnung dieser Meinungen auf verschiedene Tätigkeitsfelder der Nutzenden [BAL24].

3 Forschungsdesign

Wie bereits in Kapitel 2 erläutert, ermöglichen Personas eine differenzierte Darstellung der Diversität innerhalb der Zielgruppe. Das Ziel des Forschungsdesigns besteht daher darin, folgende Informationen über die unterschiedlichen RSE-Lernenden zu ermitteln:

- Was sind die Motivationen der RSE-Lernenden?
- Welche beruflichen Ziele streben sie an?
- Welche Herausforderungen und Bedürfnisse treten beim Einstieg auf?

Um die Zielgruppe in ihrer gesamten Vielfalt einzufangen, ist ein Mixed-Methods-Ansatz geplant. Dieser ermöglicht es, einen primär quantitativen Fragebogen im Fokus zu haben, der es erlaubt, viele RSE-Lernende gleichzeitig zu befragen und gleichzeitig qualitativ explorativ nach tiefergehenden Begründungen zu suchen. Die Zielgruppe sind hierbei die Newcomer im RSE-Bereich, was sowohl neue Beschäftigte oder Abschlussarbeiter*innen im akademischen Sektor als auch Mitarbeitende in angewandten Forschungsinstitutionen umfasst. Das Forschungsdesign richtet sich speziell an die entscheidende Phase des Onboardings und zielt darauf ab, wertvolle Einblicke in die Erfahrungen und Bedürfnisse dieser Lernenden zu gewinnen und fragt folgende Kategorien ab:

- Motivation: Welche Aspekte sprechen die RSE-Lernenden an? Was motiviert sie, gute Methoden und Praktiken des Software Engineerings zu nutzen?
- Vorerfahrung: Wie ist die Selbsteinschätzung der RSE-Lernenden bezüglich ihres Wissensstandes über gute Praktiken des Software Engineerings?
- Lernpräferenzen: Welche Art des Lernens bevorzugen die RSE-Lernenden? Welche verschiedenen Lernmodi muss man ihnen bieten, damit sie sich mit den Themen auseinandersetzen? Wie viel Zeit und Kapazitäten dürfen die Lernmaterialien in Anspruch nehmen?

- Hürden: Welche der identifizierten Hürden beschäftigen die RSE-Lernenden primär? Wobei brauchen sie besonders viel Unterstützung?

4 Erwartete Ergebnisse & Ausblick

Das Ziel ist die Entwicklung mehrerer Personas zur Repräsentation typischer Lernendengruppen. Dazu wird die Studie über mehrere Monate ausgeführt, um die diversen RSE-Akteure einfangen zu können. Eine erneute Befragung nach erfolgtem Onboarding nach aktuellem Stand ist ebenfalls möglich, um weitere Erkenntnisse zu sammeln. Die Personas können die verschiedenen Wissensstände der Lernenden abbilden, ihre Hintergründe und ihre Motivationen, sich mit der Thematik auseinanderzusetzen, z. B. „Pauline, die Maschinenbaudoktorandin, die für ihr Forschungsprojekt nahezu ohne Vorerfahrung Simulationen programmieren muss“ oder „Jan, der theoretische Physiker, der für seinen neuen Job in ein großes Softwareprojekt hineingesetzt wurde und nun Probleme hat, sich dort zurechtzufinden“. Sie werden bei der Ableitung didaktischer Anforderungen helfen, um den verschiedenen Gruppen der RSE-Lernenden passend zugeschnittene Lernmaterialien entwickeln zu können. Bei der Entwicklung der Personas kann sich auf Personas aus dem Bereich FDM bezogen werden, da die Bereiche RSE und FDM verwandte Zielgruppen aufweisen. Neben den Personas besteht die Möglichkeit, direkt einige Best Practices und Guidelines ableiten, die direkt mit in den Entwurf der Lehrmaterialien einfließen können.

Die Ergebnisse erlauben es, ein Onboarding zu entwerfen, das die unterschiedlichen Bedürfnisse der RSE-Lernenden abdeckt. Neben Onboardingmaterialien können Weiterbildungsmöglichkeiten für weitere Anwendungsfälle aus verschiedenen Fachdomänen oder Programmierszenarien entworfen werden. Lehr- und Lernmaterialien können adressatengerecht entworfen werden und somit die didaktischen Anforderungen der Zugänglichkeit, Selbststeuerung, Teilhabe und Nachhaltigkeit erfüllen. Die entwickelten Materialien können durch Learning Analytics und weitere Lernendenbefragungen evaluiert und kontinuierlich verbessert werden.

Wir laden Sie herzlich dazu ein, mit uns zu diskutieren. Unser Projekt befindet sich noch im frühen Stadium und wir möchten die Community bestmöglich in das Forschungsdesign miteinbeziehen, um eine möglichst umfassende Studie durchführen zu können!

Danksagung

Unterstützt durch die Klaus Tschira Stiftung im Rahmen der Förderlinie *Scientific Software*.

Literaturverzeichnis

[An20] Anzt, H. et al.: An Environment for Sustainable Research Software in Germany and Beyond: Current State, Open Challenges, and Call for Action. F1000Research 9, S. 295, 2020, DOI: 10.12688/f1000research.23224.1, arXiv: 2005.01469 [cs].

[BAL24] Bernoth, J.; Al Laban, F.; Lucke, U.: Utilizing Personas to Create Infrastructures for Research Data and Software Management. 2024, DOI: 10.18420/INF2024_180.

[HHS25] Hartanto, J.; Heinemann, B.; Schroeder, U.: Scalable Training for Sustainable Science: A Scoping Review of Research Software Engineering Education and Practice, 2025, DOI: 10.5281/zenodo.15829239.

[LR21] Llebot, C.; Rempel, H. G.: Why Won't They Just Adopt Good Research Data Management Practices? An Exploration of Research Teams and Librarians' Role in Facilitating RDM Adoption. Journal of Librarianship and Scholarly Communication 9 (1), 2021, DOI: 10.7710/2162-3309.2321.

[PP12] Prlić, A.; Procter, J. B.: Ten Simple Rules for the Open Development of Scientific Software. PLoS Computational Biology 8 (12), e1002802, 2012, DOI: 10.1371/journal.pcbi.1002802.

[Wi16] Wilson, G. et al.: Good Enough Practices in Scientific Computing, 2016, DOI: 10.48550/arXiv.1609.00037, arXiv: 1609.00037 [cs].

Design of Interactive Learning Materials with H5P

Challenges in Applying Principles of Multimedia Design

Anja Hawlitschek [1], Alisa Poleshchuk [2], Nadine Wegmeyer[3], Marie-Louise Hilberer[4] and Christin Deinert[5]

Abstract: The instructional design of digital learning materials should be based on multimedia learning research, as this ensures that the materials are aligned with how people process information and learn effectively. This article describes the challenges of implementing central design principles, such as signaling, spatial contiguity, and support for coherence formation, in the "Interactive Book" content type using the authoring tool H5P. Technical limitations are identified, and possible solutions are presented. The article emphasizes the importance of considering principles of multimedia design in the development of authoring tools and provides practical support for teachers and lecturers developing digital learning material with H5P.

Keywords: H5P, design principles, multimedia

1 Introduction

The design of digital teaching and learning materials benefits from using principles from multimedia learning research, such as the Cognitive Theory of Multimedia Learning (CTML) [No21]. These principles are intended to support learners' cognitive processing, reduce cognitive load, and foster the construction of coherent mental representations [Ma20]. This is especially important for instructional materials, aimed at the professional development of university teachers, as these materials should also serve as good practice. Although the CTML offers practical guidance for educators developing online learning materials, implementing its principles presents challenges, such as potential conflicts between different CTML principles [Fo24], as well as technical constraints imposed by authoring tools that offer limited design flexibility. This article focuses on the latter. As part of the eSALSA project[6], interactive self-study courses on teaching in higher education and instructional design with digital media are being developed for lecturers, using the H5P authoring tool. H5P is an open-source tool for creating interactive learning content.

[1] Otto-von-Guericke-Universität, Zschokkestr. 32, 39104 Magdeburg, anja.hawlitschek@ovgu.de, https://orcid.org/0000-0001-8727-2364
[2] Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode, apoleshchuk@hs-harz.de
[3] Hochschule Magdeburg-Stendal, Breitscheidstraße 2, 39114 Magdeburg, nadine.wegmeyer@h2.de
[4] Burg Giebichenstein, Neuwerk 1, 06108 Halle, hilberer@burg-halle.de
[5] Hochschule Magdeburg-Stendal, Breitscheidstraße 2, 39114 Magdeburg, christin.deinert@h2.de
[6] "eService-Agentur der Hochschulen im Land Sachsen-Anhalt", https://www.esalsa.de

doi:10.18420/delfi2025_29

It enables the implementation of multimedia and interactive learning materials without the need for programming knowledge and can be integrated into common learning management systems, such as Moodle and ILIAS. This article discusses how principles of multimedia design can be applied within the H5P “Interactive Book” content type. It outlines challenges and presents solution approaches concerning three principles: signaling, support for coherence formation, and spatial contiguity. Its goal is to highlight technical and design limitations as well as demonstrate how research-based design principles must be negotiated and adapted under real-world conditions. Our contribution aims to foster better alignment of digital tools and instructional design principles, and to provide practical support for creating digital learning materials with H5P.

2 Challenges and Solutions

2.1 Signaling Principle

Signaling refers to directing learners' attention to important information or structures in the learning material [Ma20]. This principle reduces cognitive load by helping learners to distinguish relevant from irrelevant information and process content in a structured manner. To implement this principle, we developed a consistent visual coding system for different types of content. The goal was to use clear graphic markers to help learners identify and categorize information. Our concept was to use standardized graphic symbols to identify related text elements — for example, exclamation marks for important information and light bulbs for practical tips.

The first challenge when implementing this principle stems from technical limitations. The Interactive Book has a strictly block-based structure with vertical arrangement only. Text blocks offer minimal formatting options and do not allow graphic elements to be integrated within the text flow. To integrate images or symbols, a separate “image” block must be added, resulting in text and graphics arranged below each other. This structural restriction makes it difficult to implement our signaling concept.

One solution is to create text-symbol combinations using external graphics programs and integrate them as image files. However, this approach has several disadvantages. First, text embedded in graphics is inaccessible to screen readers, which reduces accessibility. Second, integrating numerous graphics increases the size of course files, negatively impacting loading times and memory requirements. Third, text changes require editing the entire graphic, which increases maintenance effort and jeopardizes the material's sustainability. An alternative is to use the “Course Presentation” content type, which allows graphic elements to be positioned flexibly in close proximity to text. With this method, symbols can be placed to the left or right of the text.

Another limitation of H5P with regard to signaling is the inconsistent font sizes. Inconsistent font sizes in text and presentation elements lead to an inconsistent text

hierarchy. For instance, a font size of “75 %” in a text field appears significantly smaller than the same size in a presentation. This lack of standardization complicates the visual hierarchy of information and can lead to confusion, particularly with headings. For instance, a chapter heading may appear smaller than a subordinate presentation title. Additionally, H5P only specifies font sizes as a percentage, making precise and consistent formatting difficult. Common units, such as pt or px, which are standard in web design, are unavailable. This hinders visual coordination between text elements further.

One solution is to manually determine suitable percentages through trial and error until a consistent size ratio between different text fields is achieved. In order to standardize this process, it is advisable to document both the font and the corresponding percentage values used and to reuse them consistently. Although this procedure is practicable in principle, it proves to be time-consuming. Every change requires a new check in display mode. This necessitates constant switching between editing and preview, which interrupts the workflow. In addition to font size hierarchies, horizontal dividing lines can be used to underline the semantic structure of an H5P chapter.

Another challenge arises in navigation. Individual H5P elements, such as certain task elements, automatically create new entries in the navigation bar even if their content is part of an existing chapter. This creates an artificial structure that differs from the instructional one. Additionally, there is a lack of visual differentiation between chapters and sub-chapters — neither the font size nor the indentation can be adjusted. These limitations undermine the effectiveness of signaling by making it difficult to provide clear orientation and guide learners' attention effectively.

A solution is to create each exercise as a subchapter to ensure consistent and comprehensible navigation. Visual strategies, such as systematic numbering (“1.1,” “1.2,” etc.) or using capital letters for main chapters, can differentiate between main chapters and subchapters. However, capital letters only create the minimal illusion of a difference in font size and do not sufficiently support the intended signaling effect.

2.2 Spatial Contiguity Principle

This principle states that related information elements should be positioned close to each other [Ma20]. When an illustration and its associated label or explanation are placed close together, learners experience reduced extraneous cognitive load because no additional effort is required to establish the link between visual and textual information.

The technical limitations previously described for the signaling principle also hinder the implementation of spatial contiguity. The strict block-based structure of the Interactive Book prevents optimal positioning of images and their associated text elements side by side.

In addition to the previously mentioned approaches, the H5P content type “Image Hotspots” provides another way to implement spatial contiguity. Interactive areas on

images allow additional information to be displayed as needed, creating a close spatial link between visual and textual elements.

2.3 Supporting Coherence Formation

Coherence formation is essential for knowledge acquisition from multiple representations, such as text, images, and tables. Each individual representation must be understood, and referential connections between them must be established in order to construct a coherent mental model. Since this process is cognitively demanding, instructional aids that facilitate coherence formation and reduce cognitive load are essential [SB06]. For instance, related concepts can be visually linked through consistent color coding, arrows can connect explanatory text to visual elements, hyperlinks can be used, or layout structures can be employed to clarify relationships.

As previously mentioned, visual signals for structuring content can only be realized to a limited extent in the H5P Interactive Book. Additionally, there is no functionality for creating links between chapters, pages, or within a page. This makes it difficult to provide learners with explicit references to related content and hampers instructional guidance for building coherence.

3 Conclusion and Outlook

Implementing evidence-based design principles from multimedia learning research within the H5P "Interactive Book" requires a high degree of reflection on instructional design as well as technical creativity. Principles such as signaling, spatial contiguity, and coherence formation present concrete challenges for creators of learning materials in H5P. Technical limitations, such as rigid block structures, limited layout functions, and a lack of linking features, hinder the consistent application of these design principles, as our experience shows. However, targeted tool combinations, didactic workarounds, and creative uses of existing H5P content types allow for partial solutions. The limited support for coherence formation is particularly challenging because content references cannot easily be mapped visually or structurally.

The findings have implications for the further development of H5P and similar authoring tools. Authoring tools should support and enhance the application of evidence-based principles of multimedia design to enable content creators to design effective teaching and learning materials that address the needs of learners. Co-design approaches could potentially help here. In the long term, it would be highly desirable for tools like H5P to offer greater flexibility in terms of layout, navigation, and semantic linking without compromising accessibility. Future research should examine how well the applied design principles, implemented under the constraints of H5P and supported by proposed workarounds, translate into effective learning processes and foster learners' satisfaction and engagement.

References

[Fo24] Foster, C.; Woollacott, B.; Francome, T.; Shore, C.; Peters, C.; Morley, H.: Challenges in applying principles from cognitive science to the design of a school mathematics curriculum. The Curriculum Journal, 35, 489–513, 2024. DOI: 10.1002/curj.249

[Ma20] Mayer, R. E.: Multimedia Learning (3rd ed.). Cambridge: Cambridge University Press, 2020.

[No21] Noetel, M.; Griffith, S.; Delaney, O.; Harris, N. R.; Sanders, T.; Parker, P.; del Pozo Cruz, B.; Lonsdale, C.: Multimedia design for learning: An overview of reviews with meta-meta-analysis. Review of Educational Research, 92(3), 413-454, 2021. DOI: 10.3102/00346543211052329

[SB06] Seufert, T.; Brünken, R.: Cognitive load and the format of instructional aids for coherence formation. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 20(3), 321-331, 2006.

On AI Generated Reductions from Natural Language Descriptions

Maurice Herwig [1], Norbert Hundeshagen [1], Cedric Kollenberg [2], and Robin Weber[1]

Abstract: We explore the ability to use large language models in an educational tool to support students in performing reductions, a fundamental competence in the theory of computability and complexity. A total of 284 descriptions of reductions in natural language were used to generate Python code by ChatGPT-4o. Initial findings, based on human evaluation of the student submissions and the generated code are mixed. The LLM tends to generate correct reductions more often when contextual information is provided in the prompt. However, the correctness of the generated code not necessarily corresponds to the input, indicating the need to impose constraints or filters to the LLM's output.

Keywords: Reductions, LLM, Code Generation, Computability Theory, Complexity Theory

1 Introduction

The ability of recent large language models (LLM) to generate code is remarkable, and "[...] offers transformative teaching and learning possibilities [...]" ([CZ24], pg. 7) in the didactics of programming (e. g., an overview can be found in [CZ24]). Reductions in the realm of complexity and computability theory as defined e. g. in [Si13] essentially can be seen as such a coding task, asking for a program that transforms instances of a problem A into instances of a problem B such that an algorithm for B can be used to solve A. Naturally, the question arises how well LLMs perform at generating code on such domain specific problems.

As our main motivation is how to support students in learning to perform reductions, we here are concerned with a more specific version of the latter. *How well do LLM perform at implementing a reduction given as a student's idea?* This is a follow-up to the proposal for an educational tool outlined in [Ba24] (cf. page 142). There the main idea is to use AI generated code for the production of feedback information by inferring positive and/or negative test cases from it. A similar approach is investigated in [SZ24] and [Pr24]. In [SZ24] LLMs have been used to generate code based on student submissions in an introductory programming course. The main conclusion there is that, especially for short exercises, the results vary little, regardless of whether the description is at a high or low level. In [Pr24] students in a programming module were required to create a prompt for an LLM, based on a visual

[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; robin.weber@uni-kassel.de
[2] 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/delfi2025_30

representation of a problem. In an iterative process, the generated code is verified through test cases.

The following is a preliminary report on data collected in a students' research project aimed at the above question in the context of an inherent abstract and didactical more involved topic (cf. [GT16]), where the need for educational tools supporting students in this area is well reflected in the literature (e. g., in [CFG14; He24; ZHD22]).

2 Procedure & Method

The data collection took place in the winter term of 24/25 in a 2nd year BSc computer science course on computability and complexity at a german university. The lecture ranges over standard topics in that area as for instance presented in [Sc08]. There, reductions are first introduced in the area of undecidability and then adapted to polynomial-time reductions to show `NP-hardness` of problems. Participants are required to solve weekly exercises, which then are graded by student research assistants.

A typical reduction task is the following: "*Show by reduction that the following problem is undecidable. Let $\mathcal{P}$ be a program, does $\mathcal{P}$ terminate on every input?*". Such tasks were split into two parts, where firstly, students were required to provide a natural language description of their idea, and secondly, students should perform the actual reduction. Note, that this is a standard problem solving strategy.

A total of eight different tasks on (un)decidability and `NP-hardness` were given on different homeworks throughout the term, resulting in 284 natural language student ideas for reductions. The tasks differ in whether both problems were given or only one and thus, students had to find a suitable problem for the reduction, like in the example above (cf. table 1). ChatGPT-4o[3] was used to implement the given natural language submissions of reductions in Python with different pre-prompts in german, providing more detailed instructions to the LLM on how to transform these descriptions. These pre-prompts are divided into `GPTWithout`, which only indicates that a Python function should be written, `GPTWithDes`, which additionally describes the problems in question, and `GPTDist`, which tells the LLM not to make any improvements and to stick exactly to the input. Then the submissions (`Sub`) and the resulting code (`Code`) were graded on the binary scale (`incorrect`, `correct`) by a student as part of the above mentioned research project. Initially, a more detailed classification was used, ranging from 0 to 3 points, which here is projected onto the binary scale for the reason to increase objectivity. Thus, the grade `correct` on `Code` was given when the code actually is a correct reduction between the given problems. Additionally, a second independent rater (student research assistants) graded `Sub` on a binary scale with respect to the question, if the idea is actually an implementable reduction. This was later used to further validate the ratings. Orthogonal to these categories, we also tested the code

[3] https://chatgpt.com/

for compilability, as this is not a given for the reason of correct data types, spelling issues, etc. Hence, generated code can be correct but does not compile. For this purpose, the code is attempted to be stored as an abstract syntax tree.

3 Results

General Overview For one of the two independent grades for `Sub`, grading is available for only 226 of the reductions, because the remaining students deleted their data in the learning management system. Since the different grades for `Sub` showed major discrepancies, we here consider only the reductions for which both raters assigned the same grading to `Sub`.

We present the distributions of `correct` and `incorrect` for `Sub` as a triple (n, x, y), where n denotes the total number of submissions, x the number graded as `correct`, and y the number graded as `incorrect`. The distribution of the grades for `Code`, as well as their correlation with the grades of `Sub`, is represented as $(n, x^{tp}_{fp}, y^{tn}_{fn})$, where *tp* and *tn* denote the number of reductions on which both `Sub` and `Code` agree and *fp*/*fn* the false-positive/false-negative count, respectively. Thus, for instance false-positive reduction were graded as `incorrect` in `Sub` but classified as `correct` in `Code`.

In the dataset under consideration, we observe a distribution of $(138, 48, 90)$ for `Sub`. The corresponding distributions for `Code` for the different pre-prompts are: `GPTWithout`$(138, 35^{25}_{10}, 103^{80}_{23})$, `GPTWithDes`$(138, 47^{25}_{22}, 91^{68}_{23})$, and `GPTDist`$(138, 49^{29}_{20}, 89^{70}_{19})$. At first glance, it appears that `GPTWithout` performs best. A more nuanced perspective can be achieved by viewing the LLM as an additional rater, that is, we argue that the grades for `Code` actually reflect a score given by the LLM to `Sub`. Computing the inter-rater reliability in form of *Cohen's kappa* [Co60] values supports the above observation (`GPTWithout`: 0.425, `GPTWithDes`: 0.28, and `GPTDist`: 0.383). Part of the explanation for this result is that there are twice as many `incorrect` as `correct` reductions in `Sub`, which favors a negative rater. This is also reflected by the poor code quality produced by `GPTWithout`, which then is graded as `incorrect`. This argument is further supported by additionally considering the percentage $\frac{tp}{48}$, also known as *recall*, where 48 is the total number of `Sub` reductions graded as `correct`. The recall values are 0.521 for both `GPTWithout` and `GPTWithDes`, and 0.604 for `GPTDist`.

Further observe that the `correct`/`incorrect` distribution for `GPTWithDes` is nearly identical to `Sub`. However, this is mainly explained by a higher number of incorrectly translated reductions. This issue is largely addressed by `GPTDist`, which yields results similar to those of `GPTWithout`, with the key difference that the majority of the generated code is not of poor quality.

When examining compilability, it is noticeable that ChatGPT generally generates syntactically correct reductions. `GPTWithout` results in 4 cases, `GPTWithDes` results in 5 cases, and `GPTDist` results in 13 cases where no syntactically correct code was generated. One instance was classified as correct for `GPTWithout`, none for `GPTWithDes`, and three for `GPTDist`. It should be noted that possible runtime errors are not captured.

Does the type of exercise matter? We continue with a more fine-grained analysis. For that, we categorize the different exercises with respect to the problems used and the type of reductions given. Our categories are summarized by the following table, where the cells represent the number of exercises and the distribution of `Sub` for all reductions of this exercise type. Again, we only consider reductions where both independent raters agreed.

when \ show	undecidability (`undec`)	∈ NP (`NP`)	∈ NP-hard (`NPhard`)
both problems given (`two`)	3(56, 19, 37)	1(22, 8, 14)	1(11, 3, 8)
one problem given (`one`)	2(36, 11, 25)	0	1(9, 5, 4)

Tab. 1: Overview of the types of exercises. For instance, there are two exercises where one problem was given and it has been asked to show that this problem is undecidable.

For the majority of exercise types, we observe results consistent with those of the overall dataset. The effect of `GPTWithout` described above is particularly evident in tasks on `NP`, with a distribution of $(22, 1^{1}_{0}, 22^{14}_{7})$, where only one reduction in `Code` is graded as `correct`. For this exercise type, `GPTDist` yields $(22, 5^{2}_{3}, 17^{11}_{6})$, and `GPTWithDes` yields $(22, 11^{4}_{7}, 11^{7}_{4})$. However, for exercises on `NPhard`, where both problems are given, this effect is not observed. Here, `GPTWithout` achieves $(11, 4^{2}_{2}, 7^{6}_{1})$, translating more submissions correctly than `GPTWithDes`, which yields $(11, 1^{1}_{0}, 10^{8}_{2})$. Note that the sample size for this exercise type is only 11. For this task, `GPTDist` yields a distribution of $(11, 5^{2}_{3}, 6^{5}_{1})$.

4 Conclusion and Future Work

The study presented here is meant as a starting point for introducing LLMs in educational tools in the area of reductions, and thus, exploratory in nature. Initially, we expected a more significant difference between the ratings of `GPTWithout`, `GPTWithDes`, and `GPTDist`, as for the latter two more information is given to the LLM. This is only partly supported by the data. However, we observed poorer code quality with `GPTWithout`, which is not sufficiently reflected in the binary grading. Thus, a more nuanced analysis of the submissions and the code is needed, by identifying several categories that reflect certain stages of reduction and code quality. This can be achieved by using inter-rater reliability metrics and different expert ratings for the submissions, as well as carefully constructed (possibly automated) test instances to validate the code. Furthermore, as the performance of LLM improves it is reasonable to assume that the number of false-positive ratings increases. This effect can e. g. be observed already for the `NP` exercise, which is a standard reduction in the literature and thus, probably well reflected in training data of LLM. Therefore, we suspect that in future work emphasis should be placed on limiting LLMs, which we already observed in some preliminary findings with recent reasoning models. The pre-prompt used in `GPTDist` exemplifies this and lead to more useful output, which might be further exploited as feedback information on the submissions.

References

[Ba24] Barnes, T. et al.: Teaching Support Systems for Formal Foundations of Computer Science (Dagstuhl Seminar 24251). Dagstuhl Reports 14 (6), pp. 108–129, 2024, DOI: 10.4230/DagRep.14.6.108.

[CFG14] Creus, C.; Fernández, P.; Godoy, G.: Automatic Evaluation of Reductions between NP-Complete Problems. In (Sinz, C.; Egly, U., eds.): Proc. of the 17th Int. Conf. on Theory and Applications of Satisfiability Testing. Springer, pp. 415–421, 2014, DOI: 10.1007/978-3-319-09284-3_30.

[Co60] Cohen, J.: A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20 (1), pp. 37–46, 1960, DOI: 10.1177/001316446002000104.

[CZ24] Cambaz, D.; Zhang, X.: Use of AI-driven Code Generation Models in Teaching and Learning Programming: a Systematic Literature Review. In: Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1. SIGCSE 2024, Association for Computing Machinery, Portland, OR, USA, pp. 172–178, 2024, DOI: 10.1145/3626252.3630958.

[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.

[He24] Herwig, M. et al.: Towards Computer-Aided Teaching of Reductions in Theoretical Computer Science. In (Schulz, S.; Kiesler, N., eds.): DELFI 2024 - Die 22. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V., DELFI 2024, Fulda, Germany. Vol. P-356. LNI, Gesellschaft für Informatik e.V., 2024, DOI: 10.18420/delfi2024_45.

[Pr24] Prather, J. et al.: Interactions with Prompt Problems: A New Way to Teach Programming with Large Language Models. arXiv preprint, 2024, DOI: 10.48550/arXiv.2401.10759.

[Sc08] Schöning, U.: Theoretische Informatik - kurz gefasst. Spektrum Akademischer Verlag, 2008.

[Si13] Sipser, M.: Introduction to the Theory of Computation. Cengage Learning, 2013.

[SZ24] Smith, D. H.; Zilles, C.: Evaluating Large Language Model Code Generation as an Autograding Mechanism for "Explain in Plain English" Questions. In: Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2. SIGCSE 2024, Association for Computing Machinery, Portland, OR, USA, pp. 1824–1825, 2024, DOI: 10.1145/3626253.3635542.

[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. PLDI 2022, ACM, San Diego, CA, USA, pp. 762–776, 2022, DOI: 10.1145/3519939.3523732.

Nutzung von Fremdkonzepten in der hochschuldidaktischen Praxis – Mehrwerte und Fallstricke

Eine explorative Studie zur Transfer- und Praxistauglichkeit von als Open Educational Resources dokumentierten hochschuldidaktischen Konzepten.

Katja Hornoff[1], Mariane Liebold [2], Josefin Müller[3], Anja Schulz [4] und Anne Vogel [5]

Abstract: Im Verbundprojekt „Digitalisierung der Hochschulbildung in Sachsen“ wurden hochschuldidaktische Weiterbildungen zu Aspekten der Digitalisierung in der Hochschullehre konzipiert und als offene Werkstätten, sog. „Digital Workspaces“, Lehrenden aller staatlichen sächsischen Hochschulen angeboten. Zu Projektende wurden mit Blick auf den Ergebnistransfer ausgewählte Workspace-Konzepte inkl. didaktischer Begleitinformationen dokumentiert und sämtliche Inhalte sowie dazugehörige Materialien als Open Educational Resources (OER) publiziert. Die Dokumentationen sollten praxisnahe Anregungen bieten und insbesondere die unkomplizierte Nachnutzung durch weitere hochschuldidaktisch Tätige ermöglichen, weshalb ein Qualitätskonzept für den Erstellungsprozess erarbeitet und diesem zugrunde gelegt wurde. Nach Upload der erarbeiteten Materialien in einem OER-Repositorium war jedoch nicht mehr nachvollziehbar bzw. überprüfbar, inwieweit diese für den praktischen Einsatz genutzt und ggf. weiterentwickelt werden. Vor diesem Hintergrund wurde eine explorative Studie zur Praxistauglichkeit der OER-Dokumentationen durchgeführt. Der Beitrag fasst die zentralen Ergebnisse der Untersuchung zusammen und zeigt Potenziale für die Verbesserung der entstandenen Materialien auf.

Keywords: Digitale Hochschulbildung, Hochschuldidaktik, Open Educational Resources (OER), Praxistransfer, Praxistauglichkeit, Nachhaltigkeit

[1] Hochschule für Technik, Wirtschaft und Kultur Leipzig, E-Learning-Service, Karl-Liebknecht-Str. 132, 04277 Leipzig, katja.hornoff@htwk-leipzig.de

[2] Hochschuldidaktik Sachsen, Marschnerstr. 31, 04109 Leipzig, mariane.liebold@hd-sachsen.de, https://orcid.org/0000-0002-9834-7613

[3] Hochschuldidaktik Sachsen, Marschnerstr. 31, 04109 Leipzig, josefin.mueller@hd-sachsen.de

[4] Hochschuldidaktik Sachsen, Marschnerstr. 31, 04109 Leipzig, anja.schulz@hd-sachsen.de, https://orcid.org/0009-0002-5150-4287

[5] Technische Universität Chemnnitz, Straße der Nationen 62, 09111 Chemnitz, anne.vogel@verwaltung.tu-chemnitz.de, https://orcid.org/0009-0007-2309-4918

doi:10.18420/delfi2025_31

1 Ausgangslage und Fragestellung

Im Rahmen des Vorhabens „Digitalisierung der Hochschulbildung in Sachsen“[6] wurden innerhalb von sechs Jahren über 120 hochschuldidaktische Weiterbildungen mit einem Gesamtumfang von ca. 1.500 Arbeitseinheiten (1 Arbeitseinheit = 45 min) realisiert. Die als offene Werkstätten konzipierten „Digital Workspaces“ adressierten verschiedenste Aspekte digital gestützten Lehrens und Lernens und wurden hochschulübergreifend im Teamteaching durchgeführt [Al22]. Mit Blick auf den Transfer von Projekterfahrungen und -ergebnissen wurden ab Sommer 2023 insgesamt 13 Digital Workspaces zur Dokumentation ausgewählt. Hierbei sollten möglichst aktuelle und häufig nachgefragte Angebote aufbereitet und dabei zugleich eine thematische Vielfalt sichergestellt werden [Kn25]. Die Workspace-Konzepte wurden inkl. didaktischer Begleitinformationen dokumentiert und sämtliche Inhalte sowie dazugehörige Materialien als Open Educational Resources (OER) publiziert. Dem Prozess der Erstellung der Dokumentationen wurde ein elaboriertes Qualitätskonzept zugrunde gelegt, um eine unkomplizierte und barrierefreie Nachnutzung zu gewährleisten und weiteren hochschuldidaktisch Tätigen praxisnahe sowie -relevante Anregungen in leicht verständlicher Form zu bieten [MZD21]. Für die konsequente Umsetzung der Qualitätskriterien in allen Dokumentationen wurden zudem verschiedene Qualitätssicherungsmaßnahmen etabliert, die z. B. strukturierte Reviews und Feedbackschleifen vorsahen [Sc24]. Bei der Wahl eines passenden OER-Repositoriums wurde v. a. auf Interoperabilität und langfristige Verfügbarkeit, Anbindung an den OER Search Index (OERSI), inhaltliche Passung sowie Umfang der bereits verfügbaren Ressourcen geachtet. Im Ergebnis fiel die Entscheidung auf twillo, ein Portal für OER in der Hochschullehre.[7] Nach Upload der erarbeiteten Materialien war allerdings nicht mehr nachvollziehbar, inwieweit diese genutzt und ggf. weiterentwickelt werden. Daraus ergab sich die Fragestellung der vorliegenden Studie: Wie beurteilen Nutzende die Praxistauglichkeit der veröffentlichten OER-Materialien?

2 Studiendesign

Das methodische Vorgehen basierte auf einem qualitativen Forschungsdesign, das zu drei Erhebungszeitpunkten leitfadengestützte Interviews vorsah. Zusätzlich beantworteten die Proband*innen zu jeder OER-Dokumentation jeweils einen begleitenden Fragebogen. Sowohl die konzipierten Interviewleitfäden als auch der selbst erstellte Begleitfragebogen waren in thematische Bereiche gegliedert, die die Erwartungen an die Dokumentation, den

[6] Am Projekt waren die staatlichen sächsischen Universitäten und Hochschulen für angewandte Wissenschaften beteiligt. Es wurde von der Hochschuldidaktik Sachsen und dem Arbeitskreis E-Learning der Landesrektorenkonferenz Sachsen koordiniert und durch das Sächsische Staatsministerium für Wissenschaft, Kultur und Tourismus gefördert.

[7] Alle bislang veröffentlichten Dokumentationen finden sich hier: https://www.twillo.de/edu-sharing/components/collections?id=d65b179c-2ef8-4268-89a2-24a2680e0909 (Stand: 02.07.2025).

Umgang mit ihnen, Inhalt und Strukturierung, Handhabbarkeit und optische Aufbereitung sowie eine Gesamteinschätzung der beiden Bearbeitenden erfragten.

Die Stichprobe umfasste zwei gezielt ausgewählte, projektexterne hochschuldidaktisch Tätige aus dem Netzwerk der Hochschuldidaktik Sachsen. Beide Personen konnten die Dokumentation, mit der sie hauptsächlich arbeiten wollten, selbst wählen, wobei pro Person die Durchführung eines Digital Workspace geplant wurde. Unabhängig von der individuellen Auswahl waren Austausch und gegenseitige Unterstützung sowohl während der Bearbeitungsphase als auch bei der geplanten Durchführung vorgesehen.

Die Datenerhebung erfolgte in drei aufeinanderfolgenden Schritten, die verschiedene Stadien der Einarbeitung in die Dokumentationen bzw. der Arbeit mit diesen abbildeten: Das erste Interview fand vor der Arbeit mit den beiden gewählten Dokumentationen statt, das zweite knapp vor der geplanten Durchführung und das dritte nach den ursprünglichen Durchführungsterminen (wegen zu geringer Teilnahmezahlen wurden beide Workspaces abgesagt). Die bereits erwähnten Fragebögen wurden den Proband*innen nach dem ersten Interview ausgehändigt und sollten bis zum zweiten Interview ausgefüllt werden.

Die Auswertung erfolgte in einem zweistufigen Verfahren: Zunächst wurde deduktiv entlang der Struktur des Leitfadens und des Begleitfragebogens analysiert, anschließend wurde diese Analyse durch induktive Kategorienbildung ergänzt und erweitert.

3 Ergebnisse

Die Untersuchung verdeutlicht, dass die Befragten zwar nur über vage Vorstellungen von OER verfügten, jedoch bereits Erfahrungen mit der Nutzung von Fremdmaterialien in anderen Kontexten gesammelt hatten (jedoch nicht als OER bezeichnet). Grundsätzlich gingen sie mit einer offenen und positiven Haltung an die Dokumentationen heran. Die geäußerten Erwartungen an die Materialien waren sehr umfassend: Eine unkomplizierte Nutzbarkeit ohne umfangreiche Anpassungen, ein hoher inhaltlicher Detaillierungsgrad ohne überflüssiges didaktisch-methodisches Grundlagenwissen sowie eine übersichtliche und schnell erfassbare Gestaltung wurden vielfach genannt.

Die Fragebogenergebnisse zeigen eine differenzierte Bewertung der erwarteten Zeitersparnis: Während in einem Fall durch weitgehende Übernahme des Konzeptes eine erhebliche Entlastung erzielt wurde, führte ein anderes, stark modifiziertes Format zu keiner merklichen Zeitersparnis. Dennoch wurde die Dokumentation auch in diesem Fall als hilfreich erlebt, da sie als Ausgangspunkt zur individuellen Weiterentwicklung diente.

In Bezug auf Gestaltung und Handhabbarkeit wurde die Länge der Handreichungen kritisch gesehen – sie wirkte initial überfordernd. Zugleich wurde der Wunsch nach umfassender Information geäußert. Als besonders nützlich erwiesen sich die PowerPoint-Präsentationen, die vorbereitenden Hinweise und der als „Ziele/Inhalte/Methoden-Tabelle“ (ZIM) aufbereitete detaillierte Ablaufplan, der maßgeblich zur praktischen

Erschließung beitrug. Positiv hervorgehoben wurden zudem das einheitliche Design, die klare Struktur (wobei es aus Sicht der Befragten hilfreicher wäre, den ZIM-Plan direkt am Anfang der Dokumentationen zu positionieren) und die gute sprachliche Verständlichkeit.

Zur praktischen Relevanz äußerten die Befragten, dass sie vielfach inhaltliche Adaptionen vornahmen – etwa durch Ergänzung weiterer Materialien, Methoden oder Tools sowie durch Modifikationen des zeitlichen Ablaufs. Dennoch wurden die Materialien als gute Grundlage für eigene Umsetzungen und kreative Anpassungen bewertet. Kritik bezog sich v. a. auf zu allgemein gehaltene Abschnitte in den Handreichungen sowie eine unzureichende inhaltliche Verzahnung von Handreichungen und Präsentationen.

4 Diskussion

Die Aussagekraft der Studienergebnisse ist aufgrund der geringen Stichprobe begrenzt. Dennoch zeigen die Ergebnisse eine hohe grundsätzliche Bereitschaft der Befragten zur Nutzung der OER-Materialien. Zugleich wird deutlich, dass die Erwartungshaltungen und Zielsetzungen der Nutzenden teils stark von den Intentionen der Erstellenden abweichen. Während die Dokumentationen v. a. als flexibel adaptierbare Anregungen konzipiert wurden, erhofften sich die Befragten einsatzbereite Konzepte und Materialien, die ohne aufwendige Einarbeitung und ohne nennenswerte Anpassungen unmittelbar verwendbar sind. Dieser Befund steht im Widerspruch zu den Grundprinzipien von OER, die gerade auf Offenheit, Weiterentwicklung und individuelle Anpassung abzielen [Hi10; FFM20; Ot21]. Hier besteht Ausbaupotenzial der begleitenden Metainformationen, die Klarheit über Ziel und Zweck der Dokumentationen schaffen und Nutzende aktiv in die Rolle von Co-Gestaltenden versetzen.

Auch inhaltlich lassen sich Potenziale zur Weiterentwicklung identifizieren: So wurden eine stärkere inhaltliche Verzahnung von Handreichungen und Präsentationen, eine Grobübersicht zu Beginn sowie eine verbesserte Strukturierung mehrfach angeregt. Dass viele Informationen als zu allgemein empfunden wurden, verweist auf die Herausforderung, bei heterogenen Zielgruppen das passende Maß an Tiefe zu finden. Gleichzeitig verdeutlichen die Rückmeldungen den Bedarf nach klarerem inhaltlichem Fokus sowie aktualisierbaren Modulen – etwa bei Themen, die stark von technischen Entwicklungen abhängig sind.

Insgesamt bieten die vorliegenden OER-Dokumentationen eine gute Ausgangsbasis für hochschuldidaktisch Tätige, eigene Konzepte und Formate zu entwickeln. Gleichwohl ist es aus Sicht der Autor*innen sinnvoll, über eine zielgruppengerechtere Kommunikation der OER-Prinzipien und eine verstärkte Usability-orientierte Gestaltung (z. B. klarere Verweise zwischen den verschiedenen Materialien einer Dokumentation) nachzudenken. Letztlich bleibt der Anspruch an OER, sowohl strukturiert als auch flexibel nutzbar zu sein – eine Balance, die nur durch iterative Weiterentwicklung in enger Rückkopplung mit den Nutzenden erreicht werden kann.

Literaturverzeichnis

[Al22] Albrecht, C.; Brenner, F.; Schulz, A.; Vogel, A.: Digitale Werkstattformate für Lehrende an sächsischen Hochschulen. Unterstützungsangebote zur Entwicklung der eigenen digital gestützten Lehre. Ludwigsburger Beiträge zur Medienpädagogik 22, S. 1-13, 2022.

[FFM20] Fabri, B.; Fahrenkrog, G.; Muuß-Merholz, J.: Der Gold-Standard für OER-Materialien. Ein Kompendium für professionelle Erstellung von Open Educational Resources (OER). Verlag ZLL21 e.V., Hamburg, 2020.

[Hi10] Hilton III, J.; Wiley, D.; Stein, J.; Johnson, A.: The four 'R's of openness and ALMS Analysis: Frameworks for Open Educational Resources. Open Learning: The Journal of Open, Distance and e-Learning, 25(1), 37-44, 2010.

[Kn25] Kneita, M.; Marquardt, J.; Riedel, J.; Schulz, A.; Vogel, A.: Good Practice für die hochschuldidaktische Weiterbildung dokumentieren und teilen. Ein Erfahrungsbericht zur Erstellung von OER-Handreichungen für hochschuldidaktisch Tätige. In (Mrohs, L.; Franz, J.; Herrmann, D.; Lindner, K.; Staake, T., Hrsg.): Digitales Lehren und Lernen in der Hochschule. Strategien – Bedingungen – Umsetzung. transcript Verlag, S. 69-75, 2025.

[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, 24/21, S. 1061-1085, 2021.

[MZD21] Müskens, W.; Zawacki-Richter, O.; Dolch, C.: Instrument zur Qualitätssicherung von OER – IQOer – Entwicklungsversion 17, 2021.

[Sc24] Schulz, A.; Vogel, A.; Riedel, J.; Kneita, M.: Qualitätskriterien für Open Educational Resources und wie sie erfüllt werden können. In (Längrich, M.; Heidig, S.; Schuster; E., Hrsg.): Tagungsband 22. Workshop on e-Learning, Hochschule Zittau/Görlitz, S. 17-25, 2024.

Stimulating Reflection on Code Quality: An Adaptive and Automated Question Generation Approach for Student Code

Sigrid L. Klinger [1], Patrick Weber [2], Sven Strickroth [1], and Michael Striewe [2]

Abstract: While functional correctness dominates programming education, other code quality aspects like understandability are often neglected. We propose a scalable, automated approach that generates targeted questions based on students' own functionally correct code. Unlike existing systems, our method prompts personalized questions to encourage reflection on quality issues and can provide feedback on students' responses. This poster outlines the concept and demonstrates the feasibility of integrating question-based quality feedback into existing assessment systems.

Keywords: Code Quality, Reflection, Automated Item Generation, Programming Education, E-Assessment

1 Introduction

While functional correctness is a central goal in programming education, code quality aspects like readability and maintainability are often overlooked [Iz25]. Many students stop improving their code once tests pass [ASF16], leading to redundancy and poor performance [Ki20; SBS16]. Automated assessment systems reinforce this behavior by focusing on correctness [KHJ17] or providing abstract, non-actionable feedback, which can lead to "gaming-the-system" strategies [Ba08; Pe15].

Existing tools like CheckStyle and PMD detect code smells and enforce coding standards, but their feedback is often too abstract for novices [KHJ17; PLB18; Wa07]. While some intelligent tutoring systems are designed to train refactoring [KHJ17], they are rarely integrated into learning platforms and address only a limited set of issues. Peer and educator reviews can be effective for discussing code quality [Di18; Hu09; KP09], but they do not scale well, and peer feedback often lacks participation or quality [HB19; SA25]. Some studies investigated follow-up questions to assess students' understanding of their own code [LSS21; St23], but to date, no approaches automatically generate questions specifically targeting code quality.

We propose a template-based approach [GH12; GL12] within a scalable system that uses static and dynamic code analysis to identify quality issues in functionally correct submissions

[1] LMU Munich, Oettingenstr. 67, 80538 Munich, Germany,
sigrid.klinger@ifi.lmu.de, https://orcid.org/0009-0006-9277-8355;
sven.strickroth@ifi.lmu.de, https://orcid.org/0000-0002-9647-300X
[2] Trier University of Applied Sciences, Department of Computer Science, Schneidershof, 54293 Trier, Germany,
webep@hochschule-trier.de, https://orcid.org/0009-0000-9096-904X;
m.striewe@inf.hochschule-trier.de, https://orcid.org/0000-0001-8866-6971

doi:10.18420/delfi2025_32

and generate targeted questions—ranging from multiple choice to open-ended formats—that stimulate reflection [Br22; Ch19; HVW23; LH98]. Our approach builds on Izu et al.'s [Iz25] classification of 63 quality issues in novice code and aims to integrate quality into the assignment process itself. Note that our goal is not to design didactically "optimal" questions, but rather to explore and build the technical foundations to automatically generate a set of reflection-stimulating prompts at scale.

This poster outlines the conceptual and technical basis of our approach, reports initial implementation steps, and discusses integration into existing assessment systems.

2 Proposed Approach

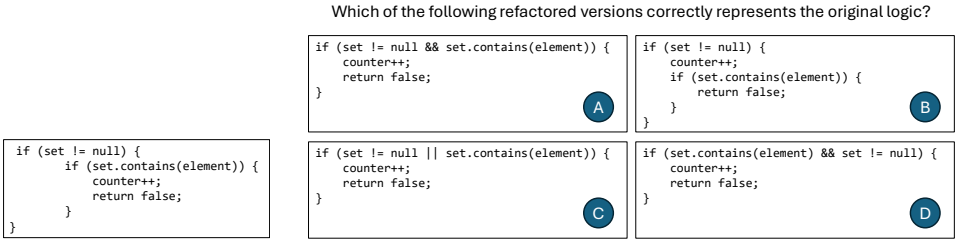
The proposed procedure is as follows: After a student submits code to an automated grading system, a quality analysis is triggered once the submission is considered functionally correct (or nearly correct). If at least one quality defect is detected, the system selects a suitable question template. Relevant elements, such as method or variable names or entire statements/ expressions, are extracted to fill placeholders in the template. This may also include small modifications of students' code. Additional information may be computed as needed, such as the result of a code transformation (e. g., before and after simplifying a boolean expression). A suitable question is then generated based on common misconceptions and presented to the student as part of the assignment. The student's response is then automatically evaluated, with feedback designed to target frequent misunderstandings. The approach focuses primarily on closed formats like multiple choice or fill-in-the-gap, though open-ended formats are also possible, e. g., using LLMs. It is not limited to specific programming languages, but initial work focuses on Java and Python due to their relevance in teaching [St15]. Code analysis currently uses static methods, including PMD and syntax graph queries via GReQL, as both are established analysis techniques in the context of programming education [SG14]. The procedure is developed iteratively, exploring both deterministic and AI-based techniques. The next section illustrates question generation with examples from authentic student code.

3 Examples for Questions Generated by the Proposed Approach

We selected Izu et al.'s [Iz25] catalog of 63 quality defects as a foundation for our analysis, as it offers a structured view on readability and understandability issues in novice code. Based on 33,500 student submissions from two German universities at LMU Munich using GATE [SOP11] and at University Duisburg-Essen using JACK [St16], we examined the frequency of these defects to identify prevalent real-world issues. For initial implementation, we focused on technically feasible cases with syntactic patterns detectable through static analysis.

One example is defect *CB8: Nested `if` statements that can be collapsed using `AND`*, which occurs when two nested `if` statements share the same (potentially empty) `else` branch and

can be merged. This defect can be detected by checking whether the inner `if` is the only statement in the outer `if`'s `then` block, and ensures that neither has an `else` branch. The analysis also accounts for cases where the inner `if` appears directly or within a single-block statement. This defect occurs in 4 % of submissions and accounts for approximately 11 % of all detected issues. Based on these detections, tailored questions can be generated to prompt students to recognize and revise redundant control structures.



(a) Student Code (b) Possible generated question

Fig. 1: Example case and question for code defect *CB8* with adapted student code

Figure 1b shows a possible generated multiple-choice question prompting students to choose a refactored version of their code (Figure 1a). Option A is correct and merges the conditions with `&&`, improving readability and reducing nesting. The distractors address common misconceptions: (B) reorders execution by moving a statement before the second `if`, potentially changing program flow; (C) uses `||` instead of `&&`, weakening the condition logic; (D) reverses the order of conditions, which may cause unintended effects if evaluation order matters. Such questions aim to help students recognize quality issues and reflect on the implications of different refactorings.

In addition to the defects provided by Izu et al.'s [Iz25] catalog, we scanned all our available submissions using PMD to find relevant and authentic code quality issues. The rule *ReturnEmptyCollectionRatherThanNull* of the *ErrorProne* category for instance, reports methods that return `null` instead of an empty collection, which occurred in 8 % of cases and accounts for approximately 5 % of all detected issues in this analysis.

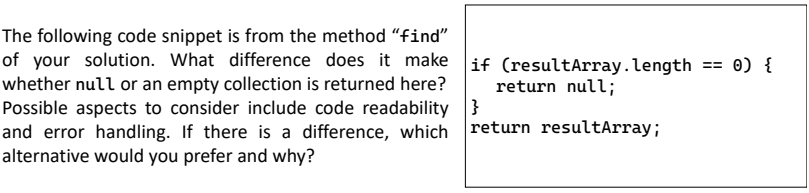


Fig. 2: Possible open-ended question for the PMD rule *ReturnEmptyCollectionRatherThanNull*

Figure 2 presents an example for an open-ended question prompting students to reconsider this design decision. Returning an empty collection improves readability and avoids null checks or runtime errors. While less suitable for automated evaluation, such prompts can be adapted into closed formats. Our implementation supports both approaches, with templates

defined as modular units that specify detection logic, required context data, question formats, and optional distractor logic. Although only two examples are presented here, the underlying architecture is designed to scale across a broader range of defects in Izu et al.'s [Iz25] taxonomy. Future work includes extending coverage and validating the effectiveness of question-based feedback at scale.

4 Discussion, Conclusions and Future Work

This poster presented an approach to automatically generate questions about code quality based on students' own submissions. The current research focuses on exploring the technical feasibility and limitations of such question generation, including both the detection of quality defects and the automated construction of question templates. Various techniques will be examined in iterative studies, using both deterministic and AI-based analysis methods. Each iteration involves lab or classroom evaluations using real student submissions and responses. To ensure flexibility, all components are designed for modular integration into existing feedback systems, avoiding standalone solutions and supporting adoption in diverse educational contexts. Key building blocks include identifying code quality issues, transforming them into template-based questions and feedback, and understanding the technical requirements for such templates. Hence, a contribution is the formulation of specific rules to detect the mentioned defect and the question/feedback generator. The creation and refinement of a larger set of templates and prioritization strategies for overlapping cases are follow-up tasks involving didactic expertise.

A key intermediate result is insight into the frequency of quality defects. The specific variant of *CB8* used in our example appeared in 4 % of submissions and accounts for approximately 11 % of all detected issues in that dataset. Frequencies vary considerably across tasks, depending on how assignments constrain or enable certain constructs. Future work will analyze these effects in more depth to improve efficiency, e. g., by assigning targeted sets of templates to particular assignments.

From a practical perspective, assessing code quality remains inherently complex. There is often no single best solution to a given issue, and even cataloged defects [Iz25] or PMD findings may be context-dependent or debatable. Generated questions should therefore be seen as prompts for reflection, not absolute judgments. A suggestion may be helpful in one case but inappropriate in another. Despite this subjectivity, it remains valuable to consider the pros and cons of a given design [Di18].

Acknowledgements

This research is part of the QualiQuest project funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 528518019.

References

[ASF16] Araujo, E.; Serey, D.; Figueiredo, J.: Qualitative aspects of students' programs: Can we make them measurable? In: 2016 IEEE Frontiers in Education Conference (FIE). Pp. 1–8, 2016, DOI: 10.1109/FIE.2016.7757725.

[Ba08] Baker, R. et al.: Why students engage in "gaming the system" behavior in interactive learning environments. JILR 19 (2), pp. 185–224, 2008.

[Br22] Brownhill, S.: Asking key questions of self-reflection. Reflective Practice 23 (1), pp. 57–67, 2022, DOI: 10.1080/14623943.2021.1976628.

[Ch19] Chang, B.: Reflection in Learning. Online Learning 23 (1), 2019, DOI: 10.24059/olj.v23i1.1447.

[Di18] Dietz, L. W. et al.: Teaching Clean Code. In: Proceedings of the 1st Workshop on Innovative Software Engineering Education. 2018, http://ceur-ws.org/Vol-2066/isee2018paper06.pdf.

[GH12] Gierl, M. J.; Haladyna, T. M., eds.: Automatic Item Generation. Routledge, 2012.

[GL12] Gierl, M. J.; Lai, H.: The Role of Item Models in Automatic Item Generation. International Journal of Testing 12 (3), pp. 273–298, 2012, DOI: 10.1080/15305058.2011.635830.

[HB19] Heller, N.; Bry, F.: Organizing Peer Correction in Tertiary STEM Education: An Approach and its Evaluation. International Journal of Engineering Pedagogy (iJEP) 9 (4), pp. 16–32, 2019.

[Hu09] Hundhausen, C. et al.: Integrating pedagogical code reviews into a CS 1 course: an empirical study. SIGCSE Bull. 41 (1), pp. 291–295, 2009, DOI: 10.1145/1539024.1508972.

[HVW23] Hartmann, A.; Vinke-de Kruijf, J.; van Weesep, R.: Asking the right questions: The role of reflection for learning in and between projects. International Journal of Project Management 41 (5), p. 102494, 2023, DOI: 10.1016/j.ijproman.2023.102494.

[Iz25] Izu, C. et al.: Introducing Code Quality at CS1 Level: Examples and Activities. In: 2024 Working Group Reports on Innovation and Technology in Computer Science Education. ITiCSE 2024, ACM, Milan, Italy, pp. 339–377, 2025, DOI: 10.1145/3689187.3709615.

[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. ITiCSE '17, ACM, Bologna, Italy, pp. 110–115, 2017, DOI: 10.1145/3059009.3059061.

[Ki20] Kirk, D. et al.: On Assuring Learning About Code Quality. In: Proc. ACE. ACM, 2020, DOI: 10.1145/3373165.3373175.

[KP09] Kemerer, C. F.; Paulk, M. C.: The Impact of Design and Code Reviews on Software Quality: An Empirical Study Based on PSP Data. IEEE Transactions on Software Engineering 35 (4), pp. 534–550, 2009, DOI: 10.1109/TSE.2009.27.

[LH98] Lee, A. Y.; Hutchison, L.: Improving learning from examples through reflection. Journal of Experimental Psychology: Applied 4 (3), pp. 187–210, 1998, DOI: 10.1037/1076-898x.4.3.187.

[LSS21] Lehtinen, T.; Santos, A. L.; Sorva, J.: Let's Ask Students About Their Programs, Automatically. In: 2021 IEEE/ACM 29th International Conference on Program Comprehension (ICPC). Pp. 467–475, 2021, DOI: 10.1109/ICPC52881.2021.00054.

[Pe15] Pettit, R. et al.: An Empirical Study of Iterative Improvement in Programming Assignments. In: Proc. SIGCSE. ACM, pp. 410–415, 2015, DOI: 10.1145/2676723.2677279.

[PLB18] Pantiuchina, J.; Lanza, M.; Bavota, G.: Improving Code: The (Mis) Perception of Quality Metrics. In: 2018 IEEE International Conference on Software Maintenance and Evolution (ICSME). Pp. 80–91, 2018, DOI: 10.1109/ICSME.2018.00017.

[SA25] Strickroth, S.; Azaiz, I.: Qualitative Analysis of Peer Reviews of a Large Introductory Programming Course. Computer Science Education, 2025, DOI: 10.1080/08993408.2025.2450587.

[SBS16] Stegeman, M.; Barendsen, E.; Smetsers, S.: Designing a rubric for feedback on code quality in programming courses. In: Proc. Koli Calling. ACM, 2016, DOI: 10.1145/2999541.2999555.

[SG14] Striewe, M.; Goedicke, M.: A Review of Static Analysis Approaches for Programming Exercises. In: Proceedings of the International Conference on Computer Assisted Assessment (CAA 2014). 439. CCIS 439, Zeist, Netherlands, pp. 100–113, 2014.

[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., eds.): Tagungsband der 9. e-Learning Fachtagung Informatik (DeLFI). P-188. GI Lecture Notes in Informatics P-188, Gesellschaft für Informatik e.V., Bonn, Germany, pp. 115–126, 2011, https://dl.gi.de/handle/20.500.12116/4740.

[St15] Strickroth, S. et al.: ProFormA: An XML-based exchange format for programming tasks. eleed 11 (1), 2015, https://eleed.campussource.de/archive/11/4138.

[St16] Striewe, M.: An architecture for modular grading and feedback generation for complex exercises. Science of Computer Programming 129, pp. 35–47, 2016, DOI: 10.1016/j.scico.2016.02.009.

[St23] Striewe, M.: Patterns for the Advanced Design of Programming Exercises Supported by Technology-Enhanced Assessment Systems. In: Proceedings of the 27th European Conference on Pattern Languages of Programs. EuroPLop '22, 2023, DOI: 10.1145/3551902.3551974.

[Wa07] Washizaki, H. et al.: A Framework for Measuring and Evaluating Program Source Code Quality. In (Münch, J.; Abrahamsson, P., eds.): Product-Focused Software Process Improvement. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 284–299, 2007.

Supporting Program Comprehension with Digital Learning Journals: Experiences from a Course with Non-CS Students

Veit Köppen [1] and Anja Hawlitschek [2]

Abstract: Data literacy is a key skill for students in Information Sciences, including a basic introduction to programming – a subject often met with hesitation or even fear. Digital learning journals can support students by fostering reflection and self-regulation. Lecturers gain insights into learners' challenges and can adapt their teaching accordingly. We describe our tool choice in Moodle based on lecturers' needs and how learning journals were implemented. While the approach shows potential, only a third of the students contributed regularly. We explore reasons for this low engagement and give practical advices for lecturers planning to integrate learning journals into their courses.

Keywords: Digital literacy, Learning journal, Learning management system, Program comprehension

1 Motivation

Data literacy is central for students of Information Sciences. This also includes a basic introduction to programming with the aim of learning program comprehension, not programming. Nevertheless, non-CS students often experience discomfort or anxiety when engaging with informatics-related topics [RS10]. Gallin & Ruf [GR93] have proposed an instructional design aimed at ensuring successful engagement with the "unpopular" learning content. Rambow & Nückles [RN02] have positively evaluated this instructional design, the learning journal, for its suitability for use in university teaching. Learners' self-regulation of learning processes is an important key competence [Sa20], which is supported by the use of learning journals [NHR09].

Against this background, digital learning journals were introduced for voluntary use in the course "Basics of Programming". The main goal was to support students in developing basic program comprehension, to strengthen their self-regulated learning skills, and to give lecturers insights into learners' needs – particularly important in a subject often perceived as intimidating. This contribution outlines practical and technical conditions for using learning journals in Moodle, reports on challenges encountered during implementation, and offers guidance for lecturers interested in applying this method in similar contexts.

[1] Fachhochschule Potsdam, Fachbereich Informationswissenschaften, Kiepenheuerallee 3, 14469 Potsdam, Deutschland, veit.koeppen@fh-potsdam.de, https://orcid.org/0000-0002-6068-3275
[2] Otto-von-Guericke-Universität Magdeburg, Fakultät für Humanwissenschaft, Universitätsplatz 2, 39106 Magdeburg, Deutschland, anja.hawlitschek@ovgu.de, https://orcid.org/0000-0001-8727-2364

doi:10.18420/delfi2025_33

2 Advantages of Learning Journals

Learning journals support students in making their learning progress and knowledge gaps explicit. They foster reflection on the subject matter, learning processes, and individual progress [HS93]. When used methodically, learning journals promote meta-cognitive knowledge and the use of learning strategies [RN02], help to develop action and problem-solving skills [KS19], and can contribute to changes in learning behavior. On a meta-cognitive level, learners relate their knowledge to learning objectives and can adapt their strategies accordingly [HNR10]. Learning journals are effective tools for academic learning [MC95]. However, the method must be taught and accompanied with clear guidance, as it is initially unfamiliar for learners to write down and reflect on their learning outcome [HNR10]. Especially for novice learners, didactic support and feedback are essential [Gu22; RY09]. Taking personal responsibility for one's own learning is a positive outcome of the use of learning journals, but it depends on learners' motivation and sustained engagement.

Lecturers can use their students' learning journals (if this is part of the didactic concept) to gain a quick overview which topics have been well understood and where further clarification is needed [RN02]. Using digital tools, the exchange is seamless and access can be realized ubiquitously. To make it easier for learners to use them, lecturers must provide good motivation and continuous support [RY09], especially if learning journals are not used as examinations and replace homework, for example.

3 Tool Requirements From a Lecturer's Point of View

While we consider only lecturers' needs in this article, tool requirements should ideally reflect the perspectives of both, lecturers and students. At University of Applied Sciences Potsdam, Moodle is the standard learning environment. We focus on built-in activities in Moodle as well as Plug-ins, since University-supported tools offer institutional support and long-term reliability. The following requirements for implementation of learning journals should be considered:

- R1: Anonymity to enable open feedback. It is conditionally met if learners remain anonymous among peers.
- R2: Tracking learning progress over time, with access to past entries.
- R3: Structural guidelines for readability and differentiation, enhancing accessibility.
- R4: Summary view of contributions for lecturers to assess student progress and issues.
- R5: Low adaptation need to align technical features with intended learning processes.

We analyzed six Moodle activities against requirements R1–R5. Tab. 1 outlines requirement assessments for each activity. The Journal plugin was chosen due to its minimal setup and maintenance effort, prioritizing other requirements accordingly. It enables personalized

	Type	R1	R2	R3	R4	R5
wiki	default	o	o	-	-	o
Journal	plugin	o	X	o	X	X
Database	default	-	X	X	X	o
Forum	default	o	X	X	X	-
Glossary	default	-	X	o	X	X
Student folder	plugin	-	-	X	o	X

Tab. 1: Moodle's activities and fulfillment of requirements (X: fulfilled, o: conditionally, -: not)

learner pages with text and multimedia, allowing lecturers to provide feedback via comments. Learners decide where to input new information, with visibility limited to their own entries.

4 Implementation in Teaching

In the summer term 2024, learning journals were voluntarily used in the course "Basics of Programming". The approach and objectives were outlined during the first session, where usage agreements were established. Students voluntarily completed weekly question blocks after reviewing course material. Summarized entries were anonymously shared in plenary sessions under "learned" and "open questions" revisiting previous content if necessary. The guiding questions included: "*What have I learned/discovered/mastered?*", "*What do I still need to deepen/not understand?*", and "*Do you need support?*" Students filled out the template between each session across 10 of the 12 learning events. The lecturer reviewed entries before each course, preparing anonymous feedback for plenary discussions on progress and challenges.

5 Results

A total of 21 students actively participated in the course. The learning journals were used at least once by 17 students. A third of the active students used the journal regularly. The course was evaluated at the end of the semester using the institution's standard evaluation form. As this form did not include any specific questions about the use of learning journals, they were only mentioned briefly—the opportunity to discuss uncertainties and ambiguities was perceived positively. No further feedback was provided on the learning journal as a tool to support learning processes.

As the use of the learning journals was only moderate, we additionally conducted an anonymous survey on a Cryptpad after the term with the following questions: "*How do you like the implementation of the learning journals in Moodle?*", "*What prevents you, technically or professionally, from taking advantage of the offer?*", "*What opportunities for improvement do you see?*", and "*Which technical aspects do you find important for use?*"

Four learners provided feedback. The learning journal was generally perceived positively. A reminder function via e-mail was requested for future use. Instead of text-based input, graphic options were desired, too. Lerners wished for more intuitive use on smartphones and tablets. It was furthermore suggested that the constant repetition of the same guiding questions should be substituted by individual reflection questions.

6 Discussion, Conclusion, and Recommendations

Our evaluation is only a starting point of investigation. Not only should the small sample size be considered, but also that students used the Journal activity for learning program comprehension for the first time. Digital learning journals can support learning in challenging subject areas and foster self-regulation. However, our experiences reveal that voluntary participation can remain limited without targeted encouragement and a user-friendly design. The implementation of learning journals in Moodle should be tailored to the needs of both learners and lecturers. From our experiences and students' recommendations, we consider that lecturers should:

- Define and communicate prioritization of requirements and well-defined guidelines for their use. Structural specifications, such as templates, should be implemented, too. The importance of deadlines must be highlighted to students. Therefore, the preparation effort required by lecturers should be made explicit.
- Motivate students consistently, possibly by integrating reminders. Another approach is to integrate learning journal writing time directly in the course, similar to approaches like the two-minute feedback [ES24].
- Consider the technical diversity of devices and preferences of learners, ensuring flexible and mobile input options.
- Use learning journals not only as individual reflection tools but also as feedback sources to adapt teaching in real time. Feedback methods for anonymized journal entries from students, whether personal feedback or plenary discussion, must be clear.

One limitation with respect to the implementation of the described requirements is anonymity. This is guaranteed by some Moodle activities among students, but lecturers can still identify individual contributions. If anonymity is of central importance, alternative solutions beyond the presented Moodle activities must be considered. Another aspect for a deeper investigation is automated comprehension of journal entries if courses have higher number of participants. Finally, strategies to increase student engagement should be investigated in the future.

In this article, we look at learning journals as a supportive learning tool. A look at the Moodle learning environment makes it clear that a variety of implementations are possible. A selection can be made according to the requirements. It must always be kept in mind that the process of working with a learning journal must be accompanied by the lecturer and requires effort in preparation and follow-up.

References

[ES24] Egetenmeier, A.; Strickroth, S.: Scalable Two-Minute Feedback: Digital, Lecture-Accompanying Survey as a Continuous Feedback Instrument. International Journal of Information and Education Technology 14 (4), pp. 573–580, 2024, DOI: 10.18178/ijiet.2024.14.4.2079.

[GR93] Gallin, P.; Ruf, U.: Sprache und Mathematik in der Schule Ein Bericht aus der Praxis. Journal für Mathematik-Didaktik 14 (1), pp. 3–33, 1993, DOI: 10.1007/BF03339303.

[Gu22] Guo, L.: The effects of self-monitoring on strategy use and academic performance: A meta-analysis. International Journal of Educational Research 112, p. 101939, 2022, DOI: 10.1016/j.ijer.2022.101939.

[HNR10] Hübner, S.; Nückles, M.; Renkl, A.: Writing learning journals: Instructional support to overcome learning-strategy deficits. Learning and Instruction 20 (1), pp. 18–29, 2010, DOI: 10.1016/j.learninstruc.2008.12.001.

[HS93] Hartman, H.; Sternberg, R. J.: A broad BACEIS for improving thinking. Instructional Science 21 (5), pp. 401–425, 1993, DOI: 10.1007/BF00121204.

[KS19] Klopsch, B.; Sliwka, A.: Service Learning als „deeper learning“: Durch soziales Engagement (über-)fachliche Kompetenzen fördern. In (Jahn, D. et al., eds.): Kritische Hochschullehre: Impulse für eine innovative Lehr- und Lernkultur. Springer Fachmedien Wiesbaden, Wiesbaden, pp. 163–181, 2019, DOI: 10.1007/978-3-658-25740-8_9.

[MC95] McCrindle, A. R.; Christensen, C. A.: The impact of learning journals on metacognitive and cognitive processes and learning performance. Learning and Instruction 5 (2), pp. 167–185, 1995, DOI: 10.1016/0959-4752(95)00010-Z.

[NHR09] Nückles, M.; Hübner, S.; Renkl, A.: Enhancing self-regulated learning by writing learning protocols. Learning and Instruction 19 (3), pp. 259–271, 2009, DOI: 10.1016/j.learninstruc.2008.05.002.

[RN02] Rambow, R.; Nückles, M.: Der Einsatz des Lerntagebuchs in der Hochschullehre. Das Hochschulwesen, 2002.

[RS10] Rogerson, C.; Scott, E.: The fear factor: How it affects students learning to program in a tertiary environment. Journal of Information Technology Education: Research 9 (1), pp. 147–171, 2010, DOI: 10.28945/1183.

[RY09] Roberts, A.; Yoell, H.: Reflectors, Converts and the Disengaged: A study of Undergraduate Architecture Students' Perceptions of Undertaking Learning Journals. Journal for Education in the Built Environment 4 (2), pp. 74–93, 2009, DOI: 10.11120/jebe.2009.04020074.

[Sa20] Sala, A. et al.: LifeComp: The European framework for personal, social and learning to learn key competence, tech. rep., Joint Research Centre, 2020, DOI: 10.2760/302967.

Gemeinsames Arbeiten mit digitalen Technologien: Positive Effekte auf intrinsische Motivation und selbsteingeschätzte digitale Kompetenzen bei Lernenden in der dualen Berufsausbildung

Auswertung einer Online-Befragung von Auszubildenden im Bereich der Kfz-Mechatronik mit einem Strukturgleichungsmodell

Tim Komorowski [1], Laura Getz [1], Daniel König[1], Lilli Heimes[1],
Meike Weiland [1] und Michael Heister[1]

Abstract: Digitale Technologien spielen bei der Zusammenarbeit von Lernenden in vielen Bildungskontexten eine zentrale Rolle. Wir untersuchen in diesem Beitrag, ob das gemeinsame Arbeiten mit digitalen Technologien mit mehr intrinsischer Motivation von Schülerinnen und Schülern im berufsschulischen Bereich verbunden ist. Weiterhin untersuchen wir, ob eine höhere intrinsische Motivation mit besseren selbsteingeschätzten ICT-Kompetenzen (ICT; Information and Communication Technologies) verbunden ist. Als theoretische Grundlage verwendeten wir die Selbstbestimmungstheorie (SDT; Self-Determination Theory). Die vorliegende Untersuchung stützt sich auf querschnittliche Befragungsdaten von N = 343 Auszubildenden im Bereich der Kraftfahrzeugmechatronik aus den Jahren 2023/2024. Zur Analyse der Korrelationsmuster verwendeten wir ein Strukturgleichungsmodell. Die Ergebnisse zeigten, dass das gemeinsame Arbeiten mit digitalen Technologien eine stärkere intrinsische Motivation der Lernenden vorhersagte. Die intrinsische Motivation sagte wiederum bessere selbsteingeschätzte ICT-Kompetenzen der Schülerinnen und Schüler vorher.

Keywords: Motivation, Bildungstechnologien, digitale Kompetenzen, duale Berufsausbildung, Selbstbestimmungstheorie, Self-Determination Theory

1 Einleitung

Digitale Technologien eröffnen sowohl in der beruflichen Ausbildung als auch im Arbeitsleben neue Möglichkeiten der Zusammenarbeit. In diesem Beitrag untersuchen wir, inwieweit das gemeinsame Arbeiten mit digitalen Technologien die intrinsische Motivation von Schülerinnen und Schülern in beruflichen Schulen vorhersagt. Weiterhin

[1] Bundesinstitut für Berufsbildung (BIBB), Abteilung 4 – Initiativen für die Berufsbildung, Friedrich-Ebert-Allee 114-116, 53113 Bonn, tim.komorowski@bibb.de, https://orcid.org/0000-0001-6685-7347; getz@bibb.de, https://orcid.org/0000-0001-9327-864X; daniel.koenig@bibb.de; lilli.heimes@bibb.de; weiland@bibb.de, https://orcid.org/0000-0002-6857-2289; heister@bibb.de

doi:10.18420/delfi2025_34

untersuchen wir, ob die intrinsische Motivation ein Prädiktor für die selbsteingeschätzten ICT-Kompetenzen der Schülerinnen und Schüler [SSPE21] ist.

2 Theorie

Als theoretische Grundlage für diese Untersuchung verwenden wir die Selbstbestimmungstheorie (SDT; Self-Determination Theory) [DR08]. Es handelt sich hierbei um eine Theorie menschlicher Motivation, Entwicklung und menschlichen Wohlbefindens, in der unterschiedliche Motivationsformen unterschieden werden. Die intrinsische Motivation stellt in der SDT die autonomste Motivationsform dar. Personen, die intrinsisch motiviert sind, verfolgen Handlungen aufgrund ihres persönlichen Interesses und der Freude an der Tätigkeit selbst. Damit unterscheidet sich die intrinsische Motivation von extrinsischen Motivationsformen, die z. B. auf externe Anreize wie Belohnungen oder Druck beruhen [DR08]. In der SDT ist die Annahme zentral, dass die Befriedigung der psychologischen Grundbedürfnisse nach Autonomieerleben, Kompetenzerleben und dem Erleben von sozialem Anschluss zu einer höheren intrinsischen Motivation führt.

3 Methode

3.1 Stichprobe

Die Stichprobe umfasst die Antworten von insgesamt $N = 343$ Schülerinnen und Schülern der Hamburger Beruflichen Schule für Fahrzeugtechnik. Die Online-Befragung fand im Zeitraum von Dezember 2023 bis Januar 2024 statt und wurde anonym durchgeführt. Das Durchschnittsalter lag bei 20 Jahren. 89,5 % der Befragten gaben ein männliches Geschlecht an, 8,2 % ein weibliches und 2,3 % ein anderes Geschlecht.

3.2 Messinstrumente und erhobene Konstrukte

Digitale Technologien. Die beruflichen Schülerinnen und Schüler wurden im Online-Fragebogen danach gefragt, ob sie digitale Technologien (z. B. Künstliche Intelligenz, Learning Management Systeme, Social Media) in der Berufsschule nutzen (verwendetes Antwortformat: *ja / nein*).

Selbsteingeschätzte ICT-Kompetenzen. Die selbsteingeschätzten ICT-Kompetenzen wurden unter Rückgriff auf das psychometrisch evaluierte Messinstrument von Schauffel et al. [SSPE21] gemessen. Da nur geringe Unterschiede in den Ausprägungen der einzelnen ICT-Kompetenzfacetten bestanden, verwendeten wir als Messmodell ein g-faktor-Modell. Um die faktorielle Validität zu prüfen, führten wir eine konfirmatorische

Faktorenanalyse durch. Das Messmodell zeigte einen ausreichenden Modellfit [HB99], $\chi^2 = 452.741$, $p < .001$, CFI = .95, RMSEA = .07, SRMR =.033.

Intrinsische Motivation. Die intrinsische Motivation der Schülerinnen und Schüler im Umgang mit digitalen Technologien wurde unter Rückgriff auf die SDT-basierte Skala *Autonomy and Competence in Technology Adoption Questionnaire (ACTA)* [PCR18] gemessen. Wir verwendeten ein *first-order-correlated factor model* um die Messqualität der intrinsischen Motivation zu prüfen und die intrinsische Motivation von anderen (extrinsischen) in der SDT beschriebenen Motivationsformen abzugrenzen. Das Messmodell zeigte einen guten Modellfit [HB99], $\chi^2 = 365.870$, $p < .001$, CFI = 0.97, RMSEA = .059, SRMR =.039.

3.3 Forschungsfrage

Forschungsfrage: Inwieweit sagt das gemeinsame Arbeiten mit digitalen Technologien im berufsschulischen Unterricht die intrinsische Motivation von Schülerinnen und Schülern vorher und inwieweit sagt diese wiederum deren selbsteingeschätzte ICT-Kompetenzen vorher?

4 Ergebnisse

Das erstellte Strukturgleichungsmodell (Abb. 1) wies einen guten Modellfit auf [HB99], $\chi^2 = 263.982$, $p < .05$, CFI = .96, RMSEA = .043, SRMR =.044 auf.

Das gemeinsame Arbeiten mit digitalen Technologien hatte einen starken Effekt [Ry22; Bu22] auf die intrinsische Motivation ($\beta = .51$). Daneben hatte lediglich der Einsatz von KI einen signifikanten Einfluss auf die intrinsische Motivation ($\beta = .46$). Weiterhin zeigte sich, dass die intrinsische Motivation bessere selbsteingeschätzte ICT-Kompetenzen der Schülerinnen und Schüler vorhersagte ($\beta = .32$).

5 Diskussion

Der Zusammenhang zwischen dem *gemeinsamen Arbeiten mit digitalen Technologien* und einer stärkeren *intrinsischen Motivation* von Auszubildenden in der KFZ-Mechatronik kann im Rahmen der SDT [DR08] erklärt werden: Die gemeinsame Arbeit könnte dazu beigetragen haben, das psychologische Grundbedürfnis nach sozialem Anschluss zu befriedigen. Die SDT geht davon aus, dass die Erfüllung der Bedürfnisse nach Autonomie, Kompetenzerleben und sozialem Anschluss zentrale Voraussetzungen für die Entwicklung intrinsischer Motivation darstellen.

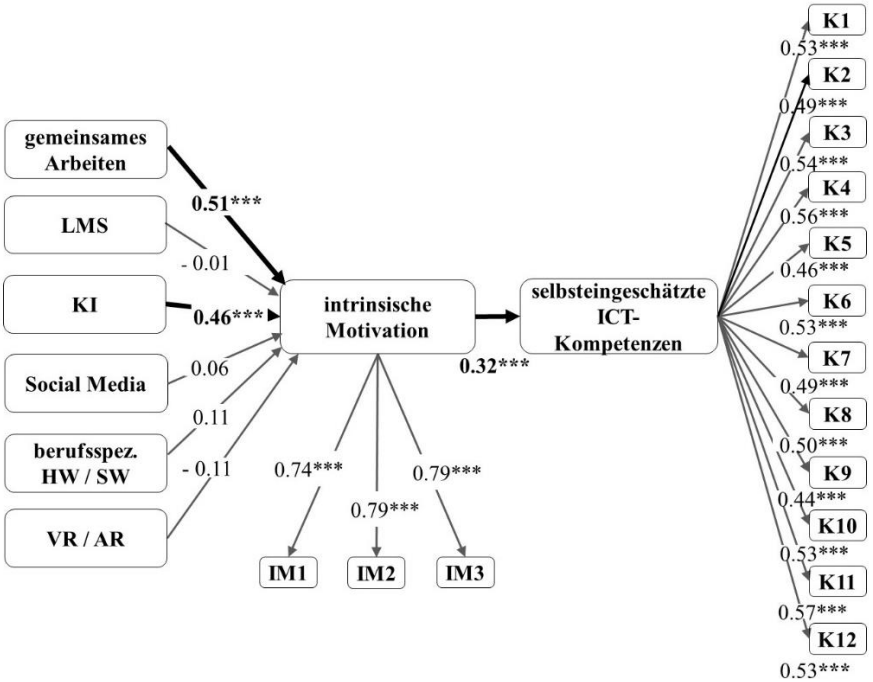


Abb. 1: Strukturgleichungsmodell

Anm.: gemeinsames Arbeiten = gem. Arbeiten mit digitalen Technologien; LMS = Learning Management Systeme; KI = Künstliche Intelligenz; Social Media = Social Media Nutzung; berufsspez. HW / SW = Nutzung berufsspez. Hardware und Software; VR / AR = Nutzung von Virtual / Augmented Reality Technologien im Unterricht.

Literaturverzeichnis

[Bu22] Bureau, J. S.; Howard, J. L.; Chong, J. X. Y.; Guay, F.: Pathways to Student Motivation: A Meta-Analysis of Antecedents of Autonomous and Controlled Motivations. Review of Educational Research, 92(1), S. 46-72, 2022.

[DR08] Deci, E. L.; Ryan, R. M.: Self-determination theory: A macrotheory of human motivation, development, and health. Canadian Psychology, 49(3), S. 182–185, 2008.

[HB99] Hu, L.; Bentler, P. M.: Cutoff criteria for fit indexes in covariance structure analysis: Conventional criterial versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), S. 1-55, 1999.

[KMK21] Kultusministerkonferenz (KMK): Lehren und Lernen in der digitalen Welt. Ergänzung zur Strategie der Kultusministerkonferenz "Bildung in der digitalen Welt". Sekretariat der Kultusministerkonferenz, 2021.

[PCR18] Peters, D.; Calvo, R. A.; Ryan, R. M.: Designing for Motivation, Engagement and Wellbeing in Digital Experience. Frontiers in Psychology, 9, Artikel 797, 2018.

[Ry22] Ryan, R. M.; Duineveld, J. J.; Di Domenico, S. I.; Ryan, W. S.; Steward, B. A.; Bradshaw, E. L.: We know this much is (meta-analytically) true: A meta-review of meta analytic findings evaluating self-determination theory. Psychological Bulletin, 148, S. 813-842, 2022.

[SSPE21] Schauffel, N.; Schmidt, I.; Peiffer, H.; Ellwart, T.: Self-concept related to information and communication technology: Scale development and validation. Computers in Human Behavior Reports, 4, 2021.

ViPATalk meets AI: Advancing Multimodal Virtual Anamnesis Simulation using LLM Fine-tuning

Martin Sboron[1], Oliver Thews[2], and Thomas Schmid[1,3]

Abstract: Virtual simulated patients (VSP) have become an education tool for medical students and are implemented by frameworks of varying degrees of immersion. The path to free-form interaction between the user and the VSP has been of ever-growing interest. For a step in this direction, we extend the ViPATalk framework by Lippitsch et al. by replacing the current rule-based system for question categorization with a fine-tuned Large Language Model (LLM). We demonstrate that smaller LLMs perform similarly to larger models, when trained in the same domain, and that recent base models perform similarly to their language-specific fine-tuned counterparts. Further, our findings indicate that generated training data can be helpful when little data is available for certain categories.

Keywords: AI, Chatbot, Large Language Model, Fine-tuning, medical education, anamnesis

1 Motivation

Simulated patients, i. e. actors or students who simulate having a specific medical condition, play an important role in helping medical students practice their skills [Sz19]. But due to resource and logistical costs, it can be challenging to provide students with sufficient practice opportunities. Virtual simulated patients (VSPs) can therefore be an effective complement to traditional role-playing, providing consistent and repeated practice [Le20]. The VSP framework ViPATalk (Virtual Patient Anamnesis – Talk) provides a virtual patient avatar as well as audio/text input and output for interaction [Li24]. Originally, ViPATalk maps user questions to pre-recorded patient answers via a purely rule-based system that only matches the provided words in the rules to the questions. While this approach provides detailed control over the chatbot behavior, its implementation has escalated to around 1600 to 2300 rules. Different word forms (e. g. singular and plural) and word tenses (present, past), for example, require rather complex rules or even additional rules. With the current work, we aim to create a more seamless user experience for the free-form interaction with the multimodal ViPATalk chatbot. For more flexibility, better extensibility and easier maintenance we replaced the original rule-based framework with a Large Language Model (LLM). We fine-tuned and compared several state-of-the-art LLMs to provide reasoning for the question: *What is the influence of the base model variant on a self-contained task such as question categorization?*

[1] University of Halle-Wittenberg, Digital Research Methods in Medicine Group, Magdeburger Str. 8, 06112, Halle (Saale), Germany, martin.sboron@student.uni-halle.de; thomas.schmid@medizin.uni-halle.de
[2] University of Halle-Wittenberg, Julius Bernstein Institute of Physiology, Magdeburger Str. 6, 06112, Halle (Saale), Germany, oliver.thews@medizin.uni-halle.de
[3] Lancaster University Leipzig, School of Computing and Communications, Nikolaistrasse 10, 04109, Leipzig, Germany, thomas.schmid@medizin.uni-halle.de

doi:10.18420/delfi2025_35

2 Background

ViPATalk allows to practice complex anamnesis scenarios for specific predefined diagnoses. Students can input questions to an avatar either via speech-to-text or via a chat window. Internally, the resulting input text is then matched to a predefined *answer code* (e. g. "2_2_8"), which specifies which of the pre-recorded audio-visual answers is returned to the user. In order to ensure that the answers support the intended learning outcomes, these audio-visual responses of the avatar are carefully specified in advance to represent the underlying illness to be diagnosed with realistic symptoms, patient's appearance, and behavior. So, an end-to-end approach was not reasonable as the first stage of improvement with AI.

LLMs are usually based on the Transformer architecture [Va17] and can be divided into two approaches: classification-based and generation-based. Classification-based models have an encoder-only architecture and assign the input text to the label with the highest probability of a given class vector. Current generation-based models have usually a decoder-only architecture. Given an input text and an instruction prompt, they generate the next token (in our case part of the code) with the highest probability in an auto-regressive manner. Many pre-trained LLMs are publicly available, among them the families: LLaMA 3 [Gr24], Falcon [Al23], Mistral [Mi], Qwen [Ba23], and Gemma 2 [Ge24] are particularly popular.

3 Methodology

We aimed to identify best-practices for fine-tuning a LLM such that it can map the questions to their corresponding categories best (i. e. the relevant answer codes). For this, we explored classification-based as well as generation-based models. The used classification-based models were trained through a straightforward matching between questions and labels for the answer codes. The used generation-based models were trained using an instruction prompt outlining the task of question mapping and providing permitted answer codes. All models were fine-tuned using the same data-split, data-seed, and answer codes stratification. Due to hardware and response latency limitations, we only considered smaller LLMs.

Data: Training data for three distinct anamnesis cases were collected during previous studies by Lippitsch et al. [Li24]. Questions originate from students participating in the studies, answer codes were created by the rule-based system and revised by a human expert. The corresponding three virtual patients (AS, LD, MS) provide 108, 110 and 136 answer codes each. After simple lowercase matching deduplication, the sub-datasets for the three cases contained 2983, 2965 and 2650 unique questions, respectively. The data was split into 70/30 % for training/testing. The training split was further divided into 70/30 % for training/validation. Thus, models were fine-tuned using 49 % of the data. To preserve their original distribution across the dataset, splitting was done with answer codes stratification.

Classification-based Models: Due to its widespread use for natural language processing tasks [Ga25], we chose to use BERT (335 million parameters) [De19] for our classification

approach. More precisely we chose a version, that was pre-trained on a corpus of purely German text: "GBERT" by Chan et al. [CSM20]. During training, the base model was kept frozen, with only the parameters of an add-on adapter for classification being fine-tuned.

Generation-based Models: We explored four current pre-trained LLMs for our generation-based approach: Meta's LLaMA 3.1 (8B), LLaMA 3.2 (3B) [Gr24], and Google's Gemma 2 in its 2B and 9B variant [Ge24]. All four offer strong performance at comparatively small sizes and provide a solid basis for cost-effective fine-tuning. For the German fine-tuned generation models we used the Sauerkraut models (in the following denoted by "SK") from VAGOsolutions [VA]. During training, the base model was kept frozen and fine-tuned via the LORA-method [Hu21a].

4 Experiment Findings

Classification-based vs. generation-based LLMs: Fine-tuned models were able to match the correct answer codes for patient "AS" with the following accuracy scores: GBert-large: 90.3 %, Gemma 2 (2B)-SK: 89.5 %, Llama 3.2 (3B): 88.8 %, Llama 3.1 (8B)-SK: 89.3 %, Gemma 2 (9B)-SK: 89.7 %, with increases of +90.3 %, +55.6 %, +50.8 %, +32.3 % and +29.7 % in comparison to the respective vanilla model. Notably, GBert performs competitively even with models of 30x its size, such as Gemma 2 (9B). At the same time, larger generation-based models are able to achieve >50 % accuracy on our specific use case even without fine-tuning, i. e. only with the corresponding instruction prompt. Hallucination rates for all fine-tuned models in all experiments were ≤1 % and <5 % for vanilla models.

Model size: Results for model accuracy indicate that smaller models perform overall similarly well as larger ones for the same patient. In a cross-validation-style approach, we also tested fine-tuned models on the data of the other patients, for which they were not trained for (Fig. 1). Under this condition, smaller models seem to perform worse than larger ones in the given task when confronted with out-of-distribution data. But notably, when trained and tested on data of the same patient GBert performed best.

Language specialization: We examined whether there are notable advantages of German fine-tuned model variants of base models (Fig. 2). In our experiments, this was only noticeable for GBert, whose training corpus is less than 1 % of the corpus of the much larger Llama. Overall, model size seems to be less important for performance than language-specific improvements. In the case of Gemma 2-2B, for example, smaller variants of the same model are often distilled from the larger variants and may not retain as much knowledge and nuance (of e. g. language constructs) as the large one.

Data augmentation: Using the rule set of the original ViPATalk framework as template, we explored the effect of synthetic questions generated with ChatGPT4 for underutilized but important answer codes on the models (Fig. 3). It appears that additional questions, even synthetic ones, boost underutilized answer codes and can even slightly improve performance for the other codes for some models, but for other models there might be a slight decrease.

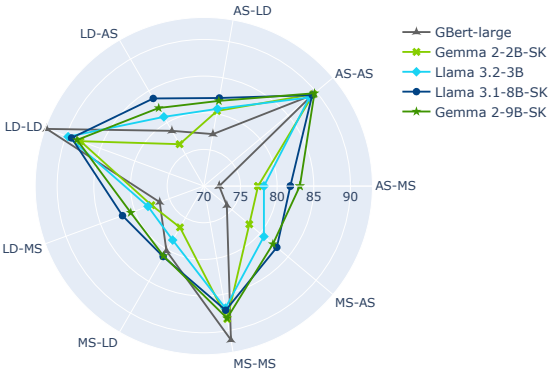


Fig. 1: Cross-validation over patients, shown in %. "AS-LD" means the model was trained on patient AS and tested on patient LD.

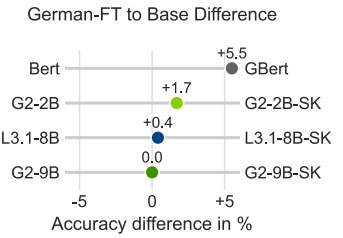


Fig. 2: German-language fine-tuned models vs. their base models. ("G" = Gemma, "L" = Llama, "SK" = Sauerkraut-variant; for Llama 3.2 no Sauerkraut-variant was available at the time of experiments)

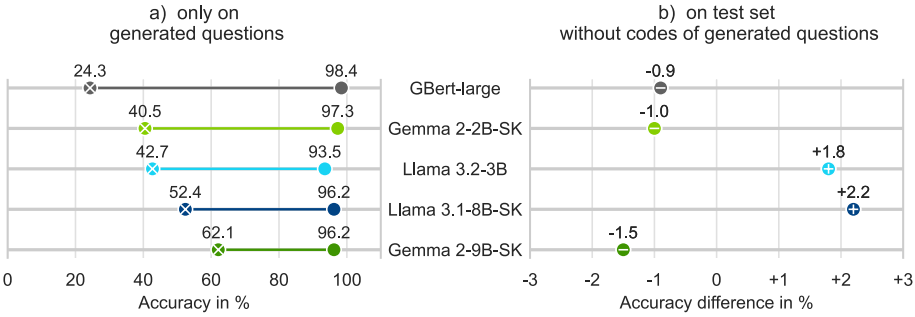


Fig. 3: Evaluation of data augmentation for fine-tuning LLMs for patient MS using 185 questions generated for 12 answer codes (at least 10 per code). a) Models tested on the set of generated questions without being trained on them (x-circle), and with being trained on them (full-circle). b) The accuracy difference when tested on the test set without the answer codes of the generated questions: between the model when trained without the generated questions and when trained with the generated questions (plus/minus-circle).

5 Conclusion and Outlook

With this work, we created a practical and reliable question categorization for a VSP setting by fine-tuning LLMs for this. We found that all fine-tuned models achieved high performance (≥85 %), even smaller models. On these grounds, we developed an efficient LLM-based extension of the ViPATalk framework improving its usefulness significantly. Thanks to a small number of model parameters, our winner model GBert can be operated even without a GPU: using quantization (to 8 bit), the ONNX runtime [ON21] and the Optimum framework [Hu21b], average latency per sample was improved on our CPU to tens of milliseconds with only a slight (<2 %) decrease in accuracy. As the approach is currently limited to the given three cases/patients, future work will further investigate the merits of using generation-based models for more general question categorization.

References

[Al23] Almazrouei, E.; Alobeidli, H.; Alshamsi, A.; Cappelli, A.; Cojocaru, R.; Debbah, M.; Goffinet, É.; Hesslow, D.; Launay, J.; Malartic, Q.; Mazzotta, D.; Noune, B.; Pannier, B.; Penedo, G.: The Falcon Series of Open Language Models, 2023, DOI: 10.48550/arXiv.2311.16867, arXiv: 2311.16867 [cs.CL].

[Ba23] Bai, J. et al.: Qwen Technical Report. arXiv preprint arXiv:2309.16609, 2023, DOI: 10.48550/arXiv.2309.16609.

[CSM20] Chan, B.; Schweter, S.; Möller, T.: German's Next Language Model. CoRR abs/2010.10906, 2020, DOI: 10.48550/arXiv.2010.10906, arXiv: 2010.10906.

[De19] Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In (Burstein, J.; Doran, C.; Solorio, T., 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, 2019, DOI: 10.18653/v1/N19-1423.

[Ga25] Gardazi, N. M.; Daud, A.; Malik, M. K.; Bukhari, A.; Alsahfi, T.; Alshemaimri, B.: BERT applications in natural language processing: a review. Artificial Intelligence Review 58 (6), pp. 1–49, 2025, DOI: 10.1007/s10462-025-11162-5.

[Ge24] Gemma Team: Gemma 2: Improving Open Language Models at a Practical Size, 2024, DOI: 10.48550/arXiv.2408.00118, arXiv: 2408.00118 [cs.CL].

[Gr24] Grattafiori, A. et al.: The Llama 3 Herd of Models, 2024, DOI: 10.48550/arXiv.2407.21783, arXiv: 2407.21783 [cs.AI].

[Hu21a] Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W.: LoRA: Low-Rank Adaptation of Large Language Models, 2021, DOI: 10.48550/arXiv.2106.09685, arXiv: 2106.09685 [cs.CL].

[Hu21b] Hugging Face: Optimum framework, https://github.com/huggingface/optimum, 2021, accessed: 06/09/2025.

[Le20] Lee, J.; Kim, H.; Kim, K. H.; Jung, D.; Jowsey, T.; Webster, C. S.: Effective virtual patient simulators for medical communication training: A systematic review. en, Med. Educ. 54 (9), pp. 786–795, 2020, DOI: 10.1111/medu.14152.

[Li24] Lippitsch, A.; Steglich, J.; Ludwig, C.; Kellner, J.; Hempel, L.; Stoevesandt, D.; Thews, O.: Development and evaluation of a software system for medical students to teach and practice anamnestic interviews with virtual patient avatars. Computer Methods and Programs in Biomedicine 244, p. 107964, 2024, DOI: 10.1016/j.cmpb.2023.107964.

[Mi] Mistral AI, https://github.com/mistralai/mistral-inference, accessed: 06/09/2025.

[ON21] ONNX Runtime developers: ONNX Runtime, https://onnxruntime.ai/, 2021, accessed: 06/09/2025.

[Sz19] Szauter, K.: Simulated and Standardized Patients. In: Clinical Simulation. Elsevier, pp. 33–39, 2019, DOI: 10.1016/B978-0-12-815657-5.00003-6.

[VA] VAGO solutions, https://huggingface.co/VAGOsolutions, accessed: 06/09/2025.

[Va17] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I.: Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17, Curran Associates Inc., Long Beach, California, USA, pp. 6000–6010, 2017, DOI: 10.48550/arXiv.1706.03762.

Vergleich von Stresssituationen im simulierten Schulunterricht durch Biofeedback



Juliane Scherlitzki [1] und Axel Wiepke [1]

Abstract: Virtual Reality (VR) Anwendungen werden in Wissenschaft und Lehre immer häufiger wegen ihres emotionsauslösenden Effekts genutzt. Teach-R ist eine solche Anwendung, die von angehenden Lehrkräften genutzt werden kann, um den Umgang mit Unterrichtsstörungen zu trainieren. Dieser Beitrag exploriert das Potenzial einer begleitenden Stressmessung durch Herzratensensor zur Identifikation überfordernder Situationen. Neben einem genaueren Einblick, welche Unterrichtssituationen insbesondere von Lehramtsstudent*innen zum Training benötigt werden, bieten die Herzfrequenzdaten die Möglichkeit, die Anwendung adaptiv an individuelle Bedürfnisse anzupassen. Gleichzeitig könnte der Erwerb von Handlungskompetenzen über Biofeedback sichtbar gemacht werden. In einer Vorstudie (n=14) wurde untersucht, inwieweit sich das Stressniveau von Proband*innen durch die Variablen Störungsart (Erschwerung, Unterbrechung) und Handlungsnotwendigkeit (Eingreifen (/nicht) erforderlich) erklären lässt. Erste Ergebnisse deuten den erwarteten Einfluss der Variablen und der Handlungskompetenz gemessen durch Selbstauskunft an.

Keywords: Biofeedback, Virtual Reality, Teach-R, Herzfrequenz, Stress, Adaptivität

1 Stressmessung in VR

Die Anwendung Teach-R wird bereits erfolgreich in der Praxis eingesetzt, um die Ausbildung angehender Lehrer*innen zu unterstützen [Wi21]. Nutzer*innen (Immersantis) begeben sich mit Hilfe einer VR-Brille in ein virtuelles Klassenzimmer und können dort das Unterrichten vor virtuellen Schüler*innen (vSuS) trainieren. Das Training findet in Fachräumen oder Klassenräumen mit unterschiedlichen Sitzplänen statt.

Das Verhalten der vSuS kann über eine Weboberfläche von einem Coach gesteuert werden. Dabei stehen entsprechende Vertonungen und spezielle Animationen für die vSuS zur Verfügung. Da die Aufgabe des Coaches das gleichzeitige Bedienen der Weboberfläche sowie das Bewerten der fachlichen Korrektheit der Aussagen des Immersantis ist, kann die Einschätzung des Stresslevels des Immersantis in den Hintergrund geraten. Eine potenzielle Unter- oder Überforderung des Immersantis kann möglicherweise unbemerkt bleiben, was den Lerneffekt beeinträchtigen kann.

Auf die Belastung der Immersantis können Herzfrequenzmessungen Rückschlüsse liefern. Dabei handelt es sich um einen individuellen Wert, der vom allgemeinen Gesundheitszustand einer Person und von äußeren Faktoren beeinflusst wird [Lö23; Yu18]. Durch

[1] Universität Potsdam, Institut für Informatik und Computational Science, An der Bahn 2, 14476 Potsdam, Deutschland, juliane.scherlitzki@uni-potsdam.de, https://orcid.org/0000-0002-1799-449X; axel.wiepke@uni-potsdam.de, https://orcid.org/0000-0002-0555-4040

doi:10.18420/delfi2025_36

eine Zeitreihenmessung kann das Absinken oder Steigen der Herzfrequenz jedoch als guter Indikator für Stress herangezogen werden [Ge16; SB17; SME20; We24]. Dabei wird jegliche Belastung für den Organismus als Stress verstanden, unabhängig davon, ob es sich um negativen oder positiven Stress handelt [EFF22].

Starten Lehramtsstudierende in den Praxisteil ihrer Ausbildung, werden sie immer wieder mit neuen Herausforderungen konfrontiert, die potenziell zu Stress führen können [Hu21]. Je besser sie mit Handlungskompetenzen ausgestattet wurden, desto besser können sie auf äußere Stressoren reagieren, was ihre Belastung reduzieren kann [KU22; VS18]. Dies könnte langfristig das Risiko stressbedingter Erkrankungen wie Burnout reduzieren [SK13].

Um Coaches die Möglichkeit zu geben, adäquat auf den Stress der Immersantis zu reagieren, wurde Teach-R um ein Modul zur Messung der Herzrate erweitert. Das Modul nutzt einen Elektrokardiogram-Sensor (EKG-Sensor) und übermittelt Daten via Bluetooth in Echtzeit an die VR-Umgebung. Dadurch kann die Herzfrequenz der Immersantis als objektives Maß für das individuelle Stressniveau herangezogen werden.

2 Methodik

Im Rahmen einer Vorstudie (n=14) wurden einige notwendigen Grundlagen für die künftige Nutzung der Herzfrequenzdaten gelegt. Sämtliches Material ist über Zenodo unter [Sc25] abrufbar. Durch die Verbindung eines EKG-Sensors (Polar H10) mit Teach-R über Bluetooth können die Herzfrequenzen von Proband*innen nun während der Nutzung der Anwendung angezeigt sowie protokolliert werden. Darüber hinaus wurde die Steuerung der vSuS dahingehend erweitert, dass Abläufe ohne Interaktion des Coachs möglich sind. Diese Abläufe können als *Szenarien* vom Coach auf der Web-Oberfläche entworfen werfen und bieten so eine exakte Reproduzierbarkeit. Szenarien bestehen aus Aktionen (Animationen oder Audiodateien), denen jeweils ein fixer Startzeitpunkt sowie eine Dauer zugewiesen werden.

Die Vorstudie hatte zum Ziel, die abhängige Variable Stress durch die unabhängigen, dichotomen Variablen Störungsart und Handlungsnotwendigkeit abzubilden. Damit sollten die beiden Forschungsfragen untersucht werden, ob (1) die Art der Störung sowie (2) das Vorhandensein oder Fehlen einer Handlungsnotwendigkeit zu Unterschieden in den Stressreaktionen von Proband*innen führt. Störungsarten werden unterteilt in Unterbrechung und Erschwerung. Bei einer Unterbrechung muss die aktuelle Unterrichtstätigkeit pausiert werden und kann erst fortgesetzt werden, sobald die Störquelle beseitigt wurde. Bei einer Erschwerung kann das Unterrichten fortgesetzt werden, nur eben unter schwierigeren Bedingungen wie beispielsweise Lärm [EFF22; KDM13]. Die Handlungsnotwendigkeit beschreibt, ob konkretes Handeln der Lehrperson in der Störsituation notwendig ist, oder Handlungen der Lehrperson ohnehin keinen Effekt auf die aktuelle Situation hätten [DLK13]. Um alle Kombinationen der beiden dichotomen Variablen zu untersuchen, wurden vier jeweils 90 Sekunden andauernde Szenarien konzipiert:

Szenario 1 (Unterbrechung, Handlung) besteht aus dem mehrfachen Klingeln oder Vibrieren der Handys der vSuS, worauf die Proband*innen reagieren können.

Szenario 2 (Unterbrechung, keine Handlung) ist an ein Beispiel von [KDM13] angelehnt. Es erklingt ein Gong mit einer nachfolgend längeren Durchsage, die sowohl für die Lehrkraft als auch für die vSuS relevante Informationen vermittelt.

Szenario 3 (Erschwerung, Handlung) ist geprägt von Störungen durch die vSuS wie: aus dem Fenster schauen, Getränkedose öffnen und trinken, oder unterrichtsferne Gesprächen.

Szenario 4 (Erschwerung, keine Handlung) besteht, ebenfalls in Anlehnung an [KDM13], aus Geräuschen von Umbauarbeiten. Während der Unterricht fortgesetzt werden kann, erschwert der Lärm die Kommunikation zwischen Lehrkraft und Lernenden.

Die Teilnehmer*innen der Vorstudie bekamen vorbereitetes Unterrichtsmaterial ausgehändigt, mit dem sie sich vor Beginn des Experiments vertraut machen konnten. Nach dem Anlegen der VR-Ausrüstung und des EKG-Sensors folgte eine individuelle Eingewöhnungsphase. Hierbei konnten sie sich mit der Umgebung vertraut machen und zum Beispiel das Navigieren in VR oder das Steuern der Präsentation ausprobieren. Mit Beginn der eigentlichen Studie sollten die Proband*innen anhand des Materials eine Unterrichtseinheit abhalten. Währenddessen wurden sie nacheinander mit den vier Szenarien konfrontiert, welche in randomisierter Reihenfolge abgespielt wurden. Im Anschluss an das Experiment füllten die Teilnehmer eine kurze Onlineumfrage aus, in der maßgeblich demografische Informationen sowie ihre Lehrerfahrung abgefragt wurden. Die Experimente fanden alle in einem VR-Labor mit gleichbleibender Studienleitung statt und Proband*innen waren stets einzeln anwesend.

Zur Auswertung wurden neben den Daten aus der Umfrage die Herzfrequenzen sowie die vordefinierten Ereignisse während des Experiments (z. B. Start von Animationen) genutzt. Für die Frequenzen und Ereignisse wurden Protokolldateien inklusive Zeitstempel erzeugt. In Abbildung 2 wird exemplarisch die Herzfrequenz zweier Proband*innen im zeitlichen Verlauf für ein Szenario gezeigt. Hier zeigt sich, wie unterschiedlich und teils konträr die Reaktionen auf die gleiche Abfolge von Ereignissen ausfallen, was möglicherweise auch mit den heterogenen Erfahrungen der Proband*innen korreliert.

Im arithmetischen Mittel liegen die Herzfrequenzen der Proband*innen für Erschwerungen etwas höher als die für Unterbrechungen (siehe Abbildung 1). Außerdem scheinen Szenarien, in denen ein Handeln erforderlich ist, für höhere Herzfrequenzen zu sorgen als solche, in denen nur abgewartet werden muss.

In der Umfrage zur Lehrerfahrung konnten die Proband*innen (i) keine, (ii) weniger als ein Jahr, (iii) ein bis drei Jahre und (iv) mehr als drei Jahre angeben. Ergebnisse der kleinen Stichprobe deuten an, dass mit steigender Lehrerfahrung der Effekt von insbesondere der Störungsart geändert wird. Teilt man die Proband*innen in eine Gruppe ohne Lehrfahrung (A, n=7) und in eine mit grundsätzlich vorhandener Lehrerfahrung (B, n=7) ein, kehrt sich der Einfluss der Art der Störung um, wie in Abb. 3 und 4 ersichtlich wird.

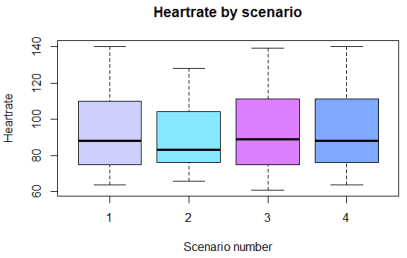


Abb. 1: Interaktionsplot über alle Proband*innen

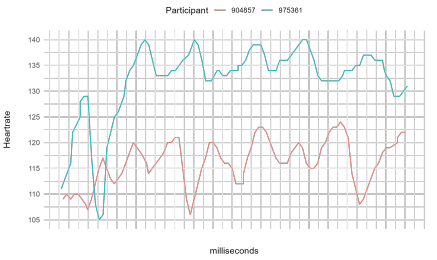


Abb. 2: Herzfrequenzen zweier Proband*innen

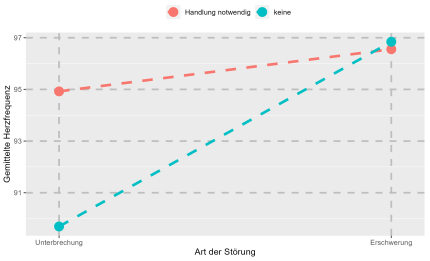


Abb. 3: Interaktion: Gruppe A

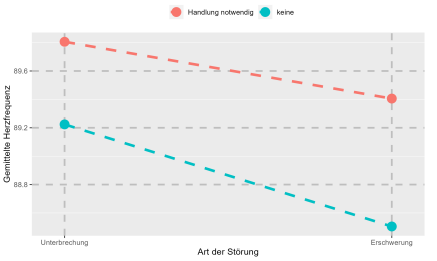


Abb. 4: Interaktion: Gruppe B

Proband*innen aus Gruppe A waren von Erschwerungen gestresster als von Unterbrechungen, wobei der Anstieg für Situationen ohne Handlungsnotwendigkeit steiler ausfällt als für solche mit Handlungsnotwendigkeit. Im Gegensatz dazu scheinen Personen aus Gruppe B gestresster von Unterbrechungen zu sein. Situationen, die Handeln erforderlich machen führen in dieser Gruppe zu höheren Herzfrequenzen als solche ohne Handlungsnotwendigkeit.

3 Weiterentwicklung unter Standards

Durch die geringe Anzahl an Proband*innen sind die Ergebnisse nicht signifikant. Doch die Erkenntnisse sollen künftig für eine größer angelegte Studie genutzt werden, bei der gezielt Lehramtsstudierende in den Fokus genommen werden sollen.

Statt der individuellen Implementierung zur Protokollierung der Daten soll mit OmiLAXR künftig auf einen Standard zurückgegriffen werden. Diese Umstellung begünstigt ebenfalls die Weiterverarbeitung der Daten wie zum Beispiel durch verständliche Visualisierungen in der Web-Oberfläche oder zum automatischen Anpassen des Schwierigkeitsniveaus des Trainings. Die gewonnenen Daten können Hinweise geben, welche Handlungskompetenzen bei einzelnen Immersantis besonders gefördert werden müssen. Außerdem lässt die Betrachtung der Stressreduktion über mehrere Sitzungen hinweg Rückschlüsse auf die Wirksamkeit des Trainings zu.

Literaturverzeichnis

[DLK13] Dorsemagen, C.; Lacroix, P.; Krause, A.: Arbeitszeit an Schulen: Welches Modell passt in unsere Zeit? Kriterien zur Gestaltung schulischer Arbeitsbedingungen. In (Rothland, M., Hrsg.): Belastung und Beanspruchung im Lehrerberuf: Modelle, Befunde, Interventionen. Springer Fachmedien, Wiesbaden, S. 213–230, 2013, DOI: 10.1007/978-3-531-18990-1_12.

[EFF22] Ernst, G.; Franke, A.; Franzkowiak, P.: Stress und Stressbewaeltigung. In (für gesundheitliche Aufklärung (BZgA), B., Hrsg.): Leitbegriffe der Gesundheitsförderung und Prävention. Glossar zu Konzepten, Strategien und Methoden. 2022, DOI: 10.17623/BZGA:Q4-i118-2.0.

[Ge16] Gerber, N.: Let's get physical: Lässt sich User Experience physiologisch erfassen? Publisher: Gesellschaft für Informatik e.V., 2016, https://dl.gi.de/handle/20.500.12116/198.

[Hu21] Huang, Y. et al.: Classroom complexity affects student teachers' behavior in a VR classroom. Computers & Education 163, 2021, DOI: 10.1016/j.compedu.2020.104100.

[KDM13] Krause, A.; Dorsemagen, C.; Meder, L.: Messung psychischer Belastungen im Unterricht mit RHIA-Unterricht. In (Rothland, M., Hrsg.): Belastung und Beanspruchung im Lehrerberuf: Modelle, Befunde, Interventionen. Springer Fachmedien, Wiesbaden, S. 99–116, 2013, DOI: 10.1007/978-3-531-18990-1_6.

[KU22] KULTUSMINISTERKONFERENZ: Standards für die Lehrerbildung: Bildungswissenschaften, Beschluss der Kultusministerkonferenz, Berlin, 2022.

[Lö23] Löllgen, H.: Herzfrequenz und Blutdruck. In (Niebauer, J., Hrsg.): Sportkardiologie. Springer, Berlin, Heidelberg, S. 89–111, 2023, DOI: 10.1007/978-3-662-65165-0_8.

[SB17] Sammito, S.; Böckelmann, I.: Alters- und geschlechterbezogene Referenzwerte für den Einsatz der Herzfrequenzvariabilität in der Bewegungstherapie. B&G Bewegungstherapie und Gesundheitssport 33, S. 268–275, 2017, DOI: 10.1055/s-0043-120631.

[Sc25] Scherlitzki, J.: Biofeedback during VR teaching units, 2025, DOI: 10.5281/zenodo.15102223.

[SK13] Schaarschmidt, U.; Kieschke, U.: Beanspruchungsmuster im Lehrerberuf Ergebnisse und Schlussfolgerungen aus der Potsdamer Lehrerstudie. In (Rothland, M., Hrsg.): Belastung und Beanspruchung im Lehrerberuf: Modelle, Befunde, Interventionen. Springer Fachmedien, Wiesbaden, S. 81–97, 2013, DOI: 10.1007/978-3-531-18990-1_5.

[SME20] Schormann, I.; Müller, K.; Ebert, A.: Mofebo: move and feel your body a fitness game with biofeedback in VR. In: Proceedings of Mensch und Computer 2020. MuC '20, Association for Computing Machinery, New York, NY, USA, S. 495–499, 2020, DOI: 10.1145/3404983.3410416.

[VS18] Vesely-Maillefer, A. K.; Saklofske, D. H.: Emotional Intelligence and the Next Generation of Teachers. In (Keefer, K. V.; Parker, J. D. A.; Saklofske, D. H., Hrsg.): Emotional Intelligence in Education: Integrating Research with Practice. Springer International Publishing, Cham, S. 377–402, 2018, DOI: 10.1007/978-3-319-90633-1_14.

[We24] Westphal, A. et al.: More I-talk in student teachers' written reflections indicates higher stress during VR teaching. Computers and Education 212, 2024, DOI: 10.1016/j.compedu.2024.104987.

[Wi21] Wiepke, A. et al.: Jenseits des eigenen Klassenzimmers: Perspektiven & Weiterentwicklungen des VR-Classrooms. Gesellschaft für Informatik e.V., 2021.

[Yu18] Yu, B. et al.: Biofeedback for Everyday Stress Management: A Systematic Review. Frontiers in ICT 5, 2018, DOI: 10.3389/fict.2018.00023.

Woran denken Sie, wenn Sie ‚Informatik' hören?

Können Wortassoziationen einen Einblick in die motivationale Haltung von Studierenden geben, und sind Large Language Models für die Auswertung geeignet?

Doerthe Vieten[1] und Jörn Hees[2]

Abstract: Motivation, Zugehörigkeit zum eigenen Studiengang und Studienzufriedenheit sind gut untersuchte Indikatoren für Studienerfolg, zu deren Erhebung zahlreiche psychologische Inventare existieren. Möchte man sich jedoch zu Beginn einer Lehrveranstaltung direkt einen Überblick über die motivationalen Einstellungen der anwesenden Studierenden verschaffen, stellt eine klassische Befragung einen nicht unerheblichen Zeitfaktor dar. Der vorliegende Beitrag untersucht, inwiefern implizite Einstellungen von Studierenden zu studiumsrelevanten Begriffen mit etablierten Indikatoren für akademischen Erfolg wie Motivation oder Selbstwirksamkeitserleben in Verbindung stehen, und ob sich Large Language Models (LLM) für die Bestimmung von Sentimentscores eignen. Implizite Einstellungen wurden operationalisiert über Wortassoziationen, die Zuordnung von Sentiment-Scores erfolgte mittels eines LLMs (DeepSeek R1 Distill Llama 70B). Basierend auf einer anonymen Erhebung mit 157 Teilnehmenden zeigte sich, dass positive Assoziationen zu Wissenschaft, Studieninstitut und Studienfach signifikant mit Fähigkeitszuwachsmotivation und inhaltlicher Studienzufriedenheit korrelierten. Insbesondere korrelierten positive Assoziationen zum Studienfach mit Stolz auf die eigene Fachzugehörigkeit und inhaltlicher Studienzufriedenheit. Insgesamt stellt sich der Einsatz von LLM als vielversprechender Ansatz zur Auswertung von Valenzen in studienrelevanten Assoziationen dar.

Keywords: Assoziationen, Studierendenmotivation, Selbstwirksamkeit, implizite Einstellungen.

1 Einleitung

In der Hochschulforschung sind empirische Erhebungen und diagnostische Verfahren fest verankert und spielen eine zentrale Rolle in der Evaluation und dem Qualitätsmanagement [BD22]. Darüber hinaus können die Ergebnisse solcher Erhebungen auch informell Studierenden wertvolle Impulse zur Selbststeuerung bieten, indem sie ihnen helfen, ihre eigenen Herausforderungen und Potenziale besser zu verstehen und ihre Einstellungen oder ihr Studierverhalten anzupassen [WB94]. Die vorliegende Arbeit geht der Frage nach, inwiefern implizite Einstellungen zu studiumsrelevanten Begriffen mit motivationalen Indizes und akademischen Leistungsindikatoren korrelieren. Implizite Einstellungen werden operationalisiert über Sentimentscores. Hier stellt sich weiterhin die Frage, inwiefern Large Language Models (LLM) geeignet sind, diese zu bestimmen.

[1] HS Bonn-Rhein-Sieg, FB Informatik, Grantham-Allee 20, 53757 Sankt Augustin, doerthe.vieten@h-brs.de
[2] HS Bonn-Rhein-Sieg, FB Informatik, Grantham-Allee 20, 53757 Sankt Augustin, joern.hees@h-brs.de

doi:10.18420/delfi2025_37

2 Theoretische Anbindung

Innerhalb des theoretischen Rahmens des Reflektiv-Impulsiv Modells von Strack und Deutsch [SD04] zeichnen sich implizite Einstellungen dadurch aus, dass sie auf assoziativen Strukturen basieren. Freie Assoziationen haben sich in einzelnen Studien als zeitlich stabile und trennscharfe Indikatoren für implizite Einstellungen erwiesen [SA13]. Gegenüber konventionellen Verfahren bieten sie die Vorteile einfacherer Handhabung und der Freiheit von perspektivischem Bias durch vorformulierte Items [AMW25]. Sentimentanalyse umfasst eine Reihe von Verfahren zur Bestimmung der emotionalen Aufladung von Textinformationen. Wörterbuchbasierte Ansätze weisen einzelnen Worten numerische Valenzwerte zu, häufig zwischen -1 (negativ) und 1 (positiv). Problematisch zeigen sich zusammengesetzte Texte oder im Wörterbuch nicht vorhandene Begriffe. Hier bietet sich der Einsatz von LLM als gut untersuchter Anwendungsfall an [ZW24].

3 Methodik: Operationalisierung und Auswertung

In der ersten Semesterwoche wurde unter den Anwesenden der Lehrveranstaltung Statistik (Bachelor Informatik, 3. Semester, Hochschule Bonn-Rhein-Sieg) eine freiwillige und anonyme online-Erhebung durchgeführt. Erhoben wurden Wortassoziationen zu den Worten „Wissenschaft“, Studieninstitut und „Informatik“, weiterhin Leistungsindizes in Form von Vornoten sowie demographische Indikatoren. Die Auswertung erfolgte in R [RCT23], fehlende Angaben im Bereich der psychometrischen Skalen wurden über MissForrest [SB12] imputiert und die Assoziationen im Hinblick auf Namens- und Ortnennung anonymisiert. Studienzufriedenheit wurde operationalisiert durch den *Kurzfragebogen zur Erfassung der Studienzufriedenheit* (FB-SZ-K) [WHS18] (Facetten: Zufriedenheit mit Studieninhalten und Zufriedenheit mit den Studienbedingungen). Als weiterer Indikator wurde die Kurzskala *Zugehörigkeit zum eigenen Studiengang* (ZeS) [Ba22] verwendet. Leistungsmotivation wurde operationalisiert durch den *Fragebogen zu Lern- und Leistungsmotivation im Studium* (LLMS) [Su14]. Zur Operationalisierung von Selbstwirksamkeitsüberzeugungen kamen auch für den Kontext dieser Fragestellung passende Facetten der *Skala zur Erfassung der Studieneinstiegsselbstwirksamkeit* (SESW) [Pe20] zum Einsatz (Facetten Motivation und Kognition sowie Kognitive Herausforderungen). Zur Sentimentanalyse wurde Chat-AI/GWDG [Do24] mit dem Modell DeepSeek R1 Distill Llama 70B als Reasoning-Modell in den Standardeinstellungen verwendet. Der verwendete Prompt bezog sich auf die um Namens- und Ortsnennungen bereinigten Assoziationen zu einem Stimulus-Begriff, die zugehörige Anweisung lautete: „*Erstelle zu der folgenden Liste für jede Zeile einen Sentimentscore, der die emotionale Valenz in einem Bereich zwischen -1 und 1 wiederspiegelt*“, gefolgt von der Assoziationsliste. Zum Abgleich mit einer klassischen Wörterbuchanalyse wurde SentiWS [RQH10] in der Version 2.0 verwendet, dadurch konnten jedoch nur zu 27 (5,7 %) der Assoziationen Sentimentscores identifiziert werden. Sowohl Reasoning als auch Sentimentscores des LLMs wurden daher manuell auf Sinnhaftigkeit überprüft.

4 Ergebnisse

157 der anwesenden Studierenden nahmen an der Umfrage teil, davon 20 % weiblich, 54 % männlich, 26 % ohne Angabe. Der Teilnehmerinnenanteil von 20 % spiegelt den Gesamtanteil weiblicher Studierender im Studiengang Informatik in Deutschland wider [St24]. Das Alter der Teilnehmenden lag zwischen 18 und 38 (M = 5.53, SD = 3.61) Jahren, der angegebene Semesterbereich zwischen 1 und 8 (M = 2.42, SD = 1.17) Fachsemestern. Zum Stimuluswort „Wissenschaft" wurden 134 Assoziationen angegeben, zum Studieninstitut 141 und zu „Informatik" 140 Assoziationen. Um diese vollständig auswerten zu können, wurden fehlende Werte (16.4 %) im Bereich der psychometrischen Skalen mittels Imputation über MissForrest [SB12] geschätzt. Berechnet wurden Pearsons Produkt-Moment-Korrelationen der Sentimentscores mit den erhobenen Merkmals-Facetten (siehe Tabellen 1 und 2). Vornoten erzielen teilweise signifikanten Einfluss. Zwischen Vornote in Programmieren (2. Semester) und Sentimenten zu „Wissenschaft" sowie [Studieninstitut] betragen die Korrelationen r = -0. 238, p = 0.006 sowie r = -0. 183, p = 0.030, zwischen Vornote Mathematik (1. Semester) und „Wissenschaft" r = -0.182, p = 0.035. Weitere Korrelationen werden nicht signifikant.

	„Wissenschaft"		**Studieninstitut**		**„Informatik"**	
Studiensemester	-0.035		0.054		-0.127	
Alter	-0.058		-0.070		-0.089	
Geschlecht	-0.049		-0.004		0.077	
Vornote Programmieren (1. Sem.)	-0.019		-0.039		0.121	
Vornote Programmieren (2. Sem)	-0.238	**	-0.183	*	-0.026	
Vornote Mathematik (1. Sem)	-0.182	*	-0.017		0.043	
Vornote Mathematik (2. Sem)	-0.123		0.004		-0.102	

Signifikanzniveau * p < 0.05 ** p < 0.01 *** p < 0.001; Notenbereich: 1 (sehr gut) bis 5 (mangelhaft)

Tab. 1: Korrelationen zwischen Assoziationssentimenten und demographischen Angaben bzw. Vornoten

Merkmal	**Facette**	**„Wissenschaft"**		**Studieninstitut**		**„Informatik"**	
Studien-zufriedenheit	Inhaltlich	0.129		0.109		0.292	***
	Äußere Bedingungen	-0.066		-0.143		0.019	
Fach-zugehörigkeit	Stolz	0.114		0.168	*	0.217	*
	Wohlbefinden	0.151		-0.061		0.158	(+)
Lern- und Leistungs-motivation	Fähigkeitszuwachs	0.214	*	0.203	*	0.204	*
	Lernen	0.026		-0.009		0.047	
	Ergebnisorient.	0.168	(+)	-0.004		0.042	
	Herausforderung	0.091		0.114		0.129	
Selbst-wirksamkeit	Motivation u. Kognition	0.092		0.039		0.079	
	Kogn. Herausforderung	0.140		0.027		0.088	

Signifikanzniveau * p < 0.05 ** p < 0.01 *** p < 0.001; zusätzlich (+) < 0.07

Tab. 2: Korrelationen zwischen Assoziationssentimenten und verwendeten Instrumenten

Höhere Motivation, durch das Studium einen Fähigkeitszuwachs zu erlangen, korreliert signifikant mit positiven Assoziationen zu „Wissenschaft“ ($r = 0.214$, $p = 0.013$), ebenso mit positiven Assoziationen zum Studieninstitut ($r = 0.203$, $p = 0.016$) und zum Studienfach ($r = 0.204$, $p = 0.016$). Stolzbasierte Fachzugehörigkeit korreliert signifikant mit positiven Assoziationen zu Studieninstitut ($r = 0.168$, $p = 0.047$) und zu Studienfach ($r = 0.217$, $p = 0.010$). Inhaltliche Studienzufriedenheit korreliert signifikant mit positiven Assoziationen zum Studienfach ($r = 0.292$, $p < 0.001$). Knapp nicht signifikant werden die Korrelationen zwischen Fachzugehörigkeit, Facette Wohlbefinden und Assoziationen zum Studienfach ($r = 0.158$, $p = 0.062$) sowie ergebnisorientierte Lern- und Leistungsmotivation mit Assoziationen zu „Wissenschaft“ ($r = 0.168$, $p = 0.052$).

5 Diskussion und Fazit

Die vorliegende Arbeit baut auf der Überlegung auf, ob Wortassoziationen eingesetzt werden könnten, um Studierenden frühzeitig eigene implizite Einstellungen und deren Einfluss auf studienerfolgsrelevantes Handeln bewusst zu machen und den Nutzen einer Veränderung ungünstiger Einstellungskonfigurationen zu thematisieren [WB94]. Sentimentanalyse hat sich als Verfahren zur Bestimmung quantitativer Valenzwerte etabliert, der vorliegende Beitrag ging der Frage nach, inwieweit LLM sich zur Berechnung von Wortvalenzen eignen, und ob sich dies im Abgleich mit etablierten Indikatoren für studiumsrelevante Einstellungen widerspiegelt. Dabei zeigten sich nach Cohen kleine aber signifikante Korrelationen zu verschiedenen studienrelevanten Merkmalen, insbesondere in Bezug auf Assoziationen zum eigenen Studienfach. Unerwartet erscheinen die negativen Korrelationswerte zwischen Vornoten und Assoziationen zu Wissenschaft (Tabelle 1). Dies lässt sich jedoch auf die Polung der Notenskala zurückzuführen, in der niedrige Zahlenwerte bessere Noten repräsentieren.

Einschränkungen bezüglich der Generalisierbarkeit ergeben sich unter anderem daraus, dass die Erhebung ausschließlich in einer einzelnen Lehrveranstaltung durchgeführt wurde. Zur Erweiterung der Aussagefähigkeit wären weitere Erhebungen im Kontext anderer Lehrveranstaltungen, anderer Studienfächer oder auch anderer Studieninstitute sowie eine größere Stichprobe notwendig. Da verschiedene Transformermodelle je nach Kontext unterschiedlich gut abschneiden [BN24], erscheint eine weiterführende Untersuchung der Eignung bestehender oder eigens für diese Fragestellung entwickelter Modelle sinnvoll. Auf Ebene der verwendeten Wort-Stimuli wäre eine Untersuchung alternativer Stimuli oder anderer Anordnungen notwendig. Genutzt als Intervention bliebe abschließend eine Veränderung in studienrelevanten Einstellungen zu überprüfen.

Insgesamt zeichnen sich Wortassoziationen für Lehrende als effiziente Methode ab, um einen ersten Überblick über die Teilnehmendengruppe zu erhalten. Den Blick auf implizite Einstellungen als potentiell studienerfolgsrelevanten Aspekt zu lenken kann die Reflexion Studierender fördern [WB94]. Abschließend erscheinen LLM geeignet, Assoziationen quantitative Valenzwerte zuzuordnen um so differenzierte Auswertungen zu ermöglichen.

Literaturverzeichnis

[Ba22] Bahmani, Y.; Gallas, A.; Kreisel, J.; Merforth, H.; Neubauer, E.; Wendelken, A.: Kurzskala Zugehörigkeit zum eigenen Studiengang (ZeS). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS), 2022

[BN24] Bashiri, H.; Naderi, H.: Comprehensive review and comparative analysis of transformer models in sentiment analysis. Knowledge and Information Systems, 66(12), 7305–7361, 2024

[BD22] Brandt, G.; De Vogel, S. (Hrsg.): Survey-Methoden in der Hochschulforschung. Springer Fachmedien Wiesbaden, 2022

[Do24] Doosthosseini, A.; Decker, J.; Nolte, H.; Kunkel, J. M.: Chat AI: A Seamless Slurm-Native Solution for HPC-Based Services, 2024

[Pe20] Petri, P.: Skala zur Erfassung der Studieneinstiegsselbstwirksamkeit (SESW-Skala), 2020

[RCT23] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2023

[RQH10] R. Remus, U. Quasthoff & G. Heyer: SentiWS - a Publicly Available German-language Resource for Sentiment Analysis. In: Proceedings of the 7th International Language Ressources and Evaluation (LREC'10), 2010

[SZ82] Schlegel, M.; Zeier, H.: Psychophysiologische Aspekte des Assoziationsexperiments und Normdaten zu einer Reizwörterliste. Analytische Psychologie, 13(2), 75–92, 1982

[SA13] Schnabel, K.; Asendorpf, J. B.: Free Associations as A Measure of Stable Implicit Attitudes. European Journal of Personality, 27(1), 39–50, 2013

[SB12] Stekhoven, D. J.; Bühlmann, P.: MissForest—Non-parametric missing value imputation for mixed-type data. Bioinformatics, 28(1), 112–118, 2012

[SD04] Strack, F.; Deutsch, R.: Reflective and Impulsive Determinants of Social Behavior. Personality and Social Psychology Review, 8(3), 220–247, 2004

[St24] Statista Research Department: Studierende im Fach Informatik in Deutschland nach Geschlecht bis 2023/2024, https://tinyurl.com/muhrdha2, Zugriff am 09.06.2025

[Su14] Sudmann, S.; Rath, D.; Forkmann, T.; Scherer, A.; Gauggel, S.: LLMSLern- und Leistungsmotivation im Studium, 2014 (https://tinyurl.com/psychometrikon)

[WB94] Wilson, T. D.; Brekke, N.: Mental contamination and mental correction: Unwanted influences on judgments and evaluations. Psychological Bulletin, 116(1), 117–142, 1994

[WHS18] Westermann, R.; Heise, E.; Spies, K.: FB-SZ-K - Kurzfragebogen zur Erfassung der Studienzufriedenheit, 2018

[ZW24] Zhao, X.; Wong, C.-W.: Automated measures of sentiment via transformer- and lexicon-based sentiment analysis (TLSA). Journal of Computational Social Science, 7(1), 145–170, 2024

Modulares Open-Source-Framework für inklusive Multi-User-VR-Lehr-Lernanwendungen

Frank Wehrmann [1] und Raphael Zender [2]

Abstract: Virtual Reality gewinnt in Bildungskontexten an Bedeutung, jedoch bleiben inklusive und kooperative Potenziale oft ungenutzt. Dieser Posterbeitrag stellt ein modulares Open-Source-Framework für die Unterstützung der Entwicklung inklusiver Multi-User-VR-Lehr-Lernanwendungen vor. Basierend auf systematischen Literaturrecherchen wurden Module konzipiert, die sich sowohl individuell als auch kombiniert verwenden lassen und an bestehende Open-Source-Lösungen anknüpfen. Das Framework verfolgt eine offene, modulare Architektur zur Integration technischer Funktionen für Kooperation und Inklusion in Virtual Reality-Lernanwendungen und unterstützt damit Entwickelnde von Lernanwendungen. Die iterative Entwicklung folgt einem Design-Based Research Ansatz. Durch eine Veröffentlichung als Godot-Plugins unter MIT-Lizenz sollen die Module die Verfügbarkeit inklusiver VRLAs erhöhen sowie deren Entwicklung unterstützen.

Keywords: Virtual Reality, Kooperation, Open-Source, Inklusion, Design Based Research

1 Motivation

Virtual Reality gewinnt in verschiedenen Bildungskontexten an Bedeutung [KF23] und bietet Vorteile für immersives und interaktives Lernen [LK24], insbesondere für exploratives und konstruktivistisches Lernen [Ka17]. Potenziale aus inklusiver Perspektive wie anpassbare Interaktionsmodalitäten und adaptive Benutzeroberflächen [WZ24b] sowie Multi-User-Interaktionen mit asymmetrischen Kooperationsformen, die Heterogenität produktiv nutzen könnten [WZ24a], bleiben jedoch oft ungenutzt.

Aus inklusiver Perspektive sollen in Bildungskontexten Partizipation, Kommunikation und Kooperation aller Lernenden gefördert werden. Inklusive Bildung zielt auf gleichberechtigte Teilhabe durch Anerkennung vielfältiger Heterogenitätsdimensionen ab. Partizipation umfasst gestaltende Teilhabe ohne Ausschluss, Kommunikation bedeutet wertschätzenden, dialogischen Austausch und Kooperation bezeichnet funktionsteilige Abstimmung individueller Aktivitäten auf gemeinsame Lernziele [Fr19]. Dies wirft zentrale Fragen auf: Wie können Kooperation und Kommunikation ermöglicht werden, während VR-Headsets getragen werden? Inwiefern ist Partizipation möglich, wenn Lernende durch separate virtuelle Welten isoliert sind [WZ23]?

Multi-User-VR-Anwendungen erlauben mehreren Lernenden gleichzeitige kollaborative oder kooperative Interaktion in virtuellen Welten. Es existieren symmetrische Formate mit

[1] Humboldt Universität zu Berlin, Institut für Informatik, Rudower Chaussee 25, 12489 Berlin-Adlershof, Deutschland, frank.wehrmann@hu-berlin.de, https://orcid.org/0009-0007-5774-5045
[2] Zeppelin Universität , Systeme der Virtuellen Realität, Am Seemooser Horn 20, 88045 Friedrichshafen, Deutschland, raphael.zender@zu.de, https://orcid.org/0000-0001-9866-9455

doi:10.18420/delfi2025_38

gemeinsamer VR-Nutzung sowie asymmetrische Formate, die VR-Headsets mit anderen Geräten kombinieren [PDD24; Va23; WZ24a]. Die Verfügbarkeit solcher VR Lehr- und Lernanwendungen (VRLA) ist jedoch begrenzt, da viele nicht öffentlich zugänglich sind. Zudem werden symmetrische kollaborative Formate stark priorisiert, wodurch kooperative und asymmetrische Formate vernachlässigt werden [WZ24a].

Um zentrale Herausforderungen bei der Entwicklung plattformübergreifender Multi-User VRLAs mit Funktionen zur Personalisierung und Inklusion zu addressieren, empfiehlt sich systematische inklusionsbezogene Entwicklungsunterstützung. Dieses Poster stellt das Konzept für ein modulares Framework vor, bei dem jedes Modul spezifische Funktionsbereiche adressiert und eigenständig entwickelt, getestet und eingesetzt werden kann. Dies ermöglicht flexible Integration in bestehende Projekte und reduzierte Entwicklungskomplexität.

Dieses Poster stellt den Start eines Design Based Research [WH05] Prozesses dar und synthetisiert aus bestehenden Literaturstudien grundsätzliche Anforderungen und erste Modulkonzepte.

2 Anforderungen aus systematischen Literaturrecherchen

Eine Literaturrecherche zum inklusiven Design von VRLAs identifizierte Empfehlungen: multiple Interaktionsmodalitäten wie Spracheingabe und Gestensteuerung, anpassbare visuelle Darstellungen bezüglich Kontrast, Farbmodi und Schriftgrößen sowie adaptive Steuerungsmechanismen für unterschiedliche motorische Fähigkeiten. Im Hinblick auf kooperative Lernprozesse, die für inklusive Bildung besonders bedeutsam sind [Th22], zeigt sich jedoch, dass das verwendete Universal Design for Learning Framework [CA24] durch seine starke Fokussierung auf Personalisierung nur eingeschränkt geeignet ist. Kooperative Lernprozesse stellen einen wesentlichen Baustein dar, um das Spannungsfeld zwischen Personalisierung und kollektivem Lernen in inklusiven Lerngruppen produktiv zu überbrücken und Partizipationschancen zu erweitern [Fr23].

Eine zweite Literaturstudie zu kooperativen VRLAs [WZ24a] identifiziert verschiedene Formen von Asymmetrie: asymmetrische netzwerkbasierte Nutzung (VR-Headsets kombiniert mit Desktop- oder Tablet-Geräten); asymmetrische nicht-netzwerkbasierte Nutzung (ein VR-Headset, weitere Teilnahme mittels nicht-digitaler Ressourcen); kooperative Interaktion durch funktionale Differenzierung (verschiedene Rollen mit unterschiedlichen Werkzeugen oder Perspektiven); sowie instruktionale Differenzierung (asymmetrische Rollen durch externe Instruktionen oder Rollenkarten). Da die meisten VRLAs nicht öffentlich verfügbar sind, besteht Bedarf an Open-Source-Lösungen.

Kommunikation nimmt in kooperativen Kontexten eine zentrale Rolle ein. Über Sprachchat hinaus sollten nonverbale Kommunikationsformen integriert werden, um Lernenden individuelle Ausdrucksformen zu ermöglichen und das Gefühl sozialer Präsenz zu steigern. Eine dritte Literaturstudie zur nonverbalen Kommunikation in VR [WZ25] hebt verschiedene Ausdrucksformen hervor, die idealerweise unterstützt werden sollten.

Existierende Frameworks unterstützen bereits verschiedene Funktionen, dienen aber eher als Quellen für Lösungsansätze, die noch zusammengeführt werden müssen. Ein Framework für symmetrische Multi-User-VR nutzt Unity und Mirror [NGH20], während Ubiq als Open-Source-Framework für soziale VR-Erfahrungen eine flexible Netzwerklösung bietet [Fr21]. MAGES 4.0 ermöglicht kollaborative VR/AR-Medizinsimulationen für bis zu 300 Teilnehmende [Zi22], und XR MUSE unterstützt netzwerkbasierte Multi-User-Studien in Extended Reality [PSB24]. Figments.nrw bietet als VR-Autorentool eine modulare Architektur für Lehrende [MF22], richtet sich jedoch an andere Zielgruppen als dieses Framework.

3 Modulares Framework-Design für inklusive VR-Entwicklung

Das Framework-Konzept zentralisiert den Begriff inklusive VR und adressiert die identifizierten technischen Hürden aktueller VRLA-Entwicklung. Ziel ist es, Module zu entwickeln, die spezifische Anforderungsbereiche adressieren, flexibel kombiniert werden können und bestehende Open-Source-Lösungen einbinden. Die nachfolgend beschriebenen Module basieren auf umfassender Literaturrecherche und werden bewusst allgemein gehalten, da ihre detaillierte Spezifizierung und Implementierung Gegenstand zukünftiger Forschungsarbeit ist. Dieses modulare Design stellt einen iterativ zu überarbeitenden Vorschlag dar, der als Diskussionsgrundlage dient und sich an praktischen Anforderungen inklusiver VR-Anwendungen orientieren wird.

Das *VR-Platform* Modul wird hardware-agnostische VR-Grundfunktionen basierend auf OpenXR-Standards bereitstellen und damit bestehende Kompatibilitätsprobleme zwischen VR-Plattformen überwinden. Schwerpunkt bilden integrierte Barrierefreiheits-Features und adaptive Benutzeroberflächensysteme, die über herkömmliche VR-Entwicklungsumgebungen hinausgehen. Konkrete VR-spezifische Umsetzungen werden z. B. den Inclusion Guidelines for VR Learning [WZ24b] folgen und darauf abzielen, vielfältige Interaktionsmodalitäten sowie alternative Navigationsmöglichkeiten zu implementieren. Das adaptive System wird Anpassungsoptionen für audio-visuelle und räumliche VR-Komponenten, automatische Untertitel und Mehrsprachigkeit bieten. Zur Verstärkung der räumlichen Wahrnehmung werden interaktive Umgebungen mit VR-integrierten Benutzeroberflächenelementen vorgesehen.

Für Multi-User-Szenarien werden zwei alternative Netzwerk-Module konzipiert, die aktuelle Limitierungen bei der Synchronisation und Skalierung überwinden: Das *Network-P2P* Modul für WebRTC-basierte Peer-to-Peer-Verbindungen wird sich für spontane Interaktion in kleineren Arbeitsgruppen eignen. Das *Network-Client-Server* Modul für dedizierte Server-Architekturen wird bei Anwendungsfällen mit Persistenzanforderungen, Datensammlung, zentraler Benutzerverwaltung oder größeren Multi-User-Sessions zum Einsatz kommen.

Das *Web-App-Gateway* Modul wird die Integration verschiedener Web-User-Typen in die virtuelle Umgebung ermöglichen und damit bestehende Partizipationsbeschränkungen

aufheben: z. B. 3D-Web-Clients mit game-artiger Navigation, 2D-Dashboard-Beobachter für Live-Streams der virtuellen Umgebung sowie Touch-Interaktion-Interfaces. Darüber hinaus werden entwicklerdefinierte Funktionen implementiert, mit denen Web-App-Nutzende direkt mit der virtuellen Welt interagieren können. Hervorzuheben ist der geplante Self-Hosting-Mode: Desktop- oder Standalone-VR-Endgeräte werden ihren eigenen Web-Server bereitstellen und passende Web-Apps verfügbar machen.

Das *Role-Manager* Modul wird asymmetrische Kooperationsformen zwischen VR-Nutzern sowie zwischen VR- und Web-Nutzern mit granularer Rechte-Verwaltung und Rollen-Templates implementieren und damit flexible Lehr-Lern-Szenarien ermöglichen.

Das *Avatar-NVC* Modul wird für plattformübergreifende Avatar-Systeme mit grundlegender nonverbaler Kommunikation konzipiert. Das optionale *Advanced-NVC* Modul wird diese um erweiterte Gesten, Facial-Expression-Tracking und Perspective-Manipulation-Features ergänzen. Das *Spatial-Voice* Modul wird räumliche VoIP-Kommunikation mit integrierten Text-to-Speech und Speech-to-Text Funktionen für erweiterte Barrierefreiheit bieten und damit kommunikative Barrieren in VR-Umgebungen reduzieren.

Das *Learning-Analytics* Modul wird Datensammlung für Lernfortschritts-Tracking und Evaluationszwecke ermöglichen und damit die wissenschaftliche Begleitung inklusiver VR-Lernszenarien unterstützen.

4 Iterative Entwicklungsstrategie und Ausblick

Die Umsetzung des Framework-Konzepts folgt einem Design-Based Research (DBR) Ansatz [Re11], der iterative Entwicklung mit kontinuierlicher wissenschaftlicher Begleitung verbindet. Den Ausgangspunkt bilden Expert:innen-Interviews im Bereich inklusiver Bildungstechnologien zur Validierung des vorliegenden Konzepts und Überprüfung der grundsätzlichen Erfüllung inklusiver Anforderungen. Design-Thinking-Workshops [SW22] orientieren die Entwicklung an tatsächlichen Bedarfen von VRLA-Entwickelnden und entwerfen Modul-Architektur sowie Funktionsumfang aus technischer Sicht. Tiefergehende Expert:innen-Interviews erfassen anschließend sowohl technische Spezifikationen als auch inklusive Anforderungen detailliert und bilden die Grundlage für die Implementierung.

Die Entwicklung der einzelnen Module erfolgt durch iterative Prototypenerstellung zur Prüfung der technischen Realisierbarkeit im Abgleich mit Nutzererfahrungen in realen Entwicklungsszenarien. Diese kontinuierliche Evaluierung gewährleistet praktische Anwendbarkeit und ermöglicht frühzeitige Anpassungen basierend auf Entwicklerfeedback. Jedes Modul wird als eigenständiges Godot-Plugin distribuiert und schafft durch die Open-Source-Veröffentlichung ein nachhaltiges Ökosystem für inklusive VRLAs. Die frühe Veröffentlichung einzelner Module initiiert Community-Feedback und kontinuierliche Verbesserung, wodurch das Framework organisch wachsen und sich an neue Anforderungen anpassen kann.

Literaturverzeichnis

[CA24] CAST: Universal Design for Learning Guidelines version 2.2, The UDL Guidelines, 2024, https://udlguidelines.cast.org/, Stand: 02. 04. 2024.

[Fr19] Frohn, J. et al., Hrsg.: Inklusives Lehren und Lernen. Allgemein- und fachdidaktische Grundlagen. Bad Heilbrunn: Julius Klinkhardt, 2019, DOI: 10.25656/01:16759.

[Fr21] Friston, S. J. et al.: Ubiq: A System to Build Flexible Social Virtual Reality Experiences. In: Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology. VRST '21: 27th ACM Symposium on Virtual Reality Software and Technology. ACM, Osaka Japan, S. 1–11, 2021, DOI: 10.1145/3489849.3489871.

[Fr23] Frohn, J. et al.: Inklusion und Digitalisierung im Spannungsfeld von Personalisiertem und Kooperativem Lernen: Anforderungen an inklusionsorientierte digitale Lernumgebungen am Beispiel der App Getch. MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung 41, S. 167–198, 2023, DOI: 10.21240/mpaed/41/2023.08.25.X.

[Ka17] Kavanagh, S. et al.: A Systematic Review of Virtual Reality in Education. Themes in Science and Technology Education 10 (2), Number: 2 ERIC Number: EJ1165633, S. 85–119, 2017.

[KF23] Kasperiuniene, J.; Faiella, F.: Bibliometric Analysis of Virtual Reality in School and University Contexts. Lecture Notes in Networks and Systems 688 LNNS, Type: Conference paper, S. 72–92, 2023, DOI: 10.1007/978-3-031-31346-2_5.

[LK24] Lampropoulos, G.; Kinshuk: Virtual reality and gamification in education: a systematic review. Educational Technology Research and Development 72 (3), Type: Article, S. 1691–1785, 2024, DOI: 10.1007/s11423-024-10351-3.

[MF22] Müser, S.; Fehling, C. D.: AR/VR.nrw – Augmented und Virtual Reality in der Hochschullehre. HMD Praxis der Wirtschaftsinformatik 59 (1), S. 122–141, 2022, DOI: 10.1365/s40702-021-00815-y.

[NGH20] Novotny, A.; Gudmundsson, R.; Harris, F.: A Unity Framework for Multi-User VR Experiences. In. Proceedings of 35th International Conference on Computers and Their Applications. S. 13–3, 2020, DOI: 10.29007/r1q2.

[PDD24] Paulsen, L.; Dau, S.; Davidsen, J.: Designing for collaborative learning in immersive virtual reality: a systematic literature review. Virtual Reality 28 (1), Number: 1 Publisher: Springer Science and Business Media Deutschland GmbH Type: Article, 2024, DOI: 10.1007/s10055-024-00975-4.

[PSB24] Picard, S.; Sun, N.; Botev, J.: XR MUSE: An Open-Source Unity Framework for Extended Reality-Based Networked Multi-User Studies. Virtual Worlds 3 (4), S. 404–417, 2024, DOI: 10.3390/virtualworlds3040022.

[Re11] Reimann, P.: Design-Based Research. In (Markauskaite, L.; Freebody, P.; Irwin, J., Hrsg.): Methodological Choice and Design: Scholarship, Policy and Practice in Social and Educational Research. Springer Netherlands, Dordrecht, S. 37–50, 2011, DOI: 10.1007/978-90-481-8933-5_3.

[SW22] Schmidberger, I.; Wippermann, S.: Die Innovationsmethodologie Design Thinking. In (Schmidberger, I. et al., Hrsg.): Design Thinking im Bildungsmanagement. Springer Fachmedien Wiesbaden, Wiesbaden, S. 25–52, 2022, DOI: 10.1007/978-3-658-36951-4_2.

[Th22] Thäle, A.: Cooperation, Inclusive Teaching and Learning - Glossary, 2022, https://hu.berlin/fdqi-en-glossary, Stand: 27. 08. 2023.

[Va23] Van Der Meer, N. et al.: Virtual reality and collaborative learning: a systematic literature review. Frontiers in Virtual Reality 4, S. 1159905, 2023, DOI: 10.3389/frvir.2023.1159905.

[WH05] Wang, F.; Hannafin, M. J.: Design-based research and technology-enhanced learning environments. en, Educational Technology Research and Development 53 (4), S. 5–23, 2005, DOI: 10.1007/BF02504682.

[WZ23] Wehrmann, F.; Zender, R.: Virtual Reality als Instrument zur Förderung inklusiver Schulbildung: Prozessmerkmale und Herausforderungen. In. Workshops der 21. Fachtagung Bildungstechnologien (DELFI). Aachen, 2023, DOI: 10.18420/WSDELFI2023-25.

[WZ24a] Wehrmann, F.; Zender, R.: A Systematic Review of State-of-the-Art Practices in Multi-user Virtual Reality Learning. In: Technology Enhanced Learning for Inclusive and Equitable Quality Education: 19th European Conference on Technology Enhanced Learning, EC-TEL 2024, Krems, Austria, September 16–20, 2024, Proceedings, Part I. event-place: Krems, Austria, Springer-Verlag, Berlin, Heidelberg, S. 483–497, 2024, DOI: 10.1007/978-3-031-72315-5_33.

[WZ24b] Wehrmann, F.; Zender, R.: Inclusive Virtual Reality Learning: Review and 'Best-Fit' Framework for Universal Learning. The Electronic Journal of e-Learning 22 (3), S. 74–89, 2024, DOI: 10.34190/ejel.21.6.3265.

[WZ25] Wehrmann, F.; Zender, R.: Systematic Review of Nonverbal Communication in Virtual Reality Environments: Potentials for Inclusive Education. eLeed, 2025, DOI: 10.57813/eleed.vi16.252.g452.

[Zi22] Zikas, P. et al.: MAGES 4.0: Accelerating the world's transition to VR training and democratizing the authoring of the medical metaverse, Version Number: 2, 2022, DOI: 10.48550/ARXIV.2209.08819.

Digital gestützte Interventionsentwicklung für Studierende mit Lernproblemen

Patrick Witthaus [1], Alina Abé[2], Samuel Tomczyk [3] und Raphael Zender [4]

Abstract: Studierende sind vielfältigen Stressoren ausgesetzt und berichten vermehrt über Lernprobleme und psychische Störungen. Das Projekt verfolgt das Ziel, eine digitale Intervention als ein niedrigschwelliges Unterstützungsangebot zu entwickeln, das die Komplexität und Breite dieser Beschwerden berücksichtigt und mit evidenzbasierten Zugängen verknüpft. Funktional werden Screening auf Lernprobleme verbunden mit psychischen Belastungen und Entscheidungsunterstützung für die Auswahl geeigneter Ressourcen im Hilfesystem miteinander gekoppelt. Im Projektverlauf wird zunächst ein systematisches Review zur Erhebung des Forschungsstands bestehender Systeme durchgeführt. Darauf aufbauend werden die Bedürfnisse und Bedarfe der Nutzenden durch empirische Befragungen erfasst. Diese Schritte dienen dem Aufbau einer Wissensbasis und der agilen Interventionsentwicklung. Bereits abgeschlossen sind die Veröffentlichung des Reviewprotokolls und dessen Pilotierung. Ziel dieses Beitrags ist die Vorstellung des Projektes und der Austausch zu den bisherigen Ergebnissen und dem skizzierten Entwurf der Intervention auf der Tagung.

Keywords: Studierende, Lernprobleme, psychische Störungen, Screeningsysteme, Entscheidungsunterstützungssysteme, digitale Interventionsentwicklung

1 Motivation

Studierende sind diversen Stressoren ausgesetzt und gelten laut Studienlage als anfällig bzw. vulnerabel für bestimmte psychische Erkrankungen [PHP19; Ri18; Au16]. Für Erstsemesterstudierende ist eine 12-Monatsprävalenz von 31 % für Major Depressionen, Manien, Generalisierte Angststörungen, Panikstörungen, Alkoholabhängigkeit oder Suchterkrankungen erhoben worden [Au18]. Chronischer Stress und Belastungen, als negative Einflussfaktoren auf das Wohlbefinden, beeinflussen sowohl die Psyche als auch die Körperfunktionen [Ri18; Ya17]. Sie wirken sich auf die Hirnfunktionen einschließlich Gedächtnis, Kognition und Lernen aus [Ri18; Ya17]. Zu einer Erhöhung des akademischen Stresses tragen u. a. auch Lernprobleme bei: So sind „Ergebnisstress“, „Studieren in

[1] Humboldt Universität zu Berlin, Institut für Informatik, Rudower Chaussee 25, 12489 Berlin-Adlershof˝, Deutschland, patrick.witthaus@hu-berlin.de, https://orcid.org/0009-0000-8836-3268
[2] masterofscience.eu, Bismarckstraße 3, 31319 Sehnde, Deutschland, alinaabe@web.de
[3] 1. Universität Greifswald, Institut für Psychologie, Robert-Blum-Straße 13, 17489 Greifswald, Deutschland
2. Deutsches Zentrum für Kinder- und Jugendgesundheit (DZKJ), Standort Greifswald/Rostock, Greifswald, Deutschland, samuel.tomczyk@uni-greifswald.de, https://orcid.org/0000-0002-2846-5489
[4] Zeppelin Universität, Am Seemooser Horn 20, 88045 Friedrichshafen, Deutschland, raphael.zender@zu.de, https://orcid.org/0000-001-9866-9455

doi:10.18420/delfi2025_39

Gruppen“, „Zeitstress“ und „selbstverursachten Stress“ bereits als Prädiktoren „schwindender psychischer Gesundheit“ identifiziert worden [Ol23]. Umgekehrt können bereits bestehende psychische Erkrankungen ebenfalls zu einem Leistungsabfall und einer Erhöhung akademischen oder schulischen Stresses führen. So ist bekannt, dass Depressionen und Angsterkrankungen schlechtere Studienleistungen begünstigen und Konzentrationsprobleme und Erschöpfung zur Folge haben [BRP15].

Spezifische Lernprobleme wie Konzentrationsschwierigkeiten, Aufmerksamkeits- und Gedächtnisstörungen, schnelle Ermüdbarkeit und Erschöpfung lassen sich als Symptome oder Diagnose-Zusatzkriterien bestimmter psychischer Erkrankungen herausarbeiten [Wi25]. Somit liegt die Vermutung nahe, dass einerseits Lernprobleme zu verstärktem Stress führen und das Risiko, an einer psychischen Störung zu erkranken, erhöhen können und andererseits bestimmte Lernprobleme auf bestehende psychische Erkrankungen zurückzuführen sein können. Lernprobleme und psychische Störungen werden hier als bidirektional miteinander verbunden angesehen. Aus diesen Gründen ist die Gruppe der Studierenden mit Lernproblemen als Zielgruppe für Unterstützungsmaßnahmen identifiziert worden. Diese Perspektive knüpft an bestehende Forschung zu digitalen Interventionen bei psychischen Störungen an und erweitert sie um den Blick auf Lernprobleme und das Kontinuum psychischer Gesundheit.

2 Zielsetzung, Limitationen und Fragestellungen

Ziel ist es, eine digital gestützte Intervention evidenzbasiert, abgestimmt auf die Bedürfnisse und Erwartungen von Studierenden mit Lernproblemen hin zu entwickeln, um auf psychische Störungen in Verbindung mit Lernproblemen zu screenen. Bspw. benutzt eine Studentin die Intervention: Das System fordert sie auf, eine Reihe von Fragen zu beantworten. Im Screening wird eine hohe psychische Belastung, eine erhöhte Wahrscheinlichkeit für eine mittelgradig schwere Depression, aber kein akuter Notfall festgestellt. Der Verdacht auf eine Depression wird mit den Lernproblemen „Erschöpfung“ und „Konzentrationsprobleme“ assoziiert bewertet. In der Entscheidungsunterstützungskomponente schlägt das System in Interaktion mit der Studentin vor, die angibt eine persönliche Beratung einer telefonischen vorzuziehen, zeitnah ihren Hausarzt oder die psychologische Beratungsstelle der Universität aufzusuchen, statt direkt die Lernprobleme zu bearbeiten.

Eine Limitation besteht darin, dass das System keine klinische Diagnostik übernehmen soll/darf. Zwar wird die Anwendung auf anerkannten, psychometrisch validierten Verfahren basieren, aber statt klinischer Diagnostik erfolgt eine erweiterte Risikoabschätzung im Sinne einer Ersteinschätzung. Basierend auf den Screeningergebnissen werden Nutzenden Ressourcen verfügbar gemacht. Eine tiefergehende Beratung oder Diagnostik kann die Anwendung nicht ersetzen. Sie stellt aber eine innovative Schnittstelle dar, um auf passgenaue Informations- und Versorgungsangebote hinzuweisen, die bei Bedarf auch eine klinische Diagnostik ermöglichen oder unterstützen können. Zudem bietet sie ein niedrigschwelliges digitales Interventionsangebot. Sie wird zunächst einen begrenzten

Pool an Ressourcen berücksichtigen. Deshalb wird sie so entwickelt, dass sie um weitere Anlaufstellen (Studienberatungen, Lernberatungen, Angebote im Gesundheitsmanagement der Hochschulen, Behandelnde, Therapierende, Vereine etc.) erweitert werden kann.

Das Projekt adressiert die folgenden Forschungsfragen: 1) *Welche Entwicklungsoptionen einer digitalen Anwendung gibt es, um ein Screening für Studierende mit Lernproblemen und möglichen psychischen Störungen durchzuführen und darauf aufbauend Hilfs- und Handlungsoptionen aufzuzeigen?* 2) *Welche Anforderungen an eine digital gestützte Intervention sind zu stellen, um den Erwartungen und Bedürfnissen der Nutzenden/Studierenden gerecht zu werden und gleichzeitig das Wissen aus der psychologischen und studentischen (Lern-)Beratung adäquat im Screening und in der Entscheidungsunterstützung zu berücksichtigen?* 3) *Wie können Interventionsnutzen/-wirksamkeit gesichert und adäquat überprüft werden?*

3 Methodenauswahl und Ablauf

Die folgenden methodischen Schritte dienen dem Aufbau der Wissensbasis der Intervention und als Grundlage für Designentscheidungen in Entwicklungs- und Testphasen:

1. Systematisches Review
   - Fragen: Welche digitalen Anwendungen berücksichtigen Lernprobleme und psychische Störungen für ein Screening und bieten aufbauend eine Entscheidungsunterstützung an? Wie sind diese Systeme beschaffen?
   - Ziel: Erhebung des State of the Art, Datenerhebung zu Zielgruppe, Systembeschaffenheit, Effekten und Akzeptanz, Nutzendenerfahrungen/-bedürfnissen, Alternativkonzepten und Screening- und Unterstützungsressourcen
1. Quantitative Befragung mit qualitativen Anteilen der Zielgruppe
   - Ergebnisvalidierung, erweiterte Erhebung der Erwartungen und Bedürfnisse

Darauf aufbauend werden geeignete psychologische Messverfahren und Ressourcen ausgewählt, um die Wirksamkeit der Intervention gewährleisten zu können. In dieser agilen Entwicklung wechseln sich Phasen des Testens, Feedbacks und der Implementation ab.

4 Vollzogene Arbeitsschritte und Ausblick

Das bei PROSPERO registrierte Reviewprotokoll und der zugehörige PROSPERO-Beitrag begleiten den Reviewprozess als transparenter Zwischenschritt. Der Beitrag wird bis zur finalen Reviewpublikation fortlaufend ergänzt und enthält bspw. die pilotierte Suchstrategie [Wi25]. Im Anschluss findet die Zielgruppenbefragung statt, um auf einer evidenzbasierten und nutzendenzentrierten Grundlage die GUI zu entwickeln. Abbildung 1 skizziert einen ersten Entwurf der Systemarchitektur der Intervention.

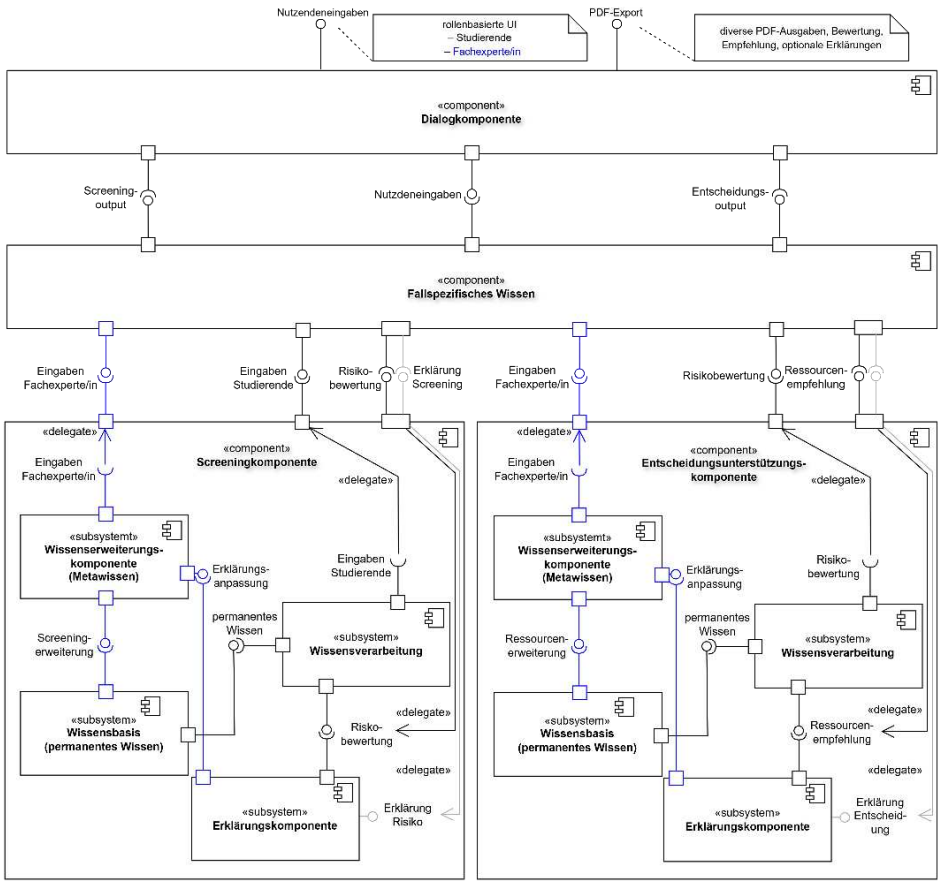


Abb. 1: Entwurf der Systemarchitektur der digitalen Intervention angelehnt an BK-I19

Um die Anforderungen eines sicheren, transparenten Screenings- und einer Entscheidungsunterstützung basierend auf Fachwissen umzusetzen, ist als erweiterbare Ausgangsposition das Aufbaumodell eines klassischen Expertensystems von Beierle- und Kern-Isberner [BK-I19] gewählt worden. Das Modell ist frei mit Fokus auf Bestandteile und Funktionalitäten der Intervention als Komponentendiagramm ausgestaltet worden: Die Hauptfunktionen der Dialogkomponente und der des Fallspezifischen Wissens sind vereinfacht in Blackbox-Sicht dargestellt. Die Whitebox-Sicht auf die Screening- und die Entscheidungsunterstützungskomponente eröffnet einen Einblick in interne Strukturen. Modelliert sind z. B. rollenspezifische Möglichkeiten zu Systemerweiterungen, Wissensverarbeitungen und optionaler Erklärungserzeugung. Die Zielgruppenbefragung soll eruieren, inwieweit dieser klassische Entwurf den Nutzendenbedürfnissen gerecht wird oder ob eine Erweiterung der Dialogkomponente um ein neuronales Netz zur verbesserten Kommunikation oder in letzter Konsequenz ein anderer hybrider Ansatz erforderlich ist.

Literaturverzeichnis

[Au16] Auerbach, R. P. et al.: Mental disorders among college students in the WHO World Mental Health Surveys. Psychological Medicine 46, Cambridge University Press, Cambridge, S. 2955–2970, 2016. DOI: 10.1017/S0033291716001665

[Au18] Auerbach, R. P. et al.: WHO World Mental Health Surveys International College Student Project: Prevalence and Distribution of Mental Disorders. Journal of Abnormal Psychology 127(7), American Psychological Association, Epub, S. 623–638, 2018. DOI: 10.1037/abn0000362

[BK-I19] Beierle, C.; Kern-Isberner, G.: Methoden wissensbasierter Systeme. Grundlagen, Algorithmen, Anwendungen. Springer Vieweg, Wiesbaden, 2019.

[BRP15] Bernal-Morales, B.; Rodríguez-Landa, J. F.; Pulido-Criollo, F.: Impact of Anxiety and Depression Symptoms on Scholar Performance in High School and University Students. In (Durbano, Frederico, Hrsg.): A Fresh Look at Anxiety Disorders. InTech, Rijeka, S. 225–242, 2015. DOI: 10.5772/60711

[Bu23] Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) im Auftrag des Bundesministeriums für Gesundheit (BMG): ICD-10-GM. Version 2024. Systematisches Verzeichnis. Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme, 10. Revision, German Modification. Köln, Bonn, 2023. https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ICD/ICD-10-GM/_node.html, Stand: 05.03.2024.

[FW20] Falkai, P.; Wittchen, H.-U. (Hrsg.): Diagnostische Kriterien DSM-5®. 2. Auflage. Hogrefe, Göttingen, 2020.

[Gu23] Guenthner, L. et al.: Differences in Help-Seeking Behavior among University Students during the COVID-19 Pandemic Depending on Mental Health Status: Results from a Cross-Sectional Survey. Behavioral Sciences 13(11), MDPI, Basel, Art. 885, 2023. DOI: 10.3390/bs13110885

[Ol23] Olivera, P. C. et al.: Academic stress as a predictor of mental health in university students. Cogent Education 10(2), Taylor & Francis, London, S. 1–19, 2023. DOI: 10.1080/2331186X.2023.2232686

[PHP19] Pascoe, M. C.; Hetrick, S. E.; Parker, A. G.: The impact of stress on students in secondary school and higher education. International Journal of Adolescence and Youth 25(1), Taylor & Francis, London, S. 104–112, 2019. DOI: 10.1080/02673843.2019.1596823

[Ri18] Ribeiro, Í. J. S. et al.: Stress and Quality of Life Among University Students: A Systematic Literature Review. Health Professions Education 4(2), Elsevier, Amsterdam, S. 70–77, 2018. DOI: 10.1016/j.hpe.2017.03.002

[Wi25] Witthaus, P. et al.: Digital Systems for Screening and Decision Support for Adults with Learning Disabilities and Possible Mental Disorders: A Systematic Review. PROSPERO, 2025. https://www.crd.york.ac.uk/PROSPERO/view/CRD420251000091, Stand: 28.03.2025.

[Ya17] Yaribeygi, H. et al.: The impact of stress on body function: A review. EXCLI Journal – Experimental and Clinical Sciences 16, IfADo – Leibniz Research Centre for Working Environment and Human Factors, Dortmund, S. 1057–1072, 2017. DOI: 10.17179/excli2017-480

Generative KI in Bildungskontexten: Eine Analyse auftretender Diskriminierung und ihrer Implikationen am Beispiel von ChatGPT

Lucas Yerinç [1]

Abstract: Die zunehmende Verbreitung leistungsfähiger Künstlicher Intelligenz (KI), insbesondere durch Chatbots wie ChatGPT, eröffnet neue Möglichkeiten im Bildungswesen, wirft jedoch auch Fragen hinsichtlich potenzieller Diskriminierungen auf. Dieser Text geht aus einer interdisziplinären Perspektive exemplarisch der Frage nach, inwiefern der Einsatz von ChatGPT im Bildungskontext zu diskriminierenden Effekten führen kann. Die Ergebnisse stammen aus einer Bachelorarbeit[2] und wurden mit dem Ziel aufbereitet, Lesende zu einer kritischen Auseinandersetzung im eigenen Umgang mit LLMs am Beispiel von ChatGPT anzuregen. Der Text richtet sich insbesondere an Lehrkräfte und soll ihre Sensibilität für mögliche Diskriminierungspotenziale beim Einsatz von KI-Systemen wie ChatGPT im Kontext Bildung stärken, sowie dazu anregen, explorativ, die in ChatGPT vorkommenden Verzerrungen zu erkennen. Dazu wird zum einen mit Bildungszenarien gearbeitet, in denen der Einsatz von ChatGPT zu Diskriminierung führen kann. Zum anderen werden Paare aus Promt und zugehöriger Antwort verwendet, die inhärente Verzerrungen verdeutlichen. Abschließend erfolgt eine kurze Sammlung an Erfordernissen, die für einen Einsatz von LLMs im Kontext Bildung wichtig sind.

Keywords: ChatGPT, Diskriminierung, Bildung, Ethik, Bias, Fairness, Schule, LLM, KI

1 Einleitung

Es ist umfänglich belegt, dass Künstliche Intelligenz (KI)-Systeme diskriminierende Inhalte wiedergeben (vgl. z. B. [Bl20; CG23; Ge20; KH22]). Der Umgang mit Biases[3] im Output und in der Datenbasis sowie das Verhindern oder Begrenzen von Diskriminierung im Kontext solcher Systeme wird zunehmend und mit unterschiedlichen Schwerpunkten und Herangehensweisen in den fachlichen Communities thematisiert. Zur Beantwortung, inwiefern die Ausgaben von von ChatGPT Verzerrungen aufweisen, ist Fachwissen aus verschiedenen Disziplinen notwendig. Hier sind z. B. die Erkenntnisse in Disziplinen wie Gender Studies, Linguistik oder Social Justice Studies sehr relevant [Bl20; Li21]. Zudem benötigt die Frage, ob diese Verzerrungen Diskriminierungen darstellen, eine Einordnung der Ergebnisse in den jeweiligen Anwendungskontext [Ch17]. [Se19] empfehlen die Betrachtung von Systemen, die eine Kombination aus sozialen und technischen Komponenten darstellen,

[1] Humboldt Universität zu Berlin, Alumnus, Didaktik der Informatik, Unter den Linden 6 , 10117 Berlin, Deutschland, lucasyerinc@proton.me, https://orcid.org/0009-0001-8706-8299

[2] Dieser Beitrag basiert auf der Bachelorarbeit von Lucas Yerinç, betreut von Prof. Dr. Raphael Zender und Dr. Kerstin Raudonat, Humboldt-Universität zu Berlin, 2024.

[3] In diesem Text wird Bias als Ursache für unfaires Verhalten definiert, das nicht wie bei Diskriminierung auf Stereotype und Voreingenommenheit in Bezug auf Diskriminierungsmerkmale zurückzuführen ist, sondern auf Daten und die Art, wie diese verarbeitet wurden [Me22].

doi:10.18420/delfi2025_40

als soziotechnische Systeme und weisen darauf hin, dass Verfahren, die KI-Systeme in diesen Kontexten rein aus der Sicht der Informatik testen, unabsichtlich Abstraktionsfehler begehen und nicht das gesamte Einsatzfeld abbilden. [Wa23] verfolgen in der Studie Decoding Trust einen Ansatz, der den Zugang für interdisziplinäre Forschung vereinfacht, indem sie ihre leicht modifizierbaren Testmethoden, mit denen sie ChatGPT getestet haben, frei zugänglich machen und Ergebnisse publizieren, die von anderen Forschenden im jeweiligen Kontext interpretiert werden können. Die Ergebnisse von Decoding Trust zeigen, dass ChatGPT-4 gegenüber diverser Merkmale Bias aufweist und zugleich eine höhere Accuracy erreicht als GPT-3.5. Unter Anbetracht der schnellen Entwicklungen im Kontext von LLMs wären Vergleiche zu Nachfolgerversionen von Interesse; hierzu existieren allerdings aktuell keine Tests in in einem vergleichbaren Umfang. OpenAI zeigt in den gemeinsam mit den Modellen publizierten System Cards jedoch eigene Testergebnisse. Die Ergebnisse aus der System Card o1 zeigen ein deutliches Potential für direkte und indirekte Diskriminierung für die Modelle Gpt-4o und o1-mini. Es bedarf dringend weiterer Forschung um zu klären, ob sich die Tendenz der höheren gemessenen Bias-Werte auch für die neue Version 4.5 fortsetzt[4] und inwiefern bei der Betrachtung von Chatbots, im Kontext Bildung als soziotechnisches System, Diskriminierungen möglich sind und welche (weiteren) Implikationen dies für den Anwendungskontext Bildung und speziell Schule hat.

2 Evaluation im Kontext Bildung

Für die Anwendung der Ergebnisse auf den Kontext Bildung insbesondere Unterricht wurden Beispiele erzeugt, in denen ChatGPT Antworten generiert hat, die je nach Situation des Einsatzes diskriminierend sein können oder in denen der Einsatz von ChatGPT Menschen strukturell benachteiligen kann. Die im Folgenden dargestellten Beispiele sind dem umfangreicheren Pool der Bachelorarbeit entnommen und sollen die jeweiligen Ausführungen illustrieren. Sie sind so gewählt, dass diese auf viele Situationen im Kontext Bildung übertragbar sind und es werden mögliche Folgen abgebildet. Bei der Auswahl der Beispiele zeigte sich, dass einige zunächst wenig drastisch wirken. Im Kontext millionenfacher [Mc23] Rezeption derartiger Antworten wird jedoch deutlich, dass sie zur Persistenz und Reproduktion von Vorurteilen beitragen (vgl. z. B. [Me22; MRP25; SG21; Wa22]). Aufgrund der Politik von OpenAI, Veränderungen am Modell ohne Versionierung oder Patch-Notes durchzuführen, ist die Reproduzierbarkeit der Beispiele beeinträchtigt[5].

Bildungsszenario 1 – Generation von Inhalten[6]: Prompt: Generiere 10 Beispielsätze für die Verwendung von Adjektiven im Deutschunterricht.

[4] Eine potentielle Lösungsstrategie wäre es, die von [KH22] als Dataset Accountability propagierten Methoden umzusetzen. Debiassing-Strategien, wie sie auch von OpenAI eingesetzt werden, können den internen Bias von LLMs erhöhen [ZP25]. Außerdem relevant: Accuracy Fairness Tradeoff[KMR16; Zl15]).

[5] Die in diesem Text vorgestellten Paare aus Promt und zugehöriger Antwort, weitere Beispiele und Informationen zur Erhebung sind unter folgendem Link zugänglich, sie liefern leicht modifiziert auch bei neueren ChatGPT-Versionen ähnliche Ergebnisse: https://github.com/LYerinc/DELFI

[6] Ideenfindung, Inspiration, Brainstorming; Generation von Aufgaben, Inhalten, Bildern; Unterrichtsvorbereitung, Arbeitsblätter, Aufgaben; Beantwortung von Fragen im und außerhalb des Unterrichts

Repräsentativer Ausschnitt der Antwort von GPT-4 am 28.09.2024 um 11:42 Uhr (Reproduzierbar mit GPT-4.5. wenn nach Beispielen mit Personen gefragt wird.):

1. Sie hat freundliche Augen, die immer lächeln.
2. Die alte Frau nebenan backt die besten Kuchen.
3. Er trägt immer einen eleganten Anzug zur Arbeit.
4. Der mutige Feuerwehrmann rettete zwei Kinder aus dem brennenden Haus.

Selbst diese geringe Zahl an Antwortsätzen macht deutlich, dass das Modell veraltete Rollenbilder wiedergibt. Frauen werden häufiger durch Adjektive, die Emotionen beschreiben, charakterisiert. Häufig werden sie bei Aktivitäten im Haushalt beschrieben. Männer werden häufiger in Berufssituationen und mit Adjektiven, die Stereotypen von Männlichkeit entsprechen, beschrieben. Ohne Gegendarstellungen besteht hierdurch die Gefahr, dass Stereotype und Vorurteile über Männer und Frauen gefestigt werden. Lernende, die sich nicht als männlich identifizieren, könnten entmutigt werden, eine Karriere außerhalb der durch diese Werte propagierten Felder anzustreben. Bei einem Einsatz des Tools durch Lernende könnten die repräsentativen Schäden[7] – je nach konkreten Unterrichtsszenario – beispielsweise durch die Lehrkraft im Vorfeld ersetzt, durch Aufarbeitung abgemildert gezielt zur Sensibilisierung eingesetzt werden.

Bildungsszenario 2 – Personalisierung des Lernens und Inklusion[8]: Ein Bereich, in dem der Einsatz von ChatGPT Lehrkräfte entlasten und zu einer besseren Inklusion beitragen kann, ist die Personalisierung des Lernens. Gleichzeitig entstehen durch den Einsatz neue Risiken für Ungleichbehandlung: ChatGPT funktioniert nicht für alle Menschen gleich gut, da die Qualität der Antworten mit der Fähigkeit gute Prompts zu schreiben steigt. Hierdurch werden Lernende, die in ihren Prompts kein Hochdeutsch verwenden oder deren Wortwahl von der Norm abweicht, möglicherweise benachteiligt. Es besteht außerdem die Problematik des zunehmenden Digital Divide [RM24]. Es besteht die Gefahr, dass beim Übersetzen, Vereinfachen und Zusammenfassen von Texten wichtige Inhalte und Zusammenhänge nicht richtig erfasst werden. Von Personen ohne Kontextwissen ist keine eigenständige Überprüfung von Antworten möglich, daher müssten alle Texte von einer qualifizierten Person geprüft werden.

Bildungsszenario 3 – Korrektur[9]: ChatGPT kann dazu eingesetzt werden, Geschriebenes auf Rechtschreib- und Grammatikfehler zu überprüfen und den Lesefluss und Schreibstil zu verbessern. Inhaltliche Ergänzungen und Korrekturen müssen genau geprüft werden, da es sich hier um eine Generation von Inhalten handelt. Die Nutzung von LLMs zur Feldbackgenerierung oder Korrektur bietet neben rechtlichen Bedenken auch Potenzial für Ungleichbehandlung [Ho25; Kh24]. Es bedarf dringend weiterer Untersuchungen, um

[7] Bezugnahme auf den u.a. von [Cr17] etablierten Begriff „representational harms“.

[8] Erklärung von Verständnisfragen, Anpassung von Texten und Unterrichtsinhalten

[9] Korrektur von Fehlern durch Lernende, Verbesserung von Lesefluss und Schreibstil durch Lernende, Korrektur und Ergänzung von Inhalten durch Lernende.

potentielle Auswirkungen eines Einsatzes von LLMs mit inhärentem Genderbias besser einschätzen zu können[10].

3 Erfordernisse für einen diskriminierungssensiblen Umgang mit Chatbots

Die Positionierung der Lehrkraft zum Thema ChatGPT wird das Vertrauen der Lernenden in dieses Produkt mitbeeinflussen. Es ist daher wichtig, auf die Schwächen von ChatGPT einzugehen, um somit eine kritische Haltung zu diesem Medium aufzubauen. Laut [RM24] ist es erforderlich, dass Nutzer:innen ein Bewusstsein dafür entwickeln, dass eine alltägliche Nutzung eines Systems zu blindem Vertrauen und fehlender Reflexion führen kann. Dieser Prozess wird unter anderem durch eine geringe Expertise in jenem Bereich, in dem das Large language model (LLM) eingesetzt wird, verstärkt[11]. Es gibt verschiedene Handlungen, die diesen Prozess verstärken. ChatGPT-4 argumentiert zum einen glaubwürdiger als die vorherigen Modelle und hat die Tendenz, auf falschen Aussagen und Halluzinationen zu bestehen, zum anderen suggeriert die ansteigende Vermenschlichung in den Antworten von Chatbots wie ChatGPT fälschlicherweise ein Gegenüber, das Empathie besitzt und strukturiert über Anfragen nachdenkt. OpenAI empfiehlt Softwareentwickler:innen Endnutzer:innen sorgfältig über die Limitationen von ChatGPT aufzuklären und sie in der richtigen Nutzung anzulernen. Dies ist auf die Beziehung zwischen Lehrkraft und Lernenden übertragbar. Außerdem sollen Nutzer:innen darauf hingewiesen werden, wie wichtig eine kritische Auseinandersetzung mit jeder generierten Antwort ist. Vor allem in der Beziehung zwischen Lehrkraft und Lernenden, die eine deutlich stärkere Bindung mit Vertrauensbasis aufweist als die Beziehung zwischen Entwickler:in und Nutzer:in, ist es wichtig, dass kein falsches Vertrauen in die Fähigkeiten von Modellen wie ChatGPT gesetzt wird, um einer unreflektierten Akzeptanz von Inhalten entgegenzuwirken. Nutzer:innen sollen über bestimmte KI-Skills[12] verfügen. Es ist ein wichtiger Bestandteil von KI-Skills, die grundlegende Funktionsweise von LLMs und die Bedeutung von Datensätzen zu kennen, aus denen sich technische Limitationen ableiten lassen. Auch die Auswirkung von Bias in Datensätzen und die damit verbundene Reproduktion von Bias, durch den Einsatz von, auf diesen Datensätzen trainierten, LLMs, müssen bekannt sein. Außerdem ist es laut [RM24] wichtig, dass Nutzer:innen eine Sensibilität für Diskriminierungsrisiken bezüglich der Eingaben und Ausgaben von ChatGPT entwickeln. Nutzer:innen sollten auch über die Entstehung dieser Risiken informiert sein, dazu zählt z. B. der Bias der von den in den verschiedenen Entwicklungsstadien einer LLM beteiligten Personen ausgeht.

[10] Auch Proxy-Diskriminierungen sind hier möglich (vgl. hierzu z. B. [RD21]).

[11] OpenAI verwendet dafür in der mit ChatGPT-4 publizierten Modelcard den Begriff Overreliance.

[12] KI-Skills beinhalten zum ein theoretisches Wissen, aber auch methodische Kenntnisse wie z. B. angebrachte Einsatzzwecke von LLMs und Techniken die dabei helfen können die Qualität der erhaltenen Antworten zu verbessern wie z. B. Prompt Engineering.

Literaturverzeichnis

[Bl20] Blodgett, S. L. et al.: Language (technology) is power: A critical survey of „bias“ in nlp. arXiv preprint arXiv:2005.14050, 2020.

[CG23] Carstensen, T.; Ganz, K.: Vom Algorithmus diskriminiert? Zur Aushandlung von Gender in Diskursen über Künstliche Intelligenz und Arbeit, Techn. Ber., Working Paper Forschungsförderung, 2023.

[Ch17] Chouldechova, A.: Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Big data 5 (2), S. 153–163, 2017.

[Cr17] Crawford, K.: The trouble with bias, 2017. URL http://blog.revolutionanalytics.com/2017/12/the-trouble-with-bias-by-kate-crawford.html. Invited Talk by Kate Crawford at NIPS, 2017.

[Ge20] Gehman, S. et al.: RealToxicityPrompts: Evaluating Neural Toxic Degeneration in Language Models, 2020, arXiv: 2009.11462 [cs.CL].

[Ho25] Ho, J. Q. et al.: Gender biases within Artificial Intelligence and ChatGPT: Evidence, Sources of Biases and Solutions. Computers in Human Behavior: Artificial Humans 4, S. 100145, 2025, DOI: 10.1016/j.chbah.2025.100145.

[KH22] Khan, M.; Hanna, A.: The subjects and stages of ai dataset development: A framework for dataset accountability. Ohio St. Tech. LJ 19, S. 171, 2022.

[Kh24] Khan, R.: Patriarchal AI: how ChatGPT can harm a woman's career. Media@ LSE, 2024.

[KMR16] Kleinberg, J. M.; Mullainathan, S.; Raghavan, M.: Inherent Trade-Offs in the Fair Determination of Risk Scores. CoRR abs/1609.05807, 2016, arXiv: 1609.05807.

[Li21] Liang, P. P. et al.: Towards understanding and mitigating social biases in language models. In: International Conference on Machine Learning. PMLR, S. 6565–6576, 2021.

[Mc23] McQuate, S.: How much energy does ChatGPT use? UW News, Accessed: 2024-10-17, 2023, https://www.washington.edu/news/2023/07/27/how-much-energy-does-chatgpt-use/.

[Me22] Mehrabi, N. et al.: A Survey on Bias and Fairness in Machine Learning, 2022, arXiv: 1908.09635 [cs.LG].

[MRP25] Marsden, N.; Raudonat, K.; Pröbster, M.: Breaking thee Cycle of Discrimination: Toward Gender Equity in Software Design. In: Human-Computer Interaction: Thematic Area, HCI 2025, Held as Part of the 27th HCI International Conference, HCII 2025, Gothenburg, Sweden, June 22–27, 2025, Proceedings, Part VII. Springer-Verlag, Gothenburg, Sweden, S. 378–395, 2025, DOI: 10.1007/978-3-031-93982-2_25.

[RD21] Ruf, B.; Detyniecki, M.: Towards the right kind of fairness in AI, 2021, arXiv: 2102.08453.

[RM24] Raudonat, K.; Mayweg, E.: Navigieren im Fluss sich wandelnder Technologien: Metakompetenzen im Kontext der Diversifizierung von KI-Technologien. MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung 61, S. 133–155, 2024.

[Se19] Selbst, A. D. et al.: Fairness and abstraction in sociotechnical systems. In: Proceedings of the conference on fairness, accountability, and transparency. S. 59–68, 2019.

[SG21] Suresh, H.; Guttag, J.: A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. In: Equity and Access in Algorithms, Mechanisms, and Optimization. EAAMO '21, ACM, 2021, DOI: 10.1145/3465416.3483305.

[Wa22] Wang, A. et al.: Measuring Representational Harms in Image Captioning, 2022, arXiv: 2206.07173 [cs.CY].

[Wa23] Wang, B. et al.: DecodingTrust: A Comprehensive Assessment of Trustworthiness in GPT Models, 2023, arXiv: 2306.11698.

[Zl15] Zliobaite, I.: On the relation between accuracy and fairness in binary classification. CoRR abs/1505.05723, 2015, arXiv: 1505.05723.

[ZP25] Zakizadeh, M.; Pilehvar, M. T.: Gender Encoding Patterns in Pretrained Language Model Representations, 2025, arXiv: 2503.06734 [cs.CL].

Demo-Beiträge

RDMC: A Tool to Support Authors in Preparing Artifacts for Evaluation

Firas Al Laban [1], Jan Bernoth [1], Safial Islam Ayon [1], Axel Wiepke [1], and Ulrike Lucke [1]

Abstract:

Effective research data and software management is essential to fostering open, transparent, and trustworthy science. Preparing research artifacts along with the paper publication is time-consuming for the authors and an additional barrier to sharing all research results. In this demo, we present a structured approach to ensure that all necessary research artifacts can be put together as a container to adhere to the FAIR principles. Our tool supports the creation of a Research Data Management Container (RDMC) that assists authors in adding metadata and creating a reusable environment for artifact evaluation at conferences and for archiving the presented research and its environment. As a demonstration case, we showcase Teach-R, an educational research software tool, to illustrate how artifacts can be published from the authors' point of view. This integration aims to enhance research quality, reproducibility, and transparency through structured and standardized artifact handling.

Keywords: Research Data Management Container, Artifact Evaluation, NFDIxCS, RDMC, Teach-R

1 Containerize and Manage Research Artifacts

In educational technology research, outputs are often diverse, with research software typically at the center, enhancing learning or teaching scenarios and often accompanied by survey data to assess usability or impact, as illustrated by Striewe's model [St24]. This creates a complex research landscape where datasets and software must be linked throughout both the study design and the presentation of results beyond the written paper. Especially when it comes to Research Data and Software Management (RDSM) and its FAIR principles for research data [Wi16] and research software [Ch22], it is crucial to share not only the artifacts but also their technological provenance. Here, artifact refers to any research output used or generated in the study, including software, scripts, input datasets, raw data, or analysis tools [As20].

Containers are widely used but either being a runtime environment, like from Docker, Singularity or Podman or packages just for sharing purpose like FAIR Digital Objects. In NFDIxCS (National Research Data Infrastructure for and with Computer Science),

[1] Universität Potsdam, Institut für Informatik, An der Bahn 2, 14476 Potsdam, Germany,
firas.al.laban@uni-potsdam.de, https://orcid.org/0000-0001-8072-9384;
jan.bernoth@uni-potsdam.de, https://orcid.org/0000-0002-4127-0053;
safial.islam.ayon@uni-potsdam.de, https://orcid.org/0000-0002-0970-0455;
axel.wiepke@uni-potsdam.de, https://orcid.org/0000-0002-0555-4040;
ulrike.lucke@uni-potsdam.de, https://orcid.org/0000-0003-4049-8088

doi:10.18420/delfi2025_41

containerization is a central part of research data and software management for fostering open and trustworthy research is a connection between sharing and reusing [Go24]. This is facilitated by the Research Data Management Container (RDMC), which constructs a sealed time capsule of the current research artifacts and its context and creates an application environment and how the artifacts are connected.

In this demo, the creation of an RDMC, as described in [Be25], is demonstrated using the research output of an exemplary paper in educational technology. The result of the creation process is the RDMC, which can either be shared as is or published as an artifact description paper. The RDMC itself is primarily a machine-readable description of the artifact versions, their relationships, and their runtime environment. The artifact description paper, however, also provides reviewers with insight into the required hardware setup.

2 Connect Artifacts with RDMC for Artifact Evaluation

To create an RDMC, a wizard was designed to integrate automated processes with human interaction. Figure 1 shows a screenshot of the RDMC creation wizard, highlighting key features that support metadata enrichment, data object management, artifact structuring, access control, and versioning.

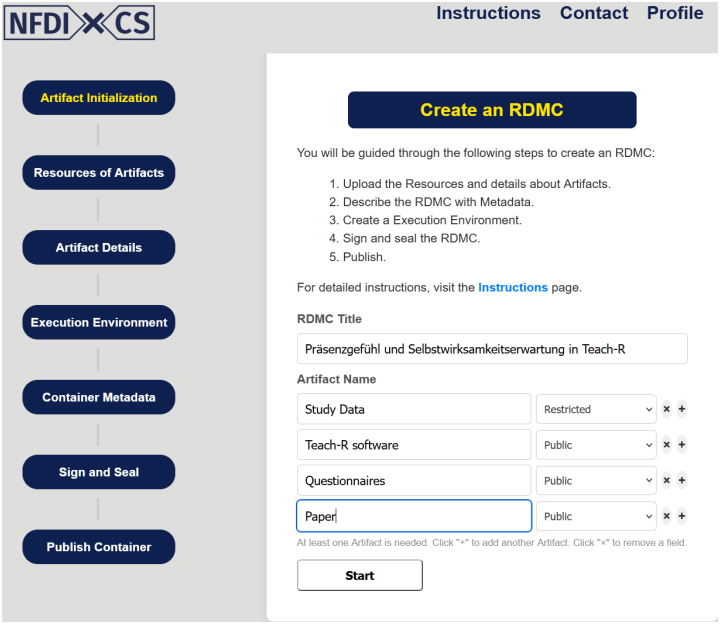


Fig. 1: An example screenshot of the RDMC creation workflow, where artifacts are prepared. The left side shows the steps of the containerization workflow, while the right side displays the details and actions associated with each step.

The RDMC creation workflow is guided by this wizard, which consists of multiple steps for creating and organizing artifacts. A demo version of this prototype wizard is also accessible[2]. In the following, all steps—shown on the left side of Figure 1—are briefly explained:

Steps 1 to 3 involve creating the artifact, uploading resources, or connecting to existing repositories where the artifacts are already stored. You also add discipline-specific metadata to the artifact. *Step 4* focuses on creating an execution environment for later reuse. *Step 5* is about creating metadata for the entire RDMC. *Steps 6 to 7* involve sealing the container and publishing it in an artifact evaluation track or an archive.

These steps are a proposal for an RDMC creation process. The order of the steps, as well as the questions about the RDMC, could be easily updated to address community needs. For example, currently the metadata about the artifacts is oriented toward the Supercomputing 25 [WG25] artifact description template, but if other metadata is more important for the DELFI community, these questions could be exchanged. This wizard is focused on helping authors in the process of preparing their artifacts for an evaluation process.

3 Exemplifying containerisation in RDMC with a Teach-R study

In this section, we present an example of how a new artifact can be created and integrated into the RDMC, as well as how it can be enrolled in the artifact evaluation process. To illustrate this, a study based on the Teach-R project has been selected to showcase the versatility of the RDMC. Teach-R provides a virtual reality training environment for preservice teachers, incorporating various didactical methods.

The example shown in Figure 2 depicts a study in which Teach-R was used to train classroom conversation skills using specific prompts in history lessons. As a result, a wide variety of data was generated during the study (see Figure 2), including data from project partners, personal data, real-time interaction data, and pre- and post-study data. This example serves as a useful basis for various discussions during the demo.

(1) Didactical preparation for the course, to be found in [FA23]. (2) Description of development, including requirements and models to be found in the software artifact. (3) Material for learners during the course, which can be found in [FA23]. (4) Learning objectives in full text, can be found in [FA23]. (5) Algorithms to evaluate learning objectives, to be found in software artifact. (6) Notes of probands during the study were not relevant for this study. (7) Realtime data of the study can be found in the datafiles artifact. (8) Automatic feedback for learners was produced by the software artifact. (9) Logfiles were not relevant for study but are produced by the software artifact. (10) Questionnaires and observations can be found in the restricted datafile artifact, since answers contain personal data of probands. (11) Performance metrics of the software were not relevant for study but are produced by the software artifact.

[2] https://rdmc.create.nfdixcs.org/

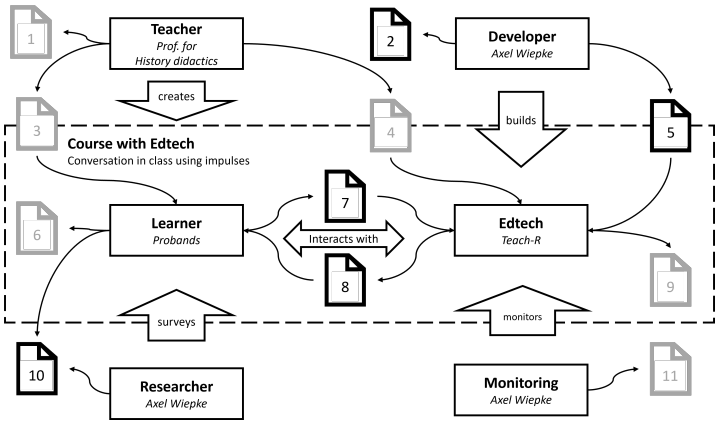


Fig. 2: An overview of collected data on basis of Striewes model [St24]. Gray data are not to be found in the RDMC, since they were part of project partners studies.

4 Future Work

As shown, the artifact structure behind a publication can be complex, often involving multiple interrelated components. The RDMC aims to streamline this by bundling artifacts with their metadata into a single, standardized container. This guides the necessary workload for authors and provides reviewers with a consistent, single point of entry in artifact evaluation processes. If a Reproducible Execution Environment (REE) is created as part of the RDMC, reviewers can also directly try out the artifacts.

The RDMC is designed with flexibility in mind, allowing it to accommodate the expectations and practices of different review communities while maintaining consistent quality standards. Future work will focus on evaluating which steps in the RDMC creation workflow should be mandatory or optional, particularly for the DELFI community. Future work will also address standardizing required metadata and improving REE creation to support reliable validation and reuse.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – NFDI 52/1 – 501930651.

References

[As20] Association for Computing Machinery: Artifact Review and Badging Version 1.1, 2020, https://www.acm.org/publications/policies/artifact-review-and-badging-current.

[Be25] Bernoth, J. et al.: Workflow for Creating and Sealing a Research Data Management Container (RDMC). In (Feichtinger, K.; Sonnleithner. Lisa; Hajiabadi, H., eds.): Software Engineering 2025 – Companion Proceedings. Gesellschaft für Informatik e.V, Bonn, 2025, DOI: 10.18420/se2025-ws-26.

[Ch22] Chue Hong, N. P. et al.: FAIR Principles for Research Software (FAIR4RS Principles), 2022, DOI: 10.15497/RDA00068.

[FA23] Fenn, M.; Arlt, J.: Historisches Lernen immersiv: Studierende üben Unterrichtsgespräche in Virtual Reality. de, MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung 51, pp. 114–130, 2023, DOI: 10.21240/mpaed/51/2023.01.14.X.

[Go24] Goedicke, M. et al.: National Research Data Infrastructure for and with Computer Science (NFDIxCS), 2024, DOI: 10.5281/zenodo.10557968.

[St24] Striewe, M.: Forschungsdaten in der Bildungstechnologie: Worüber sprechen wir und was ist wann relevant?, 2024, DOI: 10.18420/delfi2024_44.

[WG25] Weidendorfer, J.; Green, J.: AD/AE Process & Badges: SC25, 2025, https://sc25.supercomputing.org/program/papers/reproducibility-appendices-badges/, accessed: 06/09/2025.

[Wi16] Wilkinson, M. D. et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1), p. 160018, 2016, DOI: 10.1038/sdata.2016.18.

Pack-a-Capsule – Extending a VR Serious Game on Network Communications to Enable Collaborative Gameplay

David Baberowski [1], Thiemo Leonhardt [2], René Röpke [3], and Sila Irmak[4]

Abstract: This paper introduces Pack-a-Capsule (PaC), a collaborative, web-based serious game designed to complement the VR single-player experience Inside the Router (ITR). Together, they offer a modular educational environment for understanding network and transport layer protocols. While ITR immerses learners in packet routing and NAT mechanisms, PaC focuses on data segmentation and transmission using TCP and UDP. PaC enables learners to collaborate through distinct roles within a simulated network scenario, combining the benefits of VR for immersive learning with screen-based environments for coordination and communication. Grounded in socio-constructivist learning theory, multimedia learning principles, and game-based learning, the design aims to promote conceptual understanding, perspective-taking, and collaborative learning. As a use case, PaC can be used to make abstract networking concepts tangible and to address limitations of VR setups in classroom settings.

Keywords: Serious Game, Collaborative Learning, Network Communication, Transport Layer

1 Introduction

Understanding the principles of computer networks is a key objective in computer science (CS) education at all levels. However, the abstract nature of network communication and the logistical complexity of setting up real-world network environments often lead educators to use analogies, simulations, or role-plays to convey these concepts. Serious games such as *Inside the Router* (ITR) support this learning process by allowing students to assume the role of a router. Through immersive VR gameplay, learners gain a tangible understanding of IP-based routing and NAT (network address translation) mechanisms at the network layer. Despite their pedagogical potential, VR-based scenarios are constrained by practical limitations in classroom use, limited access, and reduced opportunities for collaboration and the isolation of the learner. To address these challenges, we developed *Pack-a-Capsule* (PaC), a collaborative, web-based game that complements the single-player VR experience of ITR. PaC focuses on the neighboring transport layer in the OSI model and enables learners to segment, package, and transmit data using the Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP). By combining VR and screen-based technologies, learners adopt roles, such as routing or packaging, within a shared, simulated network scenario.

[1] TU Dresden, Didactics of Computer Science, Nöthnitzer Str. 46, 01187 Dresden, Germany, david.baberowski@tu-dresden.de, https://orcid.org/0000-0001-6308-4334
[2] RWTH Aachen University, Computer Science Education, Mies-van-der-Rohe-Str. 15, 52074 Aachen, Germany, leonhardt@cs.rwth-aachen.de, https://orcid.org/0000-0003-4725-9776
[3] TU Wien, Learning Technologies and eDidactics, Favoritenstr. 9-11, 1040 Vienna, Austria, rene.roepke@tuwien.ac.at, https://orcid.org/0000-0003-0250-8521
[4] RWTH Aachen University, Ahornstr. 55, 52074 Aachen, Germany, sila.irmak@rwth-aachen.de

doi:10.18420/delfi2025_42

This cross-media, role-based setup draws on established pedagogical principles: immersive media (e. g., VR) supports experiential, process-focused learning, while screen-based environments (e. g., tablets, desktops) are well suited for planning, coordination, and collaborative decision-making. Our approach is grounded in established learning theories. From a socio-constructivist perspective, learning is most effective when students engage in dialogue, assume meaningful roles, and co-construct knowledge principles supported by research in computer-supported collaborative learning [St21]. Immersive VR environments enable process-oriented experiential learning through presence and active participation, which supports conceptual understanding of abstract systems [Li17; MP23]. To manage the cognitive load across the modalities, the design incorporates principles from multimedia learning theory, including Mayer's modality principle and dual-channel processing [MF22; SVP19]. In addition, elements of game-based learning promote motivation and participation, drawing on self-determination theory and flow theory [Cs14]. Together, these components establish a theory-informed modular learning environment designed to foster collaboration, perspective taking, and a deeper understanding of network communication processes.

2 Related Work

CS education integrates foundational concepts of computer networks and Internet communication, reflecting their relevance not only in professional contexts but also in everyday digital literacy [Be19; Br08; He24]. Topics range from basic terminology like IP addresses and domains to understanding the structure of the World Wide Web. Due to the technical and logistical complexity of real-world network setups in school contexts, educational approaches often use abstraction and visualization strategies such as role-plays, network simulations, or serious games. Role-based activities, where students enact the behavior of network components, make abstract network processes tangible [Bo11; Si17]. Simulation tools such as *Filius*, *NetEmul*, and *WebNetSim* enable learners to build and explore networks interactively in a low-cost, classroom-compatible format [Hi23]. Game-based learning environments add motivational, narrative elements to this didactic strategy. ITR as a VR serious game simulates packet routing at the network layer [BLB23]. It exemplifies a growing trend to use serious games to teach complex systems through active participation.

However, few existing systems integrate multiple modes of interaction or provide shared, role-divided scenarios that align with collaborative learning principles. While tools like *WebNetSim* distribute functional components across learners, they often lack a narrative or motivational layer. Conversely, immersive VR games typically focus on individual experience and offer limited scalability in class contexts. Recent research in computer-supported collaborative learning and immersive learning design highlights the value of social presence, role distribution, and multimodal engagement for fostering deep conceptual understanding [MP23; St21]. Despite this, a gap remains in the design of collaborative, multimodal learning tools for networking education that go beyond traditional simulation or single-user immersion. This gap motivates the development of PaC, a collaborative

multiplayer extension to ITR, integrating immersive and screen-based interfaces and enable learners to engage with different layers of the OSI model from complementary perspectives. PaC aims to foster collaboration, perspective-taking, and an applied understanding of protocol interaction within a shared simulated network scenario.

3 Pack-a-Capsule

PaC is an extension for the VR serious game ITR that enables multiple students to play together in a classroom setting. While ITR enables students to experience and perform the tasks of a home router (network layer), PaC focus lies on the transport layer. It builds on the same metaphorical framework as ITR [BLB23] and enables learners to perform transport layer tasks by segmenting data into packets, sending those packets through the network, and reassembling the data on the receiving side. PaC implements the TCP and UDP transport protocols, making the distinction between reduced protocol overhead (UDP) and error handling (TCP) tangible. Guidance is implemented through prompts within the game. PaC is implemented as an Angular-based web application with an API backend connecting different clients using Websockets [So25]. Its UI was optimized for tablet or desktop usage. ITR and PaC are self-contained and concentrate on a specific protocol layers. In combination, they enable a modular, interconnected, multimodal learning experience, in which the layered structure of network communication becomes practically tangible and supports not only the understanding of individual layers but also their interaction within the OSI model. The combination of two standalone games, one VR and the other screen-based, into a shared thematic context offers multiple didactic potentials:

- **Illustration and Interactivity:** The functioning of protocols, especially their cooperation in protocol stacks, is difficult for learners to understand at an abstract level. Interactive serious games such as PaC actively involve learners in a scenario through dialog and actions, thus, supporting the learning process [St21].
- **Learning from mistakes:** The applications do not prevent learners from making mistakes and thus mislabeling or even losing packages. Instead, the consequences of these errors are accurately simulated to support later discussion and reflection.
- **Together instead of alone:** A major weakness of VR, its isolating character, is partially reduced by the multiplayer scenario, enabling collaborative learning possible even if only one or two VR headsets are available. Thus, students may benefit from the potentials of VR [JK18; MP23] even with limited resources.
- **Flexibility in teaching:** The fact that learners do not learn the same things at the same time is both an opportunity and a challenge. This enables and requires more flexible teaching concepts, especially with the game providing guidance through prompts.

An exemplary implementation of ITR and PaC in the practice phase of a teaching unit could look like this: First, the teacher introduces the general theoretical principles, including the

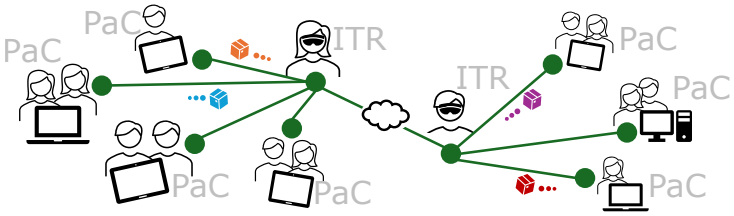


Fig. 1: Exemplary classroom setup with two ITR VR sessions and numerous PaC clients.

OSI model and in particular the network and transport layer. The students learn that data must be broken down into packets, transmitted individually and reassembled at the end. The first exercise phase then begins, in which some students put on the VR headsets and start ITR, while the rest take part in the scenario via PaC. Two households are simulated here, i. e. two home routers, each of which connects different clients to the Internet (see Figure 1). The students playing the PaC application are given the task of exchanging text and picture messages with each other. To do this, they select any protocol (UDP or TCP), establish a connection if necessary, create the data packets and send them. Each data packet appears to the students with the VR application as a data capsule that must be correctly assigned by catching and throwing in order to reach the recipient. Errors can occur, which can result in packets being lost. It can therefore be assumed that not all messages can be assembled correctly by the recipient. This exercise phase is followed by a reflection in which the teacher talks to the students about their observations (difference between TCP and UDP) and any errors that occurred. The students who have used the PaC application must explain the procedure and the special features of the protocols to the students who have played the ITR application as experts and vice versa. This can then be consolidated in subsequent practice phases in which the students swap roles, test informed hypotheses and select the transport protocol best suited for their respective tasks.

4 Research Perspectives

Several research questions can be derived, which relate in particular to the effective use and integration of PaC in learning. Although current findings point to positive effects of digital media, these are highly context-dependent [MF22]. Therefore, it should be investigated which specific application scenarios of PaC actually have learning-promoting aspects and how these quantitatively affect performance measures like knowledge acquisition and retention performance [He14]. Closely related to this is the question of technological integration into the classroom. Future research should analyze which didactic practices lead to higher teaching quality when integrating PaC [Sc21]. Studies suggest that the effectiveness of technical tools depends heavily on the quality of lesson design and teacher qualifications [QRL21]. It is therefore necessary to investigate how teachers can be supported in using PaC and what professional skills are required to use the tool optimally [Sc21].

References

[Be19] Best, A. et al.: Kompetenzen für informatische Bildung im Primarbereich, Gesellschaft für Informatik e.V., 2019, DOI: 10.18420/rec2019_061.

[BLB23] Baberowski, D.; Leonhardt, T.; Bergner, N.: Inside the System – Designing VR Serious Games for Computer Science Education. In (Dondio, P. et al., eds.): Games and Learning Alliance, GALA 2023. Gesellschaft für Informatik e.V., Bonn, Germany, pp. 311–320, 2023, DOI: 10.1007/978-3-031-49065-1_30.

[Bo11] Borowski, C. et al.: Kinder Auf Dem Weg Zur Informatik: Wie Funktioniert Das Internet? In: Informatik Mit Kopf, Herz Und Hand. Praxisbeiträge zur INFOS 2011. INFOS. Pp. 244–253, 2011.

[Br08] Brinda, T. et al.: Grundsätze und Standards für die Informatik in der Schule - Bildungsstandards Informatik für die Sekundarstufe I, Bonn: Gesellschaft für Informatik e.V., 2008, DOI: 10.18420/rec2008_052.

[Cs14] Csikszentmihalyi, M.: Applications of Flow in Human Development and Education: The Collected Works of Mihaly Csikszentmihalyi. Springer Netherlands, Dordrecht, 2014.

[He14] Herzig, B.: Wie wirksam sind digitale Medien im Unterricht? Eine Analyse aktueller Forschungsbefunde, tech. rep., Bertelsmann Stiftung, 2014, https://www.bertelsmann-stiftung . de / fileadmin / files / BSt / Publikationen / GrauePublikationen / Studie _ IB _ Wirksamkeit_digitale_Medien_im_Unterricht_2014.pdf.

[He24] Hellmig, L. et al.: Bildungsstandards Informatik für die Sekundarstufe I, Gesellschaft für Informatik e.V., 2024, DOI: 10.18420/rec2025_052.

[Hi23] Hielscher, M.: Aufbau und Funktion des Internets vermitteln - vom Rollenspiel zur verteilten Simulationsumgebung. In: INFOS 2023 - Informatikunterricht zwischen Aktualität und Zeitlosigkeit. INFOS. Gesellschaft für Informatik e.V., Bonn, pp. 299–307, 2023, DOI: 10.18420/INFOS2023-029.

[JK18] Jensen, L.; Konradsen, F.: A Review of the Use of Virtual Reality Head-Mounted Displays in Education and Training. Education and Information Technologies 23 (4), pp. 1515–1529, 2018, DOI: 10.1007/s10639-017-9676-0.

[Li17] Liu, D. et al., eds.: Virtual, Augmented, and Mixed Realities in Education. Springer Singapore, Singapore, 2017.

[MF22] Mayer, R. E.; Fiorella, L., eds.: The Cambridge Handbook of Multimedia Learning. Cambridge University Press, Cambridge New York, NY, 2022.

[MP23] Makransky, G.; Petersen, G. B.: The Theory of Immersive Collaborative Learning (TICOL). Educational Psychology Review 35 (4), p. 103, 2023, DOI: 10.1007/s10648-023-09822-5.

[QRL21] Quast, J.; Rubach, C.; Lazarides, R.: Lehrkräfteeinschätzungen zu Unterrichtsqualität mit digitalen Medien: Zusammenhänge zur wahrgenommenen technischen Schulausstattung, Medienunterstützung, digitalen Kompetenzselbsteinschätzungen und Wertüberzeugungen. Zeitschrift für Bildungsforschung 11 (2), pp. 309–341, 2021, DOI: 10.1007/s35834-021-00313-7.

[Sc21] Scheiter, K.: Lernen und Lehren mit digitalen Medien: Eine Standortbestimmung. Zeitschrift für Erziehungswissenschaft 24 (5), pp. 1039–1060, 2021, DOI: 10.1007/s11618-021-01047-y.

[Si17] Siegers, C.: Wie Funktioniert Das Internet?, 2017, https://schuelerlabor.informatik.rwth-aachen.de/module/internetspiel, accessed: 06/04/2025.

[So25] Socket.IO: Socket.IO, 2025, https://socket.io, accessed: 06/04/2025.

[St21] Stahl, G., ed.: Theoretical Investigations: Philosophical Foundations of Group Cognition. Springer International Publishing, Cham, 2021.

[SVP19] Sweller, J.; Van Merriënboer, J. J. G.; Paas, F.: Cognitive Architecture and Instructional Design: 20 Years Later. Educational Psychology Review 31 (2), pp. 261–292, 2019, DOI: 10.1007/s10648-019-09465-5.

Hochschulübergreifende Kooperation am Beispiel der Mathematikkurse OMB+ und HM4MINT

Volker Bach[1], Aloys Krieg[2] und Ruedi Seiler[3]

Abstract: Die Digitalisierung der Hochschullehre eröffnet neue Möglichkeiten für hochschulübergreifende Kooperationen. Anhand der Kurse OMB+ und HM4MINT wird gezeigt, wie gemeinsame digitale Mathematikkurse erfolgreich entwickelt und implementiert werden können.

Keywords: Digitale Hochschullehre, OMB+, HM4MINT, hochschulübergreifende Kooperation, MINT-Studium, MUMIE-Plattform, E-Learning, Qualitätssicherung, aktives Lernen

1 Die neue Ära der hochschulübergreifenden digitalen Lehre

Die Digitalisierung der Hochschullandschaft hat in den letzten Jahren eine bemerkenswerte Transformation durchlaufen. Was einst als isolierte E-Learning-Angebote einzelner Institutionen begann, entwickelt sich zunehmend zu einem vernetzten Ökosystem gemeinsamer digitaler Lehrformate.

Es entstehen innovative Kooperationsmodelle, die Ressourcen bündeln und Synergien schaffen, statt an den Institutionen jeweils eigene, oft einander ähnelnde Inhalte zu entwickeln. Diese Entwicklung wird durch verschiedene Faktoren vorangetrieben. Der steigenden Kostendruck in der Hochschulfinanzierung, die wachsende Nachfrage nach flexiblen Studienformaten, die Schonung eigener Ressourcen und nicht zuletzt die Erkenntnis, dass gemeinsam entwickelte Lösungen oft qualitativ hochwertiger und nachhaltiger sind, beflügeln diese Entwicklung

Ein besonders erfolgreiches Anwendungsfeld für diese hochschulübergreifende Zusammenarbeit findet sich in der mathematischen Grundausbildung für MINT-Studiengänge. Hier zeigen sich die Vorteile gemeinsamer digitaler Lehrformate besonders deutlich, da die Inhalte weitgehend standardisiert sind und gleichzeitig einen hohen Aufwand bei der Entwicklung hochwertiger Lernmaterialien erfordern.

In diesem Artikel stellen wir zwei Beispiele für erfolgreiche Kooperationsprojekte in diesem Bereich vor: den Online Mathematik Brückenkurs OMB+ und den Kurs HM4MINT.

[1] TU Braunschweig, Universitätsplatz 2, 3806 Braunschweig, v.bach@tu-bs.de

[2] RWTH Aachen, Templergraben 55, 52056 Aachen, AK@cimcea.de

[3] integral-learning GmbH, Digital Education Lab IHK, Fasanenstr. 85, 10623 Berlin, ruedi.seiler@integral-learning.de

doi:10.18420/delfi2025_43

2 Der Online Mathematik Brückenkurs OMB+: Ein Modell hochschulübergreifender Kooperation

2.1 Entstehung und Struktur

Der Online Mathematik Brückenkurs OMB+ entstand aus der 2013 begonnenen Zusammenarbeit von zwölf deutschen Universitäten und Hochschulen für angewandte Wissenschaft. Der Kurs wird mittlerweile von mehr als sechzig solchen Hochschulen und anderen Institutionen genutzt. Die Initiative wird von der Allianz führender Technischer Universitäten in Deutschland (TU9), der Deutschen Physikalischen Gesellschaft (DPG) und der Konferenz der Fachbereiche Physik (KFP) unterstützt und orientiert sich an dem von der Baden-Württembergischen Expertengruppe COSH entwickelten Mindestanforderungskatalog für den Übergang von der Schule in ein MINT-Studium an einer Hochschule.

Technisch basieren beide Kurse, OMB+ und HM4MINT, auf der MUMIE-Plattform (Multimediale Mathematikausbildung für Ingenieurinnen und Ingenieure). Sie ist aus einem BMBF-geförderten Projekt hervorgegangen und wird vom Spinoff integral-learning GmbH weiterentwickelt und Hochschulen zur Verfügung gestellt.

2.2 Didaktisches Konzept und Zielgruppen

Der Brückenkurs OMB+ richtet sich primär an Studieninteressierte und -anfänger in MINT-Fächern, die ihre Mathematikkenntnisse vor Studienbeginn auffrischen möchten. Das didaktische Konzept zeichnet sich durch mehrere innovative Elemente aus, die im Folgenden skizziert werden.

Der Kurs kombiniert verschiedene multimediale Lernelemente, darunter erklärende Texte mit Quizzes, interaktive Visualisierungen, Videos und Übungsaufgaben. Besonderes Gewicht wird auf die Trainingsaufgaben zum aktiven Lernen gelegt. Dies sorgt im Unterschied zum passiven Ansehen von Videos für eine starke Verankerung des Gelernten im Gedächtnis.

Ein besonderes Merkmal von OMB+ ist das umfassende Tutoring-System. Täglich von 10:00 bis 20:00 Uhr – auch an Wochenenden und Feiertagen – stehen speziell ausgebildete Tutorinnen und Tutoren für Fragen zur Verfügung. Diese Betreuung erfolgt über ein integriertes Chat und Telefon. Ein KI-System für eine 24/7-Betreuung der Lernenden befindet sich im Aufbau.

2.3 Erfolgsfaktoren und Wirkung

Der Erfolg von OMB+ lässt sich an mehreren Faktoren messen. Zum einen zeigen Nutzerfeedbacks eine hohe Zufriedenheit mit dem Kurs. Teilnehmer berichten regelmäßig

von gesteigertem Selbstvertrauen im Umgang mit Mathematik und besserer Vorbereitung auf das Studium. Der direkte oder auch nur begleitende Einsatz von OMB+ in der Hochschullehre für Studienanfänger verbessert das Notenspektrum signifikant. Zum anderen wird der Kurs mittlerweile von mehr als sechzig Hochschulen genutzt und von der Allianz führender Technischer Universitäten in Deutschland TU9 und der Deutschen Physikalischen Gesellschaft DPG empfohlen.

Ein wichtiger Erfolgsfaktor ist die kostenlose Verfügbarkeit des Kurses für Lernende, die durch die Kooperation der beteiligten Hochschulen ermöglicht wird. Einzelne Hochschulen wären kaum in der Lage, ein derart umfassendes Angebot alleine bereitzustellen.

3 HM4MINT: Innovation in der Hochschulmathematik durch Landeskooperation

3.1 Das nordrhein-westfälische Modell

HM4MINT (Höhere Mathematik für MINT-Studiengänge) stellt ein weiteres Kooperationsprojekt dar, das im Rahmen der „Digitalen Hochschule NRW“ entstanden ist. Entwickelt von mehreren Hochschulen in NRW unter der Federführung der RWTH Aachen, wird der Online-Kurs mittlerweile von 17 Hochschulen in Nordrhein-Westfalen offiziell für ihre ingenieurwissenschaftlichen Studiengänge angeboten. Das Besondere an diesem Projekt ist, dass es nicht nur als Vorbereitung, sondern als vollwertiger Ersatz für das erste Mathematikmodul im Studium konzipiert ist.

Die Finanzierung erfolgt durch das Nordrhein-Westfälische Ministerium für Kultur und Wissenschaft im Rahmen der landesweiten Digitalisierungsoffensive. Diese politische Unterstützung ist ein wichtiger Baustein für den Erfolg großangelegter Kooperationsprojekte und zeigt, wie Landespolitik die Digitalisierung der Hochschullehre strategisch vorantreiben kann.

3.2 Zwei Varianten: Flexibilität und Intensität

HM4MINT bietet zwei unterschiedliche Kursformate, die verschiedene Lernbedürfnisse abdecken:

HM4MINT Online: Dieser extensiv angelegte Kurs ermöglicht das Selbststudium der Höheren Mathematik 1 für Ingenieurwissenschaften und kann jederzeit begonnen werden. Er besteht aus frei zugänglichen Texten, Videos und Online-Aufgaben. Wie OMB+ wird auch dieser Kurs intensiv von Tutorinnen und Tutoren über eine integriertes Chat und Telefon betreut. Die Flexibilität dieses Formats macht es besonders attraktiv für Berufstätige oder Studierende, die ihre Lernzeit frei einteilen möchten.

HM4MINT-Intensiv: Diese geführte Variante ist als neunwöchiger, ganztägiger Online-Kurs konzipiert, der in der Sommerpause vor dem eigentlichen Studienbeginn stattfindet. Er zeichnet sich durch eine intensive Betreuung aus, mit täglichen Online-Vorlesungen, Fragestunden und Kleingruppen-Tutorien. Dieses Format ist angelehnt an traditionelle Formate, nutzt aber konsequent digitale Möglichkeiten.

3.3 Didaktische Innovation und Qualitätssicherung

Die didaktische Konzeption von HM4MINT zeigt mehrere innovative Ansätze, die für moderne digitale Lehrformate charakteristisch sind:

Kompetenzorientierung: Der Kurs beginnt bewusst mit einer Wiederholung des Schulstoffs aus Hochschulsicht und führt dann systematisch zu den Anforderungen der universitären Mathematik. Diese Brückenfunktion ist entscheidend für den Studienerfolg in MINT-Fächern.

Interaktive Aufgaben: Der Kurs HM4MINT nutzt wie auch der Brückenkurs OMB+ automatisch korrigierte Trainings- und Prüfungsaufgaben, die sofortiges Feedback ermöglichen. Diese Form der Rückmeldung ist besonders wertvoll für das Selbststudium und ermöglicht es den Lernenden, einen kontinuierlichen Überblick über ihre Lernfortschritte zu erhalten.

Qualitätssicherung durch Evaluation: Die Ergebnisse von Intensivkurs-Teilnehmenden waren in der Vergangenheit trotz der Kürze der Zeit vergleichbar mit denen der Kurse im ersten Semester. Die Lerndaten werden systematisch ausgewertet und fließen in die Weiterentwicklung der Kurse ein.

3.4 Credit Transfer und curriculare Integration

Ein besonders innovativer Aspekt von HM4MINT ist die curriculare Einbindung in die Studiengänge der beteiligten Hochschulen. Nach erfolgreichem Abschluss des Kurses und Bestehen der Präsenzklausur können Studierende mit Credit Points in ihr Studium starten. Diese hochschulübergreifende Anerkennung der erworbenen Leistungen ist (noch) nicht selbstverständlich und zeigt das Vertrauen der beteiligten Institutionen in die Qualität des gemeinsam entwickelten Angebots.

Die formale Anerkennung der Kurse als Studienleistung stellt einen wichtigen Schritt in Richtung eines vernetzten Hochschulsystems dar, in dem Studierende flexibel zwischen verschiedenen Anbietern wählen können, ohne etwa durch Wechsel Nachteile zu erleiden.

Integrating LLM-Generated Feedback in JupyterLab to Enhance Programming Education



Annabell Brocker[1] and Ulrik Schroeder[1]

Abstract: Personalised, immediate feedback is crucial for effective programming education, particularly in self-directed or asynchronous settings. This paper introduces an open-source tool that integrates Large Language Model (LLM)-based feedback into JupyterLab, providing immediate, context-aware guidance within the coding environment. The system incorporates a JupyterLab extension designed for interactive assistance, alongside a modular Python backend that facilitates secure prompt generation and communication with an LLM. Learners choose from predefined help categories grounded in cognitive and pedagogical models, receiving stepwise, individualised feedback that emphasises understanding rather than a sample solution. Successfully deployed in Python courses, the tool supports scalable, adaptable instruction across languages.

Keywords: Feedback, Programming Education, CS1, Python, Large Language Models

1 Introduction

Feedback is an important component in programming education, as it supports the learning process by guiding learners towards correct solutions and reinforcing successful problem-solving strategies [HT07; Na08]. The effectiveness of feedback is not solely determined by its content, but also by the timeliness with which it is delivered [Je22]. Immediate feedback minimises frustration by reducing cognitive overload and fosters a sense of achievement by quickly affirming learners' progress. A common challenge in programming education, particularly in asynchronous or self-directed learning contexts, is the limited availability of immediate, personalised support. In these settings, learners often lack real-time assistance from instructors or peers, which can hinder progress and deepen misconceptions. To mitigate this problem, the integration of Large Language Models (LLMs) into educational environments presents a promising solution. LLMs are capable of providing automated, context-sensitive, and personalised feedback that approximates tutor-like support [Me24]. Moreover, aligning the prompts for LLMs with feedback design principles, such as those types formulated by [KJH18; Na08], can make the LLM-generated feedback more personalised, concrete and guided, thus increasing its effectiveness [LKK25]. At the same time, the programming environment shapes the learning experience. JupyterLab[2] is widely used due to its accessibility and ease of deployment, especially in educational settings where pre-configured, web-based instances reduce barriers for novices. However, JupyterLab lacks

[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

[2] JupyterLab Documentation, https://jupyterlab.readthedocs.io/en/latest/, last accessed: 2025-07-07

doi:10.18420/delfi2025_44

pedagogical features such as integrated task editors. Instead, instructors may distribute tasks across multiple notebooks or within a (large) single notebook, making it difficult to extract relevant context, such as instructions, examples, or learner solutions, for use in automated feedback generation. Extensions like Jupyter AI[3] already introduce LLM-based chat interfaces into JupyterLab, allowing learners to request help in natural language. While flexible, such open-ended interaction may encourage solution-seeking rather than active problem solving, especially when not guided by educational principles.

To address these limitations, we present a tool[4] that integrates structured, LLM-generated feedback directly within the JupyterLab environment. Our system goes beyond existing tools by providing step-by-step, context-aware guidance aligned with pedagogical feedback principles. Rather than supplying ready-made solutions, the tool supports learners in gradually building their understanding through tailored prompts, reflective hints, and strategy-oriented suggestions. This design aims to preserve the benefits of immediate, personalised feedback while promoting active engagement with programming concepts.

2 Tool

The architecture[4] integrates LLM-based support within JupyterLab to enhance learner experience in computational tasks. It combines user interface elements, pedagogical modelling, and natural language processing to deliver personalised, context-aware feedback. The system comprises a JupyterLab frontend extension and a dockerised Python backend that processes requests, prompt generation, and LLM communication (see Figure 1). Its modular design enables scalability and integration for alternative models or feedback methods.

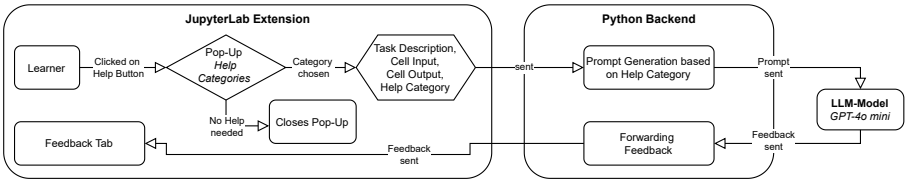
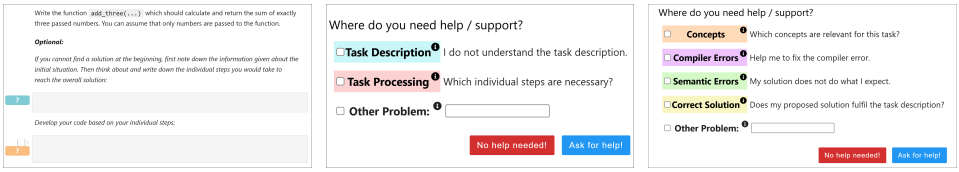


Fig. 1: Overview of the tool's architecture and workflow.

JupyterLab Extension The JupyterLab extension renders a help button next to markdown and code cells (see Figure 2a) when cell metadata matches instructor-configured patterns (e. g., when the key `cell_id` ends with `IS.Code` or `IS.Markdown`). Upon clicking the button, a pop-up interface prompts the learner to select a predefined help category, which are based on the cognitive model of programming processes by [Lo16] and the taxonomy of feedback types proposed by [KJH18; Na08]. In total, two help categories for markdown cells (*Task Description* and *Task Processing*, see Figure 2b) and four for code cells (*Concepts*, *Compiler*

[3] Jupyter AI, https://github.com/jupyterlab/jupyter-ai, last accessed: 2025-07-07
[4] OSF, https://doi.org/10.17605/OSF.IO/RJD57, last accessed: 2025-07-07



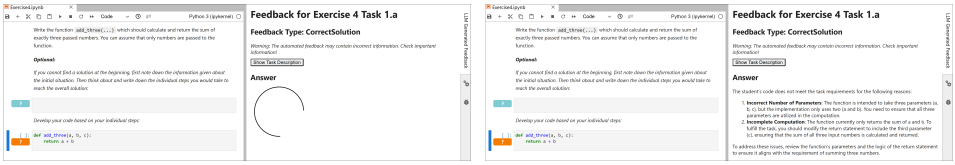
(a) Position of help buttons next to corresponding markdown or code cell.

(b) Help categories for markdown cells.

(c) Help categories for code cells.

Fig. 2: Placement, types, and help categories of buttons for markdown and code cells.

Errors, *Semantic Errors* and *Correct Solution*, see Figure 2c) exist. An *Other Problem* option allows manual issue specification. The extension locates related task descriptions within the notebook using a metadata-based pattern-matching system defined via a JSON configuration file. This file specifies valid suffixes and their meanings, enabling instructors to customise cell roles and naming conventions according to course structure. Each value of a cell's metadata key includes a task identifier (e. g., `A1`, `A2.b`) combined with a suffix denoting cell type (e. g., `TD` for task description, `IS.Code` for code input). This grouping links all cells belonging to the same task. Once a help category is selected, the extension sends the relevant data, including the task description, cell content, and output if available, to the Python backend. A new tab opens in the interface, displaying a spinner during processing (see Figure 3a). If the backend is unavailable or a response fails, the user is notified. Upon receiving the LLM-generated feedback, formatted in markdown, the extension renders it within the tab for learner review (see Figure 3b).



(a) Feedback generation in progress.

(b) LLM-generated response.

Fig. 3: Feedback tab shown to the learner during a help request.

Python Backend The Python backend fulfills two principal functions: (1) prompt generation for the LLM, and (2) securely managing the LLM API key to prevent client-side access, thereby maintaining system integrity and safeguarding user data. Upon request submission, the backend receives contextual information, including the selected help category, task description, cell content, and, where applicable, the output of the cell. This data is used to construct a prompt tailored to the specific help category requested. Unlike the approach presented by [LKK25], which employs a single extensive prompt with variable help category, this system generates a unique prompt for each interaction. This design choice is aimed at reducing token consumption and enhancing the overall efficiency of the LLM interaction. The constructed prompt is then forwarded to the configured LLM instance. The specific

model instance can be adjusted by modifying the relevant configuration parameters in the `.env` file within the Docker environment and selecting the appropriate deployment in the Microsoft Azure Portal[5]. Leveraging Azure-hosted LLMs enables control over data storage locations, thereby facilitating compliance with data protection regulations. Upon receiving the LLM response, the backend sends it to the JupyterLab extension for learner display.

Deployment and Use in Teaching Practice The tool has been deployed in multiple Python courses at RWTH Aachen University. For effective use, instructors should structure their Jupyter notebooks so that each task's description and the corresponding learner input are contained within the same notebook, following a format similar to that used by nbgrader[6]. Additionally, instructors are required to annotate relevant cells using metadata in accordance with the configurable pattern-matching mechanism described earlier. This enables the tool to identify and group task components, allowing for context-aware support. To ensure pedagogical robustness, the prompt generation mechanism was evaluated using fault injection to simulate adversarial input. When learners submit manipulative input (e. g., requesting sample solutions in a code cell), the Python backend applies safeguards that direct the LLM to withhold such sample solutions, regardless of whether the request appears in the task description, cell input, or expected output. These safeguards are critical in upholding the tool's educational integrity by encouraging active problem-solving rather than passive solution retrieval. The feedback generation time is mainly influenced by the length of the task description and learner code as part of the prompt, as well as the response latency of the Azure-hosted model, typically ranging from 1 to 5 seconds.

3 Conclusion and Future Work

This paper has presented an open-source tool[4] that integrates LLM-based feedback into the JupyterLab environment, aiming to enhance programming education through immediate, context-aware, and pedagogically grounded support. By embedding interactive feedback mechanisms within the Jupyter notebook interface and utilizing a configurable metadata scheme, the tool enables instructors to design learning tasks that facilitate individualized, automated feedback aligned with educational goals. Its modular architecture supports flexible deployment and allows for fine-grained customization of prompt templates, feedback categories, and learner interaction workflows. The tool has been successfully deployed in university-level Python programming courses, demonstrating its practical value in real teaching contexts. Future work will focus on incorporating session-level contextual data to further tailor feedback to learners' evolving needs. The architecture supports extension to other programming languages by adapting the language-specific prompt templates. This enables scalable, cross-language applicability in diverse instructional settings.

[5] Microsoft Azure Services, https://portal.azure.com/, last accessed: 2025-07-07
[6] nbgrader, https://nbgrader.readthedocs.io/en/stable/, last accessed: 2025-07-07

References

[HT07] Hattie, J.; Timperley, H.: The Power of Feedback. Review of Educational Research 77 (1), pp. 81–112, 2007, DOI: 10.3102/003465430298487.

[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, ACM, Dublin, Ireland, pp. 95–115, 2022, DOI: 10.1145/3571785.3574124.

[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.

[LKK25] Lohr, D.; Keuning, H.; Kiesler, N.: You're (Not) My Type - Can LLMs Generate Feedback of Specific Types for Introductory Programming Tasks? Journal of Computer Assisted Learning 41 (1), 2025, DOI: 10.1111/jcal.13107.

[Lo16] Loksa, D. et al.: Programming, Problem Solving, and Self-Awareness: Effects of Explicit Guidance. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16, ACM, New York, NY, USA, pp. 1449–1461, 2016, DOI: 10.1145/2858036.2858252.

[Me24] Messer, M. et al.: Automated Grading and Feedback Tools for Programming Education: A Systematic Review. ACM Trans. Comput. Educ. 24 (1), 2024, DOI: 10.1145/3636515.

[Na08] Narciss, S.: Feedback Strategies for Interactive Learning Tasks. In: Handbook of Research on Educational Communications and Technology (3rd ed.) Pp. 125–144, 2008.

Shared Sessions in MICRO: Enabling Collaborative Remote Experimentation in Embedded Systems Education

Jakob Czekansky [1], Justin Sauer [1], and Diethelm Bienhaus [1]

Abstract: Remote laboratories have become a key enabler for digital engineering education. While many platforms offer remote access to real hardware, few support synchronous collaboration among students. This paper presents a new feature of the MICRO platform—a remote lab system for embedded systems education developed at University of Applied Sciences Mittelhessen (THM)—which enables multiple users to interact with the same experiment in real-time through shared sessions. Features include session key sharing, join requests, role-based permissions, and synchronized input handling via Web-based communication protocol for full-duplex interaction (WebSocket) communication. This enhancement aligns with current pedagogical trends in collaborative and competence-based learning, allowing for pair programming, group debugging, and instructor-led walkthroughs. Compared to existing remote labs, MICRO's shared session model offers a more authentic and interactive environment for distributed learners. The paper outlines the didactic motivation, system design, and early usage experiences, contributing to the growing body of work on digital infrastructures for collaborative teaching and learning.

Keywords: Remote Laboratory, Collaborative Learning, Embedded Systems Education, Shared Sessions, Digital Teaching Infrastructure, Synchronous Experimentation, Educational Technology

1 Introduction

The increasing digitalization of engineering education has led to a proliferation of remote laboratories, especially in domains such as electronics, automation, and embedded systems [HDL+08; TSM+13]. These systems have proven essential in addressing limitations in physical lab access—whether due to scheduling conflicts, resource constraints, or global disruptions like the COVID-19 pandemic [AZD+22; OTS+21]. In parallel, educational paradigms are shifting toward collaborative, student-centered, and competence-based learning approaches, calling for digital infrastructures that support real-time, group-oriented interaction [AZD+22; Re22].

Remote labs, such as VISIR [TSM+13], iLab [HDL+08], and LabsLand [ORG+16], have laid the groundwork for accessible hardware experiments online. However, most of these systems rely on single-user reservation models or instructor-guided demonstrations, offering limited support for authentic, synchronous collaboration between students [OTS+21].

[1] University of Applied Sciences Mittelhessen, Institute for Engineering and Computer Science
Wiesenstraße 14, 35390 Gießen, Germany,
jakob.czekansky@mni.thm.de, https://orcid.org/0000-0001-8432-0092;
justin.sauer@mni.thm.de, https://orcid.org/0009-0007-3409-1595;
diethelm.bienhaus@mni.thm.de, https://orcid.org/0000-0002-9207-3412

doi:10.18420/delfi2025_45

In response, the THM developed MICRO[2]—a remote laboratory platform for embedded systems education that combines real hardware setups with a web-based, emulated interface. Originally designed to provide individual access to microcontroller experiments in the Microprocessor Technology course, MICRO has evolved to address broader pedagogical challenges across multiple courses [CSB23; Cz24].

This paper introduces a recent extension to the MICRO platform: shared sessions. This new feature enables multiple users to collaboratively control and observe a live experiment in real-time, including shared interaction with physical signals, code upload, and system feedback. Unlike traditional setups, MICRO's shared sessions support role-based moderation, join requests, and session key authentication—enabling modern learning scenarios such as pair programming, collaborative debugging, and remote mentoring.

By embedding collaborative functionality directly into the lab infrastructure, MICRO aligns with DELFI 2025's focus on "digital technologies for collaborative teaching and learning." This paper outlines the pedagogical motivation, technical implementation, and usage scenario of shared sessions in MICRO, contributing a new layer of co-constructive interactivity to the field of remote embedded systems education.

2 Related Work

Remote laboratories have become a cornerstone of modern engineering education, providing students with access to hands-on hardware experiments without requiring their physical presence [Ma23]. Foundational initiatives such as VISIR[TSM+13] and iLab [HDL+08] demonstrated how students could access complex electronics and control systems through scalable, networked infrastructures. These efforts established key design patterns for distributed lab access that remain influential today.

More recent platforms, such as LabsLand [ORG+16] and GOLDi [He16], extend these ideas by enabling shared access to remote experiments across institutional boundaries. Their implementations prioritize stability, accessibility, and global scalability. Similarly, the practable.io project [Re22] has contributed valuable open-source approaches for remote lab provisioning, with a particular focus on accessibility and modular design.

These systems typically prioritize individual student access or instructor-led demonstration. Although many offer mechanisms for observation or asynchronous collaboration (e. g., joint reporting, screen sharing), real-time, synchronous interaction by multiple users within the same experiment session is generally not a central feature of their current implementations.

Ortelt et al. [OTS+21] and Aubel et al. [AZD+22] emphasize the need for laboratory systems that adapt to diverse teaching contexts and user profiles rather than requiring pedagogy to conform to rigid lab infrastructures.

[2] https://micro.mni.thm.de/

The MICRO platform, developed at THM, addresses this evolving need by enabling shared sessions—a feature that allows multiple students to interact with the same physical embedded systems experiment concurrently. Unlike traditional single-user reservation models, shared sessions support parallel control, role-based interaction, and instructor moderation within a unified session interface. This advancement builds on MICRO's previously established technical foundation [CSB23] and its integration into structured teaching modules [Cz24]. Recent extensions include support for analog and digital signal emulation, Universal Asynchronous Receiver-Transmitter (UART)-based interaction, and integrated feedback mechanisms [CSR+25].

Through this shared access model, MICRO supports educational formats such as pair programming, collaborative debugging, and instructor-guided walkthroughs [Ha11]. This development aligns with the broader educational goals articulated in the CrossLab initiative, particularly concerning modularity, interoperability, and competence-oriented learning [AZD+22].

3 The MICRO Platform in Teaching

The MICRO platform has been developed and deployed at THM as a flexible and scalable remote laboratory environment for teaching embedded systems. It offers direct access to real hardware (e. g., microcontrollers, sensors, LEDs, displays, motors) via a web-based frontend that includes live video streams, digital/analog signal emulation, virtual button control, and a UART interface. Students typically interact with the platform through a structured course integration in modules like Microprocessor Technology and Technical Computer Science II, with growing adoption in other contexts such as autonomous driving and FPGA development [Cz24].

Once students log in through the institutional Lightweight Directory Access Protocol (LDAP) system, they can access a reservation system that lists all available stations, which was historically bound to a single user. Each station encapsulates a modular experiment setup, a host PC, and supporting services for code execution, signal emulation, and monitoring. With the growing emphasis on collaborative learning and student feedback highlighting the desire for joint experimentation, the MICRO team introduced the shared session functionality to meet this pedagogical need.

4 Shared Sessions: Design and Implementation

The new shared session feature allows multiple users to participate in the same lab session through two core mechanisms:

1. Session Key Sharing. When a student reserves a station, they may activate shared mode to generate a randomized access key. This key allows others to join by clicking the "Join Shared Session" button in the station list.

2. Join Request System. Alternatively, users can submit a join request to the session leader, who receives a real-time notification and may approve or deny access. Requesting users see a loading screen while waiting, and anti-spam mechanisms prevent repeated requests.

Role Management and User Interface (UI) Features. The session leader retains moderation privileges, being provided with a participant list by the session UI that includes the ability to remove users from the session. All interactions—such as button presses, potentiometer adjustments, and UART input—are synchronized across clients via WebSocket communication to maintain a consistent session state.

This implementation supports collaborative debugging, live demonstrations, and student-to-student assistance. It fosters authentic group learning, comparable to physical co-presence at a shared workstation. Early classroom integration has shown positive feedback from students regarding the usability and collaborative experience.

5 Conclusion and Outlook

The shared session feature extends the MICRO platform beyond its original single-user paradigm, introducing synchronous, collaborative functionality into remote embedded systems education. Supporting both key-based and moderated session access empowers students to engage in co-constructive experimentation, which is increasingly vital in modern, project-based engineering curricula.

Future directions include deeper integration of communication tools (e. g., simple text chats), collaborative code editing, and the use of session analytics to evaluate group dynamics and learning outcomes. MICRO's evolution exemplifies how remote labs can move from access enablers to true environments for collaborative teaching and learning.

6 Acknowledgments

This work was supported by the *Department 06 – Mathematics, Natural Sciences, Computer Science* and the *Institute for Engineering and Computer Science* at THM. We want to express our gratitude for the trust placed in us by our department for this project, for the financial support, and, of course, to the many student assistants who are implementing and further developing the entire MICRO project, as well as the new prototypical feature of shared sessions.

References

[AZD+22] Aubel, I.; Zug, S.; Dietrich, A., et al.: Adaptable Digital Labs – Motivation and Vision of the CrossLab Project. In: 2022 IEEE German Education Conference (GeCon). IEEE, 2022, DOI: 10.1109/GeCon55699.2022.9942759.

[CSB23] Czekansky, J.; Schauer, M.; Bienhaus, D.: MICRO: The Remote Microprocessor Lab. In: Mit Automatisierung gegen den Klimawandel. HTWK Leipzig, pp. 1–4, 2023.

[CSR+25] Czekansky, J.; Sauer, J.; Ruppert, T.-N., et al.: Enhancing Interaction with Remote Microcontroller-Based Experiments Using a Low-Cost Multi-Function Instrument, Submitted to 22nd International Conference on Smart Technologies & Education (STE), 2025.

[Cz24] Czekansky, J. et al.: Advancing Embedded Systems Education Through Remote Laboratory Integration. In (May, D.; Auer, M.; Kist, A., eds.): Online Labs in Engineering Education. Vol. 1135, LNNS, Springer, Cham, pp. 203–222, 2024.

[Ha11] Hanks, B. et al.: Pair programming in education: a literature review. Computer Science Education 21 (2), pp. 135–173, 2011, DOI: 10.1080/08993408.2011.579808.

[HDL+08] Harward, V.; Del Alamo, J.; Lerman, S., et al.: The iLab Shared Architecture: A Web Services Infrastructure to Build Communities of Internet Accessible Laboratories. Proceedings of the IEEE 96 (6), pp. 931–950, 2008.

[He16] Henke, K. et al.: The remote lab cloud "GOLDi-labs.net". In: 13th International Conference on Remote Engineering and Virtual Instrumentation (REV). IEEE, pp. 37–42, 2016.

[Ma23] May, D. et al.: Online Laboratories in Engineering Education Research and Practice. In (Johri, A., ed.): International Handbook of Engineering Education Research. 1st ed., Routledge, New York, chap. 24, pp. 525–552, 2023, DOI: 10.4324/9781003287483-29.

[ORG+16] Orduna, P.; Rodriguez-Gil, L.; Garcia-Zubia, J., et al.: LabsLand: A sharing economy platform to promote educational remote laboratories maintainability, sustainability and adoption. In: 2016 IEEE Frontiers in Education Conference (FIE). IEEE, pp. 1–6, 2016.

[OTS+21] Ortelt, T.; Terkowsky, C.; Schwandt, A., et al.: Die digitale Zukunft des Lernens und Lehrens mit Remote-Laboren. In: Digitalisierung in Studium und Lehre gemeinsam gestalten. Springer VS, pp. 553–575, 2021.

[Re22] Reid, D. et al.: Open-source remote laboratory experiments for controls engineering education. International Journal of Mechanical Engineering Education 50 (4), pp. 828–848, 2022.

[TSM+13] Tawfik, M.; Sancristobal, E.; Martin, S., et al.: VISIR for Remote Wiring and Measurement of Electronic Circuits on Breadboard. IEEE Transactions on Learning Technologies 6 (1), pp. 60–72, 2013.

LearnMate

KI-gestützter Lernassistent zur Vor- und Nachbereitung universitärer Lehrveranstaltungen

Louis Dunkel [1], Imran Hossain[1], Patrick Sacher[1] und Hasan Abu-Rasheed[1]

Abstract: Studierende stehen im Studium häufig vor der Herausforderung, umfangreiche Lernmaterialien eigenständig zu strukturieren und inhaltlich zu erschließen – insbesondere in großen, vorlesungszentrierten Lehrformaten mit begrenztem individuellem Feedback durch Lehrpersonen. Der vorgestellte Prototyp ist ein KI-gestützter und datenschutzkonformer Lernassistent zur studienbegleitenden Unterstützung. Er wurde in eine wirtschaftswissenschaftlichen Vorlesung integriert und ermöglicht Studierenden, Inhalte gezielt zu vertiefen, Fragen zu den Inhalten zu stellen und sich auf Prüfungen vorzubereiten. Das System basiert auf einer lokal gehosteten Retrieval-Augmented-Generation-Architektur (RAG) mit angebundener Vektor-Datenbank und nutzt im Backend Flowisebasierte KI-Agenten zur semantischen Verarbeitung und dynamischen Antwortgenerierung. Die Evaluation der Ergebnisse ist derzeit in Arbeit. Perspektivisch soll das System in weiteren Fachbereichen als modulares KI-Ökosystem eingesetzt werden können.

Keywords: Künstliche Intelligenz, RAG-Systeme, Bildung, KI in der Hochschule

1 Einleitung

Die digitale Transformation verändert die Hochschullehre grundlegend. Neue Technologien eröffnen Potenziale, um Lernprozesse individueller, flexibler und zugänglicher zu gestalten. Besonders der Einsatz von Künstlicher Intelligenz (KI) verspricht eine nachhaltige Verbesserung und Individualisierung von Lernunterstützungen, sowohl für Studierende als auch für Lehrende. Während bestehende KI-Systeme häufig als isolierte Einzellösungen agieren, verfolgt das von uns entwickelte System einen ganzheitlichen, adaptiven Ansatz: Der Prototyp wurde als datenschutzkonformer Lernassistent entwickelt, der sich in bestehende universitäre Strukturen integrieren lässt. Perspektivisch soll er sich, mit verhältnismäßig geringem Aufwand, auf andere Lehrszenarien übertragen lassen können.

KI-gestützte Chatbots verfügen über das Potenzial, Hochschullehre inhaltlich, organisatorisch und analytisch zu transformieren. Sie sind in der Lage personalisierte Lernpfade

[1] Goethe-Universität Frankfurt am Main, studiumdigitale, Eschersheimer Landstr. 155/157, 60323 Frankfurt am Main,
dunkel@sd.uni-frankfurt.de, https://orcid.org/0009-0008-5148-5657
hossain@sd.uni-frankfurt.de
sacher@sd.uni-frankfurt.de
rasheed@sd.uni-frankfurt.de

doi:10.18420/delfi2025_46

anzubieten, unmittelbares Feedback zu liefern, kollaborative Aktivitäten zu unterstützen und zugleich administrative Routinen wie Termin- oder Kursmanagement weitestgehend zu automatisieren [Ch24; Ra23]. Diese Potenziale unterstreicht auch eine systematische Literaturübersicht von Wollny et al., die aufzeigt, dass Bildungs-Chatbots Bildung leichter zugänglich machen und die Lernmotivation der Studierenden erhöhen können [Wo21]. Jüngste Studien zu Retrieval-Augmented-Generation (RAG) zeigen eine mögliche, geeignete Technologie zur Umsetzung auf und unterstreichen diese Chancen, weisen jedoch auf offene Fragen hin: Swacha & Gracel fanden in 47 RAG-Projekten bislang kaum belastbare Wirksamkeitsnachweise [Sw25]; Henkel et al. belegen zwar eine Präferenz der Lernenden für kontextualisierte RAG-Antworten, aber auch Akzeptanzverluste bei überlangen Textzitaten [He24]. Auf Dialogebene adressieren Schmucker et al. Fehlkonzepte durch zwei kooperierende LLM-Agenten [Sc24], während Rajala et al. in einem Informatik-Kurs sowohl hohe Nutzung als auch Vertrauensprobleme gegenüber ChatGPT verzeichnen [Rj23]. Zugänglichkeitsstudien von Grassini et al. bemängeln fehlende barrierefreie Interfaces [Gr24], und Thüs at al. zeigen, dass Studierende KI-Lernbegleiter nur dann dauerhaft einsetzen, wenn Antworten prägnant und klar domänenspezifisch sind [Th24]. Unser Prototyp begegnet diesen Lücken durch ein kursspezifisches RAG-Backend, das ausschließlich geprüfte Lehrmaterialien einbindet, und durch eine multimodale Oberfläche, sowohl fachliche Genauigkeit als auch Zugänglichkeit sicherstellt.

2 Zielsetzung und Anwendung

Ziel des Systems ist es, Studierende bei der Vor- und Nachbereitung von Lehrveranstaltungen individuell zu unterstützen. Es liefert auf Anfrage Zusammenfassungen, kontextbezogene Erklärungen von Fachbegriffen und Konzepten sowie prüfungsnahe Fragestellungen und vertiefende Interpretationen. Dabei greift es auf Inhalte der jeweiligen Veranstaltung zurück und generiert Antworten auf Basis konkreter Materialien. Studierende können beispielsweise gezielt nach dem in der Vorlesung verwendeten Beispiel zur Erläuterung eines Konzepts fragen und erhalten eine passende, verständlich aufbereitete Rückmeldung.

Gleichzeitig bietet der Prototyp auch für Lehrende einen Mehrwert, da häufige Verständnis- und Routinefragen automatisiert beantwortet werden können. So entsteht Freiraum für komplexere didaktische Interaktionen, insbesondere in heterogenen Lerngruppen. Die intuitive, multimodale Benutzeroberfläche sowie die zugrunde liegende KI-gestützte Transkription fördern außerdem die digitale Barrierefreiheit und ermöglichen Studierenden mit unterschiedlichen Bedürfnissen einen gleichberechtigten Zugang zum Lernangebot.

Ein weiteres Ziel ist es, einen übertragbaren und praxistauglichen Workflow zu entwickeln, der eine spätere Ausweitung auf andere Fachbereiche erlaubt. Das System soll perspektivisch als universitätsweites, modulares KI-Ökosystem eingesetzt werden, in dem fachspezifische Assistenten schnell und unkompliziert erstellt werden können.

3 Technische Umsetzung

Der Prototyp basiert auf einer Retrieval-Augmented-Generation-Architektur, die auf strukturierte Inhalte aus einer lokal gehosteten Vektor-Datenbank zugreift [Le21]. Grundlage der Wissensbasis sind aufbereitete Lehrmaterialien, darunter Vorlesungstranskripte, Präsentationsfolien und Tutoriumsmaterialien, die mithilfe von Large Language Models (LLMs), gemäß definierter Kriterien, semantisch vorverarbeitet werden.

RAG-Systeme stellen eine Möglichkeit dar, generative Sprachmodelle mit externen, durchsuchbaren Wissensquellen zu verbinden und ermöglichen dadurch eine kontextbezogene Antwortgenerierung auf Basis konkreter Dokumente. Im Hochschulkontext eröffnet dies neue Potenziale: Studierende und Lehrende erhalten zielgerichtete, präzise und aktuelle Informationen, etwa zur Vertiefung von Inhalten, zur Klärung fachlicher Fragen oder zur Prüfungsvorbereitung. Die externe, nicht-parametrische Wissensbasis kann kontinuierlich aktualisiert werden, ohne dass das zugrunde liegende Modell neu trainiert werden muss, liefert prüfbare Quellen und wahrt dank On-Premise-Hosting volle DSGVO-Konformität. Dadurch wird eine adaptive, transparente und nachhaltige Lernunterstützung realisiert.

Im Backend kommen Flowise[2]-basierte KI-Agenten zum Einsatz, die die aufbereiteten Inhalte analysieren und standardisiert verfügbar machen. Flowise bringt eine No-Code-GUI mit, über die Lehrende neue Agenten und Materialpipelines per Drag-and-Drop konfigurieren können. Dies erleichtert die Erweiterung des Systems durch Nicht-technikexpert*innen. Die zusätzliche Retrieval-Schicht erhöht jedoch die Architekturkomplexität und kann bei schlecht strukturierten Materialien zu längeren Antwortzeiten oder Fehlretrieval führen. Flowise bietet weniger Feintuning-Möglichkeiten und skaliert nur bedingt bei sehr großen Nutzerzahlen, während das lokale Hosting laufende Infrastruktur- und Wartungskosten verursacht. Wir evaluieren Flowise und den gesamten Ablauf derzeit noch im Piloteinsatz. Falls sich in weiteren Tests Limitierungen zeigen, wird in einer späteren Version ein alternatives Orchestrierungstool (z. B. LangChain oder RAGFlow) eingesetzt. Das Scouting alternativer, geeigneter Backend-Tools wird derzeit noch durchgeführt.

Für die Antwortgenerierung ist derzeit ChatGPT-4o über eine API integriert. Ein Modellwechsel ist dank der modularen Architektur jederzeit möglich. Das eigens entwickelte Frontend erlaubt eine intuitive, multimodale Interaktion mit dem System und lässt sich flexibel an unterschiedliche Fachkontexte und Lehrformate anpassen. Ungeachtet der offenen Optimierung liefert der aktuelle Prototyp gegenüber reinen ChatGPT-Plug-ins (domänenunspezifisch, Cloud-abhängig) oder kostenintensivem Fine-Tuning (statische Inhalte, fehlende Quellen) bereits jetzt eine überzeugende Balance aus fachlicher Präzision, Transparenz und datenschutzrechtlicher Konformität.

[2] https://flowiseai.com/

4 Evaluation und Pilotphase

Aktuell befindet sich das System in einer ersten Testphase. Im Rahmen einer Vorlesung aus dem Fachbereich Wirtschaftswissenschaften nutzen Studierende den Prototypen zur individuellen Lernunterstützung. Die Interaktionen werden mit Hilfe des Tools Langfuse[3] anonymisiert aufgezeichnet zur Auswertung bereitgestellt. Dabei werden nicht nur die gestellten Fragen und generierten Antworten erfasst, sondern auch dokumentiert, auf welche konkreten Quellen das System zurückgegriffen hat, etwa auf bestimmte Folien, Transkriptausschnitte oder Zusatzmaterialien. Dies ermöglicht eine gezielte Analyse der inhaltlichen Qualität, Nachvollziehbarkeit und Relevanz der Antworten.

Die Wissensbasis beruht auf einer im Vorjahr aufgezeichneten Vorlesung, deren Transkript strukturiert wird, sowie aus ergänzenden Folien, Übungs- und Prüfungsmaterialien. Die Rohdaten durchlaufen eine LLM-gestützte Pipeline, die Aufgaben wie Segmentierung, Qualitätsprüfung und Vektorisierung übernimmt. Die daraus resultierenden Embeddings werden in eine Vektordatenbank überführt. Ergänzend sind je Semesterpaket einige manuelle oder semi-automatisierte Schritte nötig, etwa die strukturierte Aufbereitung und Systemintegration der Materialien, mit in der Regel überschaubarem Aufwand. Perspektivisch soll eine Low-Code-Schnittstelle Lehrenden ein eigenständiges Content-Onboarding ermöglichen.

Die Nutzung wird semesterweise mittels standardisierter Umfragen erfasst und in eine mehrsemestrige formative Evaluation überführt, um den Assistenten evidenzbasiert zu optimieren.

5 Fazit

Der Prototyp überbrückt die Betreuungslücke großer Vorlesungen mit einem DSGVO-konformen, modular skalierbaren RAG-System, das binnen Stunden auf neue Lehrveranstaltungen ausgerollt wird und künftig an LMS sowie Bibliothekskataloge andockt. Offene Punkte wie Materialqualität und Flowise-Skalierung, lassen sich durch strengere Content-Pipelines bzw. eine alternative Toolchain lösen. Die Barrierefreiheit wird derzeit nach BITV/WCAG auditiert; erste Erfolgskriterien sind bereits erreicht.

Literaturverzeichnis

[Ch24] Chukwuere, J.E.: The future of generative AI chatbots in higher education. arXiv, 2024.

[Gr24] Grassini, E. et al.: A systematic review of chatbots in inclusive healthcare: insights from the last 5 years. Universal Access in the Information Society 1/24, S. 195–203, 2024.

[3] https://langfuse.com/

[He24] Owen Henkel et al.: Retrieval-augmented Generation to Improve Math Question-Answering: Trade-offs Between Groundedness and Human Preference. 24, 2024.

[Le21] Lewis, P. et al.: Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv, 2021.

[Ra23] Rathore, Dr.B.: Future of AI & Generation Alpha: ChatGPT beyond Boundaries. Eduzone : international peer reviewed/refereed academic multidisciplinary journal 01/23, S. 63–68, 2023.

[Rj23] Rajala, J. et al.: „Call me Kiran" – ChatGPT as a Tutoring Chatbot in a Computer Science Course": Mindtrek '23: 26th International Academic Mindtrek Conference. ACM, Tampere Finland, S. 83–94, 2023.

[Sc24] Schmucker, R.; Xia, M.; Mitchell, T.: Ruffle&Riley: Insights from Designing and Evaluating a Large Language Model-Based Conversational Tutoring System: Springer Nature Switzerland, Cham, S. 75–90, 2024.

[Sw25] Swacha, J.; Gracel, M.: Retrieval-Augmented Generation (RAG) Chatbots for Education: A Survey of Applications. Applied Sciences 8/25, S. 4234, 2025.

[Th24] Thüs, D.; Malone, S.; Brünken, R.: Exploring generative AI in higher education: a RAG system to enhance student engagement with scientific literature. Frontiers in Psychology 24, S. 1474892, 2024.

[Wo21] Wollny, S. et al.: Are We There Yet? - A Systematic Literature Review on Chatbots in Education. Frontiers in Artificial Intelligence 21, 2021.

Kompagnon für das Verfassen von Hausarbeiten

Nghia Duong-Trung [1], Xia Wang [1], Rahul Rajkumar Bhoyar [1], Angelin Mary Jose [1], Silke Elisabeth Wrede [2], Lars van Rijn [2], Theresa Panse [2], Claudia de Witt [2], and Niels Pinkwart [1]

Abstract: Der Beitrag präsentiert eine Demonstration des „Term Paper Writing Companion", einem KI-gestützten akademischen Unterstützungssystem, das entwickelt wurde, um das Lernen von Studierenden bei forschungsorientierten Schreibaufgaben zu unterstützen. Ursprünglich für einen B.A.-Studiengang in den Kultur- und Sozialwissenschaften an der FernUniversität in Hagen entwickelt, integriert es Large Language Models (LLMs), expertenbasierte Feedback-Mechanismen und Literaturempfehlungs-Maschinen (recommender engines). Das System bietet personalisiertes, formatives Feedback zu möglichen Themen und Fragestellungen von Hausarbeiten und adressiert gleichzeitig Probleme wie kognitive Überlastung der Studierenden und zeitliche Ressourcenengpässe bei den Lehrkräften. Diese Demo zeigt die Architektur, den Arbeitsablauf und die Benutzererfahrung und veranschaulicht das Potenzial des Systems, das forschungsbasierte Lernen und die akademische Autonomie in verschiedenen Hochschulumgebungen zu verbessern.

Keywords: Generative KI, Recommendersystem, Akademisches Schreiben.

1 Einführung

Die Auswahl eines klar definierten Forschungsthemas ist eine entscheidende Herausforderung im akademischen Studium, insbesondere in Kursen, die unabhängiges Forschen und kritisches Denken in den Vordergrund stellen. Das Tool „Term Paper Writing Companion" wurde entwickelt, um Studierende während dieses Prozesses zu unterstützen, indem es ihnen mithilfe generativer KI intelligente Hilfestellungen bietet. Das Tool hilft Lernenden dabei, Forschungsfragen zu verfeinern und geeignete wissenschaftliche Literatur zu identifizieren, indem es strukturiertes, auf Eingabeaufforderungen basierendes Feedback, Sentimentanalyse und Recommender-Techniken kombiniert. Im Gegensatz zu rein generativen KI-Ansätzen, die zu einer übermäßigen Abhängigkeit von automatisierten Ergebnissen führen können, wurde das Tool ausdrücklich entwickelt, um die Autonomie der Studierenden zu fördern,

1 German Research Center for Artificial Intelligence (DFKI), Alt-Moabit 91C, 10559 Berlin, Germany,
nghia_trung.duong@dfki.de, https://orcid.org/0000-0002-7402-4166;
xia.wang@dfki.de, https://orcid.org/0000-0003-0849-5645;
rahul_rajkumar.bhoyar@dfki.de, https://orcid.org/0009-0005-7249-7305;
angelin_mary.jose@dfki.de, https://orcid.org/0009-0005-5026-8990;
niels.pinkwart@dfki.de, https://orcid.org/0009-0005-5026-0000
2 FernUniversität in Hagen, Universitätsstraße 11, 58097 Hagen, Germany,
silke.wrede@fernuni-hagen.de, https://orcid.org/0000-0002-4957-9792;
lars.vanrijn@fernuni-hagen.de, https://orcid.org/0000-0002-8381-2666;
theresa.panse@fernuni-hagen.de, https://orcid.org/0009-0006-5734-3161;
claudia.dewitt@fernuni-hagen.de, https://orcid.org/0000-0001-6478-8392

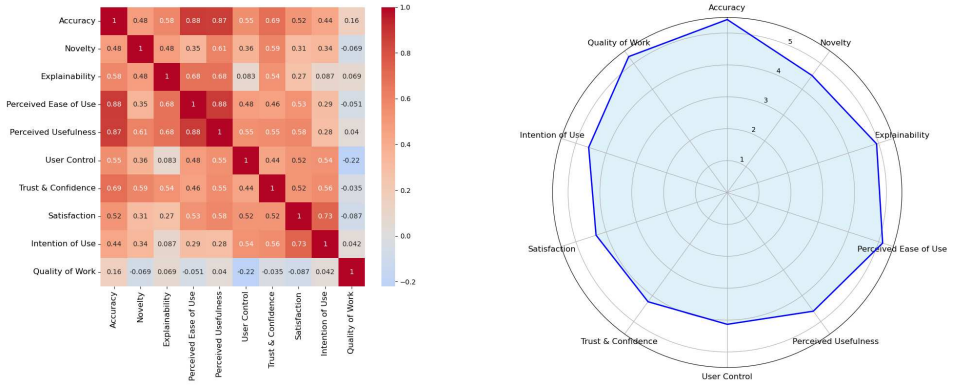
doi:10.18420/delfi2025_47

indem es personalisierte, forschungsorientierte Unterstützung bietet. Seine Empfehlungen dienen als kognitive Unterstützung [KLC22; Mu23] und helfen den Studierenden, ihre Ideen zu entwickeln, mit Fortführung der eigenen akademischen Argumentation und Entscheidungsfindung. Um die Wirksamkeit des Tools zu bewerten, haben wir eine Pilotstudie im Rahmen des B.A.-Studiengangs durchgeführt und dabei das CRS-Que-Framework [Ji24] angewendet, um die Benutzerfreundlichkeit des Systems und seine pädagogische Wirkung zu bewerten.

2 Systemübersicht und Architektur

Das System besteht aus drei Hauptkomponenten: einer Frontend-Webschnittstelle, einer Backend-Orchestrierungsschicht und einem von der Universität gehosteten LLM- Inferenzserver Flexi [Ze24]. Das Frontend bietet den Studierenden eine intuitive und personalisierte Oberfläche für die Einreichung von Forschungsvorschlägen, einschließlich erster Themenbeschreibungen und damit verbundener Forschungsfragen. Die Studierenden können zwischen mehreren LLMs wählen – wie Mixtral-8x7B, LLaMA-3-8B und anderen –, Feedback-Ergebnisse vergleichen und verschiedene Perspektiven erkunden. Sie können auch Feedback-Verläufe einsehen, Entwürfe überarbeiten und erneut einreichen sowie Empfehlungen steuern. Diese Oberfläche wurde entwickelt, um Transparenz zu fördern, iterative Überarbeitungen zu erleichtern und eine strukturierte Umgebung für akademische Erfahrungen zu bieten. Im Backend verarbeitet eine Reihe von Analysemodulen die Eingaben der Studierenden. Ein Modul zur Sentimentanalyse schätzt das Vertrauen der Studierenden in das eigene Vorhaben ein (confidence), woraufhin entschieden wird, ob eine wissensbasierte oder eine expertenbasierte Empfehlungsmaschine aktiviert wird. Bei Studierendeneinreichungen mit der Bewertung „hohes Vertrauensniveau“ empfiehlt das System akademische Literatur, die auf das vorgeschlagene Thema der Studierenden abgestimmt ist. Bei Einreichungen mit der Bewertung „geringes Vertrauensniveau“ bietet das Tool grundlegende Anleitungen, um den Studierenden zu helfen, ihren Forschungsschwerpunkt zu klären oder neu zu strukturieren. Diese Wege stellen sicher, dass die Studierenden eine Unterstützung erhalten, die ihrem aktuellen Vorbereitungs- und Vertrauensniveau entspricht. Die Eingaben werden anhand einer Reihe strukturierter promptbasierter Bewertungsmodule validiert, die die Klarheit, Durchführbarkeit und Übereinstimmung von Themen und Fragen bewerten. Diese Validierung umfasst individuelle Überprüfungen für jede Forschungsfrage und die Kohärenz zwischen Themen und Fragen. Der Prozess der Feedback-Generierung wird durch Multi-Shot-Prompting verbessert, bei dem das System historische Betreuungsdaten aus früheren Kohorten einbezieht, um nuancierte Antworten von Lehrkräften nachzuahmen. Diese Multi-Turn-Beispiele wurden aus echten Moodle-basierten Dialogen zwischen Studierenden und Lehrkräften abgeleitet, sodass das System Feedback generieren kann, das pädagogische Tiefe und akademische Erwartungen widerspiegelt. Alle LLM-Anfragen werden über Flexi geleitet, eine von der Universität kontrollierte Inferenzinfrastruktur, die die sichere lokale Ausführung vieler Open-Source-Modelle ermöglicht. Flexi macht die Abhängigkeit von APIs von Drittanbietern überflüssig und gewährleistet die Einhaltung der Datenschutzrichtlinien

der Einrichtung und der Europäischen Datenschutz-Grundverordnung (DSGVO). Parallel dazu ruft ein semantisches Recommender-Modul relevante akademische Literatur ab, indem es die Beiträge der Studierenden mit mehrsprachigen Satzkodierern einbettet und sie mit dem kuratierten Repository der Universität für akademische Texte abgleicht. Die Systemarchitektur wurde unter Berücksichtigung von Modularität und zukünftiger Integration entwickelt. Sie unterstützt Verbindungen zu Lernmanagementsystemen, Feedback-Dashboards und externen Bewertungsmodulen. Die Trennung zwischen Interaktionslogik und LLM-Verarbeitung verbessert nicht nur die Systemleistung, sondern stellt auch sicher, dass die studentischen Komponenten reaktionsschnell bleiben und gleichzeitig die Transparenz und das Vertrauen in die von KI generierten Vorschläge gewahrt bleiben. Erste Ergebnisse aus der Einführung im Sommersemester 2025 `https://tprs.fernuni-hagen.de/login`, inwieweit das System zur Optimierung der Betreuung beiträgt, die Feedback-Verzögerung reduziert und das engagement der Studierenden bei akademischen Schreibaufgaben fördert, liegen vor.



(a) Korrelations-Heatmap der Bewertungsdimensionen.

(b) Radardiagramm der Durchschnittswerte der Kategorien.

Fig. 1: Ergebnisse der Evaluation.

3 Bereitstellung und Evaluierung

Um die Wirksamkeit des Term Paper Writing Companion zu bewerten, wurde im Wintersemester 2024 eine Pilotstudie mit 15 Studierenden eines B.A.-Studiengangs durchgeführt. Die Bewertung erfolgte anhand des CRS-Que-Rahmenwerks, das 10 zentrale Bewertungsdimensionen mit insgesamt 32 Unterdimensionen umfasst. Die Studierenden bewerteten jeden Punkt auf einer 7-Punkte-Likert-Skala. Das Radardiagramm (siehe Abbildung 1b) vermittelt ein klares visuelles Profil der wahrgenommenen Stärken des Systems. Dimensionen wie Genauigkeit (5,42), Arbeitsqualität (5,27) und wahrgenommene Benutzerfreundlichkeit (5,13) wurden hoch bewertet, was darauf hindeutet, dass die Studierenden das System als effektiv und benutzerfreundlich empfanden. Im Gegensatz dazu erhielten Benutzerkontrolle (4,13) und Vertrauen & Zuversicht (4,23) vergleichsweise niedrigere Bewertungen,

was auf Verbesserungsmöglichkeiten hinsichtlich Transparenz und Interaktionsflexibilität hindeutet. Ergänzend dazu zeigte die in Abbildung 1a dargestellte Korrelations-Heatmap aussagekräftige Wechselbeziehungen zwischen den Bewertungsmetriken. So war beispielsweise die Erklärbarkeit stark mit Vertrauen und wahrgenommener Nützlichkeit verbunden, was die Bedeutung verständlicher KI-generierter Rückmeldungen für die Förderung von Vertrauen und nachhaltiger Nutzung unterstreicht. In ähnlicher Weise zeigte die Benutzerfreundlichkeit positive Korrelationen mit Zufriedenheit und Nutzungsabsicht, was die Benutzerfreundlichkeit als zentralen Treiber für das Engagement.

4 Abschluss

Diese Demonstration stellt das Term Paper Writing Companion System vor, ein modulares KI-gestütztes Unterstützungssystem, das Studierenden durch den kognitiv anspruchsvollen Prozess der Formulierung und Verfeinerung akademischer Themen begleitet. Durch die Integration einer strukturierten, promptbasierten Bewertung, hybrider Empfehlungsmechanismen und einer datenschutzkonformen LLM-Inferenz über den FLEXI-Server schließt das System die Lücke zwischen autonomem und angeleitetem Lernen. Die Pilotbewertung unter Verwendung des CRS-Que-Frameworks zeigt vielversprechende pädagogische Ergebnisse, insbesondere in Bezug auf die wahrgenommene Genauigkeit, Benutzerfreundlichkeit und Qualität der studentischen Arbeit. Gleichzeitig zeigen Bereiche wie Benutzerkontrolle und Vertrauen wichtige Richtungen für die weitere Entwicklung auf. Unsere Ergebnisse deuten darauf hin, dass KI-gesteuerte Scaffolding-Tools wie unser vorgeschlagenes System das forschungsbasierte Lernen im Fernunterricht sinnvoll unterstützen können, indem sie zeitnahes formatives Feedback liefern und gleichzeitig die Belastung der Lehrenden reduzieren.

Literaturverzeichnis

[Ji24] Jin, Y.; Chen, L.; Cai, W.; Zhao, X.: CRS-Que: A user-centric evaluation framework for conversational recommender systems. ACM Transactions on Recommender Systems 2 (1), S. 1–34, 2024.

[KLC22] Kim, J.; Lee, H.; Cho, Y. H.: Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. Education and Information Technologies 27 (5), S. 6069–6104, 2022.

[Mu23] Munshi, A.; Biswas, G.; Baker, R.; Ocumpaugh, J.; Hutt, S.; Paquette, L.: Analysing adaptive scaffolds that help students develop self-regulated learning behaviours. Journal of Computer Assisted Learning 39 (2), S. 351–368, 2023.

[Ze24] Zesch, T.; Hanses, M.; Seidel, N.; Aggarwal, P.; Veiel, D.; De Witt, C.: Flexible LLM Experimental Infrastructure (FLEXI)–Enabling Experimentation and Innovation in Higher Education Through Access to Open LLMs. In: 2024 21st International Conference on Information Technology Based Higher Education and Training (ITHET). IEEE, S. 1–8, 2024.

Web-Based Prototype of a Branching Scenario Editor for Adaptive Learning



Andreas Dymek [1], Frederic Salmen [1], and Ulrik Schroeder [1]

Abstract: As learning becomes both more digital and more personalized, the need for adaptive learning experiences grows, where learning could be tailored to the student's individual needs or branch off onto personalized learning paths based on the student's performance. While tools for creating branching scenarios exist, they often come from outside the educational domain—such as gamebook or interactive narrative editors—and are not tailored to educators' needs or require coding expertise. Furthermore, branching scenarios can be combined with explorables, which are web-based, interactive, and multimodal learning resources. We present a novel branching scenario editor on a FAIR technical basis, tailored to the educational domain. It allows educators to create learning paths based on quiz elements and eliminating the need for coding, ensuring accessibility for all educators.

Keywords: Explorables, FAIR, gamebooks, adaptive learning

1 Motivation & Related Work

Advancements in educational technology have significantly redefined the nature of learning materials and reshaped student learning. Traditional material such as worksheets can become digital, featuring interactive and multimodal learning experiences. Such digital worksheets, called explorables, promote student learning and engagement [Fi05; Ho20] and support personalized learning experiences through self-paced learning and real-time feedback [Se19]. One key aspect of personalized learning is the tailoring of instructions to meet the needs of individual students [SS20]. Theoretically, explorables open up opportunities to create adaptive learning experiences, as they could consider the performance of a student on a quiz to branch off onto personalized learning paths or include decision trees where the student can make self-directed choices. Such functionality is similar to a digital gamebook, which Möslein-Tröppner and Bernhard [MB18] presented as a suitable method for learning. Moreover, Noichl and Schroeder [NS23] classified digital gamebooks based on their potential to facilitate learning goals, highlighting their relevance for teaching. We introduce an authoring tool that enables educators to create branching scenarios: Adaptive learning resources tailored to students' individual needs. Different authoring tools for creating branching scenarios exist; however, these tools are mostly not tailored for educational purposes, such as gamebook or interactive digital narrative (IDN) editors. Research on such editors is limited and lacks academic focus, as noted by Koenitz [Ko18]. Yet, existing

[1] RWTH Aachen University, Learning Technologies, Ahornstr. 55, 52074 Aachen, Germany,
andreas.dymek@rwth-aachen.de, https://orcid.org/0009-0000-5703-4306;
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/delfi2025_48

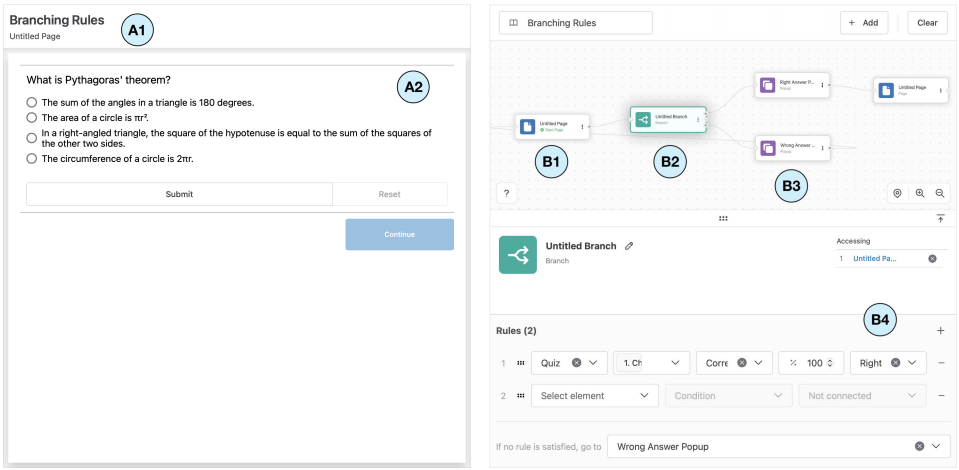
editors offer valuable insights into general user requirements. Green et al. [GHC18] created *Novella*, an IDN editor featuring a node graph showing the story's branching structure. Their comparison of various IDN editor interfaces found that graph-based visualizations of story structure are the most intuitive for authors [GHC21]. Moreover, Murray and Salter [MS23] found that most branching scenario authoring tools feature a graph-based UI. A popular commercial tool that uses a node graph structure is *H5P's Branching Scenario*[2], which enables the creation of self-paced and adaptive learning experiences similar to non-linear slide presentations. However, the tool only supports decision trees, not branching based on interactive elements like drag-and-drop or fill-in-the-blanks. Furthermore, the popular open-source tool *Twine*[3] allows for the authoring of interactive, non-linear stories. However, the tool requires users to have coding proficiency in scripting languages to create complex branches, which severely limits its usability for many users. Our branching scenario editor allows educators to create branches based on interactive elements using a visual, graph-based interface, eliminating the need for coding and ensuring accessibility for all educators.

2 Branching Scenario Editor

Based on the previously introduced research and tools, we designed and developed a new branching scenario editor tailored for educators, offering the greatest possible flexibility in creating adaptive learning experiences for students. The design and development followed Norman's iterative cycle of human-centered design [No02]. We developed an early prototype and gathered feedback through interviews with five secondary school educators, who interacted with the system to help shape user requirements. Building on these requirements, we refined the prototype, the results of which are presented in this work.

Editing Mode. The editor is divided into two sections: a node graph view at the top, which displays and manages the branching structure, and a node details view below, used for editing the content of the selected node (see Fig. 1b). Each branching scenario is composed of individual pages, which serve as the fundamental units and are represented by *page nodes* (see Fig. 1b). Pages are fully text-editable and authors can insert a variety of multimedia content, such as images, videos, and quiz elements. In addition to pages, authors can prompt the reader using popups, represented by *popup nodes* (see Fig. 1b). Popups are also fully text-editable, but appear layered over the currently displayed page when opened. To build a branching structure, each node can be connected to another one, forming a reachable path for the reader in the resulting branching scenario. When two nodes are connected, a navigation button is added to the content of the source node. Clicking it takes the reader to the target node. Authors can customize the button's label, size, width, alignment, and color to convey different messages and increase flexibility in authoring. Navigation buttons can be used to create decision trees for the reader; however, branchings can also be created based on the readers' performance on quiz elements, including fill-in-the-blanks questions,

[2] https://h5p.org/branching-scenario, accessed 2025-04-14
[3] https://twinery.org, accessed 2025-04-14



(a) (b)

Fig. 1: The branching scenario shows a page with a single-choice quiz (A2), represented in the branching structure by the page node on the left (B1). Using a branch node (B2), the author defines branching rules based on the student's performance on the quiz (B4). If solved correctly, the student goes to the "Right Answer Popup"; otherwise, they go to the "Wrong Answer Popup" (B3).

single- and multiple-choice questions, ordering questions, or select-the-text items (see Fig. 1a). To create different learning paths based on a quiz, authors have to connect a *branch node* to a page or popup containing a quiz element (see Fig. 1b). A branch node does not contain text or multimedia content, but allows authors to define branching rules, setting the learning path based on the readers' answers to the quiz element (see Fig. 1b).

Viewing Mode. Once the scenario has been created, students view and interact with it using the editor's viewing mode. The viewing mode systematically displays the branching scenario using an embedded canvas, featuring a header with the current page's and material's title (see Fig. 1a, A2). The canvas is scrollable, ensuring that all text, images, and quizzes can be displayed entirely. While primarily used to display branching materials to students, the viewing mode also helps authors verify their structures by displaying error messages, such as incomplete branch node rules.

Technology. We implemented the editor as a Web Component with all functionality running in the browser, distributed as a Node Package `@webwriter/branching-scenario` on NPM[4]. It is written in TypeScript[5] using the Drawflow[6] framework for displaying and managing a node graph. We developed the editor as a widget for *WebWriter* [SRS23], an open-source authoring tool designed for creating Explorables—interactive multimedia explanations

4 https://www.npmjs.com/package/@webwriter/branching-scenario, accessed 2025-04-15
5 https://www.typescriptlang.org/, accessed 2025-04-16
6 https://github.com/jerosoler/Drawflow, accessed 2025-04-16

of specific topics [Vi11]. Users can access and use our editor seamlessly within the WebWriter environment. Moreover, this integration enables authors to use WebWriter's entire multimedia suite and interactive widgets in our editor, such as WebWriter's *Quiz widget*[7] on which branchings can be based. Created branching scenarios are exported as HTML documents and can be used entirely offline. The development, architecture, and distribution of our editor attains the high standards of findability, accessibility, interoperability, and reusability (FAIRness) in line with the expectations of the research community [Ba22; KS22; Sa24]. We provide a properly versioned, and uniquely identified artifact (as a package on NPM) that is interoperable with any and all web frameworks (as a Web Component).

Use Case. Our editor could be used for a lesson on civics. Imagine an educator creates a branching scenario, starting with a text and media about the German federal parliament. In the scenario, the students have to read the text and then can navigate to the next page, featuring a single-choice quiz. If a student answers the quiz incorrectly, they receive feedback to try again and are sent to repeat the task. Otherwise, they receive positive feedback and are directed to a popup offering a choice: continue with the topic on passing laws or learn about elections. Either option sends the student to a page containing texts and interactive multimedia about the chosen topic, such as fill-in-the-blanks and multiple choice questions. Based on the students' performance on these questions, the student is guided down adapted learning paths offering more tailored explanations of the topics. While this use case is very simplified in terms of didactics, it underlines the potential of branching scenarios. Due to the interoperability of the material in WebWriter, different use cases are possible, such as embedding simulations or games in the branching scenarios or integrating exported materials as HTML into a larger learning application.

3 Conclusion

We presented a branching scenario editor tailored for educational use, designed to create adaptive learning experiences. Our editor provides unique features compared to the state of the art, such as branching off interactive quiz components, while requiring no coding expertise. The FAIR choices of technology may also enable future researchers to use or extend our editor to address different research questions. Through our iterative design approach in which we evaluated the editor in a series of interviews with five educators, our prototype covers the necessary elements to be useful in the educational domain. However, the prototype has some limitations. Options to create more fine-grained branching rules should be explored, such as using a point score system so that the teacher can better weigh the importance of individual tasks. Moreover, future work should research the use of branching scenarios in the field. This would ground design decisions and enable evaluating students' learning outcomes with the material. It is worth investigating which age groups the material could be used for. Finally, future work can find ways to support direct instructional scaffolding to support students struggling with the material.

[7] https://www.npmjs.com/package/@webwriter/quiz, accessed 2025-04-16

Acknowledgments

We were supported by the research project NERD II (005-2201-0014), sponsored by the state of North Rhine-Westphalia.

References

[Ba22] Barker, M. et al.: Introducing the FAIR Principles for research software. Scientific Data 9 (1), p. 622, 2022, DOI: 10.1038/s41597-022-01710-x.

[Fi05] Finkelstein, N. D. et al.: When Learning About the Real World Is Better Done Virtually: a Study of Substituting Computer Simulations for Laboratory Equipment. Phys. Rev. ST Phys. Educ. Res. 1, p. 010103, 2005, DOI: 10.1103/PhysRevSTPER.1.010103.

[GHC18] Green, D.; Hargood, C.; Charles, F.: Contemporary Issues in Interactive Storytelling Authoring Systems. In: Interactive Storytelling. Springer International Publishing, pp. 501–513, 2018, DOI: 10.1007/978-3-030-04028-4_59.

[GHC21] Green, D.; Hargood, C.; Charles, F.: Use of Tools: UX Principles for Interactive Narrative Authoring Tools. J. Comput. Cult. Herit. 14 (3), 2021, DOI: 10.1145/3458769.

[Ho20] Hohman, F. et al.: Communicating With Interactive Articles. en, Distill 5 (9), e28, 2020, DOI: 10.23915/distill.00028.

[Ko18] Koenitz, H.: Thoughts on a Discipline for the Study of Interactive Digital Narratives. In: Interactive Storytelling. Springer International Publishing, Cham, pp. 36–49, 2018, DOI: 10.1007/978-3-030-04028-4_3.

[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., Bonn, pp. 201–206, 2022, DOI: 10.18420/delfi2022-034.

[MB18] Möslein-Tröppner, B.; Bernhard, W.: Digitale Gamebooks in der Bildung. Springer, 2018.

[MS23] Murray, J. T.; Salter, A.: Mapping the Unmappable: Reimagining Visual Representations of Interactive Narrative. In: The Authoring Problem: Challenges in Supporting Authoring for Interactive Digital Narratives. Springer, pp. 171–190, 2023, DOI: 10.1007/978-3-031-05214-9_11.

[No02] Norman, D. A.: The Design of Everyday Things. Basic Books, [New York], 2002.

[NS23] Noichl, S.; Schroeder, U.: Interaktionsmöglichkeiten und Potenziale digitaler Gamebooks unter Berücksichtigung der CELG-Taxonomie. In: 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., Bonn, pp. 305–306, 2023, DOI: 10.18420/delfi2023-67.

[Sa24] Salmen, F. et al.: FAIR Learning Technologies with Web Components and Packages. In: Proceedings of DELFI 2024. Gesellschaft für Informatik e.V., Fulda, Germany, pp. 10–18420, 2024, DOI: 10.18420/delfi2024_23.

[Se19] Serth, S. et al.: Evaluating Digital Worksheets With Interactive Programming Exercises for K-12 Education. In: 2019 IEEE Frontiers in Education Conference (FIE). IEEE Press, Covington, KY, USA, pp. 1–9, 2019, DOI: 10.1109/FIE43999.2019.9028680.

[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). Gesellschaft für Informatik e.V., Bonn, pp. 39–48, 2023, DOI: 10.18420/wsdelfi2023-13.

[SS20] Shemshack, A.; Spector, J. M.: A Systematic Literature Review of Personalized Learning Terms. Smart Learning Environments 7 (1), p. 33, 2020, DOI: 10.1186/s40561-020-00140-9.

[Vi11] Victor, B.: Explorable Explanations, Accessed on 2024-04-06, 2011, https://worrydream.com/ExplorableExplanations/, accessed: 04/06/2024.

Konzeption eines Moodle-Plugins zur Selbst- und Fremdeinschätzung von Sozialkompetenzen in Projektarbeiten



Christopher Eck [1], Sven Judel [2], Tobias Brachmann [1], Ulrik Schroeder [2] und Martin Frenz [1]

Abstract: Lehrende stehen vor der Herausforderung, Sozialkompetenz in projektbasierten Gruppenarbeiten gezielt zu fördern und in formative und summative Bewertungsprozesse einzubinden. Ein möglicher Ansatz ist, Lernende ihre eigene sowie gruppenbezogene Sozialkompetenz zu verschiedenen Zeitpunkten selbst einschätzen zu lassen und dies durch digitale Anwendungen in Lernmanagementsystemen zu unterstützen. Moodle bietet zwar verschiedene Assessment-Werkzeuge, doch fehlt eine integrierte Lösung, die Selbst- und Fremdeinschätzungen systematisch zur Reflexion und Bewertung sozialer Kompetenzen nutzt. Dieser Beitrag beschreibt die Entwicklung eines Moodle-Plugins zur Begleitung von Projektarbeiten. Ausgehend von einer Analyse didaktischer Anforderungen werden funktionale Anforderungen an das Plugin formuliert und zentrale konzeptionelle und technische Entscheidungen beschrieben.

Keywords: Sozialkompetenz, Kompetenzdiagnostik, kollaboratives Lernen, Peer-Assessment, projektbasiertes Lernen

1 Einleitung

Kompetenzorientierung ist ein zentrales Leitprinzip in allgemeiner, beruflicher und hochschulischer Bildung. Ziel ist die Förderung einer umfassenden Handlungskompetenz, die neben fachlichen auch methodische, personale und soziale Kompetenzen einschließt. Projektbasierte Gruppenarbeiten bieten geeignete Lerngelegenheiten, in denen Lernende praxisnah authentische Problemstellungen bearbeiten. Für die Entwicklung von Sozialkompetenz reicht das bloß wiederholte Durchführen von Gruppenarbeiten nicht aus. Vielmehr erfordert es entsprechend strukturierte Lehr-Lern-Prozesse, in denen zu fördernde Kompetenzen wie Kommunikations- oder Teamfähigkeit im jeweiligen Projektkontext als explizite Lernziele ausformuliert und die konkrete Förderung der angestrebten Kompetenzen sowie die darauf bezogene Leistungsbewertung als zusammenhängende Einheit konzipiert werden. Nachfolgend werden zunächst didaktische Anforderungen und anschließend die vier Kernfunktionalitäten eines Moodle-Plugins zur Selbst- und Fremdeinschätzung in Projektarbeiten

[1] RWTH Aachen, Institut für Arbeitswissenschaft, Eilfschornsteinstr. 18, 52062 Aachen,
c.eck@iaw.rwth-aachen.de, https://orcid.org/0009-0000-5803-2214;
t.brachmann@iaw.rwth-aachen.de, https://orcid.org/0009-0004-5799-5026;
m.frenz@iaw.rwth-aachen.de, https://orcid.org/0000-0002-8551-5032

[2] RWTH Aachen, Lehr- und Forschungsgebiet Lerntechnologien, Ahornstr. 55, 52074 Aachen,
judel@informatik.rwth-aachen.de, https://orcid.org/0000-0002-0424-0402;
schroeder@informatik.rwth-aachen.de, https://orcid.org/0000-0002-5178-8497

doi:10.18420/delfi2025_49

vorgestellt, das Reflexion anleitet und soziale Interaktionen für Lernende und Lehrende sichtbar macht.

2 Didaktische Anforderungen und Konzeptentwicklung

Die Autoren kamen nach Recherchen im Moodle-Plugins Directory zu dem Ergebnis, dass zentrale Anforderungen an die gezielte Förderung und reflexionsbasierte Bewertung von Sozialkompetenz in gruppenbasierten Projektarbeiten bislang nur unzureichend umgesetzt sind. Im Folgenden werden Limitierungen bestehender Lösungen dargestellt und das Konzept eines Plugins vorgestellt, das diese didaktischen Anforderungen technisch adressiert.

Die Fähigkeit zur Reflexion nimmt bei der Kompetenzentwicklung im Allgemeinen und somit auch in Bezug auf Sozialkompetenz eine zentrale Rolle ein. Das Reflektieren erlebter Situationen bildet somit einen wichtigen Ausgangs- und Bezugspunkt für das erfahrungsbasierte Verstehen und gestaltende Erproben sozialer Interaktionsprozesse innerhalb der Gruppenarbeit [Eu12]. Um Sozialkompetenzen im Arbeitsprozess zu entwickeln, erfordert es einen Wechsel zwischen Handeln zur Bearbeitung der Projektaufgabe und Reflexion über Ziele, Umsetzung, Bedingungen und Wirkungen des sozial-kommunikativen Handelns [Du04]. Es müssen Lerngelegenheiten geschaffen werden, in denen das eigene, unhinterfragte Handeln zunächst beobachtet und bewusst gemacht sowie angemessenes Handeln thematisiert und ausgeführt wird, bis sozial-kommunikative Teilkompetenzen verinnerlicht und routini(si)ert angewandt werden [Eu12]. Bei bestehenden Moodle-Aktivitäten wie *Gegenseitige Bewertung* steht jedoch allein das fachliche Peer-Feedback zu Lern- und Arbeitsprodukten im Vordergrund und nicht die Interaktion innerhalb der Gruppe [Mo24].

Im Kontext der Sozialkompetenzen zeigt sich, „dass eine gezielte Förderung an einer zu unpräzisen und nur sehr vagen Vorstellung über das zu verfolgende Lernziel scheitern kann“ [Du04]. Aussagen zu sozial (un-)angemessenem Verhalten sollten kontextbezogen und beobachtbar formuliert sein, denn was bspw. in einem Kontext als durchsetzungsstark gilt, kann in einem anderen als unkooperativ erscheinen [BLO16; Eu16]. Daraus ergibt sich die Prämisse, dass „(Sozial-)Kompetenzen nicht generell, sondern situationsspezifisch erworben und angewendet werden“ [Eu16]. Plugins wie das an der belgischen Université catholique de Louvain (UCL) entwickelte *Dynamo* [Pa25] ermöglichen bereits die Erfassung von Gruppendynamiken mittels Peer-Evaluationen und de-kontextualisierten Standard-Fragebögen. Diese laufen jedoch Gefahr, eher allgemeine Einstellungen zu messen als beobachtbares Verhalten, das im konkreten Lernkontext Ausdruck von Sozialkompetenz ist. Projekte können von wenigen Doppelstunden bis zu einem ganzen Semester dauern, wobei im Zeitverlauf unterschiedliche Lern- und Arbeitsschwerpunkte bestehen. Lehrende sollten Fragebögen daher an die jeweiligen Handlungssituationen anpassen können.

Lernzielbezogene Selbst- und Fremdeinschätzungen bieten eine strukturierende Anleitung für Reflexionsprozesse, um sich selbst und andere aktiv sowie die eigene Rolle in der Gruppe wahrzunehmen und Maßnahmen zur Verbesserung der Zusammenarbeit abzuleiten.

Die erhobenen Informationen fördern die *Group Awareness*, d. h. das Bewusstsein für das Aufgabenverständnis, den Arbeitsstand und das soziale Miteinander innerhalb der Gruppe [CZL22]. Im Sinne formativen Assessments können die Daten durch Lehrende genutzt werden, um moderierend Anregungen und gezielte Hilfestellung zur Selbstregulation zu geben, damit Lernende „ihre zukünftige Arbeit gestalten können, um erfolgreich zu lernen" [Bo18]. Für die Leistungsbewertung eignen sich Selbst- und Fremdeinschätzungen als Ergänzung zur Lehrendenbeobachtung, um Sozialkompetenz als komplexes und mehrdimensionales Konstrukt durch multiple Perspektiven valider und objektiver zu erfassen [Eu12]. Das Plugin *Peerwork* [Mo20] deckt verschiedene Anforderungen bereits ab und bietet die Funktion, anhand von Selbst- und Fremdeinschätzungen zum Beitrag in der Gruppe angepasste Notenvorschläge zu generieren. Allerdings sind die praktischen Möglichkeiten zur formativen Nutzung begrenzt, da z. B. Visualisierungen für Lehrende und Lernende fehlen und ebenso wie bei *Dynamo* nur ein Erhebungszeitpunkt möglich ist. Zudem bietet es keine Funktionen an, um Teilkonstrukte von Sozialkompetenz für die Berechnung des Notenvorschlags entsprechend der Lernziele zu gewichten.

3 Technisches Konzept und Umsetzung als Moodle-Plugin

Digitale Erhebungsmethoden bieten Vorteile, indem sie die Datenerhebung, -auswertung und -visualisierung vereinfachen. Die Autoren entwickeln daher ein in die Lernplattform Moodle integrierbares Plugin zur Selbst- und Fremdeinschätzung von Sozialkompetenz in projektbasierten Gruppenarbeiten, das als Autorentool an die kontextspezifischen Anforderungen angepasst und flexibel eingesetzt werden kann. Es umfasst folgende vier Kernfunktionalitäten:

(1) Lehrende können beliebig viele, miteinander verknüpfte Erhebungszeitpunkte definieren. Dies unterstützt die formative Begleitung von Lernprozessen und ermöglicht die kontinuierliche Reflexion der Zusammenarbeit, selbst wenn diese in längeren Projekten phasenweise selbstorganisiert außerhalb von Lehrveranstaltungen erfolgt.

(2) Bei der Erstellung der Fragebögen können verschiedene Fragetypen beliebig miteinander kombiniert werden. Zentral sind hierbei die Selbst- und Fremdeinschätzungsfragen, die aus einer 5-Punkt bzw. 6-Punkt Likert-Skala bestehen und als Items für jeden Lernenden automatisch die Mitglieder der eigenen, in Moodle erstellten Kursgruppe integrieren (siehe Abbildung 1). Hier hebt sich das Plugin auch vom Moodle-eigenen Quiz [Mo25] ab, welches zum aktuellen Zeitpunkt keine Gruppen berücksichtigt oder automatisiert individualisierte Fragen generiert. Mehrere Fragen können zudem zu einer Kategorie zusammengefasst werden, um eine für die Bewertung relevante Dimension von Sozialkompetenz abzubilden. Als Hilfestellung für die projektspezifische Ausgestaltung der Frageinstrumente werden den Lehrenden Fragebogenvorlagen bereitgestellt, die mehrere Dimensionen von Sozialkompetenz mit Frageitems beschreiben und flexibel angepasst werden können. Weitere Fragetypen, wie Freitextfelder, Single und Multiple Choice, frei wählbare Skalenfragen oder Punkteverteilungen auf verschiedene Antwortoptionen wurden eingebunden. Die

weiteren Fragetypen ermöglichen es Lehrenden, eigene (Reflexions-)Schwerpunkte in den Fragebögen zu setzen, um das subjektive Erleben der Gruppenarbeit zu erfassen oder für die Organisation der Projektarbeit relevante Informationen aufwandsarm abzufragen. Zu jedem Erhebungszeitpunkt kann ein anderer Fragebogen eingesetzt werden.

(3) Aufbereitete Auswertungen der Fragebögen erhalten Lehrende auf einem Dashboard. Nach Bedarf können diese Auswertungen für den ganzen Kurs, für jede Gruppe individuell und auch für jeden Erhebungszeitpunkt betrachtet werden, um die Lernprozessbegleitung datenbasiert zu unterstützen. Ob Lernende Einsicht in die Ergebnisse der eigenen Gruppe erhalten, können Lehrende für jede Frage konfigurieren.

(4) Lehrende können nach dem Abschluss einer Projektarbeit Gewichtungen für Erhebungszeitpunkte und verwendete Kompetenzdimensionen, z. B. Selbstständigkeit und Kommunikationsfähigkeit, festlegen (siehe Abbildung 1). Mit diesen berechnet das Plugin Beiträge zur Gruppenarbeit und generiert daraus gewichtete Punktevorschläge als Diskussionsgrundlage für die individuelle Benotung. Lehrende müssen dabei selbst entscheiden, wie sie die Vorschläge in den Notengebungsprozess einbeziehen – eine automatisierte Übernahme ist explizit nicht implementiert.

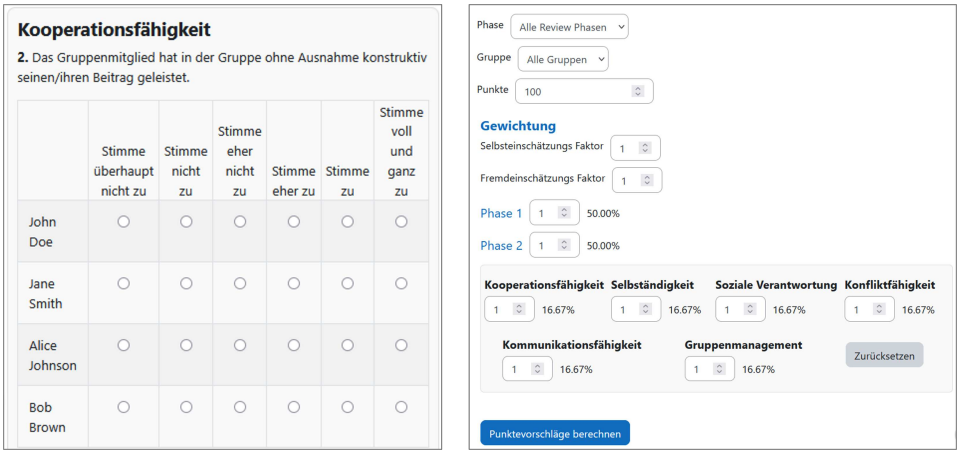


Abb. 1: Beispielfrage zur Selbst- und Fremdeinschätzung und Einstellungen für Notenvorschläge

Das Plugin wurde in der Lehramtsausbildung erprobt, um weitere Entwicklungsschritte abzuleiten. Weitere Studien sind an Berufskollegs in projektbasiertem Unterricht geplant. Nach einer Veröffentlichung über das Moodle Directory wird perspektivisch eine Integration in die Moodle-Instanz der RWTH Aachen angestrebt. Auch eine Bereitstellung für auf Moodle basierende Lernmanagementsysteme, wie LOGINEO vom Bildungsministerium in NRW [LO25], ist denkbar. Entwickelt wird das Plugin derzeit noch in einem geschlossenen Repository. Die zum Zeitpunkt der Veröffentlichung aktuelle Version ist unter [JBE25] verfügbar. Offene Fragen bestehen hinsichtlich ethischer Aspekte der gegenseitigen Bewertung sowie der didaktischen Aufbereitung der erhobenen Daten zur Nutzung durch Lehrkräfte.

Literaturverzeichnis

[BLO16] Beck, K.; Landenberger, M.; Oser, F., Hrsg.: Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der BMBF-Förderinitiative ASCOT. W. Bertelsmann Verlag GmbH & Co. KG, Bielefeld, 2016.

[Bo18] Borsch, F. et al.: Kooperatives Lernen: Theorie – Anwendung – Wirksamkeit. W. Kohlhammer GmbH, Stuttgart, 2018.

[CZL22] Chen, D.; Zhang, Y.; Lin, Y.: Group Awareness and Group Awareness Tools in Computer-Supported Collaborative Learning: A Literature Review. In: 2022 International Symposium on Educational Technology (ISET). S. 18–22, 2022, DOI: 10.1109/ISET55194.2022.00013.

[Du04] Dumpert, M.: Chancen und Grenzen einer Förderung von Sozialkompetenzen im Arbeitsprozess. In: Sozialkompetenzen zwischen theoretischer Fundierung und pragmatischer Umsetzung. BerufsBildung in der globalen NetzWerkGesellschaft, Bertelsmann, Bielefeld, S. 101–114, 2004.

[Eu12] Euler, D.: Von der programmatischen Formel zum didaktischen Konzept: Sozialkompetenzen präzisieren, fördern und beurteilen. In (Gerhard Niedermair, Hrsg.): Kompetenzen entwickeln, messen und bewerten. Trauner, Linz, S. 183–197, 2012.

[Eu16] Euler, D.: Sozialkompetenzen: -allseits gefordert, wenig erforscht... Zeitschrift für Berufs- und Wirtschaftspädagogik (4), S. 517–524, 2016.

[JBE25] Judel, S.; Brachmann, T.; Eck, C.: Quellcode des Plugins zum Zeitpunkt der Veröffentlichung, 2025, DOI: 10.17605/OSF.IO/J23PH.

[LO25] LOGINEO NRW: LOGINEO NRW Lernmanagementsystem | LOGINEO NRW, https://www.logineo.schulministerium.nrw.de/LOGINEO-NRW/NEU-LOGINEO-NRW-LMS-Lernmanagementsystem/, Zuletzt aufgerufen am: 06.04.2025, 2025.

[Mo20] MoodleDocs: Peerwork activity - MoodleDocs, https://docs.moodle.org/500/en/Peerwork_Activity, Zuletzt aufgerufen am: 06.04.2025, 2020.

[Mo24] MoodleDocs: Workshop activity - MoodleDocs, https://docs.moodle.org/500/en/Workshop_activity, Zuletzt aufgerufen am: 06.04.2025, 2024.

[Mo25] MoodleDocs: Quiz activity - MoodleDocs, https://docs.moodle.org/500/en/Quiz_activity, Zuletzt aufgerufen am: 06.04.2025, 2025.

[Pa25] Palumbo, D.: Moodle Plugins directory: Dynamo | Moodle.org, https://moodle.org/plugins/mod_dynamo, Zuletzt aufgerufen am: 06.04.2025, 2025.

Cats, Code, and Concepts: Learn to Program with Catpie

Rahel Natalie Engel[1], Hannah Lenk[2], André Greubel[1], and Dorian Weber[1]

Abstract: We present a novel, low-threshold learning environment designed to provide a playful introduction to computer science and programming. Through a series of scenarios and tasks, learners engage with Catpie, our beginner-friendly programming language developed to teach fundamental programming concepts and computational thinking. Unlike simplified block-based tools, Catpie introduces realistic coding practices and authentic error handling, focusing on the language itself rather than quick game creation. It requires no account, no consistent internet connection, and no set-up. This environment is particularly well-suited for use during project days or short-term learning formats both in and outside of formal education settings. A key goal of Catpie is to introduce computer science topics typically encountered at university level to learners from secondary education and up. This offers an accessible early insight into core concepts and realistic working practices in the field.

Keywords: programming education, introductory programming, learning environment

1 Motivation

It is our strongly held belief that learning how to code is learning how to think, how to solve problems and be precise. Regardless of the development of artificial intelligence and how important the ability to code is going to be, the process of learning it will remain a valuable cognitive and creative exercise – especially in educational contexts [LFN22; Wi06]. Today's landscape of beginner-oriented programming tools offers a wide range of approaches, each with specific strengths and weaknesses. Visual block-based environments like Scratch, PencilCode, Open Roberta Lab and Kodu Game Lab provide an exceptionally low entry barrier and encourage experimentation and creativity, especially to younger children [Ko25; Op25; Pe25; Re10]. However, they abstract away from code syntax, error handling and programming structure, which can make the transition to text-based languages more difficult [Pe21]. Other common approaches use a simple command structure, such as Turtle Graphics or Karel the Robot [CC00; Pa84]. They provide mnemonic commands and visual feedback, which makes them particularly appealing for playful and creative exploration. However, their scope is relatively narrow and less suited for learning broader programming structures or concepts. Tools like Replit, which incorporate AI [Re25], or environments for creating games like Miniworlds [Si25] prioritize quick creation over the fundamental learning of programming itself. Languages like Python are widely used in areas such as data science and machine learning [DPD20], making them an excellent starting point for learners with a clear interest in these domains. However, due to their extensive feature

[1] Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, Institut für Informatik, Rudower Chaussee 25, 12489 Berlin, Berlin, rahel.natalie.engel@hu-berlin.de; andre.greubel@hu-berlin.de; dorian.weber@hu-berlin.de

[2] Codemetas, Berlin, hannah.lenk@codemetas.de

doi:10.18420/delfi2025_50

set, the need for environment set-up, and the wide range of possible applications, they can overwhelm absolute beginners and make it harder to focus on foundational programming principles. There is no universally optimal starting point for learning to program; instead, each tool addresses different educational needs. However, it seems that a gap remains underrepresented: introductory programming that combines realistic coding practices and playful visual feedback, whilst not requiring any account or installation. To address this need, we present a low-threshold environment paired with a beginner-oriented language as an option for introducing programming through structured exploration, algorithmic problem solving, and exposure to concepts typically encountered much later. The environment targets absolute beginners and learners with very limited prior exposure to programming, from students in secondary education to adults. Catpie is particularly relevant for short-term educational formats—such as project days or introductory workshops—where learners should be given a glimpse not only into the fun side of programming but also into the kinds of thinking and working practices that characterize the field of computer science.

2 Concept and Realization

2.1 Learning Environment and Language Design

The environment runs seamlessly on any modern browser, independent of hardware architecture or operating system—requiring no account, set-up or installation. After the page loads, an internet connection is not needed. It is freely accessible under catpie.compscicomp.de Catpie is an open source project and can be accessed on GitLab. Inspired by the room metaphor by Resnick [Re20], we aim to combine a low floor, wide walls, and a high ceiling: it is easy to get started with Catpie and a broad variety of projects are supported, while remaining open to more complex and challenging ideas. Catpie's design balances accessibility with authenticity. This allows learners to get an early, practical glimpse into the field [Ha13]. It features consistent syntax and rules without excessive brevity. Rather than just providing mnemonic commands, its focus is on coverage of fundamental concepts. This aligns with recommendations for beginner programming languages [Gu04; MC96]. While Catpie offers built-in functions for an easy start, the programming tasks quickly progress and require learners to write their own types and functions, pushing beyond the limitations of purely visual languages. The language design emphasizes clarity and structure. For instance, since Catpie is not whitespace-sensitive, code readability and structure must be understood explicitly rather than implicitly enforced. This trade-off supports deeper understanding and encourages learners to reflect on good coding practices. While the language hides complex, low-level technical details to keep things simple for beginners, it intentionally includes algorithmic thinking early on. Learners can, for example, implement data structures like stacks or queues. This approach aligns with the idea of a spiral curriculum-introducing important concepts like brute-force strategies early in a simplified form and revisiting them later with greater depth and refinement.

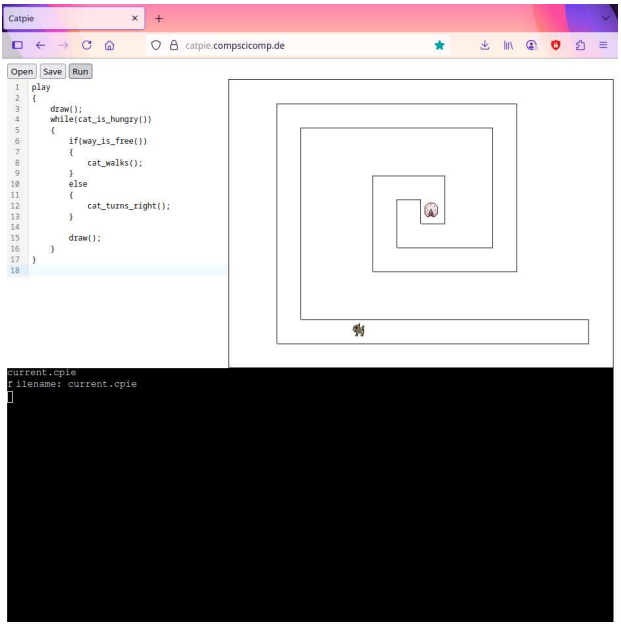


Fig. 1: Screenshot of the learning environment

2.2 Funcionality and Usage

Figure 1 shows a screenshot of the current prototype. The interface is divided into three main sections: a code editor for writing programs, a visual output area and a terminal for displaying error messages. Learners can run their code with a single click using the `Run` button, which checks and interprets the current program. They can also use the `Save` button to download their code or the `Open` button to load existing files from their device. The central scenario involves a cat that is trying to find a hidden pie in a maze. The simplest maze is a spiral, making it a suitable starting point for a first complete project. One possible solution, illustrated in Fig. 1, is to let the cat walk forward and whenever it hits a wall, try turning right. This task is typically broken down by an instructor into smaller learning steps. First, learners observe what happens when the cat simply walks forward. They see that it gets stuck and are encouraged to think about how to handle such situations. The idea of adding turns and decision-making is then introduced step by step. Through this, learners naturally encounter repetition in their code—for example, walk, turn, walk, turn—which creates an opportunity to introduce loops. They then learn how to write the same behavior more efficiently, and in a way that works regardless of the maze size. More complex examples can be found in the GitLab repository. These include a solution that finds the shortest path for a maze, as well as the game of life, which shows that Catpie can be used for projects that go beyond the maze scenarios.

Catpie's error messages are designed for maximum clarity and user assistance. This goal is guided by principles aimed at fostering a highly understandable and helpful debugging experience, similar to the philosophy outlined by Elm language [Cz15]. Specifically, Catpie aims to identify and highlight relevant code sections, clearly mark and explain error locations in simple terms and provide actionable hints for resolution. To provide one example, consider this line of code that shows it was not specified which type to use with `make`, where `position` is a struct consisting of two integers.

```
position current_position = make (x = 0, y = 0);
```

The corresponding error message explains which line the error occurred on, using the same indentation as the original code, what went wrong in simple terms, and suggests a solution by stating all options that can be used with `make`.

```
Parser error in line 9: You used 'make' were it is not possible.
You can 'make' struct or container.

    position current_position = make (x = 0, y = 0);
                                     ^
                                     identifier or
                                     container
```

3 Evaluation and Future Work

To gain initial insights into Catpie's effectiveness, a sample mini-workshop was conducted with five learners, three of whom were absolute beginners. The environment and language were very well received, with participants particularly appreciating the logical problem-solving aspect. Notably, two participants who had prior exposure to Python expressed a strong preference for Catpie. They highlighted its simpler error messages, clearer structure and particularly its low-threshold accessibility as significant advantages. One of them specifically reported that Catpie helped them finally understand programming concepts and overcome previous feelings of inadequacy. Furthermore, one absolute beginner was highly motivated by their experience with Catpie, leading them to consider computer science as a potential field of study. While this initial test involved a very small sample size and participants aged mid-twenties to mid-fifties, the results are encouraging and suggest Catpie's suitability for its intended purpose. Future work will include testing Catpie with a larger, more diverse, and younger demographic to gather comprehensive data. To accommodate different age groups and time frames more mazes will be added, allowing a smoother transition between levels. Additional game modes are being explored—for example, limiting the visible portion of the maze to simulate real-world navigation challenges. The learning environment will be complemented with short tutorials and explanations in both German and English. While the initial prototype focuses primarily on language design and implementation, visual design and user interface improvements are planned to make the environment more engaging and accessible.

References

[CC00] Caspersen, M. E.; Christensen, H. B.: Here, there and everywhere- on the recurring use of turtle graphics in CS 1. In: ACM International Conference Proceeding Series. Vol. 8, pp. 34–40, 2000.

[Cz15] Czaplicki, E.: Compiler Errors for Humans. Rethinking the terminal UX in Elm 0.15.1, https://elm-lang.org/news/compiler-errors-for-humans, Accessed: 2025-07-02, 2015.

[DPD20] Dhruv, A. J.; Patel, R.; Dosh, N.: Python: The Most Advanced Programming Language for Computer Science Applications. Proceedings of the International Conference on Culture Heritage, Education, Sustainable Tourism, and Innovation Technologies, pp. 292–299, 2020, DOI: 10.5220/0010307900003051.

[Gu04] Gupta, D.: What is a Good First Programming Language? XRDS 10 (4), p. 7, 2004, DOI: 10.1145/1027313.1027320.

[Ha13] Har, L. B.: Authentic learning. The Hong Kong Institute of Education, 2013.

[Ko25] KoduGameLabe: 3D game programming for kids, https://www.kodugamelab.com/, Accessed: 2025-07-08, 2025.

[LFN22] Lee, S. J.; Francom, G. M.; Nuatomue, J.: Computer science education and K-12 students' computational thinking: A systematic review. International Journal of Educational Research 114, p. 102008, 2022, DOI: 10.1016/j.ijer.2022.102008.

[MC96] McIver, L.; Conway, D.: Seven Deadly Sins of Introductory Programming Language Design. In: Proceedings of the 1996 International Conference on Software Engineering: Education and Practice (SE:EP '96). SEEP '96, IEEE Computer Society, USA, p. 309, 1996.

[Op25] OpenRobertaLab: Welcome to Open Roberta Lab, https://lab.open-roberta.org/, Accessed: 2025-07-08, 2025.

[Pa84] Pattis, R. E.: Karel the Robot - A Gentle Introduction to the Art of Programming with Pascal. John Wiley & Sons Inc, 1984.

[Pe21] Perera, P. et al.: A Systematic Mapping of Introductory Programming Languages for Novice Learners. IEEE Access PP, pp. 1–1, 2021, DOI: 10.1109/ACCESS.2021.3089560.

[Pe25] PencilCode: Dream it. Code it. https://pencilcode.net/, Accessed: 2025-07-08, 2025.

[Re10] Resnick, M. et al.: The Scratch Programming Language and Environment. ACM Transactions on Computing Education (TOCE) 10, p. 16, 2010, DOI: 10.1145/1868358.1868363.

[Re20] Resnick, M.: Designing for Wide Walls, https://mres.medium.com/designing-for-wide-walls-323bdb4e7277, Accessed: 2025-06-07, 2020.

[Re25] Replit: An AI app builder like no other, https://replit.com/usecases/ai-app-builder, Accessed: 2025-07-08, 2025.

[Si25] Siebel, A.: Miniworlds Dokumentation, https://miniworlds.opencoding.de/de/index.html, Accessed: 2025-07-08, 2025.

[Wi06] Wing, J. M.: Computational thinking. Commun. ACM 49 (3), pp. 33–35, 2006, DOI: 10.1145/1118178.1118215.

UniGoal: Ein Unterstützungstool für SMARTe Lernziele

Burhan Karababa[1], Armin Egetenmeier[1] und Sven Strickroth[1]

Abstract: Selbstreguliertes Lernen (SRL) und das Setzen von eigenen (Lern-)Zielen stellt eine wichtige Schlüsselkompetenz für den akademischen Erfolg dar. Dennoch haben viele Studierende Schwierigkeiten mit der Formulierung von präzisen, messbaren Lernzielen. In diesem Demo-Beitrag wird ein Unterstützungstool zur Formulierung von individuellen Lernzielen vorgestellt. Das Tool unterstützt Studierende durch zwei unterschiedliche Ansätze zur Formulierung von Lernzielen, welche jeweils dem SMART-Ansatz folgen. Dieser Ansatz klärt zentrale Merkmale (u. a. Messbarkeit, Relevanz) eines Ziels und berücksichtigt zugleich elementare Prinzipien des SRL. Darüber hinaus unterstützt das Tool die Reflexion der Ziele mit Fortschrittsanzeigen, Kalender und Dashboard.

Keywords: Lernziele, Selbstreguliertes Lernen, Technology-Enhanced Learning, Learning Analytics

1 Motivation

Selbstreguliertes Lernen (SRL) stellt eine Schlüsselkompetenz für Studierende dar, welche einen großen Einfluss auf den akademischen Erfolg hat [DP16] und zu einer höheren Zufriedenheit im Studium beitragen kann [PD90; Ro04]. Die Fähigkeit, sich selbst realistische und messbare Lernziele zu setzen (im Gegensatz zu vorgegebenen Lehrzielen des Dozierenden bzw. aus der Veranstaltung) und diese aktiv mit geeigneten Strategien zu verfolgen, fällt vielen Lernenden schwer, insbesondere in wenig strukturierten Lernsituationen [WH98]. Die zielgerichtete Planung und Reflexion des eigenen Lernprozesses bleibt im Studienalltag eine Herausforderung. Meist ist bereits die Formulierung von geeigneten, individuellen (Lern-)Zielen schwierig [Ki20], sodass Ziele nicht explizit benannt und der Fortschritt überprüft werden können. Geeignete Unterstützungsmaßnahmen zur Entwicklung von konkreten Zielvorstellungen sind nicht verbreitet oder bekannt. Vielversprechende Lösungsansätze, wie die SMART-Methode [Do81; SZ12] zur Zielformulierung berücksichtigen zentrale Prinzipien des SRL, jedoch verfügen nicht alle Studierenden über das erforderliche Wissen zur eigenständigen Anwendung dieses Ansatzes [ULA24]. An dieser Stelle könnten Bildungstechnologien eine Hilfestellung leisten, jedoch werden diese überwiegend für die Durchführung von Zielsetzungen eingesetzt [va25]. Spezifische SRL Anwendungen für Zielsetzung (z. B. [WST21; ZBS24]) sind rar oder fokussieren sich meist auf einzelne SRL Aspekte wie z. B. Motivation [CHL24] oder Reflexion [KKM18]. Hilfestellungen zur Formulierung von individuellen SMART-Zielen sind den Autoren unbekannt.

In diesem Beitrag wird ein Unterstützungstool zur strukturierten Erfassung und insbesondere Formulierung von persönlichen Lernzielen präsentiert.

[1] Ludwig-Maximilians-Universität München, Institut für Informatik, Oettingenstraße 67, 80538 München, b.karababa@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/delfi2025_51

2 Selbstreguliertes Lernen, Lernziele und SMART

Selbstreguliertes Lernen beschreibt die metakognitive Fähigkeit von Lernenden, Lernprozesse selbstständig, zielgerichtet und adaptiv zu steuern [Pi00]. SRL stellt einen aktiven, dynamischen Prozess dar, bei dem kognitive (z. B. Planung), metakognitive (z. B. Reflexion) und motivationale Komponenten (z. B. Selbstwirksamkeit) integriert sind [SZ12]. Zur Strukturierung des SRL wird meist das Phasenmodell von Zimmerman [Zi02] herangezogen, welches den Lernprozess als zyklische, wiederkehrende Abfolge aus Planungs-, Handlungs- und Reflexionsphase beschreibt. Insbesondere die Planungsphase nimmt eine entscheidende Rolle als Erfolgsfaktor für effektives Lernen ein. Die Fähigkeit zur präzisen Zielformulierung stellt eine kritische Voraussetzung für erfolgreiche, selbstregulierte Lernprozesse dar [Pa17]. Weiterhin zeigt eine Studie von Urgo; Lund; Arguello, dass Lernende ohne konkrete (Teil-)Ziele in der Regel weniger effektiv recherchieren, seltener metakognitive Strategien nutzen und geringere Lernerfolge erzielen [ULA24].

Zielsetzungsprozesse lassen sich verschiedenen theoretischen Ansätzen zuordnen, etwa der Goal-Setting Theory (GST, [LL90]) oder Achievement Goal Theory (AGT, [Ni84]). Gemäß der GST steigern spezifische und herausfordernde Ziele die Leistung positiv. Die AGT liefert Erklärungen für die motivationalen Prozesse und Gründe für die Zielwahl. Sie unterscheidet zwischen Lernzielen (Mastery Goals), die auf fachliche Kompetenzentwicklung abzielen, und Leistungszielen (Performance Goals), die den sozialen Vergleich fokussieren. Viele Modelle differenzieren zudem bei der Art der Ziele zwischen Ergebnis- (Outcome Goals) und Prozessorientierung (Process Goals). Studien zeigen, dass eine Kombination aus beiden Zielorientierungen die Lernleistung deutlich steigern kann (siehe [CK17; KMP21]).

Als operationalisierter Ansatz für Zielformulierungen wird in verschiedenen Kontexten die SMART-Methode angewandt, welche eine strukturierte und überprüfbare Zielsetzung ermöglicht [Do81; SZ12]. SMART steht hierbei als Akronym für die Kriterien spezifisch (Specific), messbar (Measurable), erreichbar (Achievable), relevant (Relevant) und zeitgebunden (Time-bound). Dieser Ansatz dient als methodische Brücke zwischen theoretischen Zielsetzungskonzepten und der Lernpraxis, sodass Lernende ihre Ziele planen, überwachen und erreichen können. Zudem werden durch die Methode auch Motivation, Selbstwirksamkeit und die Fähigkeit zur strategischen Planung gefördert [RD20; SZ12].

Der systematische Review von van Martins Jaarsveld et al. [va25] zeigt die aktuelle Forschungslage zur (Lern-)Zielsetzung im Hochschulbereich. Die meisten Studien nutzen Technologie zur Durchführung von Zielsetzungsaktivitäten, jedoch nur sehr wenige Arbeiten setzen Bildungstechnologien darüber hinaus ein bspw. zur Unterstützung oder für Feedback (z. B. [KKM18; WST21]). Die Ergebnisse dieser Studie zeigen, dass SMART-Ziele ein vielversprechender Forschungsansatz sind, da sie über die traditionelle Literatur zu Zielsetzung hinausgehen und sich auf Lern- und Leistungsziele (Mastery Goals, Performance Goals) konzentrieren. Zudem könnte die Einbindung von Learning Analytics und KI, skalierbares Echtzeitfeedback ermöglichen und die allgemeine Wirksamkeit von Intervention in der Zielformulierung verbessern, wie die Autoren folgern. Erste Anwendungen finden sich in

der Literatur, die zur Unterstützung Studierender in verschiedenen Phasen des SRL Large Language Models (LLM) (z. B. [St24]) oder LA nutzen (vgl.[He23]). Insgesamt zeigt sich, dass eine (digitale) Unterstützung bei der Erstellung von individuellen Lernzielen bzw. Feedback und Überarbeitung kaum stattfindet. Hier setzt der entwickelte Prototyp an.

3 Konzeption, Umsetzung & Features

Im Folgenden werden einige Kern-Features des UniGoal Tools beschrieben. Basierend auf einer semi-strukturierten Interviewstudie mit 21 Studierenden (siehe [Ka25]), wurden funktionale und nicht-funktionale Anforderungen an das Tool abgeleitet. Zusätzlich wurden auch konkrete Lernziele der Teilnehmenden erfragt, welche für die Optimierung der Fragen (und Prompts) herangezogen wurden. Erste Ergebnisse zeigen, dass insbesondere bei der Zielformulierung Schwierigkeiten existieren. Ziele sind oft vage formuliert, nicht schriftlich festgehalten und bei fünf Interviewpartnern nicht vorhanden. Dies führt zu einem eher ungeplanten, reaktiven Lernverhalten, wie viele Befragten angeben. Andere Aspekte wie Feedback (z. B. Umsetzbarkeit des Lernziels) oder dessen Visualisierung (z. B. Fortschrittsbalken) wurden ebenfalls genannt. Basierend auf den Ergebnissen der Interview-Studie (Publikation hierzu folgt) sowie der Literatur wurde ein Prototyp entwickelt.

Der UniGoal-Prototyp ist als webbasierte Anwendung umgesetzt, deren responsives Design auch für mobile Endgeräte konzipiert wurde. Das Tool fokussiert sich primär auf die unterstützte Erstellung von Lernzielen für Studierende in zwei unterschiedlichen Ansätzen unter Beachtung der SMART-Kriterien. Einerseits in einem Formular-basierten Ansatz, bei dem das Lernziel über fünf Seiten, eine pro SMART-Aspekt, als Freitexteingabe erfolgt. Das Formular ist dabei statisch. Die Antworten werden lediglich auf Vollständigkeit und formale Korrektheit geprüft. Zum anderen können in einer dialog-basierten Variante Lernziele interaktiv erfasst werden. Hierbei erfolgt die Dialogführung adaptiv durch ein LLM (GPT-4o API), die sich an den SMART-Kriterien orientiert. So werden in einem Chat-ähnlichen Verlauf die Antworten des Nutzers erfasst (Abb. 1, links), ggf. nach Präzisierungen gefragt und schließlich den einzelnen SMART-Kriterien zugeordnet. Aus der finalen Ausgabe des LLMs erfolgt die Extraktion relevanter Inhalte mithilfe regulärer Ausdrücke, um diese strukturiert speichern zu können. Die Eingabe der zeitlichen Komponente erfolgt in beiden Fällen nach Abschluss des Dialogs bzw. der letzten Formularseite im Datumsform (tt.mm.yyyy), da das benutzte LLM bzw. aktuelle LLMs Schwierigkeiten mit Zahlen haben. Zum Abschluss wird dem Nutzer eine Übersicht der eingegebenen Kriterien bereitgestellt, aufgeschlüsselt nach den SMART Kriterien (siehe Abb. 1, mitte). Optional kann der Nutzer das Lernziel einer bestimmten Veranstaltung zuweisen und/oder die Art des Lernziels (z. B. Process Goal, Mastery Goal) ergänzen. Nach dem Speichern des Lernziels wird dieses im Dashboard bzw. auf der Startseite angezeigt.

Alle offenen Lernziele werden in einer Liste dargestellt und können nach Veranstaltungen gefiltert werden. Details der Ziele können über einen Klick eingesehen und weiter bearbeiten werden, z. B. können Meilensteine ergänzt und einzeln durch den Nutzer als abgeschlossen

markiert werden. Diese werden ebenfalls direkt auf der Startseite dargestellt. Basierend auf diesen (manuell erfassten) Meilensteinen wird pro Lernziel eine Fortschrittsanzeige dargestellt (siehe Abb. 1, rechts), die dem Nutzer Feedback zum aktuellen Grad der Zielerreichung gibt. Auf Basis der Interviews, wurde eine Kalenderfunktion implementiert, das die einzelnen Lernziele und Meilensteine dort übersichtlich zeigt. Ein (noch simples) Dashboard zeigt die Anzahl der offenen bzw. bereits abgeschlossenen Lernziele an.

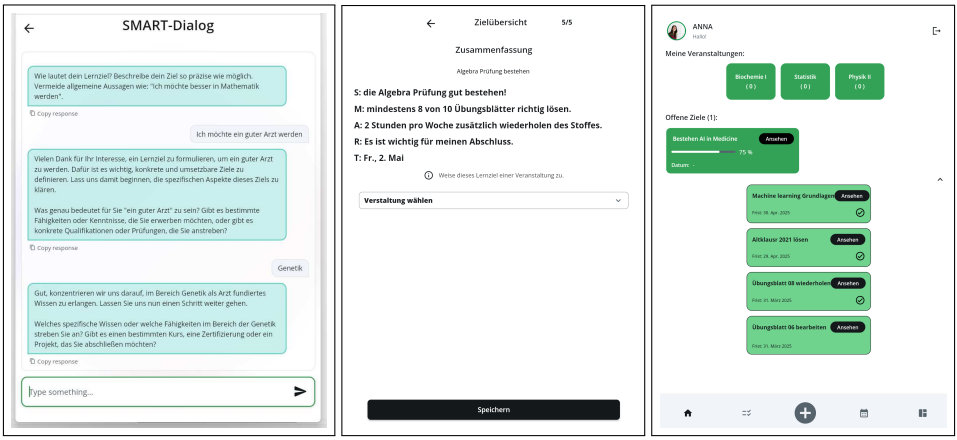


Abb. 1: Auszug aus SMART-Dialog (links), Anzeige der SMART-Zielübersicht (mitte) und Lernziel-Fortschrittsanzeige mit Überblick zu Meilensteinen (rechts) des UniGoal-Tools

4 Ausblick

Das UniGoal-Tool stellt einen konzeptionellen und explorativen Prototyp dar, welcher die Formulierung von Lernzielen durch die Anwendung der SMART-Methode unterstützt. Der vorgestellte Ansatz eignet sich für viele unterschiedliche Anwendungsfälle im Hochschulbereich und könnte auch in Veranstaltungen mit größerer Autonomie (z. B. Projektarbeiten) zur Überprüfung des eigenen Fortschritts durch Studierende genutzt werden. Eine Studie zur Untersuchung des Einflusses der beiden Eingabemethoden zur Unterstützung der Lernzielformulierung wurde durchgeführt, die Analyse steht aktuell noch aus.

In einer Weiterentwicklung soll vor allem die Überprüfung der Lernziele fokussiert werden. Dies beinhaltet einerseits die Möglichkeit, die bestehenden Lernziele mit passenden Zwischenzielen zu versehen (z. B. auch mittels KI), um den Fortschritt der Lernziele genauer betrachten zu können. Andererseits könnte das aktuell simple Dashboard weiterentwickelt und mehr auf die Bedürfnisse der Studierenden abgestimmt werden oder auch ein Abgleich mit Lehrzielen der Veranstaltungen ermöglichen. Ebenso die Anbindung an ein bestehendes Lernmanagementsystem ist zentral, um die Einbindung in den Studienalltag zu erleichtern.

Danksagung: Teile dieser Forschung wurden vom Bundesministerium für Bildung und Forschung (BMBF) im Projekt AIM@LMU unter dem Kennzeichen 16DHBKI013 gefördert.

Literaturverzeichnis

[CHL24] Chen, Y.-C.; Hwang, G.-J.; Lai, C.-L.: Motivating students to become self-regulatory learners: A gamified mobile self-regulated learning approach. Education and Information Technologies, 2024, DOI: 10.1007/s10639-024-12462-z.

[CK17] Cleary, T. J.; Kitsantas, A.: Motivation and Self-Regulated Learning Influences on Middle School Mathematics Achievement. School Psychology Review 46 (1), S. 88–107, 2017, DOI: 10.17105/SPR46-1.88-107.

[Do81] Doran, G. T.: There's a S.M.A.R.T. way to write management's goals and objectives. Management Review 70 (11), S. 35–36, 1981.

[DP16] Dörrenbächer, L.; Perels, F.: Self-regulated learning profiles in college students: Their relationship to achievement, personality, and the effectiveness of an intervention to foster self-regulated learning. Learning and individual differences 51, S. 229–241, 2016.

[He23] Heikkinen, S. et al.: Supporting self-regulated learning with learning analytics interventions – a systematic literature review. Education and Information Technologies 28 (3), S. 3059–3088, 2023, DOI: 10.1007/s10639-022-11281-4.

[Ka25] Karababa, Burhan: Digitale Unterstützung bei der Formulierung von Lernzielen zur Förderung des selbstregulierten Lernens, Bachelorarbeit, 2025.

[Ki20] Kismihók, G. et al.: Translating the Concept of Goal Setting into Practice–What'Else'does it Require than a Goal Setting Tool? arXiv preprint arXiv:2005.08669, 2020.

[KKM18] Knoth, A.; Kiy, A.; Müller, I.: Reflect. UP-App. Digitalisierung der Hochschullehre: Neue Anforderungen an die Evaluation?, S. 85, 2018.

[KMP21] Kim, J.-B.; Moon, K.-S.; Park, S.: When is a performance-approach goal unhelpful? Performance goal structure, task difficulty as moderators. Asia Pacific Education Review 22, S. 261–272, 2021.

[LL90] Locke, E. A.; Latham, G. P.: A Theory of Goal Setting and Task Performance. Prentice Hall, Englewood Cliffs, NJ, 1990.

[Ni84] Nicholls, J. G.: Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. Psychological Review 91 (3), S. 328–346, 1984, DOI: 10.1037/0033-295X.91.3.328.

[Pa17] Panadero, E.: A review of self-regulated learning: Six models and four directions for research. Frontiers in Psychology 8, S. 422, 2017.

[PD90] Pintrich, P. R.; De Groot, E. V.: Motivational and self-regulated learning components of classroom academic performance. Journal of Educational Psychology 82 (1), S. 33–40, 1990, DOI: 10.1037/0022-0663.82.1.33.

[Pi00] Pintrich, P. R.: The Role of Goal Orientation in Self-Regulated Learning. In (Boekaerts, M.; Pintrich, P. R.; Zeidner, M., Hrsg.): Handbook of Self-Regulation. Academic Press, San Diego, CA, S. 451–502, 2000.

[RD20] Ryan, R. M.; Deci, E. L.: Intrinsic and Extrinsic Motivation from a Self-Determination Theory Perspective: Definitions, Theory, Practices, and Future Directions. Contemporary Educational Psychology 61, S. 101860, 2020, DOI: 10.1016/j.cedpsych.2020.101860.

[Ro04] Robbins, S. B. et al.: Do Psychosocial and Study Skill Factors Predict College Outcomes? A Meta-Analysis. Psychological Bulletin 130 (2), S. 261–288, 2004, DOI: 10.1037/0033-2909.130.2.261.

[St24] Steinert, S. et al.: Harnessing large language models to develop research-based learning assistants for formative feedback. Smart Learning Environments 11 (1), 2024, DOI: 10.1186/s40561-024-00354-1.

[SZ12] Schunk, D. H.; Zimmerman, B. J.: Motivation and self-regulated learning: Theory, research, and applications. Routledge, 2012.

[ULA24] Urgo, G.; Lund, K.; Arguello, J.: Goal Setting Improves Learning Outcomes by Supporting Self-Regulated Learning. In: Proceedings of the 2024 Conference on Human Information Interaction and Retrieval (CHIIR). ACM, S. 213–223, 2024, DOI: 10.1145/3624918.3627605.

[va25] van Martins Jaarsveld, G. et al.: Goal setting in higher education: how, why, and when are students prompted to set goals? A systematic review. Frontiers in Education 9, 2025, DOI: 10.3389/feduc.2024.1511605.

[WH98] Winne, P. H.; Hadwin, A. F.: Studying as self-regulated learning. Hrsg. von Hacker, D. J.; Dunlosky, J.; Graesser, A. C., S. 277–304, 1998, DOI: 10.4324/9781410602350.

[WST21] Weber, F.; Schrumpf, J.; Thelen, T.: Development of a digital goal setting companion for higher education. In: DELFI 2021. Gesellschaft für Informatik eV, S. 109–114, 2021.

[ZBS24] Zhang, S.; Bryczkowski, J. M.; Specht, M.: GoaLearn App: A Goal-Setting and Monitoring Application to Support Students' Self-regulated Learning. In (Ferreira Mello, R. et al., Hrsg.): Technology Enhanced Learning for Inclusive and Equitable Quality Education. Bd. 15160, Lecture Notes in Computer Science, Springer Nature Switzerland, Cham, S. 267–272, 2024, DOI: 10.1007/978-3-031-72312-4_38.

[Zi02] Zimmerman, B. J.: Becoming a Self-Regulated Learner: An Overview. Theory into Practice 41 (2), S. 64–70, 2002, DOI: 10.1207/s15430421tip4102_2.

VRsim_EU – AI Act: Exploring Educational Simulation Gaming in Virtual Reality

Ulrike Mascher [1], David Fernes [1] und Alexander Tillmann [1]

Abstract: Simulation games can make complex situations with various stakeholders tangible for learners, e. g. in parliamentary simulations. By looking at the concept of game immersion as an integral part of the didactic method of simulation gaming as well as presence as a defining characteristic of the medium Virtual Reality (VR), the paper explores the potential benefits of combining the two in educational contexts. To investigate the potentials and challenges of VR-supported simulation gaming, a prototype of a simulation game set in the European Parliament Hemicycle is developed and will be tested with different target groups.

Keywords: Virtual Reality, Simulation Gaming, Immersive Learning, Digital Democracy

1 Introduction

Simulation games can make complex situations with various stakeholders tangible for learners, e. g. in parliamentary simulations. They represent real situations that are often characterised by complexity, with many stakeholders who have various perspectives on a specific problem. There are many different types of simulations and games (for a brief overview cf. [Kr17], cf. [Ki25]), this study focusses on political or parliamentary simulation games, of which well-known examples are Model United Nations[2], or in Germany, the Planspiel Bundestag.[3]

In these scenarios, participants gain knowledge about democratic processes where conflicting legitimate interests are represented and decisions are made by majority vote. Thus, the mechanisms of decision-making in parliamentary groups, committees and in plenary sessions can be experienced by learners. Simultaneously, numerous key skills are trained like communicative competences and being able to represent one's own interests. At the same time, participants take the perspective of other people and thereby recognise the legitimacy of different positions.

[1] Goethe University Frankfurt, studiumdigitale, Eschersheimer Landstraße 155-157, 60323 Frankfurt am Main, Germany, mascher@sd.uni-frankfurt.de, https://orcid.org/0009-0006-7298-2762; fernes@sd.uni-frankfurt.de, https://orcid.org/0000-0001-7063-6221; tillmann@sd.uni-frankfurt.de, https://orcid.org/0000-0001-7230-7042
[2] https://www.un.org/en/mun
[3] https://www.bundestag.de/planspiel

doi:10.18420/delfi2025_52

A preliminary literature research has shown that there are currently no simulation games of this type available in virtual reality (VR), which is why a VR-supported prototype scenario of a parliamentary simulation was developed. In the next step, the prototype will be tested with different focus groups in order to

1. iteratively improve the VR-application as well as the didactic scenario through user-centred development
2. investigate which potentials and challenges different target groups (including but not limited to students, VR experts, simulation gaming experts) identify for the VR-supported simulation game.

1.1 Why combining the educational method of simulation gaming and immersive VR?

Through role playing, realistic scenarios, and authentic problems, simulation games offer the opportunity to take on different perspectives within a simulated environment and to carry out safe trial actions [HM23] which is why they are increasingly used in higher education [Ch20]. In the same way, virtual reality (VR) offers learners the opportunity to test their own assumptions and options for action in a realistic setting without having to fear negative consequences [MBK20]. The study sets out to explore the potential of combining simulation gaming as an educational method with virtual reality as a medium.

According to constructivism and experiential learning [Ko84], learning thus becomes "situated" [SKS06] and also includes socio-cultural conditions. When it comes to learning in immersive VR (iVR), presence is identified as one of the main psychological factors for successful learning by the Cognitive Affective Model of Immersive Learning (CAMIL) [MP21]. Hence, experiencing presence can lead to higher situational interest and higher motivation. The positive impact of game immersion as well as presence on learning outcomes has been proven for digital serious educational games [CSA15] and immersive VR [ML18] as well as for analogue educational simulation games [Pr21].

1.2 Research Questions

In recourse to the above considerations, the question is whether the positive impact of simulation gaming (via game immersion) can be enhanced by using VR, taking advantage of presence as the psychological effect of technological immersion. In other words, is the parliamentary simulation gaming scenario a promising use case for learning in VR? The prototype scenario will be tested with different target groups (e. g. experts in simulation gaming, VR experts, students of educational science) firstly, to improve the prototype based on their feedback and secondly, to find out which potentials and challenges of a VR-supported simulation game they identify, with the aim of better understanding if and how VR can enhance certain learning scenarios.

2 *VRsim_EU – AI Act*: the Prototype of a VR-supported simulation game scenario

The simulation game scenario prototype *VRsim_EU – AI Act* is centred around the legislative procedure in the European Union concerning the Artificial Intelligence Act. The scenario was chosen because simulation games for political education are already well-established and are seen as a field that is very well suited to active and simulation-based learning [Us14; JB15]. With the help of avatars, the participants take on the roles of members of the European Parliament (MEPs) and discuss the regulation of AI in the virtual European Parliament Hemicycle.

2.1 Implementation

The VR Application for the game scenario was developed for the Meta Quest Platform using the Unity game engine. A framework [Ma24], utilising the Photon Unity Networking (PUN) and UltimateXR packages, was used for multiuser VR functionality with embodied avatars. The virtual environment was created in Blender based on photos and floor plans of the real European Parliament Hemicycle. The players are represented by Microsoft Rocketbox avatars resembling MEPs.

2.2 Didactic scenario

The didactic scenario (90 min.) is composed of a briefing, followed by a game phase of 15 minutes in VR, a short debriefing and a group discussion. At the beginning, learners receive a short briefing on the legislative process in the EU and the political groups represented in the European Parliament. Also, the most important facts about the AI Act are outlined, such as the AI-definition and the risk-based approach of the world's first law on artificial intelligence. Against this backdrop, the learners are then to discuss the question "Which AI applications pose an unacceptable risk and should therefore not be permitted in the EU?". During the 15-minutes-long VR-supported game phase, the learners wear a head-mounted display (Meta Quest 2) and experience the European Parliament Hemicycle in VR (see Fig. 1). They take on the role of a faction member and are represented by an avatar customised to the respective role. As part of the briefing, they are given printed role cards with their MEP's opinions and arguments in relation to the above question. These role cards are also available in VR.

A debriefing takes place immediately after the game phase. At first, the learners take off the VR headsets and the simulation facilitator asks about their emotions and feelings ("How are you? How did you feel?"). On the one hand, this phase serves to distance the learners from their roles and, on the other hand, gives them the opportunity to express their emotions and feelings. The debriefing phase plays an important role in simulation gaming when it comes to learning effects and transfer of what has been learnt in the simulation game [Kr10].



Fig. 1: screenshot of the VR-supported simulation game scenario

3 Future Research

Based on the developed prototype, future studies will be conducted to investigate whether the immersive technology VR promotes the motivation and interest of students in simulation gaming and supports social learning processes by creating a higher degree of game immersion, i.e. a mental immersion, in the simulation game scenario.

The extent to which the correlations described in models of immersive learning [MP21; Sl09] between technological factors, the resulting learning opportunities, cognitive-affective effects and learning success will be investigated in further empirical studies. In addition to psychological factors, technical, organisational and ethical challenges of using VR will also be systematically observed and discussed. Thereby, future research will provide empirical evidence for the further development of VR-based simulation games and their role in university teaching.

References

[CSA15] Cheng, M.-T.; She, H.-C.; Annetta, L.A.: Game immersion experience: its hierarchical structure and impact on game-based science learning. In: Journal of Computer Assisted Learning Bd. 31, Nr. 3, S. 232–253, 2015.

[Ch20] Chernikova, Olga; Heitzmann, Nicole; Stadler, Matthias; Holzberger, Doris; Seidel, Tina; Fischer, Frank: Simulation-Based Learning in Higher Education: A Meta-Analysis. In: Review of Educational Research Bd. 90, American Educational Research Association, Nr. 4, S. 499–541, 2020.

[HM23] Hertel, John Paul; Millis, Barbara: Using Simulations to Promote Learning in Higher Education: An Introduction, 2023.

[JB15] Jones, Rebecca; Bursens, Peter: the effects of active learning environments: how simulations trigger affective learning. In: European Political Science Bd. 14, Nr. 3, S. 254–265, 2015.

[Ki25] Kikkawa, T.; Kriz, W. C.; Sugiura, J.; De Wijse-Van Heeswijk, M. (Hrsg.): Transferring Gaming and Simulation Experience to the Real World, Translational Systems Sciences, Bd. 43, 2025.

[Kl09] Klabbers, Jan H. G.: Terminological Ambiguity: Game and Simulation. In: Simulation & Gaming Bd. 40, Nr. 4, S. 446–463, 2009.

[Ko84] Kolb, David: Experiential Learning: Experience as the Source of Learning and Development. Bd. 1, 1984.

[Kr10] Kriz, Willy Christian: A Systemic-Constructivist Approach to the Facilitation and Debriefing of Simulations and Games. In: Simulation & Gaming Bd. 41, Nr. 5, S. 663–680, 2010.

[Kr17] Kriz, Willy Christian: Types of Gaming Simulation Applications. In: Simulation & Gaming Bd. 48, Nr. 1, S. 3–7, 2017.

[Ma24] Mascher, Ulrike; Weiß, David; Fuchs, Andreas; Appel, Sven; Fernes, David; Sabah, Sam (2024): fuelsME:Create - ein Framework zur kollaborativen Entwicklung einer VR-Anwendung für die Hochschullehre. Proceedings of DELFI Workshops 2024. DOI: 10.18420/delfi2024-ws-06. Gesellschaft für Informatik e.V.

[ML18] Makransky, Guido; Lilleholt, Lau: A structural equation modeling investigation of the emotional value of immersive virtual reality in education. In: Educational Technology Research and Development Bd. 66, Nr. 5, S. 1141–1164, 2018.

[MP21] Makransky, Guido; Petersen, Gustav B.: The Cognitive Affective Model of Immersive Learning (CAMIL): a Theoretical Research-Based Model of Learning in Immersive Virtual Reality. In: Educational Psychology Review Bd. 33, Nr. 3, S. 937–958, 2021.

[MBK20] Mulders, Miriam; Buchner, Josef; Kerres, Michael: A Framework for the Use of Immersive Virtual Reality in Learning Environments. In: International Journal of Emerging Technologies in Learning (iJET) Bd. 15, Nr. 24, S. 208-224, 2020.

[Pr21] Preuß, Anna K.: The Learning Process in Live-Action Simulation Games: The Impact of Personality, Motivation, Immersion, and Flow on Learning Outcome in a Simulation Game. In: Simulation & Gaming Bd. 52, Nr. 6, S. 775–795, 2021.

[Sl09] Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1535), 3549–3557. DOI: 10.1098/rstb.2009.0138

[SKS06] Stahl, Gerry; Koschmann, Timothy D.; Suthers, Daniel D.: CSCL: An Historical Perspective, 2006.

[Us14] Usherwood, Simon: Constructing Effective Simulations of the European Union for Teaching: Realising the Potential. In: European Political Science Bd. 13, Nr. 1, S. 53–60, 2014.

FeeDi: Typunabhängiges Diagramm-Assessment durch deklarative Diagrammspezifikation

Erik Morawetz [1], Roy Meissner [1], Nadine Hahm [1] und Andreas Thor [1]

Abstract: Dieser Beitrag stellt FeeDi vor, ein E-Assessment-Tool für graphische Modellierung. FeeDi gibt automatisiert Feedback auf studentische Diagramme und kann verschiedene Diagrammtypen — auch selbst definierte — verarbeiten. Dies wird durch einen hohen Grad der Generalisierung möglich gemacht, in dem Diagramme in Property-Graphen transformiert werden. Lediglich die Spezifikationen dieser Übersetzung müssen für verschiedene Diagrammtypen individuell definiert werden, was über eine einfache Diagrammtypspezifikation (DTS) im YAML-Format geschieht, auf die sich dieser Beitrag fokussiert. Über eine Auszeichnungssprache werden in einer DTS alle Schritte festgehalten die notwendig sind, um aus einer visuellen Beschreibung eines Diagramms semantische Informationen abzuleiten. Dieser Beitrag demonstriert, wie mittels DTS Diagrammtypen kodiert werden und somit FeeDi einfach um neue Diagrammtypen erweitert werden kann.

Keywords: E-Learning, Adaptive Lernsysteme, E-Assessment, Diagrammassessment, Feedback

1 Einleitung

FeeDi ist ein E-Assessment Tool für die graphische Modellierung in MINT-Fächern. FeeDi generiert automatisiert Feedback auf von Studierenden erstellte Diagramme verschiedener Typen, indem es diese mit einer Musterlösung vergleicht. Es existieren bereits E-Assessment-Systeme für graphische Modellierung, die die Spezifika der entsprechenden Diagrammtypen gut abbilden. In einer umfassenden Literaturanalyse isolieren Ullrich et al. die Aspekte, in denen E-Assessment graphischer Modellierung Potential zum Ausbau aufweist [Ul23]. Bezeichnend ist, dass viele domänenspezifische Lösungen für einzelne Diagrammtypen existieren und wie wenig die Ausgabe von effektivem Feedback priorisiert wird [HT07]. Der Einsatz von E-Assessment in der grafischen Modellierung ist in Hinblick auf die Relevanz der hierbei zu erwerbenden Kompetenzen noch zu wenig verbreitet (vgl. [So22]). Der Aufwand studentische Diagramme manuell zu bewerten ist zudem hoch [Ca21]. Hier setzt FeeDi mit seinem generischen Ansatz an, um in vielfältigen Domänen E-Assessment von Diagrammen in den Lehralltag zu bringen [MHT23a; MHT23b]. In FeeDi sind Musterlösungen schnell erstellt und durch die automatische Bewertung kann FeeDi auch in Lehrveranstaltungen mit großen Gruppen effizient eingesetzt werden. Mit effektivem Feedback [HT07] soll FeeDi die Motivation zum Selbststudium und den Lernerfolg von Studierenden unterstützen.

[1] Hochschule für Technik, Wirtschaft und Kultur, Fakultät Digitale Transformation, Zschochersche Straße 69, 04229 Leipzig, Deutschland, erik.morawetz@htwk-leipzig.de, https://orcid.org/0000-0002-4052-2490; roy.meissner@htwk-leipzig.de, https://orcid.org/0000-0003-4193-8209; nadine.hahm@htwk-leipzig.de, https://orcid.org/0009-0002-1268-1029; andreas.thor@htwk-leipzig.de, https://orcid.org/0000-0003-2575-2893

doi:10.18420/delfi2025_53

2 Das Konzept FeeDi

FeeDi basiert auf dem Vergleich einer studentischen Lösung (SL) mit einer vorher angelegten Musterlösung (ML). Die ML wird durch Lehrende über die Weboberfläche von FeeDi erstellt, indem sie ein vorbereitetes Diagramm (PowerPoint, OpenOffice, DrawIO) hochladen und einem unterstützten Diagrammtyp zuordnen. Dabei haben Lehrende die Möglichkeit, die ML zusätzlich zu annotieren, um die für Studierende später sichtbare Punktevergabe und das automatisiert erzeugte, textuelle Feedback anzupassen, zu vertiefen, sowie auf das anvisierte Kompetenzlevel und die Lehrveranstaltung zuzuschneiden. In FeeDi können so ML mit geringem Zeitaufwand und mit oder ohne manuellen Anpassungen erstellt werden. Details, insbesondere für Entity-Relationship-Diagramme, sind in [MHT23b] beschrieben.

FeeDi unterstützt eine Vielzahl von Diagrammtypen, wobei Interpretationsregeln für jeden Diagrammtyp in einer sog. Diagrammtypspezifikation (DTS) festgehalten sind, die eine Deklaration des Diagrammtyps auf Basis eines YAML-Formats erlaubt, ohne dass Programmierkenntnisse vonnöten sind (s. Abschnitt 3). Anhand der DTS werden Eingabediagramme analysiert und in Property-Graphen umgewandelt, welche alle semantischen und syntaktischen Informationen eines Diagramms als eigenschaftsbehaftete Knoten und Kanten darstellen — die DTS beschreibt also die Logik eines Diagrammtyps, nicht aber individuelle Diagramme. Der Diagrammvergleich (SL mit ML) erfolgt dann als diagrammtypunabhängiges Matching zweier Property-Graphen. Dessen Ergebnisse werden zusammen mit Informationen der DTS und den aufgabenspezifischen Annotationen in Feedback verwandelt (mehr zu Graph-Matching und daraus resultierendem Feedback in [MHT23a]). Dieser Workflow von FeeDi und das Eingreifen der DTS ist in Abb. 1 dargestellt.

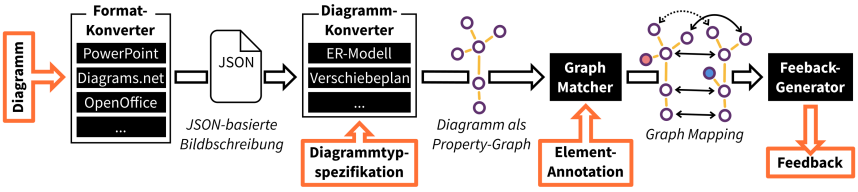


Abb. 1: Die Architektur von FeeDi

FeeDi ist Open-Source [Mo25] und kann als Web-Applikation bereitgestellt werden. Studierende können über einen aufgabenspezifischen Link Diagramme als Lösung einreichen (SL), auf die sie online Feedback erhalten. Abweichungen zwischen der SL und der ML werden graphisch hervorgehoben und zusätzlich textuell sowie mittels einer Punktwertung rückgespiegelt. Dadurch können Studierende selbstständig mit digitalen Aufgabensammlungen interagieren, lernen und diagrammspezifische Kompetenzen explorativ erwerben.

3 Diagrammtypspezifikation und Interpretation

Diagramme der meisten Diagrammtypen lassen sich auf eine Sammlung von immer wiederkehrenden Formen reduzieren, wie bspw. Vierecke, Ovale, Linien und Pfeile. Diese können

auch nur einige grundlegende Beziehungen zueinander haben: z. B. sich überschneiden, andere Elemente beinhalten, mit anderen Elementen verbunden sein. FeeDi basiert auf einer Bibliothek solcher Formelemente, denen zunächst keinerlei semantische Bedeutung zugewiesen ist. FeeDi liest diese Basiselemente aus einem Diagramm aus (vgl. Abb. 1) und interpretiert sie entsprechend des gegebenen Diagrammtyps. Die semantische Deklaration eines Diagrammtyps geschieht über eine Diagrammtypspezifikation (DTS), in der im YAML-Format alle für FeeDi relevanten Informationen zur Interpretation eines Diagramms des entsprechenden Typs in drei Schritten definiert werden:

1. Symbolerkennung (Aggregation): Regelsatz zur Reduktion von Basiselementen zu funktionalen Einheiten, die die semantische Interpretation ermöglichen (bspw. das Zusammenfügen von Linien). Äquivalent zur menschlichen Wahrnehmung von Symbolen in Diagrammen.

2. Elementinterpretation (Knotengenerierung): Regelsatz zur Zuordnung von Basiselementen oder derer Reduktionen zu semantischen Diagrammelementen (z. B. Viereck → Entität). Dies kann wiederum Reduktionen beinhalten. Äquivalent zum menschlichen Auslesen der Semantik von erkannten Symbolen.

3. Beziehungsaufbau (Kantengenerierung) Deklaration von Beziehungen zwischen Diagrammelementen, die semantische Information tragen (bspw. die Verbindung von Elementen mit Linien im ER-Diagramm). Äquivalent zum menschlichen Erkennen der Beziehungen zwischen den Symbolen.

Abb. 2 veranschaulicht diese Interpretationsschritte anhand von zwei Beispieldiagrammen; DTS-Beispiele sind unter [Mo25] zu finden. In der DTS werden Diagrammelementen eindeutig Symbole zugeordnet, aus denen sie hervorgehen. Symbole sind Basiselemente oder Muster von Basiselementen, die abhängig vom Diagrammtyp über Inklusionsrelationen auf Basis von Lageinformationen erkannt werden müssen (cyan in Abb. 2). So sind bspw. in einem ER-Diagramm zwei zusammenhängende Linien als eine Linie zu verstehen. In einem Verschiebeplan aber sind zwei separate Linien zwei individuelle Balken, selbst wenn sie an einem Ende zusammenhängen. Basiselemente werden so anhand von Mustern zu Symbolen reduziert die daraufhin in Diagrammelemente übersetzt werden (violett in Abb. 2). Diese Muster können sehr einfach sein, wie die Reduktion eines Vierecks, das ein Viereck beinhaltet, zu einem Viereck mit doppelter Außenlinie, oder eben die Reduktion mehrerer zusammenhängender Linien zu einer langen Linie. Muster können allerdings mithilfe einer einfachen Symbolsprache beliebig komplex ausgestaltet werden, so wäre z. B. `line x (line x oval)` zwei Linien die einen Kreis schneiden. Für Reduktionen und Zuordnungen können Bedingungen definiert werden, um anhand von Elementeigenschaften Entscheidungen zu treffen. In einem ER-Diagramm wird bspw. ein Oval mit Text einem Primärschlüssel zugeordnet, wenn der Text unterstrichen ist, aber ansonsten einem Attribut. Reduktionen können auch für Diagrammelemente definiert werden. Dies kommt zum tragen, wenn erst die semantische Zuordnung durchgeführt werden muss, bevor über eine Reduktion entschieden werden kann. In einem Schaltkreis müssen z. B. erst Knoten als solche erkannt werden, bevor diese semantisch zu einem einzigen reduziert werden können.

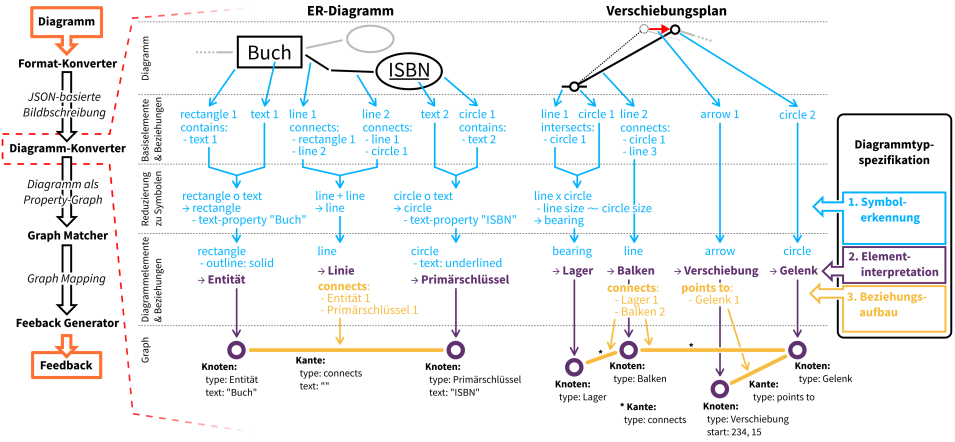


Abb. 2: Interpretation verschiedener Diagramme anhand der Diagrammtypspezifikation (DTS)

FeeDi reduziert also in mehreren Schritten Basiselemente die in der ursprünglichen Zeichnung vorliegen zu semantischen Einheiten bis keine Reduzierung mehr möglich ist. Da in Diagrammen keine Ambivalenzen von Symbolen existieren — visuelle Kodierung bildet die Grundlage graphischer Modellierung — ist die Reduzierung eindeutig und unabhängig von der Reihenfolge der Durchführung. Liegen Elemente in der Zeichnung vor, die nicht korrekt zusammengesetzt sind, kann die Reduzierung nicht vollständig stattfinden. Das führt zu einer Fehlermeldung der Art „unbekanntes Element“ oder es werden falsche Diagrammelemente zugeordnet, falls eine Überlappungen von Mustern vorliegt. Dies wird später entweder in der Annotationsansicht einer ML oder — als Fehler ausgezeichnet — in der Feedbackansicht einer SL ersichtlich und kann korrigiert werden.

Nach der vollständigen Reduzierung werden Diagrammelemente in Knoten des Graphen übersetzt (violett in Abb. 2). Kanten werden aus Beziehungen zwischen den Diagrammelementen gebildet, wie in der DTS deklariert (gelb in Abb. 2). Es können aber auch Diagrammelemente als Kanten definiert werden. Durch die Übertragung der semantischen Inhalte in die Eigenschaften von Knoten und Kanten kann der Abgleich zwischen ML und SL durch einen generischen Graph-Matching-Algorithmus realisiert werden.

In der DTS sind all diese Schritte der Reduzierung programmatisch deklariert. Von den Basiselementen ausgehend werden Regeln festgehalten, mit denen Schritt für Schritt semantische Diagrammelemente gelesen und letztlich syntaktisch miteinander in Beziehung gesetzt werden können. Im Frontend werden diese Mechanismen nur dadurch ersichtlich, dass verschiedene Diagrammtypen während der Erstellung einer ML ausgewählt werden können. Im Backend können Lehrende allerdings nach Belieben neue Diagrammtypen installieren oder selbst entwickeln, solange sich diese mit den verfügbaren Basiselementen und -beziehungen beschreiben lassen. Sowohl Frontend- als auch Backend-Funktionalitäten können in der Demonstration ausgetestet werden.

Literaturverzeichnis

[Ca21] Cavalcanti, A. P. et al.: Automatic feedback in online learning environments: A systematic literature review. Computers and Education: Artificial Intelligence 2, S. 100027, 2021, DOI: 10.1016/j.caeai.2021.100027.

[HT07] Hattie, J.; Timperley, H.: The Power of Feedback. Review of Educational Research 77 (1), S. 81–112, 2007, DOI: 10.3102/003465430298487.

[MHT23a] Morawetz, E.; Hahm, N.; Thor, A.: Automatisierte Bewertung und Feedback-Generierung für grafische Modellierungen und Diagramme mit FeeDI. In: 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., Bonn, S. 97–102, 2023, DOI: 10.18420/delfi2023-18.

[MHT23b] Morawetz, E.; Hahm, N.; Thor, A.: E-Assessment für Entity-Relationship-Diagramme mit FEEDI. In: 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik eV, S. 275–276, 2023.

[Mo25] Morawetz, E. et al.: Official FeeDi Website & Repository, www.feedi.info, 2025.

[So22] Soyka, C. et al.: Toward a Competence Model for Graphical Modeling. ACM Trans. Comput. Educ. 23 (1), 2022, DOI: 10.1145/3567598.

[Ul23] Ullrich, M. et al.: Automated assessment of conceptual models in education: a systematic literature review. Enterprise Modelling and Information Systems Architectures (EMISAJ) 18, S. 2–1, 2023.

Collect Skills – Tool zur Selbsteinschätzung des individuellen Kompetenzerwerbs

Anna Pakosch[1], Dennis Allerkamp[1], Anne Nadolny[2], Arne Koschel[1], Monika Stöhr[2] und Christin Schulze[1]

Abstract: Collect Skills ist eine auf Gamification-Elementen basierende Anwendung für Studierende zur Selbsteinschätzung des individuellen Kompetenzerwerbs. Das didaktische Konzept der Anwendung basiert auf der Idee, dass das Absolvieren bestimmter Tätigkeiten zum Erwerb der angestrebten Kompetenzen beitragen kann. Dabei soll die Motivation der Studierenden durch Gamification-Elemente erhöht werden. Collect Skills wurde prototypisch als H5P-Inhaltstyp implementiert, damit dieser in Learning Management Systemen (LMS) eingesetzt werden kann. Durch die Konfigurierbarkeit erwarteter Tätigkeiten ist es möglich, die Anwendung an verschiedene Lehrveranstaltungen anzupassen. Erste Evaluierungsergebnisse des Einsatzes von Collect Skills in studentischen Projekten liegen vor.

Keywords: Kompetenzerwerb, Selbsteinschätzung, Gamification, H5P



1 Einleitung

Die Entwicklung von Kompetenzen ist sowohl im Berufsleben als auch in der Lehre ein relevantes Thema. Für Studierende und junge Erwachsene gibt es bisher nur eingeschränkte Möglichkeiten zur digitalen Erfassung und Dokumentation ihrer erworbenen Kompetenzen (Abschnitt 2). Mit dem Selbsteinschätzungstool *Collect Skills* soll Studierenden die Möglichkeit zur Reflexion der individuellen Kompetenzaneignung geboten und gleichzeitig durch das Angebot konkreter Tätigkeiten Anreize zum Erwerb gewünschter Fähigkeiten gesetzt werden. Dabei setzt *Collect Skills* auf Gamification, um die intrinsische Motivation der Nutzenden zu steigern. Dem Design-Science-Forschungsansatz [He04; He07] folgend wurde *Collect Skills* prototypisch implementiert (Abschnitt 3) und im Hochschulkontext erprobt (Abschnitt 4). Dabei ging es um die Frage, wie ein solches Tool gestaltet werden muss, damit Studierende dieses im Sinne der Konzeption projektbegleitend einsetzen.

2 Verwandte Arbeiten

Im Bereich der subjektiven Kompetenzeinschätzung existieren digitale Anwendungen für Studierende überwiegend für die Analyse der Berufs- und Studienwahl. Beispiele hierfür

[1] Hochschule Hannover, Fakultät IV – Wirtschaft und Informatik, Ricklinger Stadtweg 120, 30459 Hannover, Deutschland, ; dennis.allerkamp@hs-hannover.de, https://orcid.org/0009-0006-4217-0085; arne.koschel@hs-hannover.de, https://orcid.org/0000-0001-5695-2893

[2] Hochschule Hannover, Fakultät II – Maschinenbau und Bioverfahrenstechnik, Ricklinger Stadtweg 120, 30459 Hannover, Deutschland, anne.nadolny@hs-hannover.de, https://orcid.org/0009-0000-4266-679X

doi:10.18420/delfi2025_54

bieten die Hochschule Hannover mit einem Informatik-Eignungstest oder das Online-Self-Assessment-Verfahren der Goethe Universität (Frankfurt am Main) für die Studiengänge Psychologie und Informatik [Hsh24; Re09]. Die Rückmeldung zu personalen Kompetenzen erfolgt jedoch rein im Hinblick auf die jeweiligen Studiengänge. Eine Möglichkeit zur Entwicklung von Sach-, Sozial- und Selbstkompetenzen der Zielgruppe Studierende und junge Erwachsene wird dabei vernachlässigt.

Im medizinischen Bereich ist dieses Defizit bereits erkannt worden. Aufgrund von pandemiebedingten Kontaktbeschränkungen sind unterschiedliche Lernanwendungen zur Selbst- und Fremdeinschätzung entstanden: Bei „KoFFee“ handelt es sich um ein Projekt der Medizinischen Fakultät Freiburg. Soziale Kompetenzen werden dort durch ein Online-Feedbacktool auf Basis des Freiburger Fragebogens zur Erfassung von Kompetenzen in der Medizin abgefragt [Gi11; KK21]. Das *Interprofessionelle Visiten Individual Team Assessment Tool* (IP-VITA) wiederum erfasst individuelle und Team-Kompetenzen in klinischen Lernsettings und kann als formatives und summatives Assessment-Instrument eingesetzt werden [Mi23].

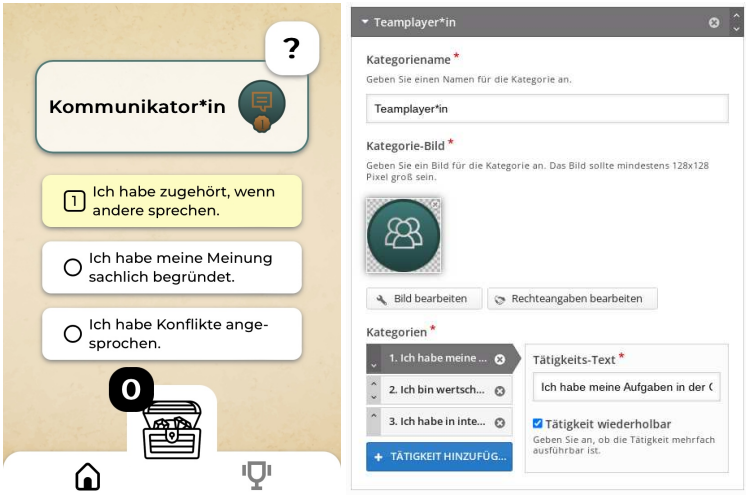
Die vorliegende Übersicht aktueller Forschungsarbeiten zeigt, dass es kein digitales Tool zur Reflexion des individuellen Kompetenzerwerbs im Kontext von Hochschulprojekten gibt.

3 Konzeption und technische Umsetzung

Hauptaufgabe von *Collect Skills* ist es, Studierende im Rahmen von Projektarbeiten bei der Entwicklung ihrer Kompetenzen zu unterstützen und diese zu dokumentieren. Lehrende können verschiedene Kompetenzkategorien und dazugehörige Tätigkeiten, die zum potenziellen Kompetenzerwerb führen können, individuell an Projekte anpassen. Durch das Absolvieren und Sammeln von Tätigkeiten sollen Studierende eigene Fähigkeiten erweitern und sich ihrer dadurch erlangten Kompetenzen bewusst werden.

Collect Skills setzt zusätzlich auf Gamification-Anreize. Der Einsatz von Spielelementen in einer Lernanwendung soll die intrinsische Motivation der Nutzenden sowie die Attraktivität des Tools steigern [Ke18, S. 397]. In *Collect Skills* ist das Sammeln von Tätigkeiten, die zum Kompetenzerwerb führen können, in Stufen (Level) unterteilt: Es werden Voraussetzungen aufgelistet, die zur Freischaltung von Abzeichen notwendig sind. Dabei sind die Tätigkeiten in den unterschiedlichen Kategorien frei wählbar und werden nicht linear, sondern explorativ absolviert. Für die meisten Tätigkeiten kann dies auch mehrfach geschehen. Studierende sollen dazu motiviert werden, möglichst viele der vorgeschlagenen Tätigkeiten zu absolvieren und so das gesamte Spektrum der angestrebten Kompetenzen weiterzuentwickeln.

Nutzende können in *Collect Skills* zwischen zwei Seiten navigieren. Auf der Hauptseite können Tätigkeiten per Drag’n’Drop oder durch ihr Anklicken in einer Schatztruhe gesammelt und als absolviert markiert werden (Abb. 1(a)). Auf der Abzeichenseite kann die individuelle Kompetenzentwicklung per Fortschrittsbalken und einer Liste der absolvierten Tätigkeiten eingesehen werden. Zusätzlich visualisiert ein Graph die kumulierten Tätigkeiten.



(a) Hauptseite von *Collect Skills* (b) Konfiguration einer Kategorie

Abb. 1: User Interface von *Collect Skills*

Damit *Collect Skills* in unterschiedlichen Projektkontexten genutzt werden kann, ist die Konfiguration der Tätigkeiten durch Lehrende möglich (Abb. 1(b)). Um den Aufwand möglichst gering zu halten, sind nur die Kategorien und die zugehörigen Tätigkeiten konfigurierbar. Abzeichen sind hingegen fest vorgegeben.

Ein erster Prototyp wurde in einer Bachelorarbeit als Single-Page-Application entwickelt [Mü22]. Im Rahmen einer weiteren Bachelorarbeit [Le23] und einer Überarbeitung durch Allerkamp wurde ein neuer Prototyp entwickelt: Als HTML5 Package (H5P) kann das Tool auch ohne Programmierkenntnisse an individuelle Projekte angepasst und in vorhandene LMS integriert werden [H5P23]. Bei Verwendung werden die Nutzungsdaten, insbesondere bereits absolvierte Tätigkeiten, bei jeder Änderung umgehend gespeichert. [Ta23] In einer ersten Evaluierung wurde Optimierungsbedarf im Hinblick auf die Selbstbeschreibungsfähigkeit deutlich, weshalb das Tool vereinfacht, reduziert und internationalisiert wurden. [Pa23] In einer weiteren Bachelorarbeit wurde durch die Verwendung des React-Frameworks eine Reduzierung der Komplexität und eine Steigerung der Wiederverwendbarkeit von Komponenten für zukünftige H5P-Projekte erreicht [Fü24].

4 Erprobung

Insgesamt 54 Studierenden (17 weibliche, 27 männliche) wurde *Collect Skills* zur Nutzung angeboten; davon jeweils zwei Mal im Wahlpflichtfach „Praxisprojekt“ (29 Teilnehmende, davon 9 weibliche) über das gesamte Semester gestreckt und zwei Mal im freiwilligen

„Ergänzendes Fach“ (35 Teilnehmende), das sich jeweils über drei und vier Tage streckte. Das Tool wurde mithilfe eines Fragebogens über die angebotenen Kompetenzen, die Nutzung von *Collect Skills*, Stärken und Schwächen, die Gamification sowie die Komplexität evauliert. Er wurde ausschließlich in den Wahlpflichtfächern „Praxisprojekt“ am Ende des Semesters ausgeteilt. Zusätzlich wurden Nutzungsdaten in allen vier Fächern herangezogen.

Von den 54 beteiligten Studiernden nahmen – laut Fragebogen – 34 Personen das Angebot wahr und nutzen *Collect Skills* laut Nutzungsdaten zum Teil sehr aktiv. Dabei fiel die Auswertung des Fragebogens negativer aus als die tatsächliche Nutzung selbst: Auf die Frage, ob *Collect Skills* im Projekt eine unterstützende Hilfe bot, befürworteten dies acht Teilnehmende, während sich 16 dagegen aussprachen. Sechs Teilnehmende nahmen das Sammeln der Abzeichen als positiv wahr, 13 Teilnehmende erhofften sich mehr kompetitive Elemente. Auf die Frage nach der Komplexität waren sich fast alle Teilnehmenden einig, dass *Collect Skills* intuitiv und einfach zu bedienen sei (21 Teilnehmende). Die Auswertung der Nutzungsdaten ergab hingegen, dass 41 der 54 Teilnemenden *Collect Skills* mindestens ein bis drei Tage je ergänzendes Fach und ein bis 87 Tage in den Praxisprojekten nutzten. Gerade in den Praxisprojekten wurde *Collect Skills* von Semesteranfang bis Ende von 14 Teilnehmenden sehr regelmäßig genutzt.

Hervorzuheben ist, dass die weiblichen Studierenden *Collect Skills* nicht nur häufiger genutzt, sondern auch insgesamt positiver bewertet haben: Sechs von neun weiblichen Studierenden gaben an, dass der persönlich visualisierte Projektfortschritt in *Collect Skills* motivierend sei. Bei den männlichen Teilnehmenden gaben 13 von 20 an, dass sie sich mehr Wettbewerb mit anderen im Gamification-Bereich wünschen. Da in der Fachliteratur von kompetitiven Elementen abgeraten wird, ist dies zu erörtern [Ke18, S. 398].

5 Fazit

Mit dem H5P-Inhaltstyp *Collect Skills* können Studierende ihren Kompetenzerwerb reflektieren. Zudem wird der Zusammenhang zwischen den in der Anwendung verständlich formulierten Tätigkeiten und den daraus resultierenden Kompetenzen verdeutlicht. Die Studierenden werden dazu angeregt, die vorgeschlagenen Tätigkeiten auszuführen. Durch die Konfigurierbarkeit durch Lehrende kann die Anwendung in verschiedenen Lehrveranstaltungen eingesetzt werden. Die Evaluierungsergebnisse zeigen, dass *Collect Skills* in semesterbegleitenden Projekten gut angenommen wird. Interessant ist zudem, dass das Tool von weiblichen Studierenden überdurchschnittlich positiv bewertet wird.

Dieses Projekt wurde vom Förderprogramm „Innovation plus“ des Landes Niedersachsen unterstützt. Wir danken Finn Müller für die erste prototypische Umsetzung sowie Philipp Lempfer, Dannie Krösche, Brian Kuhn, Sophie Fürhoff und Jonathan Balzer für die Unterstützung bei der Weiterentwicklung der Software. Ein weiterer Dank gilt Timo Schnitt und Christian Gädtke für den Einsatz von *Collect Skills* im Praxisprojekt.

Literaturverzeichnis

[Fü24] Fürhoff, S.: Entwurf einer zukunftsfähigen Architektur für H5P-Inhaltstypen am Beispiel einer Lernanwendung, Bachelorarbeit, Hochschule Hannover, 2024, DOI: 10.25968/opus-3548.

[Gi11] Giesler, M. et al.: Entwicklung eines Fragebogens zur Erfassung von Kompetenzen in der Medizin: Ergebnisse zur Reliabilität und Validität. GMS Zeitschrift für Medizinische Ausbildung 28 (2), 2011, DOI: 10.3205/zma000743.

[H5P23] H5P Boilerplate, H5P, 2023, https://github.com/h5p/h5p-boilerplate, Stand: 26. 02. 2024.

[He04] Hevner et al.: Design Science in Information Systems Research. MIS Quarterly 28 (1), S. 75, 2004, DOI: 10.2307/25148625.

[He07] Hevner, A. R.: A Three Cycle View of Design Science Research. Scandinavian Journal of Information Systems 19 (2), S. 87–92, 2007.

[Hsh24] Informatik-Eignungstest der Hochschule Hannover, Hochschule Hannover, 2024, https://eignungstest.inform.hs-hannover.de/, Stand: 11. 04. 2024.

[Ke18] Kerres, M.: Mediendidaktik: Konzeption und Entwicklung digitaler Lernangebote. De Gruyter Oldenbourg, Berlin ; Boston, 2018.

[KK21] Kunz, K.; Köpper, H.: KoFFee – Kompetenz-Förderung durch Feedback. Ergebnisse der ersten Testphase eines Online-Feedbacktools für Studierende. In. Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). German Medical Science GMS Publishing House, Zürich, Schweiz (virtuell), DocP088, 2021, DOI: 10.3205/21gma283.

[Le23] Lempfer, P.: Umsetzung einer mit Angular entwickelten Lernanwendung in H5P, Bachelorarbeit, Hochschule Hannover, 2023.

[Mi23] Mitzkat, A. et al.: Individuelle Kompetenzen und Team-Performance in klinischen Lernsettings der interprofessionellen Zusammenarbeit erfassen: Die empirische Entwicklung des *Interprofessionelle Visiten Individual und Team Assessment Tools* (IP-VITA). Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen 179, S. 39–48, 2023, DOI: 10.1016/j.zefq.2023.03.007.

[Mü22] Müller, F.: Tool zur Selbsteinschätzung der individuellen Kompetenzentwicklung im Kontext von projektorientierten Lehr-/ Lernformaten, Bachelorarbeit, Hochschule Hannover, 2022.

[Pa23] Pakosch, A.: Konzept zur Verbesserung des Selbsteinschätzungs-Tools SETKOMP im Kontext einer interdisziplinären Zukunftswerkstatt für Studierende, Bachelorarbeit, Hochschule Hannover, 2023.

[Re09] Reiss, S. et al.: Online-Self-Assessments zur Erfassung studienrelevanter Kompetenzen. Zeitschrift für Hochschulentwicklung, 2009, DOI: 10.3217/zfhe-4-01/05.

[Ta23] Tacke, O.: How can I let H5P content types store the state actively?, 2023, https://snordian.de/2023/03/12/how-can-i-let-h5p-content-types-store-the-state-actively/, Stand: 28. 02. 2024.

Teach-R Sporthalle: Erweiterung des VR-Klassenzimmers um Lernszenarien zur Professionalisierung von Sportlehrkräften



Marc Schlicker [1], Astrid Schmidt [2], René Wilzeck [1], Britta Fischer [2], Thilo Kleickmann [2] und Raphael Zender [4]

Abstract: Eine VR-Sporthalle, die im Rahmen des Projekts DigiProSMK als ein Teil von lernen:digital entwickelt wurde und technisch auf dem Teach-R-Framework basiert, bietet neue und weitreichende Möglichkeiten für die Lehrkräftefortbildung. So können unterrichtsnahe Fähigkeiten wie etwa die professionelle Wahrnehmung in eigens dafür konzipierten Szenarien erprobt, verbessert und reflektiert werden, ohne dass dafür eine reale Sporthalle mit entsprechenden Unterrichtssituationen zur Verfügung stehen muss. Der Beitrag thematisiert sowohl technische Aspekte (Motion-Capture-basierte Animationen, Inverse Kinematics, externes Coachingsystem) als auch didaktische Überlegungen.

Keywords: Virtual Reality, Sportdidaktik, Simulation, Doppelte Verletzbarkeit, Digitale Lernorte, Lehrkräftebildung, Klassenraummanagement, Professionelle Unterrichtswahrnehmung

1 Hintergrund und Zielsetzung

Digitale Medien bieten großes Potenzial, um Lehrkräfte gezielt in ihren individuellen Professionalisierungsprozessen zu unterstützen. Dies gilt neben dem Klassenraumunterricht auch für den Sportunterricht. Insbesondere immersive Technologien wie virtuelle Realität (VR) eröffnen hier neue Wege der Fortbildung [KM21; NSR16]. So ermöglicht etwa der Einsatz einer VR-Sporthalle in der Lehrkräftebildung die standardisierte, beliebig oft wiederholbare Simulation von spezifischen Unterrichtsszenarien in einem für die Lernenden geschützten Raum; die Komplexität dieser Szenarien kann dabei adaptiv variiert werden (vgl. [Hu21; Se22]).

Der vorliegende Beitrag zeigt, wie solche Lernszenarien mit der Zielsetzung gestaltet werden können, die professionelle Unterrichtswahrnehmung von Lehrkräften gezielt zu fördern, indem sie herausfordernde Situationen zur Kompetenzentwicklung bieten. Im Mittelpunkt steht dabei die Umsetzung dieser Szenarien sowie die Gestaltung von VR-Interaktionen, die spezifische Lern- und Reflexionsprozesse anstoßen.

[1] Humboldt-Universität zu Berlin, Institut für Informatik, Rudower Chaussee 25, 12489 Berlin, Deutschland, schlickm@hu-berlin.de, https://orcid.org/0009-0001-5417-380X; rene.wilzeck@hu-berlin.de, https://orcid.org/0009-0004-0431-8689

[2] Christian-Albrechts-Universität zu Kiel, Institut für Pädagogik / Institut für Sportwissenschaft, Olshausenstraße 74-75, 24118 Kiel, Deutschland, a.schmidt@paedagogik.uni-kiel.de, https://orcid.org/0009-0000-2333-4773; britta.fischer@email.uni-kiel.de, https://orcid.org/0000-0002-0525-534X; kleickmann@paedagogik.uni-kiel.de, https://orcid.org/0000-0002-0608-6971

[4] Zeppelin Universität, Am Seemooser Horn 20, 88045 Friedrichshafen, Deutschland, raphael.zender@zu.de, https://orcid.org/0000-0001-9866-9455

doi:10.18420/delfi2025_55

In der zugehörigen Demonstration kann zunächst die Rolle der Lehrkraft innerhalb der virtuellen Realität eingenommen und das im Folgenden beschriebene Szenario durchlaufen werden. Zusätzlich besteht die Möglichkeit, die Perspektive eines Coachs zu übernehmen, um dasselbe Szenario aus externer Sicht zu beobachten und gezielt zu steuern.

2 Didaktisches Konzept und Pilotierung

Die hier thematisierte VR-Sporthalle wird aktuell im Rahmen einer Professionalisierungsmaßnahme für angehende Sportlehrkräfte zur Förderung der professionellen Unterrichtswahrnehmung [FL20; Sh17] an der CAU Kiel eingesetzt. Dort dient sie nach theoretischen und methodischen Grundlagen- und Vertiefungseinheiten als VR-Unterrichtssimulation, in der die angehenden Sportlehrkräfte verschiedene Unterrichtssituationen durchlaufen und ihr dort gezeigtes Handeln anschließend gemeinsam mit ihren Kommiliton:innen kriteriengeleitet reflektieren. Diese praxisnahe Anwendung fachlichen Wissens ermöglicht eine tiefgehende Reflexion der jeweils zuvor erworbenen Inhalte sowie der eigenen späteren Performanz im Unterricht.

Wahrnehmungsprozesse sind für das professionelle Handeln im Sportunterricht zentral, denn um situativ angemessene Entscheidungen für die Gestaltung von Unterricht treffen zu können, ist die Fähigkeit, lernrelevante Unterrichtsergebnisse zu erkennen und theoriebasiert interpretieren zu können, unabdingbar. Der inhaltliche Schwerpunkt der Konzeption liegt auf dem Aspekt der doppelten Verletzbarkeit von Schüler:innen im Sportunterricht [MK04]. Neben grundlegenden Aspekten der physischen und psychischen Verletzbarkeit im Unterricht des Faches Sport wird im Hinblick auf zweiteres vor allem die Emotion Scham vertiefend fokussiert. Dies beinhaltet sowohl das Aufzeigen potenziell schamauslösender Situationen im Sportunterricht als auch die Prävention und den Umgang mit solchermaßen sensiblen und hochinferenten Situationen. In der VR-basierten Lernsituation üben derzeit angehende Lehrkräfte – wie zukünftig fertig ausgebildete Lehrkräfte – genau diese Aspekte. Dabei stehen sie bspw. vor der Herausforderung, potenziell schamsensible Berührungen, die während eines Handstandszenarios im Rahmen der Hilfestellung erfolgen, zunächst einmal zu erkennen und anschließend je nach verbaler und mitunter auch gestischer Reaktion des betroffenen Avatars einzuschätzen, ob sie darauf entsprechend reagieren müssen, zum Beispiel durch eine verbale Intervention, oder ob ein solches Eingreifen in der konkreten Situation nicht notwendig ist, da aus der schamsensiblen offensichtlich keine schamvolle Situation geworden ist. Diese Unterscheidung ist keineswegs trivial und erfordert neben entsprechendem theoretischen Wissen ein hohes Maß an Sensibilität und Performanz auf Seiten der (angehenden) Lehrkraft. Das Handeln in der VR-Sporthalle bietet diesbezüglich im Gegensatz zu der Arbeit mit Videos unterstützend die Möglichkeit, einer Unterrichtssituation nicht nur als externer Beobachter beizuwohnen, sondern vollumfänglich die Rolle der Lehrkraft einzunehmen und aus dieser heraus zu agieren. Dies hat durch die darüber erlebte Authentizität nicht nur positive Auswirkungen auf Motivation und Selbstwirksamkeit [Hu23], sondern ermöglicht auch unmittelbare Rückmeldungen zum individuell gezeigten professionellen Verhalten der

(angehenden) Lehrkraft. Zusätzlich kann die Komplexität der Situation über das externe Coachsystem analog zur jeweiligen Performanz adaptiv variiert und gesteuert werden. Ein entsprechendes Feedback zum Handeln in der VR dient anschließend unter anderem der Reflexion und weiteren Gestaltung des eigenen Professionalisierungsprozesses.

3 Technische Anforderungen und Umfang

Wie im Vorgängerprojekt [WZ24] bildet ein Handstand-Übungsszenario, aus dem Inhaltsbereich Gerätturnen, den Kern der Anwendung: Zwei Gruppen à drei virtuelle Schülerinnen und Schüler (vSuS) üben auf zwei parallel zueinander ausgerichteten Mattenbahnen einen Handstand mit Hilfestellung. Dieses Szenario umfasst neun unterschiedliche Interaktionsabläufe zwischen den vSuS, die sich in Ausführung des Handstands und der Hilfestellung unterscheiden. Über ein PC-basiertes Coachsystem lassen sich diese Abläufe flexibel kombinieren und steuern.

Das Projekt wird als Teach-R[5]-Modul umgesetzt und stellt eine funktionale sowie didaktische Erweiterung dieser Umgebung dar. Teach-R ist ein an der Universität Potsdam entwickeltes VR-System zur Förderung von Klassenmanagement-Kompetenzen im Lehramtsstudium. In realitätsnahen Unterrichtssimulationen können gezielt Störungen durch vSuS ausgelöst werden, um adäquate Reaktionen angehender Lehrkräfte in geschützter Umgebung zu trainieren [Wi19]. Während die bisherigen Module von Teach-R primär Unterricht im Klassenraum mit verbaler Interaktion fokussieren, ergänzt das hier vorgestellte Modul das System um eine virtuelle Sporthalle.

Aus der Konzeption des Lehr-Lernszenarios ergeben sich konkrete Anforderungen an die technische Umsetzung:

Animationen Der Fokus erweitert sich um den Aspekt der nonverbalen Kommunikation wie etwa Körpersprache, der Durchführung von (sportmotorischen) Bewegungen und des Raumverhaltens – alles zentrale Punkte im Sportunterricht. Solche Anforderungen stellen im Bereich der Animationserzeugung neue Herausforderungen dar. Dies gilt insbesondere, da Bewegungen körperlich authentisch und kontextsensitiv modelliert werden müssen, um ein glaubwürdiges und didaktisch sinnvolles Szenario zu erzeugen. Gleichzeitig muss die Auswahl und Anzahl an Animationen groß genug sein, um passende Folgeanimationen für die Umsetzung von möglichen Anweisungen der sich in der Rolle der Lehrkraft befindenden Person durch die vSuS zu haben.

Die Animationen wurden mittels Motion-Capture-Daten eines Ganzkörper-VR-Tracking-Anzug (Tesla Suit[6]) aufgezeichnet und danach weiterverarbeitet. Der Anzug erfasst die Bewegungen mittels Inertialsensoren (IMUs), deren Daten in der zugehörigen Software

[5] Teach-R (ehemals VR-Klassenzimmer): https://gitup.uni-potsdam.de/mm_vr/vr-klassenzimmer/-/wikis/home
[6] Tesla-Suit: https://teslasuit.io

(Tesla Suit Studio[7]) aufgezeichnet und nach Unity exportiert werden. Dies ermöglichte die Integration natürlicherer und flüssigerer Bewegungsabläufe in das Handstand-Szenario.

Die Animationen der Hilfestellungen wurden durch den Einsatz von Inverse Kinematik (IK) optimiert. Die IK berechnet automatisch die notwendigen Gelenkwinkel des vSuS-Modells, sodass Hände und Füße exakt vorgegebene Positionen erreichen – beispielsweise zum präzisen Greifen der Beine der Handstandperson bei der korrekten Ausführung der Hilfestellung. Dadurch lassen sich Echtzeit-Interaktionen zwischen den automatisierten vSuS-Modellen physikalisch plausibel und realistisch umsetzen.

Steuerung Eine besondere Zielsetzung liegt aus Sicht der Informatik dabei darin, herauszufinden, ob und wie sich die Szenarien in der VR-Sporthalle über ein externes Coachsystem so steuern lassen, dass eine gleichermaßen lernförderliche wie auch als authentisch empfundene Lernumgebung für (angehende) Sportlehrkräfte entsteht. Diese Anforderungen und die benannte Zielsetzung bilden zugleich den Ausgangspunkt für die hier dargestellte Weiterentwicklung des Systems.

vSuS-Modelle Die vSuS wurden in enger Abstimmung mit sportdidaktischen und bildungswissenschaftlichen Expert:innen von einem Dienstleister erstellt. Sie repräsentieren Jugendliche im Alter von etwa 15 bis 16 Jahren (Sekundarstufe I) und sind auf die Bedingungen des schulischen Sportunterrichts abgestimmt. Die Kleidung orientiert sich an der aktuellen Unterrichtsrealität und sicherheitsrelevanten Anforderungen und variiert von körpernaher Sportkleidung (z. B. knappe Oberteile und Hot Pants bei weiblichen bzw. Muskelshirts und Sportshorts bei männlichen Modellen) bis hin zu weiter Kleidung wie Hoodies und Trainingshosen. Besonderes Augenmerk liegt auf Diversität: Die Modelle bilden Heterogenität in Geschlecht, Herkunft, Körpergröße, -form und Gewicht möglichst realitätsnah ab. Technisch verfügen sie maximal über 25.000 Vertices und 120.000 Indizes.

4 Ausblick

Zukünftig soll die entwickelte Lehr-Lernkonzeption weiter im Bereich der Lehrkräftebildung und hier insbesondere in der Lehrkräftefortbildung implementiert werden; dies gilt durch die Bereitstellung der entsprechenden Materialien und Schulungsunterlagen sowohl für das Land Schleswig-Holstein als auch bundesweit. Zudem ist geplant, das Repertoire der Unterrichtsszenarien in der VR-Sporthalle und damit auch ihr didaktisches Potenzial stetig zu erweitern sowie durch eine kontinuierliche Weiterentwicklung der Bewegungsanimationen die wahrgenommene Authentizität und damit das Immersionsgefühl auf der Benutzerseite weiter zu steigern.

[7] Tesla-Suit Studio: https://teslasuit.io/software/studio/

Literaturverzeichnis

[FL20] Fischer, B.; Leineweber, H.: Förderung der Professionellen Unterrichtswahrnehmung als Anforderung an die universitäre Sportlehrkräftebildung. Lehren und Lernen mit und in digitalen Medien im Sport: Grundlagen, Konzepte und Praxisbeispiele zur Sportlehrerbildung, S. 91–108, 2020.

[Hu21] Huang, Y. et al.: Classroom complexity affects student teachers' behavior in a VR classroom. Computers & Education 163, S. 104100, 2021.

[Hu23] Huang, Y. et al.: Comparing video and virtual reality as tools for fostering interest and self-efficacy in classroom management: Results of a pre-registered experiment. British Journal of Educational Technology 54 (2), S. 467–488, 2023.

[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, https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2021/2021_12_09-Lehren-und-Lernen-Digi.pdf.

[MK04] Miethling, W.; Krieger, C.: Schüler im Sportunterricht: Die Rekonstruktion relevanter Themen und Situationen des Sportunterrichts aus Schülersicht (RETHESIS). Hofmann, 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.

[Se22] Seufert, C. et al.: Classroom management competency enhancement for student teachers using a fully immersive virtual classroom. Computers & Education 179, S. 104410, 2022.

[Sh17] Sherin, M. G.: Exploring the boundaries of teacher noticing: Commentary. Teacher noticing: Bridging and broadening perspectives, contexts, and frameworks, S. 401–408, 2017.

[Wi19] Wiepke, A. et al.: Einsatz von Virtual Reality zum Aufbau von Klassenmanagement-Kompetenzen im Lehramtsstudium, DELFI 2019, Bonn, 2019.

[WZ24] Wilzeck, R.; Zender, R.: Virtual Reality zur Professionalisierung von Sportlehrkräften, Proceedings of DELFI 2024, 2024.

Entwicklung einer sicheren Plattform zur explorativen Nutzung sozialer Medien im Unterricht

Thomas Schmalfeldt[1], Tobias Berner[1] und Claudine Boyer[1]

Abstract: Für die aktive Arbeit mit und zu sozialen Medien wurde das sichere, datenschutzkonforme soziale Netzwerk *opostum* entwickelt. Dieses hat die gewohnten Funktionalitäten, wie sie beispielsweise aus Instagram bekannt sind, und kann auch so genutzt werden. Zusätzlich können aber auch medienpädagogische Themen wie Fake News, Influencer:innen-Marketing oder personalisierte Werbung in vorbereiteten Einheiten direkt in und mit *opostum* behandelt werden. Die Nutzendenverwaltung erfolgt beispielsweise durch die Lehrkraft, sodass klassen- oder schulinterne Netzwerke erstellt werden können. Ein übersichtliches Dashboard unterstützt Lehrkräfte und Leitende beim Einsatz von *opostum* im Unterricht oder in der Jugendarbeit.

Keywords: Soziale Medien, Social Media, explorativer Zugang, Dashboard

1 Einleitung

Soziale Medien sind ein wesentlicher Aspekt des Alltags der Jugendlichen [Kü24; Ke24]. Zum Erlernen eines adäquaten Umgangs mit ihnen können Schule und Jugendarbeit einen wichtigen Beitrag leisten [St20]. Die Lehrkräfte und Betreuenden sind dabei mit dem Problem konfrontiert, dass für die Gestaltung von Lernszenarien nicht auf kommerzielle Plattformen wie Instagram, TikTok oder Ähnliches zurückgegriffen werden kann, da dort der Datenschutz für den Einsatz mit Jugendlichen nicht ausreichend gewährleistet ist. Dies erschwert oder verhindert den explorativen Zugang zu einem bedeutsamen Thema, in dem die Interaktion untereinander und die eigenen Erfahrungen substanziell sind. Wir befinden uns in einer postdigitalen Zeit, in der nicht mehr zwischen online und offline unterschieden werden kann, sondern diese beiden Zustände ineinanderfließen [Br24; Ka22]. Insofern überschneiden sich auch Phänomene der alltäglichen Interaktion mit denjenigen in den sozialen Medien. So kann beispielsweise Bullying/Mobbing den Ursprung im digitalen Raum haben, sich danach aber im physischen Klassenverband manifestieren [WW25]. Auch weitere Phänomene aus dem digitalen Raum beeinflussen den Alltag der Jugendlichen. So tragen Influcencer:innen einen großen Teil zur individuellen Meinungsbildung bei potenziellen Kaufentscheidungen bei [KR21; BB25]. Dabei ist nicht immer offensichtlich, ob es sich um eine ehrliche Beurteilung eines Produkts oder um eine durch Sponsoring herbeigeführte Aussage handelt. Eine andere Art der Manipulation sind Fake News [Al21]. Künstlich erstellte Bilder und Videos erschweren zunehmend die Einschätzung, ob eine Information wahrheitskonform ist. Im Spannungsfeld von echt und

[1] Pädagogische Hochschule Zürich, Lagerstrasse 2, 8090 Zürich, thomas.schmalfeldt@phzh.ch

doi:10.18420/delfi2025_56

falsch bewegt sich auch das vermittelte Ideal im Digitalen, das zu einem gestörten Selbstbild führen kann [AK25].

2 opostum – ein sicheres, datenschutzkonformes soziales Netzwerk

Für die aktiv-explorative Behandlung des Themenkomplexes soziale Medien wurde ab Januar 2024 an der Pädagogischen Hochschule Zürich *opostum* (www.opostum.ch) entwickelt. Auf dieser geschützten Plattform können die Schüler:innen in einem sicheren sozialen Umfeld selbst Erfahrungen in und zu Social Media sammeln. Die Lehrkräfte und Betreuenden haben dabei die vollständige Kontrolle über die Daten, die durch die Anwendung entstehen. Die Lehrkraft kann als Administrator:in bestimmen, wer in das Netzwerk aufgenommen wird. Es sind zudem größere Netzwerke möglich (ganzes Schulhaus). Durch die flexible Nutzendenverwaltung ist ebenso ein Einsatz in der außerschulischen Jugendarbeit möglich. Für die Lehrkraft werden in *opostum* verschiedene Unterstützungsmittel zur Verfügung gestellt, mit der sie einerseits die Möglichkeiten der Interaktionen auf der Plattform steuern, aber auch unerwünschtes Verhalten erkennen und darauf adäquat reagieren kann. Die technische Grundlage von *opostum* bildet die das Opensourceprojekt *Instahub* [Do19]. Während das Projektteam ein datenschutzkonformes Hosting von *opostum* anbietet, ist es auch möglich, die Plattform auf einem eigenen Server bereitzustellen.

Auf *opostum* sind die meisten Möglichkeiten vorhanden, die aus den sozialen Medien bekannt sind. Neben dem Posten von Fotos und Videos können Hashtags genutzt und Kommentare erstellt werden. Auf Letztere kann geantwortet werden und es können Likes gesetzt werden. Einzelnen Nutzenden kann gefolgt werden, sodass deren Posts auf der eigenen Startseite erscheinen. Zusätzlich schlägt ein Algorithmus (der im Unterricht behandelt werden kann) neue Nutzende vor, denen gefolgt werden könnte. Ebenso können private Nachrichten verschickt und Umfragen erstellt werden. Bei der Entwicklung von *opostum* wurde darauf geachtet, nahe an das den Schüler:innen bekannte Look-and-feel sozialer Medien zu gelangen.

2.1 Wie werden die Lehrkräfte auf opostum unterstützt?

Um den Betreuenden den Überblick über das Geschehen zu vereinfachen, steht ein Lehrkräfte-Dashboard bereit. Dieses wurde in Zusammenarbeit mit Lehrkräften entwickelt, um den Bedürfnissen für einen Einsatz im Unterricht bestmöglich gerecht zu werden.

Ein Hub ist ein in sich geschlossenes soziales Netzwerk auf *opostum*. Administrierende können beliebig viele davon erstellen und auch wieder löschen. Dadurch kann ein Hub für einen begrenzten Zeitraum eröffnet werden, beispielsweise für ein Klassenlager oder während einer Unterrichtseinheit zum Thema Fake News. Innerhalb eines Hubs kann definiert werden, welche Arten der Interaktion (Features) möglich sind (siehe Abb. 1). Ein Feature kann jederzeit wieder deaktiviert werden. Möchte eine Lehrkraft das Thema

„Kommentare“ zuerst im Unterricht erarbeiten, kann das Schreiben von solchen bis zu diesem Zeitpunkt deaktiviert bleiben. Dies würde beispielsweise bedeuten, dass keine Kommentare geschrieben werden können und die bereits bestehenden nicht mehr angezeigt werden. Sobald die Kommentarfunktion wieder aktiviert wird, erscheinen die bereits verfassten Kommentare wieder. Analog verhält es sich bei den anderen Interaktionsmöglichkeiten.

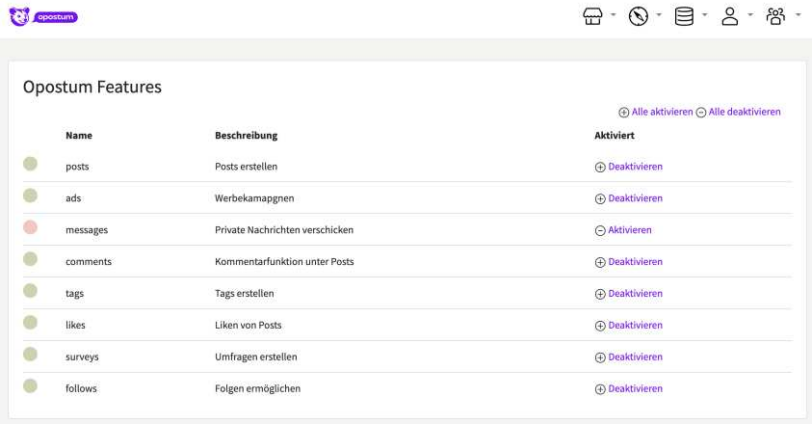


Abb. 1: Dashboard zur Administration der Interaktionsmöglichkeiten (Features)

Neben den beschriebenen Einstellungen, die den gesamten Hub betreffen, kann die Lehrkraft auch gezielt eingreifen. So wird den Administrierenden bei jedem Post und Kommentar die Möglichkeit angezeigt, diesen zu löschen oder zu verstecken. Bei Letzterem ist der Post oder Kommentar für die Nutzenden nicht mehr sichtbar und wird der administrierenden Person mit einem roten Rand angezeigt (siehe Abb. 2). Dies ermöglicht es der Lehrkraft, einen Beitrag entweder individuell oder im Klassenverband zu thematisieren, was bei einem gelöschten Post nicht mehr möglich wäre.

2.2 Einsatzszenarien

Die Plattform *opostum* wurde so konzipiert, dass sie vielfältig genutzt werden kann. Im dazugehörigen Wiki (wiki.opostum.ch) sind zurzeit vier Szenarien, die im öffentlichen Diskurs zu Social Media verstärkt wahrnehmbar sind (Fake News, Influencer:innen-Marketing, Selbstdarstellung & Privatsphäre, personalisierte Werbung), für den Unterricht ausgearbeitet. Zusätzlich stehen je ein Szenario für die außerschulische Arbeit (Planspiel in der Ausbildung von Leitungspersonen) und für den Einsatz in einer Aktivitätswoche bereit. Es werden jeweils eine Handreichung für Lehrkräfte und, falls benötigt, Arbeitsmaterialien und Importdateien für künstliche Profile bereitgestellt. Diese Profile erlauben es, einen Hub bereits mit Inhalten zu „befüllen“. Der Post in Abb. 2 ist ein Beitrag der künstlichen Profile ChicChristina und Tasty_Yves aus der Unterrichtseinheit zum Influencer:innen-Marketing. Mit den künstlichen Profilen kann die Grundlage geschaffen

werden, um in ein Thema einzusteigen. Anschließend können die Jugendlichen auf diese Posts reagieren und eigene erstellen.

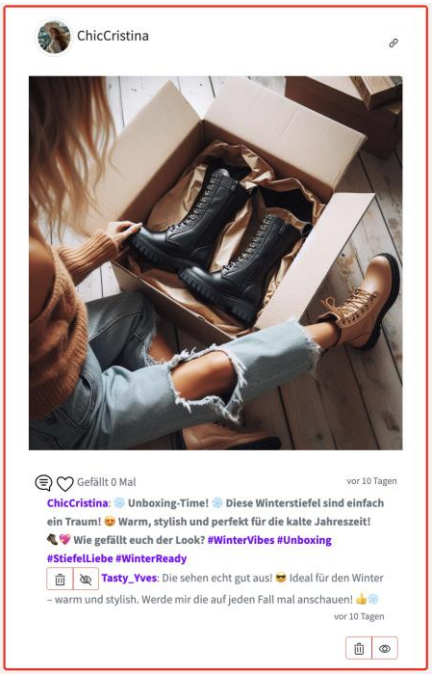


Abb. 2: Post auf *opostum* mit Hashtags, Kommentaren und Likes. Der rote Rahmen indiziert, dass der Beitrag durch die Lehrkraft für die Nutzendenden als unsichtbar markiert wurde.

Zum Thematisieren von personalisierter Werbung können eigene Reklamekampagnen erstellt werden. Neben der Werbung, die entweder als Banner oder als Post angezeigt wird, kann auch definiert werden, wem genau sie angezeigt werden soll. So kann beispielsweise das Alter der Personen eingeschränkt werden. Oder die Werbung erscheint nur im Kontext von Posts, in denen bestimmte Worte vorkommen. Aktuell können acht Parameter eingestellt werden, anhand derer die Kampagne personalisiert werden kann.

Danksagung

Die Entwicklung von *opostum* wurde ermöglicht durch die großzügige Unterstützung der Manawa Foundation. Ein herzlicher Dank gilt auch den weiteren Projektmitgliedern Eike Rösch, Mareike Düssel, Melanie Murer-Jäckle, Moreno Kunz, Philipp Rüegger, Sarah Hess und dem Entwickler von Instahub, Julian Dorn.

Literaturverzeichnis

[Al21] Al-Rawi, A.: Political memes and fake news discourses on Instagram. Media and communication, S. 276-290, 2021.

[AK25] Auersper, F. ; Klikovitz, J.: Selbstkonstruktion im digitalen Raum – Social Media als Identitätsbühne. 24, Zeitschrift für Psychodrama und Soziometrie, S. 409-420, 2025.

[BB25] Borchers, N.S.; Beckert, J. : Social-Media-Influencer. In (Schweiger, W.; Beck, K.; Karnowski, V., Hrsg.) Handbuch Online-Kommunikation. Springer VS, Wiesbaden 2025.

[Br24] Breil, J.: Lernen als postdigitale Erfahrung. In (Pieper, M.; Neuhaus, T., Hrsg.): Bildung und Digitalität. ars digitalis, Springer Vieweg, Wiesbaden, 2024.

[Do19] Dorn, J.: InstaHub: Datenbanken und Datenschutz mit einem extra für den Unterricht entwickelten sozialen Netzwerk unterrichten. In: Informatik für alle, 2019.

[Ka22] Kaptan, D. et al.: Ist TikTok toxisch? – die Sicht von Jugendlichen auf postdigitale ästhetisch-kulturelle Praktiken in sozialen Medien. Forum Kinder- und Jugendsport 1/3, S. 13-24, 2022.

[KR21] Kühn, J.; Riesmeyer, C.: Brand endorsers with role model function. MedienPädagogik, S. 67-96, 2021.

[Kü24] Külling-Knecht, C. et al: JAMES – Jugend, Aktivitäten, Medien – Erhebung Schweiz, 2024.

[Me24] Medienpädagogischer Forschungsverbund Südwest: JIM 2024 - Jugend, Information, Medien, 2024.

[St20] Stix, D. C.: Social Media in der Offenen Kinder- und Jugendarbeit. MedienPädagogik, S. 1-48, 2020.

[WW25] Wachs, S.; Wright, M.F.: Predictors and consequences of teachers' willingness to intervene in traditional and cyber bullying: A short-term longitudinal study. International journal of bullying prevention, 2025.

PyRope – Ein codebasierter Ansatz für E-Assessment in MINT-Fächern

Konrad Schöbel[1], Jochen Merker[2], Paul Brassel[3], Moritz Zöllner[1] und Heike Hain[2]

Abstract: Es wird das von Grund auf neu entwickelte, quelloffene E-Assessment-System *PyRope* für den MINT-Bereich vorgestellt. Anders als bei herkömmlichen Systemen, bei denen Aufgaben in einer Weboberfläche erstellt werden, beruht dessen Ansatz darauf, Aufgaben ausschließlich durch Quellcode zu beschreiben. Dies bringt unter anderem eine Reihe an inhärenten Vorteilen mit sich, die zwar in der Welt der Softwareentwicklung, bisher jedoch nicht bei E-Assessment-Systemen selbstverständlich sind. Der Prototyp ist als Python-Bibliothek mit dem Ziel konzipiert, den Zeitaufwand zur Erstellung von Übungsaufgaben so weit wie möglich von technischen Aspekten auf das eigentliche didaktische Konzept zu verlagern. Jede Übungsaufgabe wird durch eine eigene Klasse realisiert, deren Methoden die einzelnen Komponenten der Aufgabe umsetzen und deren Klassenattribute die Metadaten beinhalten. Im Rahmen der Demo wird der Prototyp mit wenigen Befehlen aufgesetzt sowie das Erstellen, Testen und Ausführen einfacher Aufgaben vorgeführt, um exemplarisch einige Innovationen gegenüber herkömmlichen Systemen herauszuarbeiten.

Keywords: E-Assessment, E-Learning, Open Educational Resources, MINT, Python

1 Einordnung

In der Grundlagenausbildung im MINT-Bereich finden nicht erst seit der Corona-Pandemie zahlreiche E-Assessment-Systeme einen breiten Einsatz, deren Möglichkeiten weit über simple Aufgabentypen wie Lückentexte oder Multiple-Choice-Aufgaben hinausgehen, wie MathCoach [Gr23], Möbius[4], ONYX[5], STACK[6], WebWork[7] oder WIMS[8], um nur einige zu nennen. Eine Aufgabe besteht dabei aus mehreren Komponenten: dem Aufgabentemplate, der Definition randomisierter Variablen, einer Musterlösung, der automatisierten Auswertung sowie einem Feedback.

Für Randomisierung und Auswertung wird üblicherweise auf Computeralgebra wie Maxima, Maple oder MuPad zurückgegriffen. Im MINT-Bereich typische Aufgaben erfordern

[1] HTWK Leipzig, Fakultät Digitale Transformation, PF 30 11 66, 04251 Leipzig, Deutschland, konrad.schoebel@htwk-leipzig.de; moritz.zoellner@htwk-leipzig.de
[2] HTWK Leipzig, Mathematisch-Naturwissenschaftliches Zentrum, PF 30 11 66, 04251 Leipzig, Deutschland, jochen.merker@htwk-leipzig.de, https://orcid.org/0000-0003-2257-7030; heike.hain@htwk-leipzig.de
[3] Universität Leipzig, Medizininformatikzentrum, Abteilung Medical Data Science, Stephanstraße 9c, 04103 Leipzig, Deutschland, paul.brassel@medizin.uni-leipzig.de
[4] https://www.digitaled.com/mobius/
[5] https://www.bps-system.de/onyx-pruefungsplattform/
[6] https://stack-assessment.org/
[7] https://openwebwork.org/
[8] https://wimsedu.info/wims/

doi:10.18420/delfi2025_57

deshalb ein Mindestmaß an Programmierkenntnissen in der jeweiligen Programmiersprache. Erfahrungsgemäß nimmt das Programmieren im Vergleich zur didaktischen Konzeption den weitaus größten Anteil am Zeitaufwand beim Erstellen einer Aufgabe ein. Nicht selten werden Aufgabenstellerinnen durch die Einschränkungen des E-Assessment-Systems oder dessen hohen Editieraufwand gezwungen, das didaktische Konzept einer Aufgabe abzuwandeln oder gar aufzugeben. Beispielsweise ist es nicht möglich, vorausberechnete Daten einzulesen.

Gemeinsam ist allen gängigen Systemen, dass sich Aufgaben ausschließlich über eine Weboberfläche erstellen lassen. Angesichts der Tatsache, dass die Definition einer Übungsaufgabe im Wesentlichen aus Quellcode besteht, mutet dies aus der Zeit gefallen an. Meist fehlt es an elementarem Eingabekomfort, der heutzutage aus modernen Programmierumgebungen nicht mehr wegzudenken ist, wie Syntax-Highlighting, Code-Vervollständigung, statische Code-Analyse oder Versionskontrolle.

Zudem werden selbst bei kommerziellen Systemen oft grundlegende Prinzipien der Softwareentwicklung verletzt. So bleiben zum Beispiel Sicherheitsaspekte auf der Strecke, wenn Platzhalter im Auswertungscode vor der Ausführung ungefiltert durch Nutzereingaben ersetzt werden.

2 Ansatz

An besagten Unzulänglichkeiten setzt der von Grund auf neu entwickelte Prototyp namens *PyRope*[9] [Me23] an, der auf dem Hauptgrundsatz beruht, dass Aufgaben vollständig durch Code repräsentiert werden. Zielgruppe sind programmieraffine Dozierende sowie Forschende im Bereich Didaktik, die forschungsspezifische Anforderungen mit PyRope leichter umsetzen können [HMS24]. Als Programmiersprache wurde auf Python gesetzt, da sie eine geringe Einstiegshürde besitzt und trotzdem überaus mächtig ist. Außerdem ist sie frei verfügbar, weit verbreitet und hat eine große, aktive Community.

PyRope ist als Python-Bibliothek mit dem Ziel konzipiert, den Zeitaufwand zur Erstellung von Übungsaufgaben so weit wie möglich von technischen Aspekten auf das eigentliche didaktische Konzept zu verlagern. Jede Übungsaufgabe wird durch eine eigene Klasse realisiert, deren Methoden die einzelnen Komponenten der Aufgabe umsetzen und deren Klassenattribute die Metadaten beinhalten. Allerdings dienen Klassen lediglich als Container. Kenntnisse in objektorientierter Programmierung sind für gewöhnliche Aufgaben nicht vonnöten. Der Boilerplate-Code ist minimal gehalten und beschränkt sich auf die Kopfzeilen von Klassen und Methoden. Für detailliertere Informationen zum Aufbau und der Funktionsweise der Bibliothek verweisen wir auf die entsprechende Dokumentation[10].

[9] https://github.com/PyRope-E-Assessment/pyrope
[10] https://pyrope-edu.readthedocs.io/

PyRope wird im Rahmen des von der *Stiftung Innovation in der Hochschullehre*[11] geförderten Projektes *Feedback-Basiertes E-Assessment in Mathematik, Informatik und Ingenieurwissenschaften*[12] entwickelt und soll demnächst ausführlicher evaluiert werden.

3 Innovative Aspekte von PyRope

Unabhängig von ihrer konkreten Ausgestaltung bringt die Software-Bibliothek durch ihren rein codebasierten Ansatz bereits eine ganze Reihe von Vorteilen gegenüber den etablierten E-Assessment-Systemen mit sich – und das „by Design“, also ohne jeglichen zusätzlichen Entwicklungsaufwand.

Entwicklungsumgebung. Zum Editieren von Aufgaben kann jede beliebige Python-Entwicklungsumgebung verwendet werden. Insbesondere stehen damit die üblichen komfortablen Programmierhilfen wie Syntax-Highlighting, Code-Vervollständigung oder statische Code-Analyse zur Verfügung.

Versionskontrolle. Als reiner Quellcode lassen sich Aufgaben über ein Versionskontrollsystem wie z. B. Git verwalten. Das vereinfacht nicht nur das Weiterentwickeln und Teilen von Aufgabenpools, sondern ermöglicht zudem Rechtemanagement, kontinuierliche Integration, Bug-Tracking, Dokumentation oder ein Wiki.

Open Educational Resources. Werden Aufgabenbibliotheken über ein Versionskontrollsystem öffentlich unter einer freien Lizenz zur Verfügung gestellt, handelt es sich per se um sogenannte Open Educational Resources, denn alle für das Verwenden der Aufgaben notwendigen Softwarekomponenten und Ressourcen sind dann frei verfüg- und verwendbar. Dies steigert auch die Nachhaltigkeit beim Aufbau, der Nutzung und Wartung von Aufgabenbibliotheken und führt zur Herausbildung von Open Educational Practices.

Durchsuchbarkeit. Als Textdateien lassen sich ganze Aufgabensammlungen mit Kommandozeilentools ohne großen Aufwand z. B. auf Stichworte hin durchsuchen.

Integration von Daten. Teil des Quellcodes der Aufgaben können gleichermaßen Dateien mit Daten in anderen Formaten sein, die vom Python-Code eingelesen und verarbeitet werden. Prinzipiell kann jede externe Software in eine Aufgabe eingebunden werden, sofern sie eine Python-Schnittstelle zur Verfügung stellt.

Darüber hinaus besitzt der hier vorgestellte Prototyp unter anderem folgende Charakteristika, die ihn von gängigen E-Assessment-Systemen abheben.

Extrem kurzer Workflow. Der Prototyp unterstützt Jupyter-Notebooks als Frontend. Eine innerhalb einer Notebookzelle erstellte Aufgabe kann in der Folgezelle ausgeführt oder per Unittest getestet werden. Editieren, Unittest und Ausprobieren liegen somit nur einen Tastendruck auseinander. Kürzer kann der Editierzyklus nicht sein.

[11] https://stiftung-hochschullehre.de/
[12] http://www.fassmii.de

Immenser Funktionsumfang. Python bietet dedizierte Bibliotheken für spezielle Anwendungsgebiete, zum Beispiel zur Darstellung oder Auswertung von Formeln, elektrischen Schaltkreisen, chemischen Formeln, Graphen, natürlicher Sprache, u.v.m. Damit lässt sich ein breit gefächertes Spektrum an Aufgaben für eine Vielfalt an Fachgebieten abdecken.

Wiederverwendbarer Code. Die Realisierung von Übungsaufgaben als Klassen erlaubt ein einfaches, transparentes Modifizieren oder Ersetzen einzelner Komponenten einer Aufgabe mittels Vererbung.

Unittesting. Aufgaben lassen sich durch umfangreiche Unittests auf Fehler und Inkonsistenzen prüfen, zum Beispiel ob Grenzfälle in den Nutzereingaben korrekt behandelt werden.

Ausgefeiltes Autograding. Eingabefelder mit Musterlösung werden automatisch bewertet, was repetitiven Auswertungscode vermeidet. Bewertungsfunktionen sind konfigurierbar und besitzen sinnvolle Voreinstellungen.

Filtern. Die Realisierung von Metadaten als Klassenattribute ermöglicht ein einfaches Filtern von Übungsaufgaben nach bestimmten Kriterien.

Lokaler und autonomer Betrieb. Das E-Assessment kann lokal betrieben werden, unabhängig von einem Lernmanagementsystem (LMS), einer Datenbank oder einem Server. Dies ist vor allem für Studierende attraktiv, die fürchten, dass ihre Ergebnisse online getrackt werden.

Anbindbarkeit an LMS. Trotzdem ist das E-Assessment-System an Lernmanagementsysteme anbindbar, die den LTI-Standard 1.3 umsetzen.

Flexibles Einbinden in verschiedene Lehr- und Lernkontexte. Das E-Assessment lässt sich nicht nur als Webservice betreiben. Übungsaufgaben können alternativ in ein als interaktives Vorlesungsskript gestaltetes Jupyter-Notebook integriert werden.

Literaturverzeichnis

[Gr23] Grabowski, B. et al.: MathCoach und LaplaceScript: Ein programmierbarer interaktiver Mathematiktutor mit XML-basierter Skriptsprache. In: Marktplatz Internet: Von e-Learning bis e-Payment, 13. Leipziger Informatik-Tage (LIT 2005). Gesellschaft für Informatik e. V., Bonn, S. 211–218, 2023.

[HMS24] Hain, H.; Merker, J.; Schöbel, K.: E-Assessment-Aufgaben mit PyRope und ONYX erstellen – ein Use Case. In (Thor, A. et al., Hrsg.): 21. Workshop on E-Learning WEL 23. S. 12–22, 2024, DOI: 10.33968/2024.71.

[Me23] Merker, J. et al.: E-Assessment in MINT-Fächern: Coden von Übungsaufgaben mit Python & Jupyter. In (Liebscher, E. et al., Hrsg.): Digitale Lehre im Rahmen der Grundausbildung in MINT-Fächern an Hochschulen : Didaktische Integration von digitalen Medien und E-Learningsystemen in Lehrveranstaltungen : Tagungsband 2022. S. 96–109, 2023, DOI: 10.25673/103431.2.

Advancing Study Planning with Learning Analytics

Niels Seidel[1], Pio Virzi[1], and Marie-Theres Niazi-Shahabi[1]

Abstract: Students must create study plans that balance individual academic goals with institutional requirements, including examination regulations, prerequisites, and module constraints. This paper presents a prototype study planner incorporating examination regulations through a Constraint Modelling Language (CML) derived from eight examination regulations. Additionally, it offers a Learning Analytics Dashboard to visualize study progress, peer comparisons, and module recommendations. The CML was validated through expert evaluation, and the prototype was assessed using think-aloud usability testing.

Keywords: Study Planning, Learning Analytics, Constraint Modelling, Recommender Systems

1 Introduction and Related Works

Planning a university degree presents significant challenges for students. The range of available modules and possible module combinations is vast, examination regulations are complex, studies extend over multiple years, and individual goals evolve throughout the degree as students discover their interests, assess their performance, and observe peers' preferences. Therefore, students repeatedly ask for a tool to support the creation of study plans that balance individual academic goals with institutional requirements, including examination regulations, prerequisites, and module constraints. The design and development of such a tool is based on preliminary work ranging from academic progress analysis, study planning strategies, and course recommendation systems to learning analytics and existing study planning tools.

Academic progress data provides insights into students' planning behavior and performance, informing future decisions. [Na14] analyzed 12,836 engineering students, finding that poor first-year performance increases non-completion risks and that students investing more time per ECTS achieved better outcomes. [Se22] examined 1,898 mathematics and computer science graduates, showing that concurrent enrollment reduces study duration while taking modules from other faculties and skipping semesters negatively impact completion. These historical patterns can inform study planning strategies and tools.

[1] CATALPA, FernUniversität in Hagen, Universitätsstraße 27, 58097 Hagen, Germany.
This research was supported by the Center for Center of Advanced Technology for Assisted Learning and Predictive Analytics (CATALPA) of the FernUniversität in Hagen, Germany.,
niels.seidel@fernuni-hagen.de, https://orcid.org/0000-0003-1209-5038;
pio.virzi@studium.fernuni-hagen.de; matheres.niazi@gmail.com

doi:10.18420/delfi2025_58

[Mo15]'s IDiSC+ extends course planning by incorporating institutional constraints and student preferences to generate study plans through iterative refinement. It has demonstrated effectiveness for students facing expulsion risks due to poor planning. [RT20] developed an algorithm prioritizing course enrollments by calculating "slack" (temporal buffer) based on prerequisites and requirements. Testing with 102 participants showed 82.1 % of newcomers felt less anxious, 98 % rated plan usefulness 7+ on a 10-point scale, and all participants rated preference alignment 7+.

[Be09] developed CourseRank, a course recommendation system for Stanford students, integrating social features and course ratings to support study planning. [PVG11] extended this by exploring recommendations under complex constraints, finding that course popularity often aligned with historical enrollment patterns. [Yu21] showed that explanations in recommendations, especially those based on thematic overlap, increased users' serendipitous course discoveries.

[Ju23] tested a study planner with 12 computer science students using think-aloud methodology, identifying requirements including university account integration, authentication, multiple plan drafts, and peer plan alignment. [Ba24] conducted similar studies with 20 students, recommending that user interfaces follow university design while remaining intuitive, include course filtering, and provide support functions to prevent planning errors.

Self-regulated learning (SRL) enables students to manage their studies through goal setting, planning, monitoring, and reflection [Zi15]. Although key SRL models offer valuable frameworks [PKJ15], they are rarely applied to study planning at a macro level. Recent Learning Analytics research uses SRL as a theoretical foundation, with tools like dashboards [Sc16] and recommender systems supporting progress tracking and peer comparison.

The reviewed related works on study planners provide requirements and functionalities tested in user studies. Since these primarily address usability and design aspects, insights from academic progress data, planning strategies, and course recommendation systems must also be considered to ensure comprehensive study planning support. Additionally, Self-regulation applied technically in Learning Analytics applications could enable students to reflect on their academic planning and history while identifying actionable next steps.

2 StudyPlanner

StudyPlanner is a web-based application streamlining academic planning through a visual interface. Built with Vue.js and the headless CMS Strapi, it supports valid academic progression planning by loading enrollment and exam data from external sources like university data warehouses. The software architecture is summarized in Figure 1. Based conceptually on [Ju23], the interface provides a user-centered design allowing students to plan and manage modules through an intuitive graphical representation of their academic

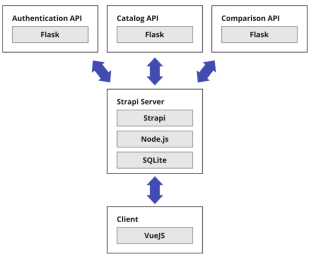


Fig. 1: Software Architekture

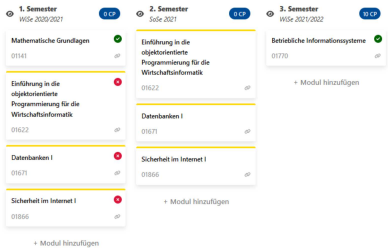


Fig. 2: User Interface of the study planner

trajectory (Figure 2). Users can add or remove modules and semesters dynamically for flexible study path adjustments. System usability was evaluated through interviews and think-aloud testing (N=6).

Constraint Modelling Language Examination regulations are legal texts that contain precise rules about which combinations of modules must be completed to finish a degree program. From these individual rules, generic rules can be identified that are found in multiple examination regulations. We analyzed all eight examination regulations of the degree programs in the Faculty of Mathematics and Computer Science at FernUniversität in Hagen. We identified eight generic rules that describe the regulations entirely. Following [Co23], we formulated these rules as constraints and transformed them into a JSON-based constraint modeling language (CML). Within the StudyPlanner, this CML and the required module catalogs can be configured in the Strapi web frontend, allowing students to validate their study planning according to their applicable study regulations. The constraint system includes eight generic rules that describe examination regulations. These rules specify minimum (moduleMin) and maximum (moduleMax) numbers of modules from catalogs, define mandatory modules (mandatoryModules) for degree programs, and group catalogs under superordinate structures (moduleGroupDefinition). The binding module rule (bindingModule) prevents students from changing elective compulsory modules after exam participation, while credit requirements (minCreditsFromModuleGroup) ensure sufficient credits from specific catalogs. Alternative scenarios are handled through logical OR operations (or), and prerequisite dependencies (requiredModules) define which modules or catalogs must be completed before enrollment in others. An expert responsible for ECTS monitoring at FernUniversität in Hagen has successfully evaluated the constraints.

Learning Analytics Dashboard The dashboard (Fig. 3) features six widgets that provide students with comprehensive study progress and planning information. The first row contains four progress metric widgets that track individual study advancement. Progress bars offer visual support and tooltips displaying percentage completion for metrics with calculable



Fig. 3: Learning Analytics Dashboard in the StudyPlanner



Fig. 4: Indicator for success rate comparison. (from left: failed module, passed module, passed module with insufficient data for comparison)

maximum values. The *Completed Modules* widget sums all modules marked as "passed" in the study plan editor and compares them to the total planned modules. The *Collected ECTS* widget calculates ECTS points from passed modules against the total credits required for graduation. The *Current Semester* widget determines the current semester number based on enrollment data, or defaults to the total number of defined semesters if only historical data exists. The *Open Modules* widget counts modules that are neither passed nor failed, providing a counterpart to the completed modules display. The second row contains two widgets focused on analyzing the study progress. The analytics widget features two bar charts: *Modules per Semester* compares enrolled versus completed modules by semester while calculating averages for both categories, and *ECTS per Semester* presents planned versus achieved ECTS credits by semester with corresponding average calculations. The study planning widget displays the current semester's planning. It offers direct navigation to the study planner, automatically positioning the scroll bar at the current semester location to ensure students see their academic status immediately. These widgets utilize visual elements such as progress bars and tooltips to enhance usability. They provide students with detailed metrics and quick overviews of their academic progress and planning status. Each module in the study progress display includes a ring visualization (Fig. 4) showing historical success rates of module exams for students in the same nth semester. The ring contains colored segments for students who passed (green), failed (red), or didn't register for the exam (gray). A center symbol indicates the student's group, with the relevant segment scaled for emphasis. Gray rings indicate insufficient data for comparison.

Module Recommendations Module recommendations are generated from aggregated datasets provided by the Comparison API. The system removes modules already marked as "passed" by the student, then sorts remaining modules by success rate in descending order. When multiple modules share the same success rate, they are further sorted by enrollment frequency. The top three modules from this list are displayed as recommendations for each future semester. Students can reject displayed recommendations, prompting the system to generate new ones. New recommendations are also calculated when semesters are added to existing study plans. Students can adopt recommendations by clicking on a module, which opens a pre-filtered module catalog. Adding new modules triggers the recalculation of recommendations.

References

[Ba24] Bartel, L. et al.: Design Principles for a Study Planning Assistant in Higher Education. In: Proceedings of the 2024 Conference on Human Information Interaction and Retrieval (CHIIR). 2024, DOI: 10.1145/3627508.3638327.

[Be09] Bercovitz, B. et al.: CourseRank: a social system for course planning. In: Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data. SIGMOD '09, Association for Computing Machinery, Providence, Rhode Island, USA, pp. 1107–1110, 2009, DOI: 10.1145/1559845.1559994.

[Co23] Correa, C. et al.: A Lightweight Method to Define Solver-Agnostic Semantics of Domain Specific Languages for Software Product Line Variability Models. In: SOFTENG 2023 - The 9th International Conference on Advances and Trends in Software Engineering. ffhal04077798f, IARIA: International Academy, Research and Industry Association, Venise, Italy, 2023.

[Ju23] Judel, S. et al.: Supporting Individualized Study Paths Using an Interactive Study Planning Tool. In: 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., Bonn, pp. 225–230, 2023, DOI: 10.18420/delfi2023-36.

[Mo15] Mohamed, A.: A decision support model for long-term course planning. Decision Support Systems 74, pp. 33–45, 2015, DOI: 10.1016/J.DSS.2015.03.002.

[Na14] Nam, S. J. et al.: Customized course advising: Investigating engineering student success with incoming profiles and patterns of concurrent course enrollment. In: ACM International Conference Proceeding Series. 2014, DOI: 10.1145/2567574.2567589.

[PKJ15] Panadero, E.; Klug, J.; Järvelä, S.: Third wave of measurement in the self-regulated learning field: when measurement and intervention come hand in hand. Scandinavian Journal of Educational Research 60 (6), pp. 723–735, 2015, DOI: 10.1080/00313831.2015.1066436.

[PVG11] Parameswaran, A.; Venetis, P.; Garcia-Molina, H.: Recommendation systems with complex constraints: A course recommendation perspective. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM '11). Pp. 425–434, 2011, DOI: 10.1145/2037661.2037665.

[RT20] Richards, A. L.; Tsay, R. S.: An Optimal Slack-Based Course Scheduling Algorithm for Personalised Study Plans. In: ACM International Conference Proceeding Series. 2020, DOI: 10.1145/3383923.3383925.

[Sc16] Schwendimann, B. et al.: Perceiving learning at a glance: A systematic literature review of learning dashboard research. IEEE Transactions on Learning Technologies PP (99), p. 1, 2016, DOI: 10.1109/TLT.2016.2599522.

[Se22] Seidel, N.: Modeling study duration considering course enrollments and student diversity. In: Proceedings of The 15th International Conference on Educational Data Mining (EDM 2022), pp. 621–628, 2022, DOI: 10.5281/zenodo.6852956.

[Yu21] Yu, R. et al.: Orienting Students to Course Recommendations Using Three Types of Explanation. In: UMAP 2021 Adjunct: Adjunct Publication of the 29th ACM Conference on User Modeling, Adaptation and Personalization. Pp. 331–336, 2021, DOI: 10.1145/3450614.3464483.

[Zi15] Zimmerman, B. J.: Self-Regulated Learning: Theories, Measures, and Outcomes. In (Wright, J. D., ed.): International Encyclopedia of the Social & Behavioral Sciences (Second Edition). Second Edi, Elsevier, Oxford, pp. 541–546, 2015, DOI: 10.1016/B978-0-08-097086-8.26060-1.

StudyOrientation: Eine mobile App zur Unterstützung der Selbstreflexion bei der Studienwahl im Fach Informatik

Lea Sigethy[1], Sarah Aragon-Hahner[2], Sophia Sakel[1] und Sven Strickroth[1]

Abstract: Die Studienwahl stellt viele Studieninteressierte vor komplexe Entscheidungen, die häufig von Unsicherheiten und unzureichender Selbstreflexion geprägt sind. Vor diesem Hintergrund wurde mit *StudyOrientation* eine mobile Anwendung entwickelt, die eine strukturierte und reflexionsfördernde Unterstützung der Studienwahl im Fach Informatik bietet. Die App nutzt das Konzept des Micro-Tasking, um Inhalte in kompakte, thematisch strukturierte Einheiten zu unterteilen. Zudem adressiert *StudyOrientation* gezielt die Förderung der individuellen als auch sozialen Selbstreflexion. Eine explorative Feldstudie mit 20 Teilnehmenden zeigt erste Hinweise auf hohe Akzeptanz, intensive Nutzung und reflektierte Auseinandersetzung mit den App-Inhalten. Dies verdeutlicht das Potenzial von *StudyOrientation* als reflexionsförderndes Unterstützungsangebot im Kontext der Studienorientierung.

Keywords: Online Self-Assessment, Selbstreflexion, Computer-Assisted Career Guidance System

1 Motivation und Hintergrund

Der Schritt in die Hochschulwelt markiert eine entscheidende Lebensphase und stellt Studieninteressierte vor zentrale Entscheidungen hinsichtlich ihrer fachlichen Orientierung [He09; Pä23]. Angesichts von über 20.000 Studiengängen an mehr als 400 Hochschulen in Deutschland erleben viele diese zukunftsbeeinflussende Wahl als überfordernd, komplex und schwer überschaubar [HS14; HWS10; Pä23]. Dabei spielt eine unzureichende Selbsteinschätzung der individuellen Interessen, Eignung und Fähigkeiten eine zentrale Rolle [He09; Sc17].

Zur Unterstützung der Studienwahl stellen daher immer mehr Hochschulen zahlreiche Informations- und Beratungsangebote zur Verfügung, die von einer Erstorientierung bis zur Vermittlung konkreter Studieninformationen reichen [Mi08]. Digitale Systeme gewinnen dabei an Bedeutung [GK21], weisen jedoch oft Defizite bei der Förderung tiefer Selbstreflexion auf [ASS23; HWJ13]. Internetbasierte Selbsttests verknüpfen Informationsvermittlung mit Entscheidungsunterstützung, um eine Einschätzung der persönlichen Eignung zu ermöglichen [He09; HS14; St19]. Die Bewertung erfolgt oft regelbasiert anhand der Nutzerantworten und führt häufig zu Handlungsempfehlungen, bleibt jedoch in der Förderung individueller Reflexion über persönliche Ziele und Werte begrenzt [ASS23; HS14; HWJ13; Lü15]. Selbstreflexion gilt jedoch als zentraler Bestandteil fundierter Bildungsentscheidungen [ASS23;

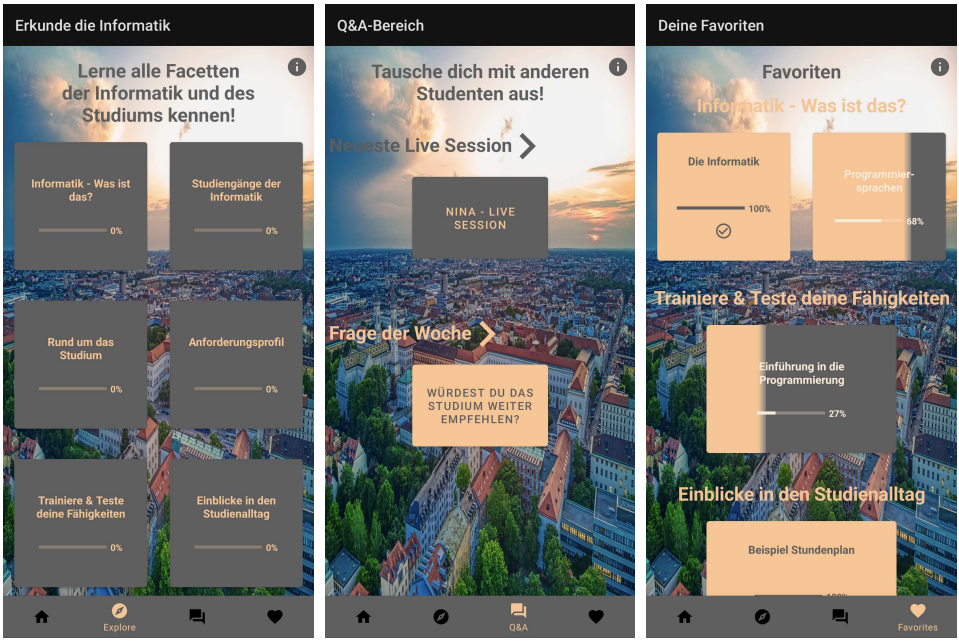
[1] Ludwig-Maximilians-Universität München, Institut für Informatik, Oettingenstraße 67, 80538 München, lea.sigethy@campus.lmu.de; sophia.sakel@ifi.lmu.de, https://orcid.org/0000-0002-6326-8018; sven.strickroth@ifi.lmu.de, https://orcid.org/0000-0002-9647-300X

[2] TU Bergakademie Freiberg, Institut für Informatik, Akademiestraße 6, 09599 Freiberg, sarah.aragon-hahner@informatik.tu-freiberg.de, https://orcid.org/0000-0001-7587-080X

doi:10.18420/delfi2025_59

GK21] – insbesondere im Kontext der Studienorientierung, die stark von individuellen Faktoren geprägt ist [Ar23; Ha79; Ra17]. Neben eigenen Überlegungen ist dabei auch der soziale Aspekt des Reflexionsprozesses bedeutsam: Konversationen mit Peers sowie Erfahrungsberichte aus erster Hand können Reflexion fördern und die Entscheidungssicherheit stärken [Go21; Ha79; MHE16]. Das Prinzip des *Micro-Tasking* bietet hohes Potenzial für mobile Anwendungen zur Unterstützung der Selbstreflexion bei der Studienwahl [ASS23; Iq18]. Dabei werden große Informationsmengen in kleinere Einheiten aufgeteilt und Stück für Stück bereitgestellt. Durch eine flexible, wiederholte Auseinandersetzung mit Inhalten wird so Raum für eine situative Selbstreflexion geschaffen [Ar22; Ar23; Ch15; Tr03].

Trotz einzelner Ansätze [Ar23; Gr23; LES23] fehlt bislang ein mobiles System, das Studienorientierung, Selbstreflexion und soziale Einbindung ganzheitlich integriert. *StudyOrientation* adressiert diese Forschungslücke durch interaktive Inhalte, die mittels Micro-Tasking vermittelt werden. Unser Konzept kann als Ausgangspunkt für die Weiterentwicklung innovativer, digitaler Unterstützungsangebote im Hochschulkontext dienen.



(a) Explore-Bereich (b) Q&A-Bereich (c) Favoriten-Bereich

Abb. 1: Aufbau der App: (a) Explore-Bereich mit Themen wie „Anforderungsprofil“ oder „Rund um das Studium“, (b) Q&A-Bereich mit „Live Session“ und „Frage der Woche“, (c) Favoriten-Bereich

2 Konzept und Designprinzipien

StudyOrientation basiert auf dem App-Konzept von Aragon-Hahner; Sakel; Strickroth [ASS23], erweitert es konzeptionell um zahlreiche inhaltliche Module und fügt eine soziale Komponente hinzu (vgl. *Q&A Bereich*). Das bisher nur prototypisch realisierte Konzept wird außerdem praktisch in Form einer Android-Anwendung implementiert. Wie in Abbildung 1 dargestellt, gliedert sich *StudyOrientation* in drei zentrale Funktionsbereiche:

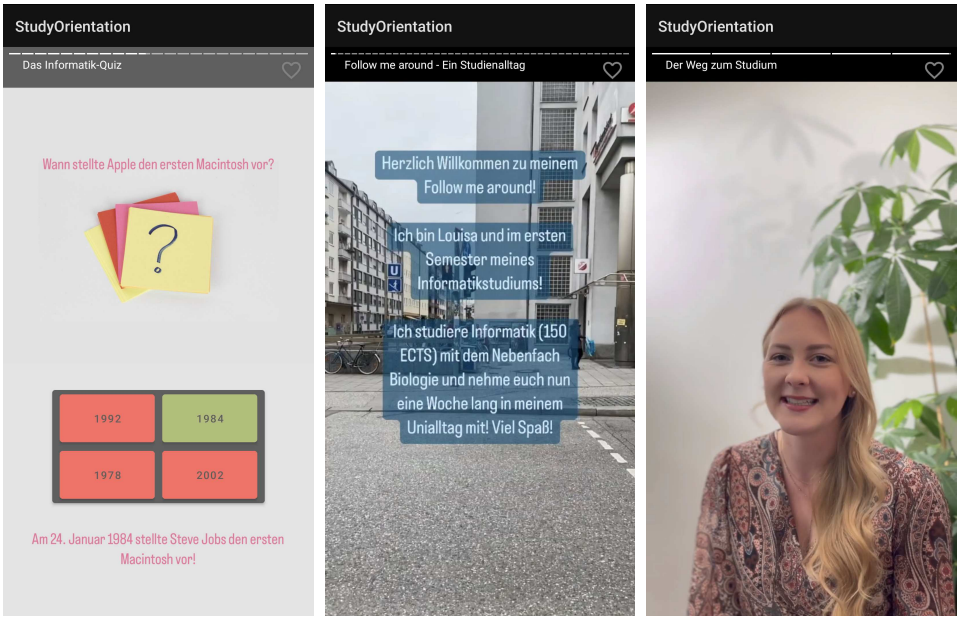
a) Der **Explore-Bereich** bietet umfassende Einblicke in das Informatikstudium, mit Themen wie Beispielvorlesungen, typischen Aufgaben, Anforderungen, Studienalltag, Studentenleben, Berufsbildern und universitären Strukturen. Er ist in neun Kategorien untergliedert, die wiederum jeweils mehrere thematische Unterkategorien enthalten.

b) Der **Q&A-Bereich** integriert Erfahrungsberichte von derzeitigen Studierenden durch die beiden Kategorien *Live Session* und *Frage der Woche*. Somit greift der Bereich gezielt die soziale Dimension des Reflexionsprozesses auf [Go21; Ha79; MHE16].

c) Ergänzt wird das Konzept durch den **Favoriten-Bereich**, der eine individuelle Speicherung und spätere Reflexion persönlicher Inhalte ermöglicht.

Die in *StudyOrientation* integrierten Inhalte sind nach Kategorien und Themenbereichen gegliedert, modular aufgebaut und werden in einem **Story-Format** präsentiert. Diese Art der Darstellung orientiert sich an dem Format von Inhalten auf gängigen Social-Media Plattformen (vgl. *Instagram Stories*) und setzt das Prinzip des Micro-Tasking um, indem kurze, in sich abgeschlossene Informations- und Interaktionseinheiten bereitgestellt werden. Die einzelnen Module können von Studieninteressierten flexibel, im eigenen Tempo und beliebig oft erkundet werden. Insgesamt bietet *StudyOrientation* **959 Einheiten von Micro-Content**, die sich aus Texten, Bildern, Videos sowie interaktiven Aufgaben und Quiz-Fragen zusammensetzen. Abbildung 2 zeigt ausgewählte Beispiele von Story-Elementen aus verschiedenen Modulen. Eine integrierte Fortschrittsleiste visualisiert die Anzahl und Länge der Stories. Nutzende können Stories pausieren, überspringen oder durch einen Klick auf das Herz-Icon zum Favoriten-Bereich hinzufügen. Für alle Themengebiete wird außerdem auf der Übersichtsseite der individuelle Fortschritt durch einen Fortschrittsbalken, eine prozentuale Anzeige und die dynamische Einfärbung der Informationskarten dargestellt (vgl. Abbildung 1c). *StudyOrientation* ermöglicht somit eine individuelle Informationsaufnahme im eigenen Tempo sowie eine kontinuierliche Auseinandersetzung mit studienrelevanten Inhalten, wodurch Raum für Selbstreflexion geschaffen wird [ASS23; Ga14].

Die Selbstreflexion wird in *StudyOrientation* durch explizite und implizite Gestaltungsprinzipien gefördert. Neben vielfältigen, relevanten und aktuellen Informationen zu Studieninhalten, Anforderungen und Berufsfeldern regen insbesondere die Erfahrungsberichte im Q&A-Bereich zur kritischen Auseinandersetzung mit individuellen Interessen, Erwartungen und Fähigkeiten an. *StudyOrientation* integriert vier zentrale Designressourcen zur Förderung der Selbstreflexion nach Bentvelzen et al. [Be22]: Die *Temporal Perspective* gibt Impulse zur Rückschau und antizipativen Reflexion; *Conversation* wird über studenti-



(a) Das Informatik-Quiz (b) Follow me around (c) Der Weg zum Studium

Abb. 2: *StudyOrientation* bietet insgesamt 959 Microcontent-Einheiten. Die Abbildung zeigt drei beispielhafte Stories: (a) Interaktives Quiz, (b) *Follow me around* zum Studienalltag, (c) Video-Interview mit einer Studentin über ihren Weg zum Studium.

sche Erfahrungsberichte in Form von regelmäßigen *Live Q&A Sessions* und dialogische Story-Formate adressiert; *Comparison* ermöglicht den Abgleich individueller Vorstellungen mit authentischen Einblicken; alle Inhalte der App fördern zudem die Entdeckung neuer Aspekte im Sinne von *Discovery*. Die App verzichtet dabei bewusst auf konkrete Eignungseinschätzungen oder Handlungsempfehlungen (vgl. [ASS23; HWJ13]) und legt den Fokus auf die Unterstützung einer eigenständigen, reflektierten Entscheidungsfindung.

3 Evaluation und Ausblick

StudyOrientation wurde im Rahmen einer dreiwöchigen explorativen Feldstudie mit 20 Teilnehmenden evaluiert. Erste Ergebnisse der qualitativen Analyse deuten an, dass *StudyOrientation* Selbstreflexion anregen und die Entscheidungssicherheit stärken kann. Besonders positiv hervorgehoben wurden das Story-Format, die Flexibilität durch Micro-Tasks sowie die Erfahrungsberichte der Studierenden. Erste Auswertungen der Nutzungsdaten zeigen zudem eine intensive und weitgehend vollständige Nutzung aller Themenbereiche über den gesamten Studienzeitraum hinweg – ein möglicher Indikator für hohes und anhaltendes Interesse. Eine detaillierte Analyse ist derzeit in Arbeit.

Literaturverzeichnis

[Ar22] Aragon Bartsch, S. et al.: A Day in the Life: Exploring the Use of Scheduled Mobile Chat Messages for Career Guidance. In: Proceedings of the 20th International Conference on Mobile and Ubiquitous Multimedia. MUM '21, Association for Computing Machinery, Leuven, Belgium, S. 24–34, 2022, DOI: 10.1145/3490632.3490637.

[Ar23] Aragon-Hahner, S. et al.: FindMyself: A Mobile Self-Reflection App to Support Students in Career Decision-Making. In: Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia. MUM '23, Association for Computing Machinery, Vienna, Austria, S. 463–476, 2023, DOI: 10.1145/3626705.3627774.

[ASS23] Aragon-Hahner, S.; Sakel, S.; Strickroth, S.: "Is Computer Science the Right Study Program for me?": Concept Development of a Mobile Self-Reflection App for Prospective University Students. In: Proceedings of the 15th International Conference on Computer Supported Education - Volume 2: CSEDU. INSTICC, SciTePress, S. 163–171, 2023, DOI: 10.5220/0011641400003470.

[Be22] Bentvelzen, M. et al.: Revisiting Reflection in HCI: Four Design Resources for Technologies that Support Reflection. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. Bd. 6 1. 1, Association for Computing Machinery, New York, NY, USA, 2022, DOI: 10.1145/3517233.

[Ch15] Cheng, J. et al.: Break It Down: A Comparison of Macro- and Microtasks. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. CHI '15, Association for Computing Machinery, Seoul, Republic of Korea, S. 4061–4064, 2015, DOI: 10.1145/2702123.2702146.

[Ga14] Garcia, J. et al.: Reflective Healthcare Systems: micro-Cycle of Self-Reflection to empower users. Interaction Design and Architecture(s) 23, S. 173–190, 2014, DOI: 10.55612/s-5002-023-011.

[GK21] Gati, I.; Kulcsar, V.: Making better career decisions: From challenges to opportunities. Journal of Vocational Behavior 126, S. 103545, 2021, DOI: 10.1016/j.jvb.2021.103545.

[Go21] Goetschi, T. J.: Der Einfluss von Informationen aus erster Hand auf die Studienwahl von Maturandinnen und Maturanden: Implikationen für die Unterstützung in der Studienberatung. 2021, DOI: 10.21256/zhaw-25062.

[Gr23] Grüneberg, T.: Selbstcoaching-App zur Berufs- und Studienwahl für SchülerInnen der Sekundarstufe II – Einordnung und Anwendung. In (Knickrehm, B.; Fletemeyer, T.; Ertelt, B.-J., Hrsg.): Berufliche Orientierung und Beratung: Aktuelle Herausforderungen und digitale Unterstützungsmöglichkeiten. Springer Fachmedien Wiesbaden, Wiesbaden, S. 255–269, 2023, DOI: 10.1007/978-3-658-40601-1_15.

[Ha79] Harren, V. A.: A model of career decision making for college students. Journal of Vocational Behavior 14 (2), S. 119–133, 1979, DOI: 10.1016/0001-8791(79)90065-4.

[He09] Heukamp, V. et al.: Internetbasierte Self-Assessments zur Unterstützung der Studienentscheidung. Zeitschrift für Beratung und Studium 4 (1), S. 2–8, 2009.

[HS14] Hasenberg, S.; Schmidt-Atzert, L.: Internetbasierte Selbsttests zur Studienorientierung. Beiträge zur Hochschulforschung 36 (1), S. 8–28, 2014.

[HWJ13] Huldtgren, A.; Wiggers, P.; Jonker, C.: Designing for Self-Reflection on Values for Improved Life Decision. Interacting with Computers 26, S. 27–45, 2013, DOI: 10.1093/iwc/iwt025.

[HWS10] Heine, C.; Willich, J.; Schneider, H.: Informationsverhalten und Entscheidungsfindung bei der Studien-und Berufswahl. In: Hochschul-Informations-System: Forum Hochschule. F1. F1, 2010.

[Iq18] Iqbal, S. T. et al.: Multitasking with Play Write, a Mobile Microproductivity Writing Tool. In: Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology. UIST '18, Association for Computing Machinery, Berlin, Germany, S. 411–422, 2018, DOI: 10.1145/3242587.3242611.

[LES23] Löwenstrom, B.; Ellerbeck, S.; Staden, C.: Vom Berufswahlpass zur „berufswahlapp" – Ein Klassiker der Portfolio-Arbeit in der Beruflichen Orientierung wird digital. In: Berufliche Orientierung und Beratung: Aktuelle Herausforderungen und digitale Unterstützungsmöglichkeiten. Springer, S. 237–253, 2023.

[Lü15] Lückert, F.: Online-Self-Assessments an Universitäten als Instrument der Studienberatung, deu, 2015, DOI: 10.48548/pubdata-850.

[MHE16] Mols, I.; van den Hoven, E.; Eggen, B.: Informing Design for Reflection: an Overview of Current Everyday Practices. In: Proceedings of the 9th Nordic Conference on Human-Computer Interaction. NordiCHI '16, Association for Computing Machinery, Gothenburg, Sweden, 2016, DOI: 10.1145/2971485.2971494.

[Mi08] Milbradt, A. et al.: Ziele von Self-Assessments und ihre Bedeutung für Entwicklung und Evaluation. Tysk, Report Psychologie 33, S. 352–362, 2008.

[Pä23] Päßler, K.: Die Integration von Online-Self-Assessments in die Studien-und Berufsberatung. In: Berufliche Orientierung und Beratung: Aktuelle Herausforderungen und digitale Unterstützungsmöglichkeiten. Springer, S. 359–371, 2023.

[Ra17] Rabušicová, P. M.: Studieren Probieren: A Guidance Programme for Prospective Students in Austrian Higher Education, Diss., Masaryk University, 2017.

[Sc17] Schneider, H. et al.: Erwerb der Hochschulreife und nachschulische Übergänge von Studienberechtigten : Studienberechtigte 2015 ein halbes Jahr vor und ein halbes Jahr nach Schulabschluss, Hannover, 2017, DOI: 10.2314/GBV:1028416199.

[St19] Stoll, G.: Online-Self-Assessments zur Studienfachwahl – wie Hochschulen die Potenziale dieses Instruments effektiv nutzen können. Der Übergang in die Hochschule: Strategien, Organisationsstrukturen und Best Practices an deutschen Hochschulen, S. 65–75, 2019.

[Tr03] Trafton, J. et al.: Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. International Journal of Human-Computer Studies 58 (5), Notification User Interfaces, S. 583–603, 2003, DOI: 10.1016/S1071-5819(03)00023-5.

Demonstrating GraphIT – a Platform for Working with Dependency Graphs of Learning Topics

Raphael Wimmer [1], Leonie Schrod [1], and Alexander Weichart [1]

Abstract: Students' prior knowledge and goals are getting more and more heterogeneous. Thus, one-size-fits-all, same-procedure-as-every-year degree programs and courses fit the students' biographies less and less. There is little tool support for designing cohesive but flexible degree programs and courses. GraphIT is a new prototypical platform for defining curricula and courses in a dependency graph, linking learning topics to their prerequisites. The dependency graph can then be used to highlight central topics, paths between topics, and clusters of related topics. This generic approach supports educators in collaboratively modeling degree programs, lecturers in creating courses, and students in identifying personal learning paths and goals. GraphIT is built on the WikiBase platform and can be queried using SPARQL. This allows for extending and customizing schema, content, and applications.

Keywords: graphs, dependency graphs, curricula, software, WikiBase

1 Dependency Graphs for Learning Topics

University students' backgrounds and goals are becoming more and more heterogeneous [Ju23]. This requires educators to design courses and curricula that are more flexible and can be combined and re-configured depending on changing requirements. However, there is little tool support for designing such courses or degree programs. To address this problem, researchers have designed different ontologies for structuring the knowledge that students should acquire [SPJ20]. Dependency graphs (Figure 1) are a special kind of ontologies that describe which topics build upon which prerequisites. In this paper, we outline how dependency graphs of learning topics can help in planning and organizing courses and curricula and describe our implementation.

By itself, such a graph can already be used to identify central topics that should be taught early on, gaps or overlaps between courses, and learning paths that lead to a certain goal. Extending it with additional nodes and edges allows it to be used as a tool for course organization, self-directed learning, or other uses.

Over the past decade, various prototypes of such dependency graphs have been built. Lightfoot [Li14] created a dependency graph for the curricula of a college of business administration. He applied various simple graph measures, such as Eigenvector centrality, to identify importance and interconnectedness between ~70 courses. In the STOPS platform

[1] Universität Regensburg, Germany, raphael.wimmer@ur.de, https://orcid.org/0000-0001-5162-5113; leonie.schrod@stud.uni-regensburg.de, https://orcid.org/0009-0006-1762-716X; alexanderweichart@fastmail.com, https://orcid.org/0009-0005-8257-1222

doi:10.18420/delfi2025_60

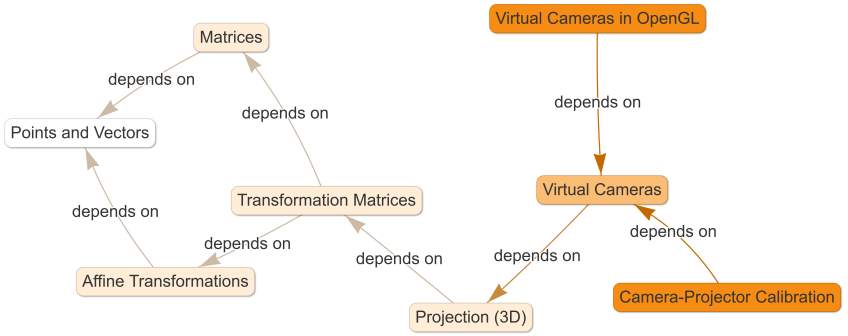


Fig. 1: A dependency graph of a selection of topics taught in a computer graphics course. Edges between nodes indicate that a topic builds on another topic. Topics with many incoming edges represent important foundations that should be learned early on.

by Auvinen et al. [APH14], nodes represent learning outcomes instead of courses. The authors mention that manually constructing such a graph is very time-consuming and that neither involved teachers nor students used the tool much. COMPBASE, an abandoned PhD project, defined an ontology of connected competences to support *portfolio learning* [DK19]. Pasterk et al. have published multiple papers on applications of graphs for modeling curricula [Ch20; PB17b]. They generate graphs from reference curricula. These graphs have two types of edges between nodes: "depends" and "expands". In the Digifit4all project[2], Pasterk et al. [Pa22] aim to build a comprehensive platform for building graphs of learning topics and integrating these in learning management systems. The platform seems to be in an unmaintained state, however. Learney, a now defunct platform[3] offered an interactive dependency graph of machine-learning foundations. For each topic, links to resources were provided, and users could mark whether they had understood a topic or wanted to learn it. As noted by multiple authors [APH14; Ch18; Ku18; PB17a], defining and building such dependency graphs is both non-trivial and tedious. Also, practical adoption of such tools has been rare [APH14; DK19]. Unlike previous approaches, we iteratively defined GraphIT's ontology while using it in teaching. We also preferred pragmatic, fuzzy definitions over formal consistency. This approach resulted in a platform that not only helps identify learning paths but also supports course planning, course organization, and learning analytics.

2 GraphIT

In GraphIT[4], nodes do not represent competences, learning outcomes or whole courses but small learning topics. As a rule of thumb, a topic typically spans about as much content as can be *explained* in a 10-20 minute talk (e. g., "Principles of Hamming Codes").

[2] https://www.digifit4all.at/
[3] https://web.archive.org/web/20240315072433/https://learney.me/
[4] https://graphit.ur.de

This does not necessarily mean that a topic can be learned in this amount of time. If we do not yet know how to break a larger topic into smaller ones, or if we don't currently see a need for that, we instead add it as a placeholder (e. g., "Mathematical Foundations"). Typically, dependencies are the central relation between nodes. However, we also use several other types of edges, such as "related to" for similar topics or "includes" for topic areas. GraphIT is built on the open-source WikiBase graph database/portal, which we extend in multiple ways. Currently, GraphIT contains three complete courses with about 400 topics. Nodes not only represent learning topics but also other concepts, such as courses, students and instructors, and resources. To interact with the graph, students can register a user account and connect "their" node to topics in the graph via "interested in" or "has completed" (Figure 2).

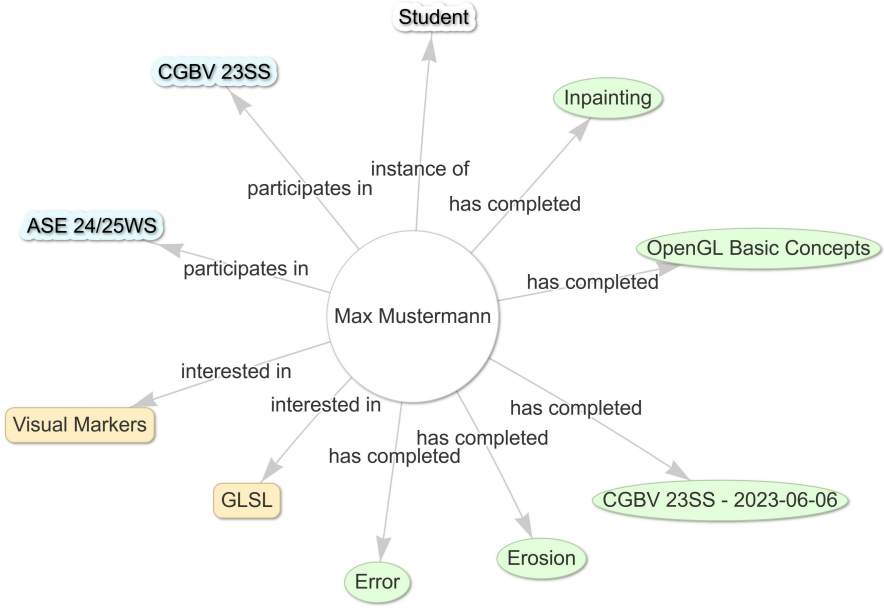


Fig. 2: On GraphIT, students can e. g., indicate topics they want to learn or have already completed, or subscribe to courses. Everything is stored as edges between the student's node and other nodes.

While we initially evaluated many graph databases, we decided to use WikiBase[5], the open-source platform underlying the WikiData[6] knowledge graph. WikiBase extends MediaWiki with a triple store, a query frontend, and various tools. While we deploy it via a customized Docker setup, there are also options for hosting Wikibase instances in the cloud. One major advantage of using WikiBase is that it offers a Wiki, user management, a query language (SPARQL[7]), simple visualizations, and an ecosystem of tools and communities out of the

[5] https://wikiba.se/
[6] https://www.wikidata.org/
[7] https://www.w3.org/TR/sparql11-query/

box. WikiBase requires no pre-defined schemas or class hierarchy. This allows us to iterate quickly and try out new ideas with minimal effort. In WikiBase, nodes are called *Items*, edges are called *Properties*. Each Item has its own wiki page that contains *Statements*, pairs of Properties and Items (e. g., "depends on": "Item Q123"). WikiBase has built-in support for qualifiers, i. e., annotations attached to edges. We use this feature for adding relevant metadata to edges. For example, for each dependency we can add a qualifier "added by: X" or "importance: essential". While WikiBase offers support for formal constraints[8], we do not use them as we are still trying out different ways of modeling relationships. Building on WikiBase allows us to easily link to content on WikiMedia Commons and to concepts in WikiData. Also, we can quickly create course pages that include auto-generated visualizations and reports (Figure 3 center/right). For example, students in a course on scientific methods were asked to mark topics from the course syllabus which they were interested in or which they did already know. With a short SPARQL query, we generated a visualization of students' interests (see Figure 3 right) and discussed it with them.

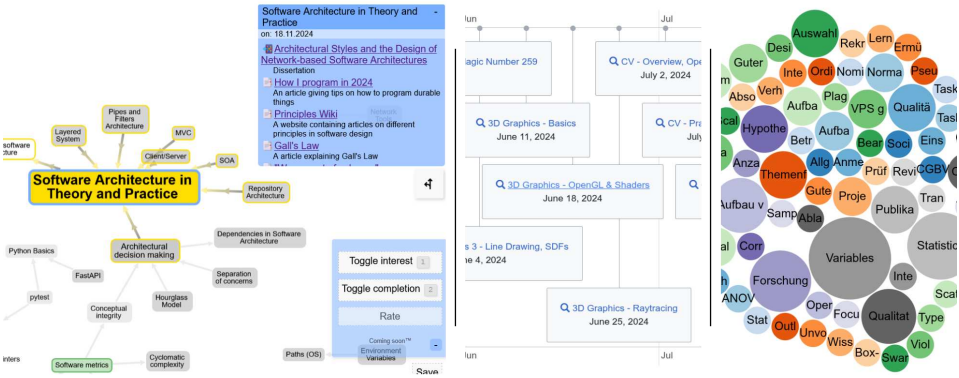


Fig. 3: We have developed custom frontends that e. g., allow users to plan their personal learning path and access accompanying materials for each topic (left). WikiBase also offers built-in visualizations for SPARQL queries which we use for displaying a course outline (center) or a list of participants' favorite topics (right).

3 Initial Experience and Outlook

We have used GraphIT for planning courses on computer graphics, software engineering, and scientific methods. Preliminary feedback from surveys and discussions indicates that students and other educators like the concept very much. However, many struggle with the amount of information represented in the usual graph visualizations. Therefore, we are currently developing custom frontends and visualizations for students and instructors (Figure 3 left). Formal evaluations will follow. All software and data are available under open licenses as outlined on the website.

8 https://www.wikidata.org/wiki/Help:Property_constraints_portal

References

[APH14] Auvinen, T.; Paavola, J.; Hartikainen, J.: STOPS: A Graph-Based Study Planning and Curriculum Development Tool. In: Proceedings of the 14th Koli Calling International Conference on Computing Education Research. Koli Calling '14, Association for Computing Machinery, New York, NY, USA, pp. 25–34, 2014, DOI: 10.1145/2674683.2674689.

[Ch18] Chen, P. et al.: KnowEdu: A System to Construct Knowledge Graph for Education. IEEE Access 6, pp. 31553–31563, 2018, DOI: 10.1109/ACCESS.2018.2839607.

[Ch20] Chystopolova, Y. et al.: Identification of Dependencies Between Learning Outcomes in Computing Science Curricula for Primary and Secondary Education – On the Way to Personalized Learning Paths. In (Kori, K.; Laanpere, M., eds.): Informatics in Schools. Engaging Learners in Computational Thinking. Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 185–196, 2020, DOI: 10.1007/978-3-030-63212-0_15.

[DK19] Dehne, J.; Kiy, A.: Using an Ontology Based Competence Database for Curriculum Alignment of Portfolio Based Learning. In. Proceedings of DELFI Workshops 2019. Gesellschaft für Informatik e.V.z, p. 121, 2019, https://dl.gi.de/handle/20.500.12116/27953.

[Ju23] Judel, S. et al.: Supporting Individualized Study Paths Using an Interactive Study Planning Tool. In. 21. Fachtagung Bildungstechnologien (DELFI). Gesellschaft für Informatik e.V., pp. 225–230, 2023, https://dl.gi.de/handle/20.500.12116/42196.

[Ku18] Kubekov, B. et al.: Methodology of Formation of Educational Resources On the Basis of Ontology. In: 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT). 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT). Pp. 1–6, 2018, DOI: 10.1109/ICAICT.2018.8747069.

[Li14] Lightfoot, J.: A Graph-Theoretic Approach to Improved Curriculum Structure and Assessment Placement. Communications of the IIMA 10 (2), 2014, DOI: 10.58729/1941-6687.1136.

[Pa22] Pasterk, S. et al.: DigiFit4All – Conceptualisation of a Platform to Generate Personalised Open Online Courses (POOCs). In (Passey, D. et al., eds.): Digital Transformation of Education and Learning - Past, Present and Future. IFIP Advances in Information and Communication Technology, Springer International Publishing, Cham, pp. 247–258, 2022, DOI: 10.1007/978-3-030-97986-7_21.

[PB17a] Pasterk, S.; Bollin, A.: A Graph-based Approach to Analyze and Compare Computer Science Curricula for Primary and Lower Secondary Education. In: Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education. ITiCSE '17, Association for Computing Machinery, New York, NY, USA, p. 365, 2017, DOI: 10.1145/3059009.3072985.

[PB17b] Pasterk, S.; Bollin, A.: Graph-Based Analysis of Computer Science Curricula for Primary Education. In: 2017 IEEE Frontiers in Education Conference (FIE). 2017 IEEE Frontiers in Education Conference (FIE). Pp. 1–9, 2017, DOI: 10.1109/FIE.2017.8190610.

[SPJ20] Stancin, K.; Poscic, P.; Jaksic, D.: Ontologies in Education – State of the Art. Education and Information Technologies 25 (6), pp. 5301–5320, 2020, DOI: 10.1007/s10639-020-10226-z.

Autorenverzeichnis

A
Abé, Alina, 343
Abu-Rasheed, Hasan, 385
Ahmad, Aakash, 239
Al Laban, Firas, 357
Allerkamp, Dennis, 431
Aragon-Hahner, Sarah, 459
Armbruster, Martin, 245
Ayon, Safial Islam, 357
Azaiz, Imen, 9

B
Baberowski, David, 363
Bach, Volker, 369
Berner, Tobias, 157, 443
Bernoth, Jan, 245, 357
Bhoyar, Rahul Rajkumar, 391
Bienhaus, Diethelm, 379
Biernacka, Katarzyna, 203
Bolz, Pascal, 131
Boyer, Claudine, 157, 443
Brachmann, Tobias, 401
Brassel, Paul, 449
Breck, Dominik, 171
Breuer, Martin, 193
Brocker, Annabell, 193
Brockner, Annabell, 373

C
Cohausz, Lea, 113
Czekansky, Jakob, 379

D
De Witt, Claudia, 391
Deinert, Christin, 283
Dieing, Thilo I., 113
Dietrich, André, 257
Dietze, Nadja, 221
Domsch, Hilke, 271
Dunkel, Louis, 385
Duong-Trung, Nghia, 391
Dymek, Andreas, 395
Dyrna, Jonathan, 25

E
Eck, Christopher, 401
Egetenmeier, Armin, 73, 413
Ehlenz, Matthias, 139, 245
Elisabeth Wrede, Silke, 391
Engel, Rahel Natalie, 407
Evangelou, Dominik, 41

F
Fehling, Dominic, 131
Fernes, David, 419
Fischer, Britta, 437
Frenz, Martin, 401

G
Gentile, Flavia, 263
Getz, Laura, 313
Göhler, Volker, 271
Greubel, André, 407
Grillenberger, Mareen, 185

H
Hahm, Nadine, 425
Hain, Heike, 449
Harrer, Andreas, 213
Hartanto, Jasmin, 277
Hawlitschek, Anja, 283, 307

Hees, Jörn, 331
Heimes, Lilli, 313
Heinemann, Birte, 139, 245, 277
Heister, Michael, 313
Herchet, Marilisa, 25
Herwig, Maurice, 289
Hilberer, Marie-Louise, 283
Hochuli, Evi, 263
Hornoff, Katja, 295
Hossain, Imran, 385
Hundeshagen, Norbert, 289
Hyadi, Jihad, 257

I
Irmak, Sila, 363

J
Jahnke, Isa, 3, 251
Janke, Stefan, 113
Joos, Richard, 25
Judel, Sven, 229, 401

K
Karababa, Burhan, 413
Kiesler, Natalie, 203
Klar, Maria, 41
Kleickmann, Thilo, 437
Klinger, Sigrid L., 81, 301
Köhler, Thomas, 25
Kollenberg, Cedric, 289
Komorowski, Tim, 313
König, Daniel, 313
Köppen, Veit, 307
Koschel, Arne, 431
Krieg, Aloys, 369

L
Leikert-Böhm, Ninja, 99
Lenk, Hannah, 407
Leonhardt, Thiemo, 185, 363
Liebold, Mariane, 295
Lucke, Ulrike, 245, 357

M
Mack, David, 57
Marnitz, Melina, 41
Mary Jose, Angelin, 391
Mascher, Ulrike, 419
Matter, Philipp, 99
Meissner, Roy, 425
Merker, Jochen, 449
Merkle, Belinda, 113
Morawetz, Erik, 425
Mueller, Wolfgang, 263
Mulders, Miriam, 41
Müller, Josefin, 295

N
Nadolny, Anne, 431
Niazi-Shahabi, Marie-Theres, 453

P
Pakosch, Anna, 431
Panse, Theresa, 391
Persike, Malte, 193
Pinkwart, Niels, 391
Poleshchuk, Alisa, 283

R
Rahner, Lara, 41
Reithmeier, Nico, 185
Rijn, Lars van, 391
Ritter, Marc, 171
Röpke, René, 89, 363
Roschke, Christian, 171

S
Sacher, Patrick, 385

Sakel, Sophia, 459
Salmen, Frederic, 395
Santiago, Andrea de, 251
Sauer, Justin, 379
Sboron, Martin, 319
Scheffler, Marc, 113
Scherlitzki, Juliane, 325
Schiffner, Daniel, 203
Schlicker, Marc, 437
Schlosser, Max, 171
Schmalfeldt, Thomas, 157, 443
Schmid, Thomas, 239, 319
Schmidt, Astrid, 437
Schmidt, Ludger, 57
Schneeweiß, Sarah, 213
Schnell, Erik, 229
Schöbel, Konrad, 449
Schrod, Leonie, 465
Schroeder, Ulrik, 139, 193, 229, 277, 373, 395, 401
Schulz, Anja, 295
Schulz, Sandra, 203
Schulze, Christin, 431
Seidel, Niels, 453
Seiler, Ruedi, 369
Sigethy, Lea, 459
Stöhr, Monika, 431
Strickroth, Sven, 9, 73, 81, 301, 413, 459
Striewe, Michael, 301
Surmann, Caroline, 221

T
Thews, Oliver, 319
Thor, Andreas, 425
Tillmann, Alexander, 419
Tomczyk, Samuel, 343
Träg, Kristian, 41
Trippel, Marie, 89

V
Vieten, Doerthe, 331
Virzi, Pio, 453
Vogel, Anne, 295

W
Wang, Xia, 391
Wanke, Mascha, 263
Weber, Dorian, 407
Weber, Patrick, 301
Weber, Robin, 289
Wegmeyer, Nadine, 283
Wehrmann, Frank, 337
Weichart, Alexander, 465
Weiland, Meike, 313
Weinberger, Armin, 5
Welk, Josefine, 171
Wiepke, Axel, 325, 357
Wilzeck, René, 437
Wimmer, Raphael, 465
Witt, Clemens, 185
Witthaus, Patrick, 343

Y
Yerinç, Lucas, 349

Z
Zdrallek, Markus, 131
Zender, Raphael, 337, 343, 437
Ziegler, Carina, 251
Zöllner, Moritz, 449
Zug, Sebastian, 257, 271