

Jacqueline Marie-Charlotte Schmidt

Grundlagenwissen zu Künstlicher Intelligenz von angehenden Lehrkräften

Modellbasierte Testentwicklung und Validierung

Grundlagenwissen zu Künstlicher Intelligenz von angehenden Lehrkräften

Modellbasierte Testentwicklung und Validierung

Jacqueline Marie-Charlotte Schmidt

Die Reihe **Berufsbildung, Arbeit und Innovation** bietet ein Forum für die grundlagen- und anwendungs-orientierte Forschung zu den Entwicklungen der beruflichen Bildungspraxis. Adressiert werden insbesondere berufliche Bildungs- und Arbeitsprozesse, Übergänge zwischen dem Schul- und Beschäftigungssystem sowie die Qualifizierung des beruflichen Bildungspersonals in schulischen, außerschulischen und betrieblichen Handlungsfeldern.

Hiermit leistet die Reihe einen Beitrag für den wissenschaftlichen und bildungspolitischen Diskurs über aktuelle Entwicklungen und Innovationen. Angesprochen wird ein Fachpublikum aus Hochschulen und Forschungseinrichtungen sowie aus schulischen und betrieblichen Politik- und Praxisfeldern.

Die Reihe ist gegliedert in die **Hauptreihe** und in die Unterreihe **Dissertationen/Habilitationen**.

Reihenherausgebende:

Prof.in Dr.in habil. Marianne Frieze

Justus-Liebig-Universität Gießen

Institut für Erziehungswissenschaften

Professur Berufspädagogik/Arbeitslehre

Prof. Dr. paed. Klaus Jenewein

Otto-von-Guericke-Universität Magdeburg

Institut I: Bildung, Beruf und Medien

Arbeitsbereich Gewerblich-technische Berufsbildung

Prof.in Dr.in Susan Seeber

Georg-August-Universität Göttingen

Professur für Wirtschaftspädagogik und Personalentwicklung

Prof. Dr. Lars Windelband

Karlsruher Institut für Technologie (KIT)

Institut für Berufspädagogik und Allgemeine Pädagogik

Professur Berufspädagogik

Wissenschaftlicher Beirat

- Prof. Dr. Matthias Becker, Hannover
- Prof.in Dr.in Karin Büchter, Hamburg
- Prof. Dr. Frank Bünning, Magdeburg
- Prof. Dr. Hans-Liudger Dienel, Berlin
- Prof. Dr. Uwe Faßhauer, Schwäbisch-Gmünd
- Prof. Dr. Karl-Heinz Gerholz, Bamberg
- Prof. Dr. Philipp Gonon, Zürich
- Prof. Dr. Dietmar Heisler, Paderborn
- Prof. Dr. Torben Karges, Flensburg
- Prof. Dr. Franz Ferdinand Mersch, Hamburg
- Prof.in Dr.in Manuela Niethammer, Dresden
- Prof.in Dr.in Karin Reiber, Esslingen
- Prof. Dr. Thomas Schröder, Dortmund
- Prof.in Dr.in Michaela Stock, Graz
- Prof. Dr. Tade Tramm, Hamburg
- Prof.in Dr.in Ursula Walkenhorst, Osnabrück

Weitere Informationen finden
Sie auf wbv.de/bai

Jacqueline Marie-Charlotte Schmidt

Grundlagenwissen zu Künstlicher Intelligenz von angehenden Lehrkräften

Modellbasierte Testentwicklung und Validierung

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin.

Dissertation zur Erlangung des akademischen Grades eines Doctor rerum politicarum (Dr. rer. pol.) an der Wirtschaftswissenschaftlichen Fakultät der Universität Leipzig

Titel der Dissertation: „Das Grundlagenwissen zu Künstlicher Intelligenz von (angehenden) Lehrkräften im berufsbildenden Bereich – Modellbasierte Testentwicklung und Validierung“

Disputationsdatum: 20.09.2023

Gutachter: Prof. Dr. Roland Happ

Gutachterin: Prof.in Dr.in Michaela Stock

Berufsbildung, Arbeit und Innovation –
Dissertationen/Habilitationen, Band 79

2024 wbv Publikation
ein Geschäftsbereich der
wbv Media GmbH & Co. KG, Bielefeld

Gesamtherstellung:
wbv Media GmbH & Co. KG, Bielefeld
wbv.de

Umschlagmotiv: 1expert, 123rf

Bestellnummer: 76522
ISBN (Print): 978-3-7639-7652-2
ISBN (E-Book): 978-3-7639-7653-9
DOI: 10.3278/9783763976539

Printed in Germany

Diese Publikation ist frei verfügbar zum Download unter
wbv-open-access.de

Diese Publikation mit Ausnahme des Coverfotos ist unter
folgender Creative-Commons-Lizenz veröffentlicht:
creativecommons.org/licenses/by-sa/4.0/deed.de



Für alle in diesem Werk verwendeten Warennamen
sowie Firmen- und Markenbezeichnungen können
Schutzrechte bestehen, auch wenn diese nicht als solche
gekennzeichnet sind. Die Verwendung in diesem Werk
berechtigt nicht zu der Annahme, dass diese frei verfü-
gbar seien.

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Die freie Verfügbarkeit der E-Book-Ausgabe dieser Publikation wurde ermöglicht durch ein Netzwerk wissenschaftlicher Bibliotheken und Institutionen zur Förderung von Open Access in den Sozial- und Geisteswissenschaften im Rahmen der *wbv OpenLibrary 2023*.

Die Publikation beachtet unsere Qualitätsstandards für Open-Access-Publikationen, die an folgender Stelle nachzulesen sind:

https://www.wbv.de/fileadmin/importiert/wbv/PDF_Website/Qualitaetsstandards_wbvOpenAccess.pdf

Großer Dank gebührt den Förderern der OpenLibrary 2023 im Fachbereich Berufs- und Wirtschaftspädagogik:

Otto-Friedrich-Universität **Bamberg** | Humboldt-Universität zu **Berlin** | Universitätsbibliothek **Bielefeld** | Bundesinstitut für Berufsbildung (BIBB, **Bonn**) | Rheinische Friedrich-Wilhelms-Universität **Bonn** | Technische Universität **Braunschweig** | Vorarlberger Landesbibliothek (**Bregenz**) | Staats- und Universitätsbibliothek **Bremen** | Universitäts- und Landesbibliothek **Darmstadt** | Universitäts- und Landesbibliothek **Düsseldorf** | Sächsische Landesbibliothek – Staats- und Universitätsbibliothek (SLUB, **Dresden**) | Goethe-Universität **Frankfurt am Main** | Pädagogische Hochschule **Freiburg** | Justus-Liebig-Universität **Gießen** | Fernuniversität **Hagen** | Staats- und Universitätsbibliothek **Hamburg** | TIB **Hannover** | Universitätsbibliothek **Kassel** | Karlsruhe Institute of Technology (KIT, **Karlsruhe**) | Pädagogische Hochschule **Karlsruhe** | Universitätsbibliothek **Kiel** | Universitäts- und Stadtbibliothek **Köln** | Universitätsbibliothek **Leipzig** | Zentral- und Hochschulbibliothek (ZHB, **Luzern**) | Hochschule der Bundesagentur für Arbeit (**Mannheim**) | Fachhochschule **Münster** | Universitäts- und Landesbibliothek **Münster** | Landesbibliothek **Oldenburg** | Pädagogische Hochschule **Schwäbisch Gmünd** | Universitätsbibliothek **St. Gallen** | Zürcher Hochschule für Angewandte Wissenschaften (ZAHW, **Winterthur**)

Für meinen Vater

Egon

(1953–2021)

Vorwort*

Ende 2023 wird niemand mehr bestreiten, dass Anwendungen der Künstlichen Intelligenz (KI) das Lernen in Schule und Hochschule entscheidend verändern werden. Das war 2020 noch anders. Es gab (und gibt) viele kritische Stimmen, die darauf hinwiesen, dass die Integration von KI-Anwendungen auch Gefahren berge und deshalb in Klassenzimmern und Seminarräumen nichts zu suchen habe. Lehrkräfte, sowohl aus dem Schul- als auch aus dem Hochschulbereich, haben von der Integration von KI-Anwendungen in den Unterricht abgeraten. Es gab sogar Länder (wie Italien), die den KI-basierten Chatbot ChatGPT verboten haben. Umso bemerkenswerter ist es, dass sich Frau Schmidt bereits im Jahr 2020 aktiv und konstruktiv mit der Integration von KI-Anwendungen in die Hochschullehre auseinandersetzt.

Die Leistung der Dissertation von Frau Schmidt ist in mehrfacher Hinsicht positiv zu bewerten und liefert hilfreiche Anknüpfungspunkte für die weitere Forschung. *Erstens* ist es Frau Schmidt gelungen, ein Messinstrument zu konstruieren und zu validieren, das sowohl Grundlagenwissen über Anwendungen der KI als auch nichtkognitive Facetten der KI-Anwendungen (wie Motivation, Interesse und Überzeugungen) erfassen kann. Neben Lehrbuchanalysen sind hier auch Experteneinschätzungen eingeflossen, die zeigen, dass Frau Schmidt im Rahmen ihrer Dissertation Brücken zwischen der beruflichen Bildung und der (Wirtschafts-)Informatik schlägt. *Zweitens* ist es Frau Schmidt gelungen, ein Fellow-Programm des Stifterverbandes einzuwerben, das es ermöglicht hat, Lerninhalte des KI-Campus in die Hochschullehre zu integrieren. Das Fellow-Programm hat den Austausch mit anderen Disziplinen (u. a. Psychologie, Publizistik, Jura) gefördert. *Drittens* hat Frau Schmidt ein Lehrkonzept entwickelt, das Anwendungen der KI innovativ in die Hochschullehre integriert. Dabei geht es nicht nur darum, dass die Studierenden Kompetenzen in Anwendungen der KI erwerben, sondern das Lehrkonzept hat zwei Perspektiven: Die Studierenden evaluieren digitale Lernangebote des KI-Campus (1) und erwerben quasi nebenbei Inhalte zu Anwendungen der KI (2). Das Lehrkonzept wurde von Frau Schmidt auf mehreren Tagungen vorgestellt und kam so gut an, dass es mittlerweile an 9 weiteren Hochschulen in der Lehre eingesetzt wird. Damit hat Frau Schmidt *viertens* die Vernetzung des Standortes Leipzig mit der PH Freiburg, der PH Schwäbisch Gmünd, der TU Kaiserslautern, den Universitäten Mainz, Stuttgart, Hamburg und Göttingen sowie im europäischen Kontext mit Breslau, Zürich und Graz weiter vorangetrieben. Schließlich hat Frau Schmidt mit ihrer Dissertation wichtige Vorarbeiten geleistet, die nun *fünftens* in weitere Forschungsprojekte am Institut für Wirtschaftspädagogik der Universität Leipzig münden. Eines davon ist das Projekt KIWi-MOOC (Massive Open Online Course zur Förderung und Erfassung von KI-Kompetenzen in der Domäne Wirtschaft), das auf dem Messinstrument der Dissertation aufbaut und für alle drei Phasen der Lehrerbildung

* Dieser Text wurde sprachlich mithilfe des KI-basierten Schreibassistenten DeepWrite überarbeitet.

(Studium, Vorbereitungsdienst, Lehrerfortbildung) Lernangebote zu KI-Anwendungen für (angehende) Lehrkräfte bereitstellt.

Es gibt noch viele weitere gute Gründe, die die Dissertation von Frau Schmidt äußerst lesenswert machen. Bei der Lektüre wird deutlich, wie anspruchsvoll das Ziel der Arbeit (theoretische Modellierung und empirische Prüfung) von Frau Schmidt verfolgt wird. Es bleibt zu wünschen, dass die vielen wertvollen Hinweise und Anregungen aus dieser Arbeit in die Lehre sowohl im schulischen als auch im universitären Bereich einfließen. Weitere wissenschaftliche Studien und Projekte, die auf der Arbeit von Frau Schmidt aufbauen, sind erwartbar.

Leipzig, Dezember 2023

Roland Happ

Danksagung

Die vorliegende Dissertation ist zwar im Ergebnis eine Einzelleistung, doch ist sie viel mehr das Resultat von gemeinsamer Anstrengung, kollektivem fachlichen Diskurs und der Unterstützung zahlreicher Menschen, die mich auf diesem Weg begleitet haben. Einigen von ihnen möchte ich an dieser Stelle meinen besonderen Dank aussprechen.

Mein aufrichtiger Dank gilt zunächst meinem Doktorvater Prof. Dr. Roland Happ. Mit Ihrer herausragenden Betreuung und der beispiellosen Unterstützung während der gesamten Bearbeitungszeit haben Sie nicht nur maßgeblich zum Erfolg dieses Projekts beigetragen, sondern auch die Weichen für meine persönliche und berufliche Weiterentwicklung gestellt. Danke für das Vertrauen, das Sie mir als Ihrer ersten betreuten Promovendin entgegengebracht haben!

Mein weiterer Dank richtet sich an Prof. Dr. Michaela Stock, die sich nicht nur bereit erklärt hat, das Zweitgutachten für die Dissertation zu übernehmen, sondern auch extra für die Disputation den weiten Weg aus Graz nach Leipzig auf sich genommen hat. Den fachlichen Diskurs habe ich als sehr bereichernd empfunden und Ihre ermutigenden Worte zum Abschluss (*„Ab einem gewissen Punkt machen Sie eine Arbeit nicht mehr besser, sondern nur noch anders!“*) werde ich wohl auch zukünftig noch gern zitieren.

Darüber hinaus danke ich Prof. Dr. Bogdan Franczyk, der den Vorsitz der Promotionskommission als Zeremonienmeister übernommen hat und mir stets ermöglicht (hat), meine Forschung durch die Schnittstelle zur Wirtschaftsinformatik auf eine breitere Basis zu stellen.

Ein herzliches Dankeschön gilt außerdem meinen (ehemaligen) Kolleg:innen an der Universität Leipzig. Allen voran dabei Dr. Ute Moschner, von der ich fachlich sowie persönlich so wahnsinnig viel lernen durfte und die mich insbesondere in der Endphase des Schreibprozesses mit ihrem akribischen Lektorat und den motivierenden Sprachnachrichten am Morgen so sehr bestärkt und unterstützt hat. Daneben bedanke ich mich von Herzen beim besten Barista im ganzen Zwischengeschoss, Sebastian Heidel, bei Dr. Carsten Vierke, Robert Zander, Stefan Rodegast, Dr. Luisa Fernanda Manrique Molina, Prof. Dr. Juliana Schlicht und Dr. Carsta Militzer-Horstmann. Ich bin sehr dankbar für die gemeinsame Zeit mit euch und für die wertvollen Freundschaften, die ich während dieser Zeit knüpfen durfte.

Bedanken möchte ich mich außerdem bei den studentischen und wissenschaftlichen Hilfskräften des Instituts für Wirtschaftspädagogik. Mein ganz besonderer Dank geht dabei an Tobias Gries und Lisa Gries, die mich tatkräftig in meinem Promotionsprojekt unterstützt haben. Ein weiterer Dank gilt allen Kooperationspartner:innen, die es mir ermöglicht haben, meine Untersuchung an den jeweiligen Studienstandorten durchzuführen.

Ein riesengroßer Dank gilt meinen Eltern Kerstin und Egon Schmidt. Ohne euch wäre es mir niemals möglich gewesen, überhaupt ein Studium aufzunehmen. Auch wenn die universitäre Welt so neu und unbekannt für euch war, habt ihr mich nie daran zweifeln lassen, wie stolz ihr auf mich seid. Umso trauriger macht es mich, dass mein bester Freund, mein Vater Egon, den Abschluss der Promotion nicht mehr miterleben konnte. Ihm ist diese Arbeit in tiefer Dankbarkeit gewidmet.

Darüber hinaus danke ich meinen Geschwistern, insbesondere Patrick Schmidt und Maren Südekum, und meinen Freund:innen, insbesondere Karolin Höfert, Yvonne Hein und Corinna Preiß für die vielen unbeschwerten Momente und dafür, dass ihr einfach für mich da seid!

Nicht zuletzt gilt mein ganz besonderer Dank meinem Ehemann Aaron Schmidt. Ich danke dir von ganzem Herzen für deinen unerschütterlichen Rückhalt, deine Geduld und deine grenzenlose Unterstützung.

Jacqueline Schmidt, im Januar 2024

Inhalt

Abkürzungsverzeichnis	15
Abbildungsverzeichnis	19
Tabellenverzeichnis	20
1 Entwicklung des Forschungsvorhabens	23
1.1 Relevanz der Arbeit	23
1.1.1 Digitalisierung und KI-gestützte digitale Transformation im berufsbildenden Bereich	23
1.1.2 Künstliche Intelligenz in der beruflichen Bildung	28
1.1.3 Zwischenfazit und Implikationen für die Arbeit	31
1.2 Zentrale Zielstellung und Forschungsfragen	32
1.3 Aufbau der Arbeit	35
2 Theoriebasierte Modellierung des KI-Grundlagenwissens für Lehrkräfte	37
2.1 Verortung des Untersuchungsgegenstandes	37
2.1.1 Zur Professionellen Kompetenz von Lehrkräften	37
2.1.2 Digitalisierungsbezogene Kompetenzen von Lehrkräften	41
2.1.3 Technologieakzeptanz	45
2.1.4 Zwischenfazit und Implikationen für die Untersuchung	49
2.2 Kognitive und non-kognitive Kompetenzfacetten	51
2.2.1 Wissen	51
2.2.2 Einstellungen	53
2.2.3 Überzeugungen	54
2.2.4 Motivation	56
2.2.5 Zwischenfazit	57
2.3 Inhaltliche Anforderungen an das Wissen zu Künstlicher Intelligenz von (angehenden) Lehrkräften	58
2.3.1 Künstliche Intelligenz als Fachgebiet der Informatik	58
2.3.2 Inhaltliche Strukturierung des Fachgebiets Künstliche Intelligenz	61
2.3.3 Zwischenfazit	64
2.4 Forschungsstand zur Ausprägung KI-bezogener Kompetenzfacetten bei (angehenden) Lehrkräften	66
2.4.1 Konzeptionelle Ansätze zu KI-Kompetenzen	66
2.4.2 Empirische Studien zu KI-bezogenen Kompetenzen von (angehenden) Lehrkräften	69
2.4.3 Zwischenfazit und Implikationen für die Untersuchung	73
2.5 Strukturmodell KI-bezogener Kompetenzfacetten von (angehenden) Lehrkräften im berufsbildenden Bereich	75

3	Operationalisierung KI-bezogener Kompetenzfacetten	77
3.1	Messgegenstand und Validierungsaspekte	77
3.1.1	Bestimmung des Messgegenstandes	77
3.1.2	Aspekte der Validierung	78
3.2	Item- und Testkonstruktion	85
3.2.1	Festlegung allgemeiner Testmerkmale	85
3.2.2	Itemformulierung und Itemrevision	88
3.2.3	Erste Testversion und Rahmenbedingungen der Testung	88
3.3	Objektivität des Tests	98
3.4	Dokumentation der Items zur Erfassung der KI-bezogenen Kompetenzfacetten	99
4	Empirische Erprobung des Testinstruments	101
4.1	Design	101
4.2	Datenaufbereitung und Umgang mit fehlenden Werten	106
4.3	Beurteilung des Wissenstests anhand von Kriterien der klassischen Testtheorie	107
4.3.1	Beurteilungsverfahren und Annahmen	107
4.3.2	Verteilungsanalyse	108
4.3.3	Itemkennwerte	113
4.3.4	Analyse der Antwortalternativen	120
4.3.5	Reliabilität	124
4.4	Validierungsaspekt: Interne Struktur	126
4.5	Validierungsaspekt: Beziehung zu anderen Merkmalen	127
4.6	Analyse der quantitativen Daten	130
4.6.1	Betrachtung im Querschnitt	130
4.6.2	Analyse der Daten im Längsschnitt	133
5	Das Grundlagenwissen zu Künstlicher Intelligenz bei angehenden Lehrkräften: neue Erkenntnisse und Perspektiven	137
5.1	Zusammenfassung der zentralen Ergebnisse	137
5.2	Limitationen und anschließende Forschungsdesiderata	144
5.3	Implikationen	147
5.3.1	Implikationen für die theoretische Weiterentwicklung des Modells	147
5.3.2	Implikationen für die Weiterentwicklung des Instruments	148
5.3.3	Implikationen für die Aus- und Weiterbildung von (beruflichen) Lehrkräften	149
	Literaturverzeichnis	151
	Anhang A: Fragebogen im Pretest (Version Universität Leipzig, 2022)	185
	Anhang B: Fragebogen im Posttest (Version Universität Leipzig, 2022)	205
	Anhang C: Verteilung der Summenscores (Wissenstest) der Subgruppen zu beiden Messzeitpunkten	231

Abkürzungsverzeichnis

AERA	American Educational Research Association
APA	American Psychological Association
AR	Augmented Reality
ASCOT	Förderrichtlinie des BMBF zur technologieorientierten Kompetenzmessung in der beruflichen Bildung (2011–2014)
B. Sc.	Bachelor of Science
BMBF	Bundesministerium für Bildung und Forschung
BPäd	Berufspädagogik
BWP	Berufs- und Wirtschaftspädagogik
CFA	Konfirmatorische Faktorenanalyse (Confirmatory Factor Analysis)
CK	Content Knowledge
COACTIV	Cognitive Activation in the Classroom (Studie zu Fachwissen und fachdidaktischem Wissen von Mathematiklehrkräften)
DC	Digitale Kompetenz (Digital Competence)
DGfE	Deutsche Gesellschaft für Erziehungswissenschaft
DigComp	Digital Competence Framework for Citizens
DigCompEdu	European Framework for the Digital Competence of Educators
DigEuLit	European Framework for Digital Literacy
DL	Digital Literacy
DSGVO	Datenschutzgrundverordnung
EFA	Explorative Faktorenanalyse
EQR	Europäischer Qualifikationsrahmen
FrüSe	Frühjahrssemester (Schweiz)
GPT	Generative Pre-Trained Transformer
ICT	Informations- und Kommunikationstechnologien
ILLEV	Innovativer Lehr-Lernortverbund in der akademischen Hochschul- ausbildung

IRT	Item Response Theory
ITS	Intelligente Tutorielle Systeme
KI	Künstliche Intelligenz
KiWiPäd	KI in der Wirtschaftspädagogik (Projekttitle)
KMK	Kultusministerkonferenz der Bundesrepublik Deutschland
KMO	Kaiser-Meyer-Olkin-Kriterium der Stichprobeneignung
KT	Klassische Testtheorie
M. A.	Master of Arts
M. Ed.	Master of Education
M. Sc.	Master of Science
ML	Machine Learning (Maschinelles Lernen)
MOOC	Massive Open Online Course
MW	Mittelwert
NCME	National Council on Measurement in Education
NBPTS	National Board for Professional Teaching Standards
NLP	Natural Language Processing
OECD	Organisation for Economic Cooperation and Development
OER	Open Educational Resources
PCA	Hauptkomponentenanalyse (Principal Component Analysis)
PCK	Pedagogical Content Knowledge
PH	Pädagogische Hochschule
PK	Pedagogical Knowledge
PPäd	Pflegepädagogik
SC	Single Choice
SoSe	Sommersemester
SPSS	Statistical Package for the Social Sciences
SWK	Ständige Wissenschaftliche Kommission der Kultusministerkonferenz
TAM	Technology Acceptance Model
TBA	Task-Based-Approach
TK	Technology Knowledge

TCK	Technological Content Knowledge
TPB	Theory of Planned Behavior (Theorie des geplanten Verhaltens)
TPCK	Technological Pedagogical Content Knowledge
TRA	Theory of Reasoned Action (Theorie des überlegten Handelns)
TU	Technische Universität
UL	Universität Leipzig
UTAUT	Unified Theory of Acceptance and Use of Technology
VCL	Virtual Collaborative Learning
VR	Virtual Reality
WINF	Wirtschaftsinformatik
WiPäd	Wirtschaftspädagogik
WiSe	Wintersemester

Abbildungsverzeichnis

Abb. 1	Aufbau der Arbeit	36
Abb. 2	Professionelle Kompetenzen von Lehrkräften	38
Abb. 3	Binnendifferenzierung des Professionswissens	39
Abb. 4	Fachkompetenz in der Domäne Wirtschaftswissenschaften	40
Abb. 5	Das TPACK-Modell	44
Abb. 6	Theorie der überlegten Handlung	47
Abb. 7	Technology Acceptance Model 2 (TAM2)	48
Abb. 8	Einflussfaktoren auf die Akzeptanz	49
Abb. 9	Einordnung der Begriffe Künstliche Intelligenz, Machine Learning und Deep Learning	59
Abb. 10	Inhaltliche Subdomänen der KI	63
Abb. 11	Einordnung der AI Literacy in die Taxonomie nach Bloom (1976)	68
Abb. 12	Theoretisches Rahmenmodell als Grundlage für die Operationalisierung des Messgegenstandes	75
Abb. 13	Messmodell der kognitiven und non-kognitiven Kompetenzfacetten	99
Abb. 14	Verteilung der Summenscores der Zielgruppe zum Messzeitpunkt t1 (N = 153)	109
Abb. 15	Normalverteilungsdiagramm (Q-Q-Diagramm) zum Summenscore der Zielgruppe zum Messzeitpunkt t1	110
Abb. 16	Verteilung der Summenscores der Zielgruppe zum Messzeitpunkt t2 (N = 81) ..	112
Abb. 17	Normalverteilungsdiagramm (Q-Q Diagramm) der Zielgruppe zum Messzeitpunkt t2	112
Abb. 18	Lösungshäufigkeiten der Zielgruppe zu Messzeitpunkt t1	115
Abb. 19	Lösungshäufigkeiten in der Zielgruppe zu Messzeitpunkt t2	117

Abb. C1	Verteilung der Summenscores Subgruppe M. Sc. WiPäd t1	231
Abb. C2	Verteilung der Summenscores Subgruppe M. Sc. WiPäd t2	231
Abb. C3	Verteilung der Summenscores Subgruppe M. Ed. WiPäd t1	232
Abb. C4	Verteilung der Summenscores Subgruppe M. Ed. WiPäd t2	232
Abb. C5	Verteilung der Summenscores Subgruppe M. A/M. Sc. Berufspädagogik t1	233
Abb. C6	Verteilung der Summenscores Subgruppe M. A/M. Sc. Berufspädagogik t2	233
Abb. C7	Verteilung der Summenscores Subgruppe Lehrdiplom Berufsmaturität t1	234
Abb. C8	Verteilung der Summenscores Subgruppe Lehrdiplom Berufsmaturität t2	234
Abb. C9	Verteilung der Summenscores Kontrastgruppe B. Sc. WiPäd t1	235
Abb. C10	Verteilung der Summenscores Kontrastgruppe B. Sc. WiPäd t2	235
Abb. C11	Verteilung der Summenscores Kontrastgruppe Wirtschaftsinformatik t1	236

Tabellenverzeichnis

Tab. 1	Übersicht der analysierten Grundlagenwerke	62
Tab. 2	AI Literacy-Aspekte	67
Tab. 3	Forschungsstand zum KI-Wissen von (angehenden) Lehrkräften	69
Tab. 4	Forschungsstand zu non-kognitiven Kompetenzfacetten von (angehenden) Lehrkräften zu KI	71
Tab. 5	Abgeleitete Annahmen für die empirische Überprüfung	74
Tab. 6	Matrix zur Systematisierung des Grundlagenwissens	78
Tab. 7	Validierungsaspekte	80
Tab. 8	Validierungsschritte nach den internationalen Teststandards der AERA et al. (2014)	85
Tab. 9	Zuordnung der Wissensitems zu den Inhaltsbereichen	89
Tab. 10	Einordnung der Items in die Matrix zur Systematisierung des Messgegenstandes	90

Tab. 11	Übersicht der adaptierten Konstrukte zur Erfassung der non-kognitiven Kompetenzfacetten	91
Tab. 12	Operationalisierung der Einstellungen zu Künstlicher Intelligenz	92
Tab. 13	Operationalisierung der Überzeugungen zu Künstlicher Intelligenz	93
Tab. 14	Operationalisierung der motivationalen Dispositionen	94
Tab. 15	Einzelhebungen in der Zielgruppe	103
Tab. 16	Übersicht der erfassten Studiengänge (Subgruppen der Zielgruppe)	104
Tab. 17	Einzelhebungen in der Kontrastgruppe „Studierende der Wirtschaftsinformatik“	105
Tab. 18	Zusammenfassung von Ziel- und Kontrastgruppe(n)	106
Tab. 19	Codierung der Likert-Skala	106
Tab. 20	Bestimmung der zentralen Tendenz der Messung in der Zielgruppe	108
Tab. 21	Prüfung der Verteilungsannahmen bei den Subgruppen zu t1	111
Tab. 22	Prüfung der Verteilungsannahmen bei den Subgruppen zu t2	113
Tab. 23	Schwierigkeitsindex (P) der Wissensitems in der Zielgruppe zu Messzeitpunkt t1	114
Tab. 24	Schwierigkeitsindex (P) der Wissensitems in der Zielgruppe zu Messzeitpunkt t2	116
Tab. 25	Korrigierte Trennschärfen der Fachitems	119
Tab. 26	Distraktorenanalyse zu t1 und t2	121
Tab. 27	Mittlere Testscores der Zielgruppe (gesamt) und der Kontrastgruppe WINF ...	128
Tab. 28	Mittlere Testscores der Zielsubgruppe (WiPäd) und der Kontrastgruppe WINF	128
Tab. 29	Mittlere Testscores der Zielgruppe (gesamt) und der Kontrastgruppe Bachelor WiPäd	129
Tab. 30	Mittlere Testscores der Zielsubgruppe (WiPäd) und der Kontrastgruppe Bachelor WiPäd	130
Tab. 31	Deskriptive Statistik zum Testscore der Zielgruppe zu t1	130
Tab. 32	Deskriptive Befunde zu den Einstellungen der Zielgruppe zu t1	131

Tab. 33	Reliabilität der adaptierten Skalen zur Erfassung der Einstellungen	132
Tab. 34	Deskriptive Befunde zu den Überzeugungen der Zielgruppe zu t1	132
Tab. 35	Reliabilität der adaptierten Skala zur Erfassung der epistemologischen Überzeugungen	132
Tab. 36	Deskriptive Befunde zu den motivationalen Dispositionen der Zielgruppe zu t1	133
Tab. 37	Reliabilität der adaptierten Skalen zur Erfassung der Motivation	133
Tab. 38	Testscore der Zielgruppe im Pre-Post-Vergleich	134
Tab. 39	Einstellungen der Zielgruppe im Pre-Post-Vergleich	135
Tab. 40	Überzeugungen der Zielgruppe im Pre-Post-Vergleich	135
Tab. 41	Motivation der Zielgruppe im Pre-Post-Vergleich	135

1 Entwicklung des Forschungsvorhabens

1.1 Relevanz der Arbeit

1.1.1 Digitalisierung und KI-gestützte digitale Transformation im berufsbildenden Bereich

Seit mehreren Dekaden werden unter dem Schlagwort „Digitalisierung“¹ die Veränderungen von Lebens- und Arbeitsbereichen durch die Nutzung von Computern und digitalen Medien diskutiert (vgl. Schumann et al., 2022). Dieser Diskurs wird seit nunmehr 20 Jahren maßgeblich von der Verbreitung von Internettechnologien vorangetrieben (Niegemann, 2020, S. 17; Seufert et al., 2020a, S. 10; Seufert et al., 2021, S. 10). Dabei ist der Einfluss von Digitalisierung, insbesondere durch die Nutzung von digitalen Kommunikationstechnologien, sowohl im beruflichen als auch im privaten Alltag mittlerweile als omnipräsent zu bewerten (Euler & Wilbers, 2020, S. 428; Gerholz et al., 2022b, S. 11) und die Digitalisierung hat sich damit zu einem Mega-Trend der Gegenwart etabliert (Euler & Wilbers, 2020, S. 428). Die digitalisierungsbedingten Veränderungen auf der Ebene der Arbeits- und Geschäftsprozesse beeinflussen in immer stärkerem Maße die beruflichen Tätigkeiten und damit auch die berufliche Aus- und Weiterbildung (vgl. Geiser, 2022; Geiser et al., 2021; Winther, 2019). Dabei wurde lange primär die Perspektive der Substitution eingenommen, in deren Zusammenhang der Wegfall von Berufen und Tätigkeitsfeldern durch Automatisierungsprozesse kritisch (auch mit Sorgen behaftet) diskutiert wurde (siehe dazu Bonin et al., 2015; Dengler & Matthes, 2018; Frey & Osborne, 2013, 2017). In den letzten Jahren verlagerte sich diese Perspektive stärker auf die Frage nach substanziellen Änderungen von beruflichen Tätigkeiten² im Rahmen konkreter Arbeits- und Geschäftsprozesse und der damit assoziierten Fragestellung, inwieweit sich dadurch die Kompetenzanforderungen an (zukünftige) Fachkräfte verändern (Teichmann et al., 2020, S. 513). Winkler und Schwarz (2021, S. 84) konstatieren, dass sich die Veränderungen durch die Digitalisierung „insbesondere auf Fachkräfte in der mittleren Qualifikationsebene auswirken“. Unter dem Schlagwort *Berufsbildung 4.0* wird in diesem Zusammenhang diskutiert, was (zukünftige) Fachkräfte vor dem Hintergrund der dynamischen Transformationsprozesse³

1 Im aktuellen Diskurs ist zu beobachten, dass die Begriffe der Digitalisierung und der digitalen Transformation häufig synonym verwendet werden (Bindner & Cramer, 2021; SWK, 2022, S. 14). Dieser synonymen Begriffsverwendung schließt sich die vorliegende Arbeit an.

2 Diese werden für verschiedene Tätigkeitsfelder differenziert betrachtet. Für den industriellen Bereich kann dabei auf Baethge-Kinsky (2019) verwiesen werden und für den kaufmännisch-verwaltenden Bereich liefert Dubs (2018) eine Prognose über digitalisierungsbedingte Veränderungen und Ansätze zur Reform der beruflichen Bildung. Riebenbauer et al. (2022) präzisieren die Veränderungen von digitalisierungsbedingten Kompetenzanforderungen im kaufmännisch-verwaltenden Bereich für (angehende) Beschäftigte im Finanz- und Rechnungswesen.

3 Gössling et al., (2019, S. 549) argumentieren, dass die Akteurinnen und Akteure der beruflichen Bildung nicht nur auf die technologischen Veränderungen reagieren, sondern die Bildung auch ein Faktor ist, der weitere Entwicklungen und damit einhergehende veränderte Kompetenzanforderungen erst initiiert und ermöglicht. Es besteht damit eine wechselseitige Beziehung zwischen beruflicher Bildung und digitalen Transformationsprozessen, wodurch sich eine Gestaltungsaufgabe manifestiert (Gerholz et al., 2022b, S. 11; Roppertz, 2021, S. 19).

wissen⁴ und können müssen (vgl. Müller, 2018) und wie die berufliche Aus- und Weiterbildung daran auszurichten ist.

Bereits seit der Einführung des Lernfeldkonzepts im Jahr 1996 ist das übergeordnete Ziel der beruflichen Bildung der Erwerb von Handlungskompetenz (Achtenhagen & Pätzold, 2010, S. 140). Diese wird definiert als „Bereitschaft und Befähigung des Einzelnen, sich in beruflichen, gesellschaftlichen und privaten Situationen sachgerecht durchdacht sowie individuell und sozial verantwortlich zu verhalten“ (KMK, 2007, S. 10). Insbesondere durch die Strukturierung nach beruflichen Handlungsfeldern⁵ unterscheidet sich der Unterricht an berufsbildenden Schulen maßgeblich von allgemeinbildenden Schulen (Gerholz et al., 2022b, S. 12). Dadurch ist „die Berufsbildung [...] besonders früh und intensiv von der Innovationsrasanz in Wirtschaft und Technik betroffen“ (Euler & Wilbers, 2020, S. 428).

Im Kontext der Digitalisierung haben sich insbesondere digitale Medienkompetenzen (z. B. Baacke, 1996) als essenziell für (zukünftige) Fach- und Führungskräfte etabliert (vgl. Dubs, 2018; KMK, 2017; Wilbers, 2012, 2019). Die Europäische Kommission ordnet diese als eine von acht elementaren Schlüsselkompetenzen ein (Europäische Kommission, 2010, S. 29). Bei der Betrachtung von sich verändernden Kompetenzprofilen als Folge der Digitalisierung wird dabei häufig die Perspektive der Lernenden eingenommen (vgl. Eickelmann et al., 2019). Aus der ICIL-Studie⁶ geht hervor, dass die digitalisierungsbezogenen Kompetenzen von deutschen Schülerinnen und Schülern signifikant über dem internationalen Mittelwert liegen, seit der letzten Kohorte im Jahr 2013 allerdings keine signifikante Verbesserung festgestellt werden kann. Diese für den allgemeinbildenden Schulbereich generierten Befunde und die im internationalen Vergleich festgestellte unterdurchschnittliche Technikausstattung der Schulen in Deutschland können nach Einschätzung von Gerholz et al. (2022b, S. 12) auch für den berufsbildenden Bereich angenommen werden. Das ist besonders vor dem Hintergrund relevant, dass die Förderung digitalisierungsbezogener Kompetenzen nicht nur durch das Lernen über Technologien (*als Lerngegenstand*), sondern auch das Lernen mit und durch⁷ Technologien (*als Lernmedien*)⁸ ein wichtiger Faktor für die Vorbereitung der Lernenden auf eine digitalisierte (Arbeits-)Welt ist (vgl. Gerholz et al., 2020b, S. 11; Guggemos & Seufert, 2020). Im Zentrum steht dabei primär⁹ die Frage, wie der Kompetenzerwerb mit digitalen Medien unterstützt werden kann (vgl. Fischer & Barabasch, 2020; Roll & Ifenthaler, 2020; Schmidt et al., 2020). Digitale Me-

4 Einen Überblick über digitalisierungsbezogene Wissensanforderungen im kaufmännisch-verwaltenden Bereich liefert Geiser (2022, S. 39–43).

5 Klausner (1999, S. 414) spricht in diesem Zusammenhang von einer Tätigkeitsorientierung.

6 ICIL steht für *International Computer and Information Literacy* (Bach, 2016, S. 107; Eickelmann et al., 2019, S. 7).

7 Zur Systematisierung von Lernaktivitäten mit digitalen Medien hat sich z. B. das SAMR-Modell (vgl. Puentedura, 2006, 2012) etabliert. Eine Standortbestimmung zum Lernen und Lehren mit digitalen Medien liefert Scheiter (2021).

8 Das Lehren und Lernen mit digitalen Medien wird über alle Stufen des Bildungssektors ausgehend vom Kindergarten (vgl. Cohen & Hemmerich, 2020; Nenner et al., 2021) über die Primarstufe (vgl. Botturi, 2019), die Sekundarstufe (vgl. Gentilin, 2019) bis hin zur akademischen Hochschulbildung (vgl. Seufert et al., 2020b) adressiert. Einen Überblick liefert das Gutachten der SWK (2022, S. 14).

9 Zum Teil wird auch der Frage nachgegangen, wie die (Lern-)Motivation durch die Nutzung digitaler Medien gefördert werden kann (siehe dazu Siegfried & Hermkes, 2020). Darüber hinaus sind mit dem Einsatz digitaler Medien in Lehr-Lern-Prozessen Fragen der Akzeptanz assoziiert. Diese werden sowohl aus der Perspektive der Schüler:innen (vgl. Ifenthaler & Schweinbenz, 2016) als auch aus der Perspektive der Lehrenden untersucht (vgl. Ifenthaler & Schweinbenz, 2013).

dien beinhalten in diesem Kontext insbesondere Potenziale für die authentische und realitätsnahe Gestaltung von Aus- und Weiterbildung (siehe z. B. die Bürosimulation LUCA, Deutscher et al., 2022; Rausch et al., 2021) sowie für die weniger zeit- und ortsgebundene Verlagerung von Lehr-Lern-Prozessen auf E-Learning-Plattformen (z. B. MOOCs, vgl. Kögler et al., 2020) oder das Lernen mit Smartphones (vgl. De Witt, 2013; Korbach & Niegemann, 2016; Niegemann, 2020).

In Anlehnung an die *Curriculum-Instruction-Assessment-Triade* (vgl. Pellegrino, 2010) nimmt die Digitalisierung auch Einfluss auf die Prüfungsgestaltung. Die Relevanz von technologiebasierter Kompetenzmessung wird u. a. durch die zahlreichen Projekte im Rahmen der ASCOT-Initiative verdeutlicht, die bereits innerhalb der ersten Förderphase für breitenwirksame Resultate gesorgt haben (vgl. Abele et al., 2016; Baethge-Kinsky et al., 2016; Beck et al., 2016; Dietzen et al., 2016; Döring et al., 2016; Eberle et al., 2016; Seeber et al., 2016; Seifried et al., 2016; Walker et al., 2016; Weber et al., 2016; Winther et al., 2016; Ziegler et al., 2016).

Eine neue Dynamik haben Digitalisierungsprozesse im Bildungssektor auch durch die Covid-19-Pandemie erhalten (vgl. Martín-Núñez et al., 2023; Wuttke et al., 2020, S. 9). In diesem Kontext haben z. B. Videokonferenzen ihren *Peak* erreicht (Niegemann, 2020, S. 17). Die Umstellung von Präsenz- auf Distanz- und Hybridunterricht impliziert für die Lehrenden und Lernenden Herausforderungen (vgl. Slepcevic-Zach et al., 2022; Wagner-Herrbach et al., 2022). Es haben sich vor allem in der Zeit der Pandemie Formate bewährt, die bereits vorher digital unterstützt waren (z. B. das *virtual collaborative learning (VCL)*, vgl. Schoop et al., 2021).

Die inhaltliche und didaktisch-methodische Planung und Durchführung der Lehr-Lern-Prozesse am Lernort Schule ist Aufgabe der berufsbildenden Lehrkräfte (Berding & Lamping, 2014, S. 1). Die Lehrkräfte sind in diesem Zusammenhang in doppelter Hinsicht gefordert. Einerseits sollen sie die digitalen Kompetenzen der Auszubildenden in der jeweiligen (dynamischen Veränderungen unterliegenden) beruflichen Handlungsdomäne fördern und andererseits selbst digitale Medien für die lernwirksame Ausgestaltung von Bildungsprozessen und damit der Erreichung von Kompetenzzielen einsetzen können (Meiners et al., 2022, S. 124; Gerholz et al., 2022b, S. 12). Die Lehrkräfte sind in diesem Zusammenhang demnach mit einer fachwissenschaftlichen und fachdidaktischen Aufgabe konfrontiert (Busse et al., 2022; Gerholz et al., 2022b, S. 12) und nehmen eine zentrale Rolle im Kontext der digitalen Transformation der beruflichen Bildung ein.

Aus dem Diskurs haben sich zahlreiche Forschungsperspektiven ergeben, die die digitalisierungsbezogenen Kompetenzfacetten von Lehrkräften untersuchen (siehe z. B. das *TPACK-Modell*, Mishra & Koehler, 2006; *DigCompEdu*, Redecker, 2017; *Business Digital Literacy*, z. B. Schlottmann et al., 2021). Im Rahmen der Aus- und Weiterbildung von Lehrkräften im berufsbildenden Bereich werden außerdem zunehmend digitale Technologien genutzt, um die damit assoziierten Potenziale auszuschöpfen und den digitalen Kompetenzerwerb zu unterstützen (z. B. im Rahmen digital gestützter komplexer Lehr-Lern-Arrangements (siehe dazu Schlicht et al., 2017; Schlicht et al., 2022)). Dies geschieht u. a. mit der Zielstellung, authentische Situationen der zukünfti-

gen beruflichen Praxis der (angehenden) Lehrkräfte bereits in die universitäre Ausbildung zu integrieren und damit eine situative Förderung und Erfassung von notwendigen Kompetenzfacetten zu ermöglichen (vgl. Faath-Becker & Walker, 2020; Gerholz et al., 2022c; Kuhn et al., 2018; Saas et al., 2020; Schäfer et al., 2020). Ziel dieser Ansätze ist es, die häufig von (angehenden) Lehrkräften berichteten Schwierigkeiten, die Theorie aus dem Studium in praktischen Situationen in der beruflichen Tätigkeit umzusetzen, zu minimieren (vgl. Böhner, 2009; Gruber et al., 1999).

Aktuell werden die Diskussionen im Kontext der Digitalisierung und digitalen Transformation zunehmend durch den Bezug zu Anwendungen der Künstlichen Intelligenz (KI) geprägt (vgl. Wilbers, 2021). Dabei wird KI häufig in der sogenannten zweiten Welle der Digitalisierung verortet (Seufert et al., 2020a, S. 5; Wahlster, 2017; Winkler & Schwarz, 2021, S. 88). Die wachsende Bedeutung von KI in zahlreichen Lebens- und Arbeitsbereichen wird nicht zuletzt durch die „Strategie Künstliche Intelligenz der Bundesregierung“ (Die Bundesregierung, 2018) untermauert. KI wird dabei als „Treiber der Digitalisierung“ (Die Bundesregierung, 2018, S. 10) identifiziert, der zahlreiche Potenziale¹⁰ für die Wirtschaft, den Arbeitsmarkt, die Gesellschaft und die Bildung bringt (vgl. Chiu et al., 2022; Holmes et al., 2019). Die Entwicklungen der KI und ihre Folgen werden dabei hoch kontrovers diskutiert und die Einschätzungen bewegen sich auf einem Kontinuum zwischen Euphorie und Dystopie (Pargmann & Berding, 2022, S. 216). Auf der einen Seite des Spektrums stehen Vertreter:innen, die große Chancen und Potenziale für Arbeit und Gesellschaft sehen. Auf der anderen Seite wird mit Risiken durch den Datenschutz und Kontrollverlust sowie mit Diskriminierungsproblematiken argumentiert. Charakteristisch für den Diskurs ist somit die präsente „Binarität der Argumentation“ (Pargmann & Berding, 2022, S. 219) und zumindest teilweise ist der kontroverse Diskurs mit dem aktuellen *Hype* um die neuen Möglichkeiten zu erklären (Holmes et al., 2019, S. 1), der zum Teil mit Unsicherheiten behaftet ist.

Es ist unstrittig, dass sich durch den Einsatz von KI-basierten Systemen Arbeits- und Geschäftsprozesse zahlreicher Tätigkeitsfelder bereits geändert haben und noch weiter ändern werden (Seifried & Ertl, 2021, S. 341). Die Gestaltung dieses Strukturwandels wird dabei als große Herausforderung herausgearbeitet (Die Bundesregierung, 2018, S. 10). Es wird erwartet, dass der zunehmende Einsatz von KI-Technologien die Arbeit von Menschen „grundlegend und stetig“ (S. 27) verändern wird. Der künftigen (neu zu ordnenden) Zusammenarbeit von Menschen und Maschinen wird dabei ein großes Potenzial zugesprochen. Durch die Übernahme von zeitintensiven Aufgaben (z. B. Recherche) durch KI-basierte Systeme sollen deutliche Effizienzsteigerungen ermöglicht werden (Seufert et al., 2021, S. 17). Die digitalisierungsbezogenen Veränderungen beruflicher Tätigkeiten werden daher in aktuellen Studien auch mit konkretem Bezug zu Anwendungen der KI diskutiert (vgl. Hauser & Klotz, 2022). Dabei ist zu beobachten, dass der bereits angesprochene Paradigmenwechsel von einer

10 Die Entwicklung von KI wird mit dem Potenzial assoziiert, die Gesellschaft grundlegend zu verändern, so wie es einst die Elektrizität oder die Dampfmaschine taten (vgl. Holtel, 2016; McKinsey Global Institute, 2018).

Substitutions- auf eine Augmentationsperspektive¹¹ durch die Entwicklungen der KI verstärkt wird und insbesondere die Zusammenarbeit von Mensch und System im Fokus der Betrachtung steht (André et al., 2021, S. 11; Bittner et al., 2019; Hoc, 2000; Huchler, 2016; Jarrahi, 2018, S. 583; Peissner et al., 2019, S. 11; Steil & Wrede, 2019). Seufert et al. (2021, S. 17) konstatieren, dass die damit einhergehende Steigerung der menschlichen Fähigkeiten „eine der entscheidenden Herausforderungen in der gegenwärtigen Übergangsphase der digitalen Transformation geprägt durch die Entwicklungen der KI“ ist. Durch die Veränderung von Berufsbildern, Tätigkeitsbereichen und Arbeitsinhalten sind auch die (Kompetenz-)Anforderungen an Mitarbeitende ähnlich zur ersten Welle der Digitalisierung einer gewissen Dynamik ausgesetzt (vgl. André et al., 2021, S. 9; Pfeiffer, 2020; Schmitt et al., 2021; Schumann & Lange, 2019; Seufert et al., 2021, S. 19; Winkler & Schwarz, 2021, S. 100; Teichmann et al., 2020, S. 522). Daugherty & Wilson (2018) liefern zur Konkretisierung der veränderten Kompetenzanforderungen mit den sogenannten „*Fusion Skills*“ einen konzeptionellen Ansatz. Diese Fähigkeiten sollen das produktive Potenzial der Mensch-Maschinen-Interaktion voll erschließen (Seufert et al., 2021, S. 20) und sollten daher aus der Perspektive der Berufsbildung in den Fokus genommen werden. Seufert et al. (2021, S. 20) konstatieren, dass diese Augmentationsansätze eine Verlagerung von einem „statischen“ Kompetenzverständnis zu dynamischeren Entwicklungsperspektiven begünstigen.

Die bisherigen Ausführungen machen deutlich, dass in der Vorbereitung der Schüler:innen als zukünftige Fach- und Führungskräfte auf das Leben in einer digitalisierten Welt (Michaeli et al., 2020, S. 72) ein zentrales Ziel der beruflichen Bildung liegen muss (siehe auch SWK, 2022, S. 67). Das erfordert adäquate Qualifikationsprozesse (Seifried & Ertl, 2021, S. 341). Dafür sollen im Rahmen einer nationalen Weiterbildungsstrategie die notwendigen Kompetenzen von Erwerbstätigen gefördert werden (Die Bundesregierung, 2018, S. 7). Die Bundesregierung (2018, S. 31) fordert daher explizit eine stärkere Integration von KI in die berufliche Aus- und Weiterbildung.

Es steht außer Frage, dass KI bereits jetzt einen Einfluss auf das Lehren und Lernen in verschiedenen Institutionen hat (vgl. Chiu et al., 2022; Holmes, 2019; Su et al., 2023). Insbesondere neuere KI-gestützte Technologien wie die Anwendung *Chat-GPT* (vgl. OpenAI, 2023) verstärken die Diskussion um KI-Systeme im Kontext des (beruflichen) Lehrens und Lernens (vgl. Steppuhn, 2023). Seufert (vgl. HSG, 2023) bezeichnet diese in einem Interview als sogenannte „Volks-KI“ und betont, dass es immer mehr zur Allgemeinbildung gehören werde, Systeme dieser Art kompetent nutzen zu können. Bei den dargelegten Entwicklungen und der Präsenz von KI in vielen Arbeits- und Lebensbereichen ist es verwunderlich, dass im Jahr 2022 noch immer 12 % der Probandinnen und Probanden einer repräsentativen Befragung angeben, noch nie etwas von dem Begriff der KI gehört zu haben (Overdiek & Petersen, 2022, S. 15). Ob und in welchem Umfang die restlichen 88 % tatsächlich über Wissen zu dem Fachgebiet verfügen, geht aus der Studie nicht hervor. Sie impliziert allerdings, dass die Probandinnen und Probanden die Systeme eher akzeptieren, wenn sie mehr darüber wissen

11 Mit dem Begriff der Augmentation wird in der Literatur in Anlehnung an Davenport und Kirby (2016, S. 2) die Ergänzung menschlicher und maschineller Fähigkeiten diskutiert.

(Overdiek & Petersen, 2022, S. 22). Das Wissen über KI scheint damit ein wichtiger Indikator für die Akzeptanz und damit die potenzielle Nutzung von Technologien dieser Art zu sein. Der Begriff KI ist auf Grundlage der erzielten Befunde gleichermaßen negativ (mit den Stichworten „Unsicherheit“, „Gefahr“ und „Kontrollverlust“) als auch positiv (mit den Begriffen „Fortschritt“, „Genauigkeit“ und „Effektivität“) konnotiert, wobei eine Tendenz zu eher positiven Begriffskonnotationen beobachtbar ist (Overdiek & Petersen, 2022, S. 24).

Die Relevanz von KI in der Bildung manifestiert sich u. a. durch zahlreiche Publikationen und Förderprogramme der letzten Zeit (Hasse et al., 2019; Lytras et al., 2020; Seufert & Guggemos, 2022; Yueh & Chiang, 2020). Holmes et al. (2019, S. 180) postulieren in diesem Zusammenhang:

„Whether we welcome it or not, AI is increasingly being used widely across education and learning contexts. We can either leave it to others – the computer scientists, AI engineers and big tech companies – to decide how artificial intelligence in education unfolds, [...] or whether we adopt a critical stance, to help ensure that the introduction of AI into education reaches its potential and has positive outcomes for all.“

Aus dieser Einschätzung geht deutlich die Notwendigkeit eines Forschungsfeldes rund um die KI-bezogenen Kompetenzanforderungen an (berufsbildende) Lehrkräfte hervor (Seufert & Guggemos, 2022, S. 224). Das könnte z. B. bedeuten, dass die Facette „Digitale Skills fördern“, wie sie in zahlreichen Modellierungen für digitalisierungsbezogene Kompetenzen von Lehrkräften inkludiert ist, um KI als Lerninhalt erweitert werden sollte oder unter „Mediendidaktik“ auch der Umgang mit KI-Systemen verstanden wird (S. 224). Bei der Betrachtung von KI in der Bildung ergeben sich maßgeblich zwei Fragen: was über KI gelehrt und gelernt werden soll und wie diese Lehr-Lern-Prozesse ausgestaltet sein sollen (vgl. Holmes et al., 2019). Dabei sind sowohl die Bildungsinhalte, die sich aus dem Gegenstandsbezug der beruflichen Handlungsdomäne ergeben, als auch die Methoden des Lehrens und Lernens von KI-bedingten Veränderungen betroffen (vgl. Krämer et al., 2017).

1.1.2 Künstliche Intelligenz in der beruflichen Bildung

1.1.2.1 KI als Inhalt

Arbeitsinstrumente im beruflichen Handlungsfeld werden aufgrund der Organisation der beruflichen Bildung in Deutschland zu Lerninhalten in berufsbildenden Schulen (Roppertz, 2021, S. 4). So finden digitalisierungs- und KI-bezogene Veränderungen in den beruflichen Tätigkeiten, wenn auch mit einem entsprechenden zeitlichen Verzug, auch Einzug in die schulische Aus- und Weiterbildung. Die Änderung der betrieblichen Wertschöpfungs- und Geschäftsprozesse durch die Digitalisierung (vgl. Winther, 2019) wird durch KI-basierte Technologien noch weiter vorangetrieben. In diversen Berufsbildern werden Anwendungen der KI etabliert, um z. B. Wettbewerbsvorteile zu generieren (Teichmann et al., 2020, S. 513). In den Curricula der Ausbildungsberufe in Deutschland ist das Themenfeld der KI bisher allerdings (noch) nicht präsent. Die Viel-

zahl an Publikationen und Projekten zum Thema KI in der beruflichen Bildung zeigt indes die wachsende Relevanz von KI für zahlreiche Berufsbilder (siehe z. B. Ertl & Seifried, 2021). Das verdeutlicht, dass die Curricula dementsprechend einer Revision unterzogen werden müssen.

Die in Kapitel 1.1.1 dargelegten Veränderungen von Arbeits- und Geschäftsprozessen können anhand konkreter Beispiele für verschiedene berufliche Fachrichtungen weiter spezifiziert werden. KI-gestützte Systeme finden sowohl im gewerblich-technischen Bereich (vgl. Becker et al., 2021), im kaufmännisch-verwaltenden Sektor (vgl. Wilbers, 2021) als auch im Bereich der Gesundheit und Pflege Anwendung (vgl. Bendel, 2021). Becker et al. (2021) nehmen für den Bereich der gewerblich-technischen Berufsbildung eine Differenzierung in Fertigkeitsbasierte Systeme, Wissensbasierte Systeme, Lernerorientierte Systeme und Simulationsorientierte Systeme vor (S. 35–37). In Anlehnung an Becker (2016, S. 74) nehmen die Autoren eine Modellierung von „Facharbeit an und in KI-beeinflussten (Produktions-)systemen“ (Becker et al., 2021, S. 43) vor. Die in den Tätigkeitsfeldern sichtbar werdenden Veränderungen führen damit zu neuen Aufgaben und Kompetenzanforderungen, die „nicht bloß durch die Beschreibung von Digitalisierungskompetenzen charakterisierbar sind“ (Becker et al., 2021, S. 45). Das bestätigt die Annahme, dass die Anwendung von Modellen zu digitalisierungsbezogenen Kompetenzen für die Spezifika von KI-gestützten Systemen nur bedingt hinreichend ist.

Wilbers (2021) diskutiert für die kaufmännische Aus- und Weiterbildung ein Spektrum aus einem Substitutions-, Diffusions- und Augmentationsansatz (S. 57–59). Analog zu Becker et al. (2021) für den gewerblich-technischen Bereich arbeitet Wilbers (2021) auch für den kaufmännisch-verwaltenden Bereich die „Erweiterung und Vertiefung des menschlichen Vermögens“ (S. 59) im Augmentationsansatz als besonders relevant für die Diskussion um veränderte Kompetenzanforderungen heraus (S. 70). In diesem Ansatz geht es entgegen bisheriger Annahmen nicht um die einseitige Anpassung des Menschen an die Maschine, sondern es ergibt sich ein neues Interaktionsmodell mit Wechselwirkungen zwischen den menschlichen und maschinellen Akteuren (Ertl & Seifried, 2021, S. 155).

Im Bereich der Gesundheit und Pflege wird aktuell z. B. der Einsatz von Pflegerobotern diskutiert (vgl. Bendel, 2021). Neben organisatorischen, rechtlichen und ethischen Fragestellungen, die mit dem Einsatz assoziiert sind, sind auch die notwendigen und neu zu erwerbenden Kompetenzen der Beschäftigten in dieser Domäne relevant (Bendel, 2021, S. 147–148).

Obwohl der Einsatz von KI-Systemen noch wenig verbreitet ist (einen Überblick liefern Winkler & Schwarz, 2021), stellen diese Anwendungen eine wichtige Zukunftsperspektive dar. Ertl und Seifried (2021, S. 159) postulieren daher, dass „relevante Ausbildungs- und Fortbildungsbedarfe jedoch frühzeitig erkannt und gedeckt werden“ müssen, um adäquat auf die Entwicklungen vorbereitet zu sein. Wie die Befundlage zeigt, gilt dies für alle Bereiche der Berufs- und Wirtschaftspädagogik, d. h. es sind so-

wohl die kaufmännisch-verwaltenden als auch die gewerblich-technischen Berufe als auch der Bereich Gesundheit und Pflege betroffen¹².

1.1.2.2 KI in der Lehr-Lern-Prozessgestaltung

KI-gestützte Systeme können in verschiedenen Szenarien als Lerninstrument in der beruflichen Bildung eingesetzt werden (Roppertz, 2021, S. 4; Seifried & Ertl, 2021; Wölfel, 2021). Sie bergen dabei das Potenzial, Lernprozesse zu individualisieren und zu personalisieren und damit besonders für heterogene Gruppen einen Mehrwert zu generieren (Holmes et al., 2018, S. 4; Luan et al., 2020; Luckin et al., 2016; Montebello, 2018, Pargmann & Berding, 2022, S. 220). Die Gruppe der Lernenden in der beruflichen Bildung kann bspw. im Vergleich zum allgemeinbildenden Bereich als besonders heterogen bewertet werden (vgl. Förster-Kuschel & Fürstenau, 2020). Das ist u. a. mit der Breite der beruflichen Tätigkeiten und Ausbildungswege sowie mit unterschiedlichen Zugangsvoraussetzungen zu begründen (vgl. Förster-Kuschel, 2022).

Im Kontext der Lehr-Lern-Prozessgestaltung mit Anwendungen der KI werden z. B. Game-based-Ansätze und Chatbots (Holmes, 2019, S. 11) sowie *Virtual Reality* (VR) und *Augmented Reality* (AR)-Ansätze zur Schaffung authentischer Lehr-Lern-Szenarien diskutiert (vgl. Niegemann, 2020; Pletz & Zinn, 2018; Wölfel, 2021).

Söllner et al. (2021) diskutieren Möglichkeiten der Individualisierung von Lehr-Lern-Prozessen unter Berücksichtigung von sogenannten *Hybrid-Intelligence-Ansätzen*. *Hybrid-Intelligence* meint dabei die Kombination aus menschlicher und maschineller Intelligenz, die durch Synergieeffekte bestmöglich geeignet ist, um komplexe Problemstellungen zu bewältigen (Söllner et al., 2021, S. 167). Damit wird sich im Rahmen der Diskussion um Lehr-Lern-Prozesse analog zum Diskurs um Einsatzmöglichkeiten von KI in Arbeits- und Geschäftsprozessen gegen den Ansatz „Mensch gegen Maschine“ im Rahmen einer Substitutionsperspektive gewandt und der Fokus auf die gewinnbringende Synergie von menschlichen und maschinellen Fähigkeiten im Sinne eines Augmentationsansatzes gesetzt.

Es mangelt insgesamt noch an substanziellen Ansätzen und „fundierte[n] Studien zu den Einsatzmöglichkeiten bzw. Grenzen sowie zum didaktischen Mehrwert von KI-Anwendungen in der Berufsbildung“ (Seifried & Ertl, 2021, S. 346). Diese seien zwingend notwendig, um das Lehren und Lernen in der beruflichen Bildung unter Nutzung von KI-Anwendungen zu unterstützen. Es ist zu beobachten, dass die Nutzung von KI-Technologien mit vielen Unsicherheiten behaftet ist (Steppuhn, 2023, S. 122; Söllner et al., 2021, S. 170). Insbesondere die Diskussion um Technologien wie *Chat-GPT* (*GPT-3*) äußert sich hochgradig ambivalent. Einerseits wird die Technologie als sogenannter „Meilenstein“ oder „*Gamechanger*“ betitelt (Steppuhn, 2023, S. 123), andererseits ist die Rede von einer disruptiven Wirkung auf das Bildungssystem (Wessels, 2023, S. 126). Der Umgang mit KI-gestützten Tools wie *Chat-GPT* im Bildungsbereich stellt Lehrkräfte damit vor ähnliche Herausforderungen wie es einst die flächen-

¹² Daraus ergibt sich insbesondere für den empirischen Teil der Arbeit der Anspruch, idealerweise alle drei Bereiche der Berufs- und Wirtschaftspädagogik abzudecken. Das erfordert eine standortübergreifende Datenerhebung (siehe ausführlicher Kapitel 4).

deckende Nutzung mit Suchmaschinen im Internet (Steppuhn, 2023, S. 123) oder die Nutzung von programmierbaren Taschenrechnern (Wessels, 2023, S. 126) taten. Diese beiden Beispiele gelten seitdem als essenzielle Unterstützungssysteme in Bildungsprozessen. Pilotprojekte und Studien liefern bereits erste Hinweise für positive Effekte durch die Nutzung von KI-Technologien im Unterricht auf die Lernmotivation und damit auch auf den individuellen Lernerfolg (Steppuhn, 2023, S. 124).

Ein weiteres zentrales Anwendungsfeld von KI-gestützten Technologien in der beruflichen Bildung stellen Learning Analytics dar (vgl. Berding et al., 2021; Ifenthaler & Widanapathirana, 2014; Ifenthaler & Yau, 2021; Lipp & Stock, 2022; Pargmann & Berding, 2022; Seufert et al., 2020b). Ziel von *Learning Analytics* ist es dabei nicht, die automatisierten Rückmeldungen unreflektiert auf die Schüler:innen zu übertragen, sondern primär die eigene Arbeit als Lehrkraft zu reflektieren (Pargmann & Berding, 2022, S. 224). Für die Auswertung der Befunde und die Nutzung im Rahmen der eigenen beruflichen Praxis benötigen die Lehrkräfte entsprechende Kompetenzen im Umgang mit KI (Pargmann & Berding, 2022, S. 224). Dazu zählt z. B. ein grundlegendes Verständnis über die Funktionsweisen der Technologien. Diese Forderung wird von den aktuell etablierten Modellen zu digitalisierungsbezogenen Kompetenzen bei Lehrkräften (z. B. *DigCompEdu*) nicht explizit abgedeckt.

1.1.3 Zwischenfazit und Implikationen für die Arbeit

Die Relevanz von KI-bezogenen Kompetenzen manifestiert sich einerseits mittelbar aus dem Gegenstandsbezug der beruflichen Bildung und der damit verbundenen Aufgabe, Auszubildende adäquat¹³ für eine digitalisierte Welt auszubilden, und andererseits unmittelbar aus der Bedeutung von KI-gestützten Technologien für die Gestaltung von Lehr-Lern-Prozessen und die Förderung des Kompetenzerwerbs (Roppertz, 2021, S. 2). Die (angehenden) Lehrkräfte im berufsbildenden Bereich stellen damit zentrale Akteurinnen und Akteure in der digitalen Transformation dar, die in diesem Kontext in doppelter Hinsicht gefordert sind (vgl. Meiners et al., 2022).

Obwohl die Förderung der notwendigen Kompetenzen bei (angehenden) Lehrkräften ein deklariertes Forschungsdesiderat ist (Seufert & Guggemos, 2022, S. 224; Steppuhn, 2023, S. 125), mangelt es bisher an einer grundlegenden Konzeptionalisierung der betreffenden Kompetenzfacetten. Lehrkräfte sollen das gleiche Wissen¹⁴ haben, das sie bei ihren Schülerinnen und Schülern fördern wollen (Guggemos & Seufert, 2020, S. 3). Es ist fraglich, wie Lehrkräfte dieses Niveau erreichen sollen, wenn die betreffenden Inhalte bisher nicht in ihrer Aus- und Weiterbildung implementiert sind (Pargmann & Berding, 2022, S. 222).

Da sich Anwendungen der KI maßgeblich von klassischen Computerprogrammen unterscheiden (vgl. Buxmann & Schmidt, 2021; Ertel, 2016; Taulli, 2019), ist die Anwendung von Modellen zu digitalisierungsbezogenen Kompetenzen von Lehrkräften auf den Kontext KI nur bedingt möglich. Um angemessene und adressaten-

13 Prägend sind dabei nach Reetz und Seyd (1995) die drei Relevanzprinzipien: Wissenschaftsprinzip, Situationsprinzip und Persönlichkeitsprinzip.

14 Neuweg (2014) spricht in diesem Zusammenhang vom „Wissen der Wissensvermittler“.

gerechte Aus- und Weiterbildungsangebote für das (angehende) berufliche Bildungspersonal zu entwickeln, bedarf es zunächst einer Analyse der Anforderungen an ein solches Angebot. Durch die bisher fehlende Verankerung von KI-bezogenen Inhalten in den Curricula (Pargmann & Berding, 2022, S. 222) kann davon ausgegangen werden, dass (angehende) Lehrkräfte kaum Vorkenntnisse zu KI haben. Es existieren bisher keine gesicherten empirischen Befunde über die Ausprägung des Wissens von (angehenden) Lehrkräften im berufsbildenden Bereich zu KI. Erste Studien zu diesem Themenkomplex wurden primär im allgemeinbildenden Bereich oder mit Informatiklehrkräften durchgeführt, denen aufgrund der fachlichen Nähe ein größeres Verständnis von KI-Inhalten unterstellt werden kann. Die Befunde aus den Studien lassen ein defizitäres Wissen zu KI vermuten (vgl. Lindner & Berges, 2020). Aufgrund der hergeleiteten Relevanz von KI für berufsbildende Lehrkräfte sowie der defizitären Studienlage wird in der vorliegenden Arbeit ein Grundlagenwissen fokussiert. In bisherigen Studien wurde das Wissen zu KI zudem häufig durch eine subjektive Selbsteinschätzung der Probanden erfasst. Diese Methodik sieht sich häufig mit der Kritik konfrontiert, keine validen Testwertinterpretationen (vgl. Kane, 1992, 2006a, 2006b, 2012, 2013) zuzulassen¹⁵, da „Selbsteinschätzungen häufig schwer von Selbstwirksamkeitserwartungen abzugrenzen“ sind (SWK, 2022, S. 111). Die Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (SWK) postuliert daher die Messung von digitalisierungsbezogenen Kompetenzen von (angehenden) Lehrkräften als Forschungsdesiderat, „da es bislang wenig Instrumente gibt, die performanz- oder testbasiert Kompetenzen dieser Zielgruppe erfassen“ (SWK, 2022, S. 121). Um diesem Anspruch gerecht zu werden, ist zunächst eine differenzierte und zielgruppenspezifische Analyse der Anforderungen im Zuge der KI-gestützten Digitalisierung an das (angehende) Lehrpersonal in der beruflichen Aus- und Weiterbildung notwendig.

Auf Grundlage einer theoretischen Modellierung kann anschließend ein Testinstrument¹⁶ entwickelt werden, mit dem die KI-bezogenen Kompetenzfacetten (angehender) Lehrkräfte im berufsbildenden Bereich objektiv, reliabel und valide erfasst werden können. An dieser Stelle setzt die vorliegende Dissertation zur theoretischen Modellierung und empirischen Überprüfung des Grundlagenwissens zu Künstlicher Intelligenz bei (angehenden) Lehrkräften im berufsbildenden Bereich an.

1.2 Zentrale Zielstellung und Forschungsfragen

Das übergeordnete Ziel der vorliegenden Arbeit ist es, das Konstrukt des Grundlagenwissens zu Künstlicher Intelligenz bei (angehenden) Lehrkräften im berufsbildenden Bereich theoretisch zu modellieren und empirisch erfassbar zu machen.

¹⁵ Für eine differenziertere Betrachtung der Validität von Selbsteinschätzungen kann auf Schmid et al. (2020) und Seufert et al. (2020a, S. 2–3) verwiesen werden.

¹⁶ Ausschlaggebend für diese Einordnung ist die Einschätzung nach Chomsky (1976), dass nur aufgrund von beobachtbarem Verhalten (Performanz) auf eine mögliche Kompetenz geschlossen werden kann. Auch Blömeke (2007, S. 195) sieht in der Performanz-Kompetenz-Diskrepanz eine zentrale Herausforderung für die Bildungsforschung.

In Anlehnung an Modelle zur professionellen Kompetenz von Lehrkräften (vgl. Baumert & Kunter, 2006; Brunner et al., 2006; Zlatkin-Troitschanskaia et al., 2013b), digitalisierungsbezogenen Kompetenzen (TPACK, Mishra & Koehler, 2006) und Ansätzen der Technologieakzeptanzforschung (vgl. Davis, 1989; Venkatesh & Davis, 2000) wird ein Theoriemodell entwickelt, das die relevanten KI-bezogenen kognitiven und non-kognitiven Kompetenzfacetten für (angehende) Lehrkräfte im berufsbildenden Bereich abbildet. Das Modell bildet die Grundlage für die Präzisierung des Messgegenstandes sowie die Operationalisierung der inkludierten Kompetenzfacetten. Im Anschluss wird das entwickelte Testinstrument empirisch überprüft und die erzielten Befunde in Anlehnung an die „Standards for Educational and Psychological Testing“ (vgl. AERA et al., 2014) umfassend validiert. Da für die (angehenden) Lehrkräfte im berufsbildenden Bereich ein geringes Grundlagenwissen vermutet wird, soll im Rahmen der vorliegenden Dissertation eine Pretest-Posttest-Studie durchgeführt werden, um im Rahmen einer pädagogischen Intervention Wissenszuwächse im Semesterverlauf (vgl. Ifenthaler, 2006) zu erfassen.

Das Vorgehen der Arbeit ist dabei strukturell angelehnt an die Entwicklung und Validierung eines Tests zur Erfassung des fachdidaktischen Wissens von Lehrkräften im kaufmännisch-verwaltenden Bereich (vgl. Kuhn, 2014). Neben der Validität werden in der Arbeit auch die Aspekte der Reliabilität und der Objektivität geprüft. Diese sind durch eng definierte Standards als eher kontrollierbar zu bewerten und bedürfen daher einer weniger umfassenden Betrachtung als das Kriterium der Validität (Krohne & Hock, 2007, S. 65). Letztere wird aufgrund ihrer Bedeutsamkeit für die weitere Verwendung des Testinstruments über alle Phasen der Theorieentwicklung (Kapitel 2), Operationalisierung (Kapitel 3) und empirischen Erprobung (Kapitel 4) hinweg adressiert. Die Analysen zur Einschätzung der Validität erlauben ein Urteil darüber, inwieweit die empirisch erzielten Befunde mit dem theoretisch modellierten Konstrukt erklärt werden können (siehe dazu Hartig et al., 2020).

Auf Grundlage des entwickelten Testinstrument sollen erste empirische Befunde über die Ausprägung des Grundlagenwissens zu KI sowie zu non-kognitiven Kompetenzfacetten bei angehenden Lehrkräften im berufsbildenden Bereich generiert werden. Dabei stehen insbesondere die kollektive Betrachtung der erzielten Testwerte sowie die Entwicklung im Längsschnitt über ein Semester im Fokus.

Daraus ergibt sich die für die vorliegende Arbeit übergeordnete zentrale Frage:

Wie kann das Grundlagenwissen zu Künstlicher Intelligenz von (angehenden) Lehrkräften im berufsbildenden Bereich theoretisch modelliert und empirisch erfasst werden?

Aus der übergeordneten Fragestellung lassen sich entlang der einzelnen Phasen der theoriebasierten Testentwicklung und empirisch orientierten Validierung die folgenden Leitfragen ableiten:

Leitfragen in der Phase der theoretischen Modellierung:

1. Wie lässt sich das Grundlagenwissen zu KI in Modelle der Lehrprofessionalität (1.1), der digitalisierungsbezogenen Kompetenzen (1.2) und der Technologieakzeptanzforschung (1.3) einordnen und wie können diese Ansätze für den Untersuchungsgegenstand adaptiert werden?
2. Welcher Wissensbegriff kann der Arbeit zugrunde gelegt werden (kognitionsbezogene Modellierung)?
3. Welche non-kognitiven Kompetenzfacetten müssen in der Untersuchung berücksichtigt und für den Untersuchungsgegenstand definiert werden?
4. Welche Inhaltsbereiche sind für das Grundlagenwissen von (angehenden) Lehrkräften zu Künstlicher Intelligenz relevant (inhaltliche Modellierung)?
5. Wie ist der Forschungsstand zu KI-bezogenen Kompetenzfacetten bei (angehenden) Lehrkräften?
6. Wie lässt sich der Untersuchungsgegenstand in einem theoretischen Modell abbilden?

Leitfragen in der Phase der Operationalisierung und Testentwicklung

7. Wie können die Validierungsaspekte der „Standards for Educational and Psychological Testing“ (vgl. AERA et al., 2014) auf die vorliegende Untersuchung angewendet werden?
8. In welcher Form lässt sich der Messgegenstand „Grundlagenwissen zu KI von (angehenden) Lehrkräften“ erfassen?
9. Welche Skalen können für die Erfassung der non-kognitiven Kompetenzfacetten adaptiert werden?
10. In welchem organisatorischen Rahmen kann die Testung bei der Zielgruppe durchgeführt werden?
11. Wie ist die Objektivität des Tests zu beurteilen?

Leitfragen in der empirischen Erprobung

12. Wie kann das entwickelte Instrument anhand der Kriterien der klassischen Testtheorie auf Test- (12.1), Item- (12.2) und Antwortebene (12.3) beurteilt werden?
13. Wie ist die Reliabilität des Tests zu beurteilen?
14. Wie ist die Validität der Testwerte auf Basis des quantitativen Zugangs zu beurteilen?
15. Wie ist das Grundlagenwissen zu KI in der Zielgruppe ausgeprägt?
16. Wie kann die Ausprägung der non-kognitiven Kompetenzfacetten bei der Zielgruppe bewertet werden?
17. Welche Veränderungen lassen sich in der Ausprägung des Grundlagenwissens (17.1) und in der Ausprägung der non-kognitiven (17.2) Kompetenzfacetten im Verlauf eines Semesters mit dem entwickelten Testinstrument erfassen?

Leitfragen für die kritische Würdigung und Implikationen der Arbeit

18. Welche Limitationen weist die Studie aufgrund der theoretischen (18.1) und empirischen (18.2) Herangehensweise auf?
19. Welche Implikationen lassen sich auf Basis der Befunde für die Weiterentwicklung des Modells und Instruments (19.1) sowie für die Aus- und Weiterbildung berufsbildender Lehrkräfte (19.2) ableiten?

1.3 Aufbau der Arbeit

In Kapitel 2 wird das Grundlagenwissen zu KI von (angehenden) Lehrkräften im berufsbildenden Bereich theoretisch modelliert. Zunächst erfolgt dazu die Einordnung des Untersuchungsgegenstandes in Modelle zur professionellen Kompetenz von Lehrkräften (Kapitel 2.1.1, Leitfrage 1.1), digitalisierungsbezogenen Kompetenzen von Lehrkräften (Kapitel 2.1.2, Leitfrage 1.2) sowie in etablierte Ansätze der Technologieakzeptanzforschung (Kapitel 2.1.3, Leitfrage 1.3). Die sich daraus ergebenden Implikationen für das theoretische Modell werden in einem Zwischenfazit (Kapitel 2.1.4) zusammengefasst. Aus dieser Betrachtung ergibt sich die Notwendigkeit der Berücksichtigung von kognitiven sowie non-kognitiven Kompetenzfacetten im Rahmen der vorliegenden Arbeit. Die Facetten Wissen (Kapitel 2.2.1, Leitfrage 2), Einstellungen (Kapitel 2.2.2), Überzeugungen (Kapitel 2.2.3) sowie Motivation (Kapitel 2.2.4) werden für den Untersuchungskontext definiert und die der vorliegenden Arbeit zugrunde liegenden Begriffseinordnungen in einem Zwischenfazit (Kapitel 2.2.5, Leitfrage 3) zusammengefasst. Im Anschluss erfolgt die Analyse der inhaltlichen Anforderungen an das Grundlagenwissen zu KI von (angehenden) Lehrkräften (Kapitel 2.3). In diesem Zusammenhang wird KI zunächst in das Fachgebiet der Informatik eingeordnet (Kapitel 2.3.1), bevor die inhaltliche Strukturierung anhand einer Analyse von Grundlagenwerken (Kapitel 2.3.2.2) sowie der Befunde aus Experteninterviews (Kapitel 2.3.2.3) durchgeführt wird. In einem weiteren Zwischenfazit lassen sich die für den Untersuchungskontext als relevant identifizierten Inhaltsbereiche zusammenfassen (Kapitel 2.3.3, Leitfrage 4). Auf dieser konzeptionellen inhaltstheoretischen Grundlage erfolgt die Analyse des Forschungsstandes zu KI-bezogenen Kompetenzen von (angehenden) Lehrkräften (Kapitel 2.4, Leitfrage 5). Das Kapitel 2 endet mit dem für die vorliegende Arbeit entwickelten Strukturmodell einer KI-bezogenen Kompetenz von (angehenden) Lehrkräften im berufsbildenden Bereich (Kapitel 2.5, Leitfrage 6).

Kapitel 3 dient der Operationalisierung des Untersuchungsgegenstandes auf Basis des entwickelten Strukturmodells in der Phase der theoretischen Modellierung. Dazu wird zunächst der Messgegenstand präzisiert (Kapitel 3.1.1) und anschließend werden die für die vorliegende Arbeit anwendbaren Aspekte der Validierung in Anlehnung an die „*Standards for Educational and Psychological Testing*“ der AERA et al. (2014) ausgewählt (Kapitel 3.1.2, Leitfrage 7). In Kapitel 3.2 wird die Item- und Testkonstruktion erläutert, in deren Verlauf die einzelnen Elemente des entwickelten Strukturmodells anhand von Konstruktions- und Adaptionsprozessen in Items überführt werden (Leitfragen 8 und 9). In Kapitel 3.2.3.4 werden die Rahmenbedingungen der Testung

und damit insbesondere das als Treatment in der Pre-Post-Untersuchung fungierende Lehr-Lern-Konzept beschrieben und begründet (Leitfrage 10). Anschließend erfolgt die Beurteilung der Objektivität der Testung (Kapitel 3.3, Leitfrage 11). Eine Übersicht des entwickelten Messmodells zur Erfassung der kognitiven und non-kognitiven Kompetenzfacetten wird in Kapitel 3.4 präsentiert.

In Kapitel 4 erfolgt die empirische Erprobung des Testinstruments anhand eines quantitativen Forschungsansatzes. Das Design (Kapitel 4.1) und die Datenaufbereitung (Kapitel 4.2) werden zunächst dargelegt und im Anschluss erfolgt die Beurteilung des Testinstruments anhand von Kriterien der klassischen Testtheorie (Kapitel 4.3), die erstens auf Testebene (Kapitel 4.3.2, Leitfrage 12.1), zweitens auf Itemebene (Kapitel 4.3.3, Leitfrage 12.2) und drittens auf Ebene der Antwortalternativen (Kapitel 4.3.4, Leitfrage 12.3) durchgeführt wird und in einer Einschätzung der Reliabilität der Testung mündet (Kapitel 4.3.5, Leitfrage 13). Danach erfolgt die Beurteilung der Validität anhand der Aspekte der internen Struktur (Kapitel 4.4) sowie der Beziehung des Messgegenstandes zu anderen Merkmalen (Kapitel 4.5, Leitfrage 14). Kapitel 4 endet mit einer Analyse der quantitativen Daten (Kapitel 4.6) zur Ausprägung der kognitiven (Leitfrage 15) und non-kognitiven Kompetenzfacetten (Leitfrage 16) im Querschnitt sowie einer Betrachtung der erhobenen Pre-Post-Daten im Längsschnitt (Leitfrage 17).

Die Arbeit endet mit Kapitel 5, in dem zunächst die zentralen Erkenntnisse zusammengefasst werden (Kapitel 5.1) und anschließend auf Limitationen (Leitfrage 18) und nachfolgende Forschungsdesiderata eingegangen wird (Kapitel 5.2). In Kapitel 5.3 werden abschließend die Implikationen präsentiert (Leitfrage 19), die sich aus der vorliegenden Dissertation für die theoretische (Kapitel 5.3.1) und empirische Arbeit (Kapitel 5.3.2) sowie für die berufliche Lehrkräftebildung (Kapitel 5.3.3) ergeben.

Der strukturelle Aufbau der Arbeit kann aus Abbildung 1 entnommen werden.

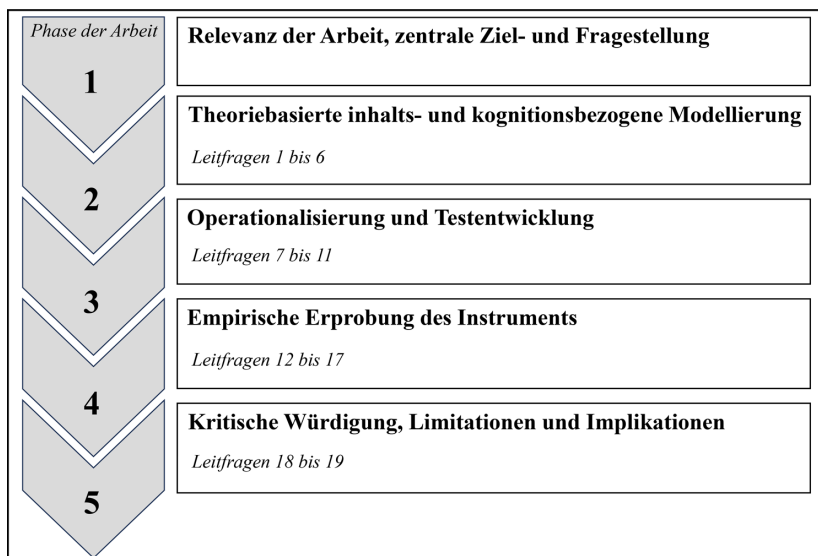


Abbildung 1: Aufbau der Arbeit

2 Theoriebasierte Modellierung des KI-Grundlagenwissens für Lehrkräfte

2.1 Verortung des Untersuchungsgegenstandes

2.1.1 Zur Professionellen Kompetenz von Lehrkräften

Das primäre Ziel von beruflicher Bildung besteht im Erwerb beruflicher Handlungskompetenz¹⁷ der Lernenden (Roppertz, 2021, S. 2), die durch von den Lehrkräften gestaltete Lehr-Lern-Prozesse gefördert werden soll (Baumert & Kunter, 2006, S. 470). Um die Frage zu beantworten, welche Fähigkeiten und Fertigkeiten Lehrkräfte benötigen, um dieser Anforderung gerecht zu werden, etablierte sich seit den 1980er-Jahren die Lehrerexpertiseforschung (Seifried, 2009, S. 33). Die Betrachtung professioneller Kompetenz von Lehrkräften manifestiert sich dabei durch die Annahme, dass diese das Lehrerhandeln und damit auch den Lernerfolg der Schüler:innen beeinflusst (Seifried & Wuttke, 2016, S. 1).¹⁸ Die Qualität der beruflichen Aus- und Weiterbildung wird demnach implizit durch die Aus- und Weiterbildung des pädagogischen Personals beeinflusst (vgl. Frommberger & Lange, 2020). Wie die erforderlichen Kompetenzen in der Aus- und Weiterbildung von Lehrkräften im berufsbildenden Bereich bestmöglich gefördert werden können, ist Gegenstand mehrperspektivischer Forschungsanliegen (Frommberger & Lange, 2020, S. 520; Seifried & Wuttke, 2015, S. 126) und spätestens seit der Reform der allgemeinen Lehrkräftebildung ein zentrales Diskussionsthema (Baumert & Kunter, 2006, S. 478). Es wird angenommen, dass die Professionalisierung von Lehrkräften einen lebenslangen und komplexen Entwicklungsprozess darstellt (siehe z. B. Berding & Lamping, 2014, S. 15¹⁹).

International haben sich dabei nach Baumert und Kunter (2006, S. 478) die Berichte des AERA Panel on Research and Teacher Education „*Studying Teacher Education*“ (vgl. Cochran-Smith & Zeichner, 2005) sowie der National Academy of Education „*Preparing Teachers for a Changing World*“ (vgl. Darling-Hammond & Bransford, 2005) als Durchbrüche erwiesen. Für den deutschsprachigen Raum identifizieren der Autor und die Autorin die Werke von Terhart (2000) zu den „Perspektiven der Lehrerbildung in Deutschland“ sowie Oser und Oelkers (2001) zur „Wirksamkeit der Lehrerbildungssysteme“ als wegweisend (Baumert & Kunter, 2006, S. 478). Oser (2001) liefert ein pädagogisch-psychologisches Kompetenzmodell, welches Kompetenzprofile und dazuge-

17 Die Kompetenzorientierung hat sich seit den 1970er-Jahren etabliert. Die seit den 1990er-Jahren geführte Diskussion um die Handlungskompetenz hat diesen Grundsatz manifestiert (vgl. Seeber & Nickolaus, 2010). Zur Differenzierung beruflicher Handlungskompetenzen als Bündel aufeinander bezogener Teilkompetenzen wird auf Rebmann et al. (2011) verwiesen.

18 Trotz der weitverbreiteten Annahme konstatieren Baumert und Kunter (2006, S. 469) „einen erheblichen Mangel an empirischer Evidenz hinsichtlich der Bedeutung professioneller Kompetenzen für die Qualität von Unterricht sowie den Lernfortschritt und die Persönlichkeitsentwicklung von Schülerinnen und Schülern“.

19 Dieser Einschätzung liegt die Annahme zugrunde, dass es sich beim Können nicht um ein Ergebnis der Erstausbildung, sondern um ein Entwicklungsziel in der zweiten Ausbildungsphase handelt (Baumert & Kunter, 2006, S. 479).

hörige Gütemaßstäbe beinhaltet. Der Autor definiert dabei die vier Kriterien *theoretische Fundierung*, *empirische Bewährung*, *Graduierbarkeit* und *praktische Relevanz* (darin eingeschlossen Lehr- und Lernbarkeit²⁰) als obligatorisch, um entwickelte Kompetenzprofile zur Definition von Ausbildungsstandards heranziehen zu können (vgl. Oser, 2001). Diesem Modell fehlt es allerdings nach Baumert und Kunter (2006, S. 479) an einem „Rahmen eines professionellen Handlungsmodells für Lehrkräfte“, weshalb die Autorinnen und Autoren den Entwurf eines heuristischen Modells anschließen.

Der Kompetenzbegriff wird im Diskurs einerseits inflationär (Seifried & Wuttke, 2015, S. 126) und andererseits inkonsistent genutzt (vgl. Zlatkin-Troitschanskaia & Kuhn, 2010). Etabliert hat sich das mehrdimensionale Begriffsverständnis (siehe dazu Zlatkin-Troitschanskaia, 2013b, S. 113) nach Weinert (2001, S. 27)²¹, in dem sowohl kognitive Leistungsdispositionen als auch motivationale, soziale und volitionale Bereitschaften modelliert werden. Diese Dimensionen bilden sich auch im Modell professioneller Kompetenz nach Baumert und Kunter (2006) ab, das auf Grundlage der Kernaussagen des „*National Board for Professional Teaching Standards*“ (NBPTS) formiert wurde. Darin subsumieren die Autorin und der Autor die in Abbildung 2 illustrierte Kombination aus Wissen und Können (deklaratives und prozedurales Wissen), Werthaltungen und Überzeugungen, motivationale Orientierungen sowie metakognitive Fähigkeiten der Selbstregulation (Baumert & Kunter, 2006, S. 481).

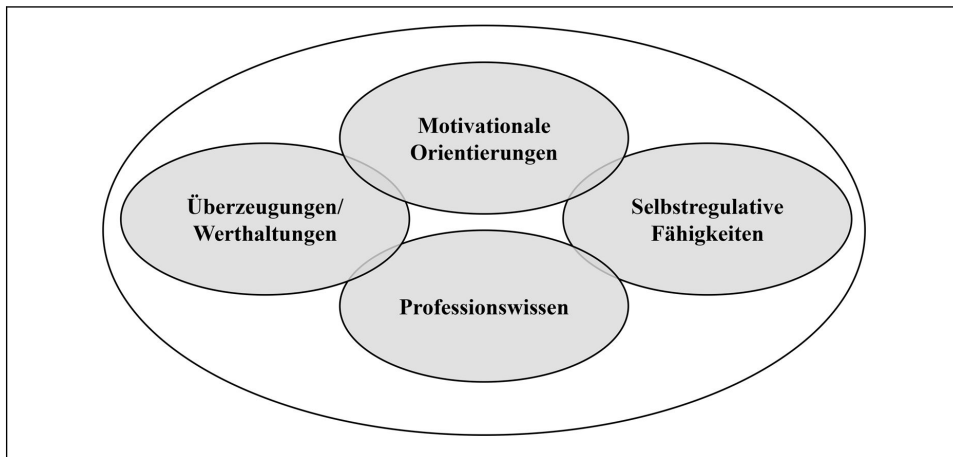


Abbildung 2: Professionelle Kompetenzen von Lehrkräften (adaptiert aus Baumert und Kunter, 2006, S. 482)

Das Professionswissen kann dabei als zentrales Element von professioneller Kompetenz postuliert werden (Baumert & Kunter, 2006, S. 481; Brunner et al., 2006, S. 59;

20 Kunter (2011, S. 535) postuliert in diesem Zusammenhang: „Leitend für die Anwendung des Kompetenzbegriffs auf die Lehrerforschung ist die Annahme, dass es sich bei den Voraussetzungen nicht um unveränderbare, allgemeine persönliche Attribute handelt, sondern um vermittelbare, speziell im Rahmen der beruflichen Ausbildung und Karriere erworbene und vertiefte Kenntnisse, Fertigkeiten, Einstellungen und Orientierungen“.

21 Dieses etablierte Kompetenzverständnis sieht sich allerdings mit der Kritik konfrontiert, es sei inhaltsleer und kontextunspezifisch (Hartig et al., 2008; Klieme & Hartig, 2007). Diese Kontextspezifität wird durch Klieme & Leutner (2006) ergänzt, wodurch eine Abgrenzung der kontextspezifischen Kompetenz als kognitive Leistungsdisposition zur kontextunspezifischen Intelligenz erreicht wird (siehe auch McClelland, 1973, S. 1).

Seufert & Guggemos, 2022, S. 216; Zlatkin-Troitschanskaia et al., 2013b, S. 113). Das Wissen stellt dabei eine „erlern- und veränderbare“ (Kuhn, 2014, S. 16) Facette dar und bietet sich deshalb aus forschungspragmatischer Sicht (u. a. aufgrund der inhaltlichen Eingrenzbarkeit und der Messbarkeit) für eine empirische Studie an. Dabei ist die Dimension des Professionswissens nicht eindimensional zu betrachten, sondern kann z. B. in Anlehnung an die von Shulman (1986, 1987) eingeführte und von Bromme (1997) expandierte Differenzierung des Lehrerberufswissens in die Subdimensionen Pädagogisches Wissen (*pedagogical knowledge*, PK), Fachwissen (*content knowledge*, CK) und fachdidaktisches Wissen (*pedagogical content knowledge*, PCK) binnendifferenziert werden (siehe z. B. Baumert & Kunter, 2006, S. 479, 482; Brunner et al., 2006, S. 59; Shulmann, 1986, 1987). Diese Wissensbereiche lassen sich in weitere Wissensfacetten untergliedern (Baumert & Kunter, 2006, S. 482). Die konzeptionelle Ordnung der Wissensbereiche und -facetten kann aus Abbildung 3 entnommen werden.

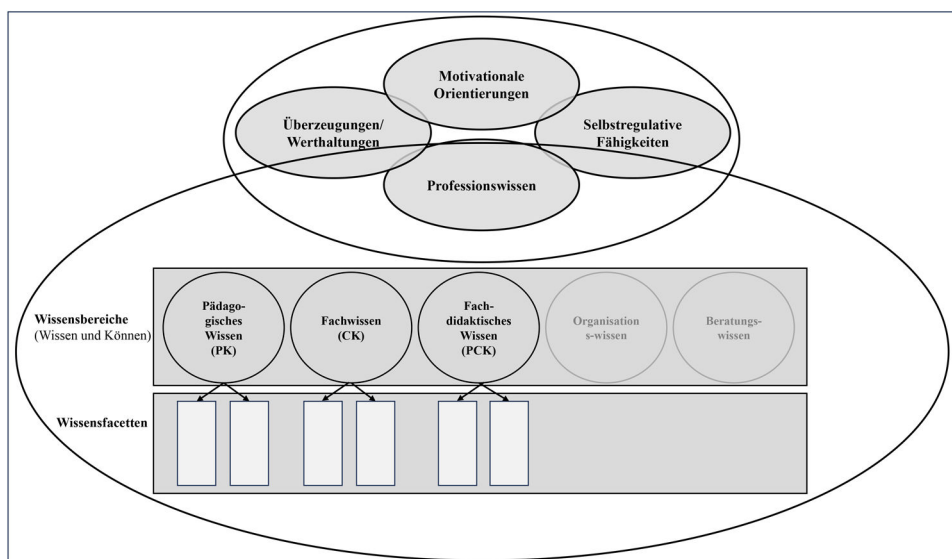


Abbildung 3: Binnendifferenzierung des Professionswissens. Adaptiert aus Baumert und Kunter (2006, S. 482)

Während es in den konzeptionellen Ansätzen gelingt, ein umfassendes Verständnis von verschiedenen Konstituenten in einem Kompetenzmodell zu vereinen, birgt diese Modellierung für empirische Analysen Herausforderungen. Weitgehend ungeklärt bleiben die Beziehungen sowie potenzielle reziproke Effekte zwischen den einzelnen Facetten (Seifried & Wuttke, 2015, S. 126). Die mangelnde Evidenz wird u. a. damit begründet, dass die Effekte nur schwer empirisch überprüfbar sind (Seifried & Wuttke, 2016, S. 1; Zlatkin-Troitschanskaia & Seidel, 2011, S. 226). Das führt in der empirischen (Berufs-)Bildungsforschung häufig dazu, dass die in einem Kompetenzmodell modellierten Facetten einzeln konzeptualisiert und operationalisiert werden (vgl. z. B. Kunter

et al., 2011). Die separate Modellierung und Erfassung wird dabei fortlaufend empfohlen (Klieme & Hartig, 2007, S. 25; Kuhn, 2014, S. 16).

Baumert und Kunter (2006, S. 470) liefern insgesamt ein „allgemeine[s] Modell zur Handlungskompetenz von Lehrkräften [...], das es erlaubt, empirische Befunde zur Qualifikation, professionellen Kompetenz und Persönlichkeit von Lehrpersonen in ihrer Bedeutung für Unterricht und Lernen zu ordnen und theoriebezogen zu diskutieren“. Auf dieser Grundlage haben sich zahlreiche domänenspezifische Ansätze entwickelt. Für angehende Lehrkräfte im berufsbildenden Bereich (die Zielgruppe der vorliegenden Studie) in der universitären Ausbildung wurde ein Modell zur Fachkompetenz in der Domäne Wirtschaftswissenschaften konstruiert (vgl. Zlatkin-Troitschanskaia et al., 2013b). Das Projekt ILLEV²² fokussiert dabei „die Modellierung und valide Erfassung der Ausprägung und Entwicklung der fachbezogenen Kompetenz in der Domäne Wirtschaftswissenschaften sowie den systematischen Vergleich von Studierenden in unterschiedlichen wirtschaftswissenschaftlichen Studiengängen mit oder ohne Lehramtsoption“ (Zlatkin-Troitschanskaia et al., 2013b, S. 70). Die Modellierung im Projekt ILLEV erfolgt auf Grundlage der Befunde aus der COACTIV Studie (vgl. Brunner et al., 2006), die eine analoge Modellierung professioneller Kompetenz für Mathematiklehrkräfte beinhaltet. Das Ergebnis ist ein domänenspezifisches und empirisch erprobtes (vgl. Zlatkin-Troitschanskaia et al., 2013b; Schmidt et al., 2015) Rahmenmodell (Abbildung 4), das als Grundlage für die kontextualisierte Modellierung in der vorliegenden Studie fungiert.

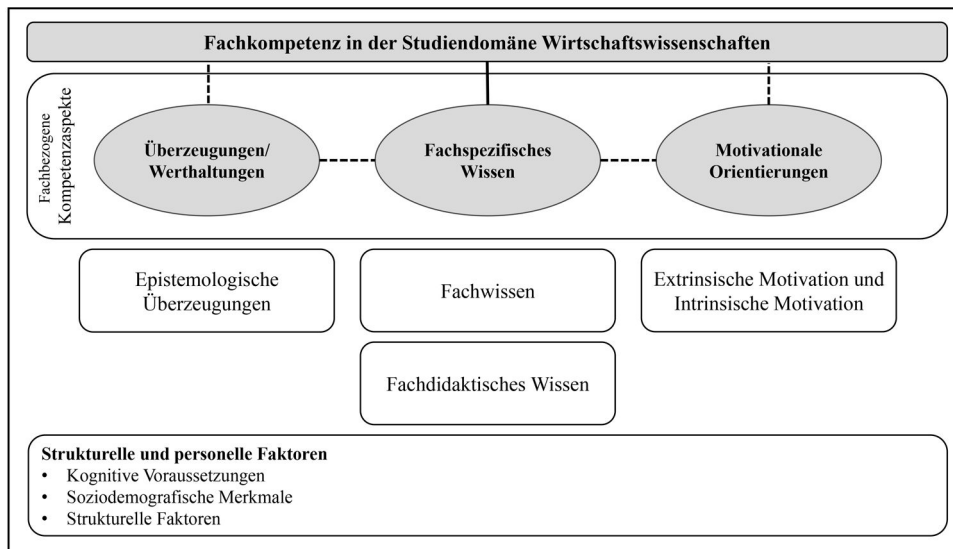


Abbildung 4: Fachkompetenz in der Domäne Wirtschaftswissenschaften. Adaptiert aus Zlatkin-Troitschanskaia et al. (2013b, S. 73)

22 ILLEV steht für „Innovativer Lehr-Lernortverbund in der akademischen Hochschulausbildung“.

Da im Rahmen dieser Modelle bisher keine technologiebezogenen Aspekte betrachtet werden, schließt sich die Analyse von Modellen zu digitalisierungsbezogenen Kompetenzen an, um sich dem Untersuchungsgegenstand anzunähern.

2.1.2 Digitalisierungsbezogene Kompetenzen von Lehrkräften

In Kapitel 1 wurde bereits die Bedeutung der Digitalisierung für den berufsbildenden Bereich dargelegt. Die Relevanz manifestiert sich dabei zusammengefasst einerseits insbesondere durch die digitalisierungsbezogenen Veränderungen in Arbeits- und Geschäftsprozessen, die aufgrund des Gegenstandsbezugs die inhaltliche Ebene der beruflichen Aus- und Weiterbildung maßgeblich beeinflussen (vgl. Gerholz et al., 2022a, 2022b), und andererseits durch die Potenziale digitaler Technologien für die Ausgestaltung von Lehr-Lern-Prozessen (vgl. Meiners et al., 2022). Daraus ergibt sich, dass die digitale Transformation für die berufliche Bildung sowohl eine fachwissenschaftliche als auch eine fachdidaktische Aufgabe darstellt (vgl. Kamsker & Riebenbauer, 2022). Lehrpersonen nehmen eine entscheidende Rolle bei der Integration von Technologien in den Unterricht ein und sollen gleichermaßen digitale Kompetenzen bei den Lernenden fördern (Meiners et al., 2022, S. 124; Seufert & Guggemos, 2022, S. 214). Digitale Kompetenzen sind damit ein „zentraler Bezugspunkt der digitalen Transformation beruflicher Bildung“ (Wilbers, 2019, S. 11). Der Aktionsrat Bildung postuliert die digitale Kompetenz als „vierte Kulturtechnik“ und fordert nachdrücklich deren systematische Förderung (VBW, 2018, S. 20).²³

Um die Nachfrage nach sogenannten „*digital professionals*“ (Botturi, 2019, S. 147) zu bedienen und zukünftige Fach- und Führungskräfte in Anlehnung an die *Future Work Skills* (vgl. Institute for the Future (IFF), 2011; Sczogiel et al., 2019) und die *21st Century Skills* (vgl. Partnership for 21st Century Skills (P21), 2016), in deren Kern digitale Kompetenzen verankert sind, auszubilden, sind die Akteure in der beruflichen Bildung gefragt. Durch die digitale Transformation verändern sich nicht nur Arbeits- und Geschäftsprozesse (siehe Kapitel 1.1), sondern damit einhergehend auch Kompetenzanforderungen an (zukünftiges) Fachpersonal (z. B. Schlottmann et al., 2021, S. 1). Seit 2001 hat sich der Begriff der „*digital natives*“ etabliert (vgl. Kirschner & Bruyckere, 2017; Prensky, 2001, Schulmeister, 2012; Tóth et al., 2022). Darunter werden jene Personen verstanden, die mit aktuellen Technologien aufgewachsen sind und denen daher weitreichende Kompetenzen im Umgang damit zugesprochen werden. Im aktuellen Diskurs handelt es sich dabei sowohl um die Generation der Lernenden als auch die Generation der angehenden Lehrkräfte. Neuere Forschungsergebnisse revidieren diese Einschätzung und legen nahe, dass diese Kompetenzen nicht *per se* vorhanden sind, sondern erst erworben werden müssen und der Begriff der *digital natives* somit nicht tragfähig ist (Redecker, 2017, S. 12; Stock et al., 2022, S. 177). Trotz der wahrgenommenen Relevanz konstatieren Kamsker und Riebenbauer (2022, S. 29) ein defizitäres Aus- und Weiterbildungsangebot für (angehende) Lehrkräfte. Die systematische Förderung digitalisierungsbezogener Kompetenzen von Wirtschaftspädagoginnen und -pädago-

23 Die digitalen Kompetenzen sollen demnach mit der gleichen Priorität gefördert werden wie das Schreiben, Lesen und Rechnen (siehe auch Kerres, 2017, S. 86).

gen als angehende Lehrkräfte im berufsbildenden Bereich erweist sich aktuell als schwierig, da für diesen Zweck zunächst curriculare Grundlagen geschaffen werden müssen (Schlottmann et al., 2021, S. 5). Die Forderung nach systematischen Aus- und Weiterbildungsangeboten verstärkt sich aus den Ergebnissen empirischer Studien, die Hinweise dafür liefern, dass die Lehrkräfte sich dieser Aufgabe nur bedingt gewachsen sehen (Seufert & Guggemos, 2022, S. 221).

Dabei werden die als relevant erachteten Kompetenzfacetten im Diskurs u. a. unter Begriffen wie „Digitale Mündigkeit“ (Hoffmann et al., 2019, S. 83), „(digitale) Medienkompetenz“ (vgl. Blömeke, 2003, 2004; Meiners et al., 2022; KMK, 2017), „Digitale Kompetenzen“ (vgl. Seufert & Guggemos, 2022; Stock et al., 2022), *Digital Literacy* (vgl. Gerholz et al., 2022a; Lipp & Stock, 2022; Martin, 2006; Schlottmann et al., 2021; Schlottmann & Gerholz, 2022) oder *ICT-Literacy* (SWK, 2022, S. 15) subsumiert. In diesem Zusammenhang unterscheiden sich die konzeptionellen Ansätze u. a. auch darin, welche Facetten mit dem jeweiligen Kompetenzbegriff abgedeckt werden. Darüber hinaus finden sich auch bei gleicher Begriffsverwendung unterschiedliche konzeptionelle Rahmenmodelle. Das legt die Vermutung nahe, dass die mit digitalisierungsbezogenen Kompetenzen assoziierten Begriffe analog zum Kompetenzbegriff²⁴ nicht konsistent verwendet werden.

Digitalisierungsbezogene Kompetenzen werden in der internationalen Literatur hauptsächlich unter den Stichworten *Digital Literacy* (DL) und *Digital Competence* (DC) diskutiert (vgl. Sánchez-Caballé et al., 2020; Spante et al., 2018). Dabei wird der Frage nachgegangen, welches Wissen und Können Mitglieder der Gesellschaft haben sollen und wie dieser Kompetenzerwerb unterstützt werden kann (Spante et al., 2018, S. 2). Eine Untersuchung von englischsprachigen Quellen zwischen 1997 und 2017 zeigt, dass die Begriffe DC und DL überwiegend synonym oder zumindest ohne konkrete Abgrenzung voneinander genutzt werden (Spante et al., 2018, S. 2). Die Autorinnen und Autoren zeigen dabei auf, dass vermehrt der Begriff *Digital Literacy* genutzt wird. Im Kontext der Lehrkräftebildung ist allerdings eine Bevorzugung des Begriffs *Digital Competence* sichtbar. Auffällig ist dabei, dass in 47 % der analysierten Publikationen keine Definition der Konzepte vorgenommen wird (Spante et al., 2018, S. 5). Es zeigt sich außerdem, dass der Begriff *Digital Literacy* stärker wissenschaftlich fundiert ist als der Begriff der *Digital Competence* (Spante et al., 2018, S. 6). Für die Definition von *Digital Literacy* wird dabei häufig auf den Ansatz des „European Framework for Digital Literacy“ (*DigEuLit*) rekurriert (vgl. Martin, 2006).²⁵ Diesem Konzept mangelt es jedoch an der geforderten Kontextualisierung des Literacy-Begriffs (Spante et al., 2018, S. 7).

²⁴ Zur inkonsistenten Verwendung des Kompetenzbegriffs siehe z. B. Zlatkin-Troitschanskaia & Kuhn, 2010.

²⁵ Digital Literacy wird mit Bezug zu Martin (2006, S. 155) definiert als „das Bewusstsein, die Einstellung und die Fähigkeit eines einzelnen, digitale Werkzeuge und Ausstattung angemessen zu nutzen, um neue Wissensstrukturen zu identifizieren, darauf zuzugreifen, zu verwalten, zu integrieren, zu bewerten, zu analysieren und zu synthetisieren, mediale Ausdrucksformen in bestimmten Lebensbereichen zu schaffen und mit anderen zu interagieren“ definiert. Diese Definition verdeutlicht, dass das Verständnis von Literalität weit über die kognitive Facette hinausgeht. Stock et al. (2022, S. 172) subsumieren unter dem Begriff dagegen die Komponenten „Wissen, Fertigkeiten und Fähigkeiten“ und klammern die non-kognitiven Facetten aus. Digital Literacy ist nach diesem Verständnis die Grundlage für digitale Kompetenz.

Im internationalen Kontext hat sich außerdem das „*European Framework for Digital Competence*“ (*DigComp*) (vgl. Carretero et al., 2017) als wegweisend etabliert, da es Qualifikationen und Kompetenzen europaweit transparent und verständlich darstellbar macht (Seufert & Guggemos, 2022, S. 216). Mit diesem Framework (aktuell *DigComp*, Version 2.1) wurde ein Strukturmodell (weiter-)entwickelt, das die notwendigen Kompetenzen für alle Mitglieder der Gesellschaft umfassend beschreibt (vgl. Carretero et al., 2017). Das Modell wurde insbesondere für Schüler:innen und Lehrkräfte weiter spezifiziert. Obwohl die Ergebnisse aus empirischen Studien vermuten lassen, dass einige Lehrkräfte digitale Kompetenzen als nicht notwendig für ihre berufliche Praxis ansehen (Seufert et al., 2020a, S. 4), sind diese zwingend erforderlich, um digitalisierungsbezogene Kompetenzen bei den Schülerinnen und Schülern zu fördern (Sánchez-Caballé et al., 2020, S. 64). Für die differenzierte Beschreibung von digitalen Kompetenzen von Lehrkräften wurde das „*European Framework for the Digital Competence of Educators*“ (*DigCompEdu*) erarbeitet (vgl. Redecker, 2017). Das Rahmenmodell umfasst die Bereiche der professionellen Kompetenzen der Lehrkräfte, pädagogische Kompetenzen der Lehrkräfte sowie Kompetenzen der Schüler:innen und zeigt deren Verknüpfungen auf (Redecker, 2017, S. 8). Insgesamt werden somit 22 Kompetenzen in sechs Bereichen im Modell abgebildet (Redecker, 2017, S. 9). Das Framework bietet nicht nur einen konzeptionellen Rahmen, sondern auch ein Entwicklungsmodell, mit dem Lehrkräfte ihre digitalen Kompetenzen erfassen und auf Grundlage dessen weiterentwickeln können (Redecker, 2017, S. 9). Dabei sind zur Einordnung die Stufen A1 (*Newcomer*) bis C2 (*Pioneer*) vorgesehen (Redecker, 2017, S. 9). Kamsker und Riebenbauer (2022, S. 25) postulieren in Anlehnung an das *DigCompEdu-Framework* (vgl. Redecker, 2017) die notwendigen Kompetenzen als „allgemeine und fachspezifische digitale Kompetenzen, um neben der Bereitschaft auch Wissen, Fähigkeiten und Fertigkeiten zur Nutzung digitaler Technologien in der (generellen und berufsspezifischen) Kommunikation, Kollaboration und Problemlösung mitzubringen“.

Die von Seufert und Guggemos (2022, S. 215) synonyme Verwendung von „digitalen Kompetenzen“ und „professionelle Kompetenzen von Lehrpersonen im Kontext des digitalen Wandels“ liefert überdies einen weiteren Hinweis dafür, dass die Modelle der Professionellen Kompetenz von Lehrkräften (Kapitel 2.1.1) für die digitale Transformation kontextualisiert werden und demnach konzeptionelle Similaritäten zu erwarten sind. Diese Verzahnung manifestiert sich z. B. im *TPACK-Modell*, in dem die kognitiven Facetten im etablierten Verständnis professioneller Kompetenz nach Baumert und Kunter (2006, siehe Abbildung 3 in Kapitel 2.1.1) um technologiebezogene Aspekte erweitert werden (vgl. Mishra & Koehler, 2006). Das Modell hat sich zu einem der populärsten Modelle im Kontext der technologiebezogenen (Berufs-)Bildungsforschung etabliert (vgl. Davies & West, 2014; Guggemos & Seufert, 2020; Hew et al., 2019; Meiners et al., 2022, S. 128; Niederhauser & Lindström, 2018; Petko, 2020).

Das *TPACK-Modell* bildet eine Expansion der nach Shulman (1986, 1987) etablierten Differenzierung des Wissens in die Subdimensionen *PK*, *CK* und *PCK* um technologiebezogene Facetten ab (Abbildung 5) und wird als Erklärungsansatz für die Integration von Technologien in den Unterricht durch Lehrkräfte herangezogen (Gugge-

mos & Seufert, 2020, S. 2). Im Einzelnen handelt es sich dabei nach Mishra und Koehler (2006, S. 1026) um die Bereiche des *technology knowledge* (TK), *technological content knowledge* (TCK), *technological pedagogical knowledge* (TPK) sowie die Verknüpfung der Facetten im Zentrum des Modells als *technological pedagogical content knowledge* (TPCK).

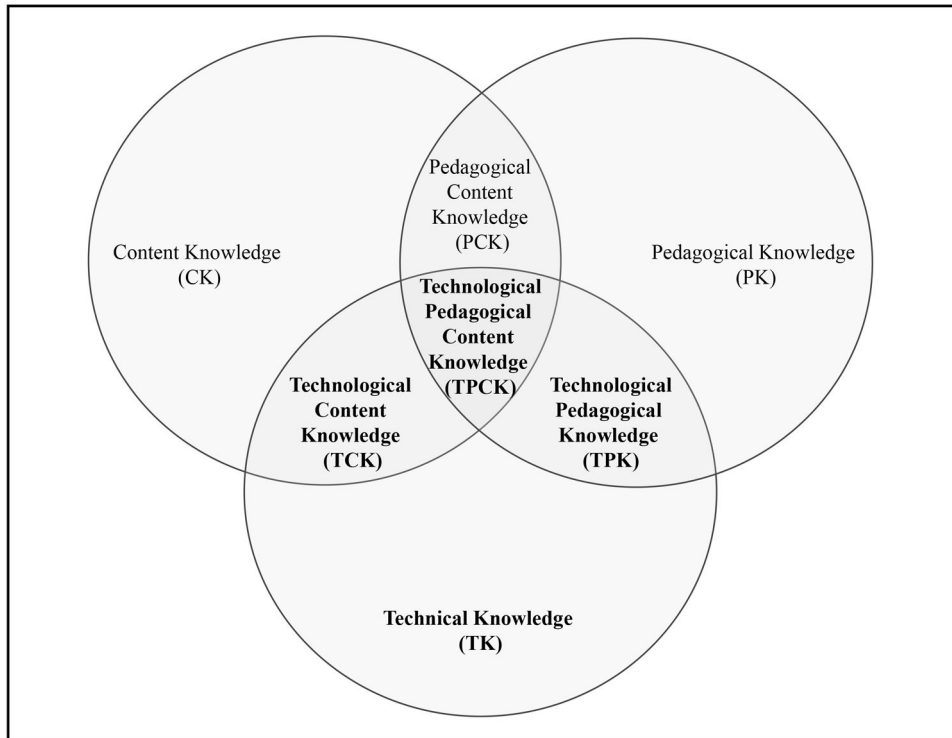


Abbildung 5: Das TPACK-Modell. Adaptiert aus Mishra und Koehler (2006, S. 1025)

Trotz oder gerade wegen seiner weiten Verbreitung ist das Modell nicht frei von Kritik (siehe dazu z. B. Graham, 2011, 2012). Die Beurteilung wird häufig mit der wenig trennscharfen definitorischen Abgrenzung der einzelnen Facetten und deren Operationalisierung begründet (Petko, 2020, o. S.; Seufert et al., 2020, S. 2). Darüber hinaus fehlt es an empirischer Evidenz, die den Modellcharakter bekräftigt.²⁶ Die einzelnen Facetten müssen demnach für den jeweiligen Untersuchungsgegenstand spezifiziert werden.²⁷ Guggemos und Seufert (2020, S. 3) expandieren das Modell darüber hinaus

²⁶ Empirische Studien liefern Hinweise dafür, dass die Facetten bei angehenden Lehrkräften eher transformativ zu betrachten sind (vgl. Schmid et al., 2020) und bei praktizierenden Lehrkräften einen eher integrativen Charakter aufweisen (vgl. Chai et al., 2016).

²⁷ Für die Facette des TCK liefern Schlottmann et al. (2021) ein für den kaufmännisch-verwaltenden Bereich kontextualisiertes Modell der *Business Digital Literacy*.

mit Bezug zum *DigComp*-Framework aufgrund seiner praktischen Relevanz um eine Facette zur digitalen Kollaboration (*Technical Collaboration Knowledge, TCok*).

Mishra und Koehler (2006, S. 1027) definieren das *TK* dabei als das Wissen über Technologien und subsumieren darunter neben Büchern und Tafeln auch das Internet und Videotechnologien. *TK* umfasst dabei das Wissen über die Anwendung sowie die Anwendung selbst. Die Autoren weisen darauf hin, dass das Verständnis von *TK* dem technologischen Wandel unterliegt und demnach dynamisch angepasst werden müsse (Mishra & Koehler, 2006, S. 1028). Für die vorliegende Untersuchung bedeutet das, dass das Wissen über und die Anwendung von KI-gestützten Systemen in den Bereich des *TK* kontextualisiert werden kann. Dieser Facette kann darüber hinaus das von Seufert und Guggemos (2022, S. 222) formulierte Forschungsdesiderat zugeordnet werden, in dem die Frage aufgeworfen wird, welches Informatikbasiswissen (angehende) Lehrkräfte benötigen. Damit ist gleichzeitig die Frage assoziiert, wie die Aus- und Weiterbildung adäquat gestaltet werden kann.

Die Facette *TCK* umfasst laut Mishra und Koehler (2006, S. 1028) das Wissen über die wechselseitige Beziehung zwischen Inhalt und Technologie. Dies beinhaltet z. B. das Wissen darüber, inwieweit sich der inhaltliche Bezugsgegenstand durch die Anwendung von Technologien ändert. In Bezug zum Gegenstand der vorliegenden Arbeit kann diese Facette mit der dynamischen und durch technologische Entwicklungen initiierten Veränderung von Arbeits- und Geschäftsprozessen spezifiziert werden (siehe ausführlicher dazu Kapitel 1.1.1).

Die adressatengerechte Auswahl von Technologien und (digitalen) Systemen für die Gestaltung von Unterricht wird im Bereich des *TPK* abgebildet (Mishra & Koehler, 2006, S. 1028). An der zentralen Schnittstelle des Modells manifestiert sich damit die von den Lehrkräften eingenommene Doppelrolle, die das Lehren mit (als Lernmedium) und über (als Lerngegenstand) digitale Technologien plausibilisiert (vgl. Guggemos & Seufert, 2020).

Zusammenfassend werden bei der Vielzahl der konzeptionellen Ansätze zu digitalisierungsbezogenen Kompetenzen meist analog zu den Modellen der professionellen Kompetenz (siehe Kapitel 2.1.1) mehrere kognitive und non-kognitive Kompetenzfacetten modelliert, bei denen das Wissen und Können die Kernfacetten bilden. Für die Kontextualisierung der Facetten wird dabei häufig das *TPACK*-Modell verwendet (vgl. Geiser, 2022; Seufert & Guggemos, 2022). In der vorliegenden Studie wird die Orientierung an den technologiebasierten Facetten des *TPACK*-Modells zur Konkretisierung der kognitiven Facette einer KI-bezogenen Kompetenz als zweckmäßig erachtet.

2.1.3 Technologieakzeptanz

Die Bereitschaft zur Auseinandersetzung mit digitalen Technologien und deren Anwendung ist eng verbunden mit deren Akzeptanz. Die Akzeptanz wird dabei als die positive Nutzungsentscheidung durch den Nutzenden verstanden (Nistor et al., 2012, S. 345) und stellt einen zentralen Erfolgsfaktor bei der Einführung neuer Systeme oder Technologien dar (Huchler, 2016, S. 119; Kollmann, 1998, S. 37). Die Akzeptanz hat sich

damit zu einem Schlüsselbegriff in Entwicklungs-, Implementations- und Evaluationsprozessen etabliert (vgl. Bürg & Mandl, 2004; Kollmann, 1998, 2000).

Modelle der Technologieakzeptanz wurden in der empirischen (Berufs-)Bildungsforschung bisher häufig adaptiert, wenn es um die Nutzung(-sbereitschaft) von E-Learning-Ressourcen geht (vgl. Bürg & Mandl, 2004; Klauser, 2006). Die Akzeptanz kann dabei „als Bedingung, Indikator und Kriterium für pädagogische Effektivität und ökonomische Effizienz von Lernangeboten“ gesehen werden oder auch als „Ausgangs- und Bezugspunkt für Investitionsentscheidungen [...], für die Planung und Evaluation von Bildungsmaßnahmen sowie für die Konstruktion oder den Erwerb von Lernangeboten“ (Klauser, 2006, S. 111) deklariert werden. Im Fokus steht dabei stets die Frage nach der Ausprägung von Akzeptanz, da in jedem Fall eine hohe Akzeptanz angestrebt wird (Klauser, 2006, S. 111).

In der Akzeptanzforschung hat sich die Definition von Akzeptanz als zweidimensionales Konstrukt etabliert (Bürg & Mandl, 2004, S. 5). Müller-Böling und Müller (1986, S. 26–27) unterscheiden hierbei die (nicht direkt beobachtbare) Einstellungsakzeptanz und die (beobachtbare) Verhaltensakzeptanz. Dementsprechend werden unter dem Begriff sowohl die Einstellungen zum Verhalten als auch das Verhalten selbst subsumiert. Die Einstellungsakzeptanz beinhaltet kognitive, affektive und konative Komponenten (Klauser, 2006, S. 114). Während frühe Studien keine empirisch gesicherten Erkenntnisse über einen direkten Zusammenhang zwischen Einstellungs- und Verhaltensebene generieren konnten (z. B. Rosch & Frey, 1994; Bürg & Mandl, 2004), ist dieser direkte Zusammenhang mittlerweile hinreichend mit Evidenz unterlegt (vgl. Pletz & Zinn, 2018). Essenziell ist auf Basis der Befunde, dass für eine Analyse, die belastbare Befunde generieren soll, sowohl die Untersuchung der Einstellungen als auch die Untersuchung des Verhaltens zwingend die gleiche Abstraktions- bzw. Konkretisierungsstufe aufweisen müssen (Bürg & Mandl, 2004, S. 8; Klauser, 2006, S. 114).

Etabliert haben sich in diesem Kontext sog. Drittvariablen-Ansätze (Bürg & Mandl, 2004, S. 8). In diesen wird der Zusammenhang zwischen Einstellungen und Verhalten durch Persönlichkeits- und Situationsvariablen moderiert (vgl. Rosch & Frey, 1994). Ein etabliertes Drittvariablenmodell stellt die Theorie der überlegten Handlung dar (*Theory of Reasoned Action* (TRA), vgl. Fishbein & Ajzen, 1975). In diesem Modell wird ein direkter, gerichteter Zusammenhang zwischen Nutzungsintention (Verhaltensintention) und Nutzung (Verhalten) angenommen. Die Nutzungsintention wird dabei von den Variablen „Einstellung zum Verhalten“ und „Subjektive Norm“ beeinflusst, zwischen denen reziproke Effekte angenommen werden (siehe Abbildung 6). Neben der Theorie des geplanten Verhaltens (*Theory of Planned Behavior* (TPB), vgl. Ajzen, 1985, 1991) stellt die TRA eine fundamentale Grundlage für etablierte Konzepte der Technologieakzeptanzforschung dar.

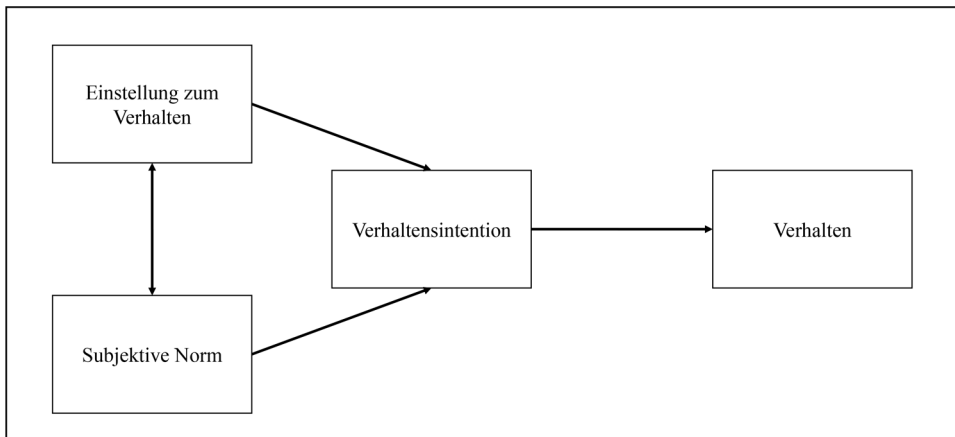


Abbildung 6: Theorie der überlegten Handlung (Adaptiert nach Fishbein und Ajzen [1975])

Zahlreiche Modelle der Technologieakzeptanz, wie das *Technology Acceptance Model* (TAM) (vgl. Davis, 1989), knüpfen an diesen Forschungsergebnissen an und differenzieren die Einflussfaktoren auf die Nutzungsintention weiter aus. Das TAM (Davis, 1989) kann dabei als international anerkanntes Modell angesehen werden (SWK, 2022, S. 17). Im TAM werden die beiden Konstrukte *Perceived Usefulness* (Wahrgenommener Nutzen) und *Perceived Ease of Use* (Wahrgenommene Einfachheit der Benutzung) als direkte Einflussfaktoren auf die Nutzungsbereitschaft einer Technologie operationalisiert und empirisch erprobt (Davis, 1989, S. 324). Das TAM gilt als Grundlage für weitere Technologieakzeptanzmodelle, in denen die (direkten und indirekten) Einflussfaktoren auf die Nutzungsbereitschaft und die tatsächliche Nutzung weiter ausdifferenziert wurden (vgl. Venkatesh & Davis, 2000; Venkatesh & Bala, 2008; Venkatesh et al., 2003). Venkatesh und Davis (2000) stellen im Rahmen der Expansion des Modells eine signifikante Interrelation zwischen dem wahrgenommenen Nutzen und der Einschätzung über die Relevanz für die eigene berufliche Praxis fest. Aus dieser Weiterentwicklung heraus hat sich die *Unified Theory of Acceptance and Use of Technology* (UTAUT) (vgl. Venkatesh et al., 2003) im aktuellen Diskurs etabliert und bildet die theoretische Basis für empirische Studien zum Technologieeinsatz im (Berufs-)Bildungsbereich (vgl. z. B. Bach, 2016; Guggemos et al., 2020).

Insbesondere vor dem Hintergrund der Entwicklung von KI stellt die Akzeptanz eine wichtige Bezugsgröße dar (vgl. Klimczak et al., 2019). Scheuer (2020) entwickelte auf Basis der Besonderheiten von Anwendungen der KI ein kontextualisiertes Akzeptanzmodell auf Grundlage der UTAUT, das auf der Annahme beruht, Nutzer:innen nehmen die KI humanoid wahr. Entsprechend findet das Modell Anwendung, wenn es um die Interaktion mit einem konkreten System (z. B. einem Roboter) geht.

Zentral für die Einschätzung der Akzeptanz haben sich demnach explizit kognitive und soziale Einflussgrößen etabliert. Je nach Untersuchungskontext können auch kontextspezifische Faktoren und Rahmenbedingungen in die Modellierung inkludiert werden. Grundlegend besteht in den Modellen auf Basis des TAM demnach die An-

nahme, dass die Verhaltensintention (und damit das konkrete Verhalten) durch die zwei Faktoren Wahrgenommener Nutzen und Wahrgenommene Einfachheit der Benutzung direkt beeinflusst wird (siehe Abbildung 7). Auf Grundlage der Ergebnisse statistischer Analysen werden diese maßgeblich durch die Faktoren *Job Relevance* und *Subjective Norm* beeinflusst (Venkatesh & Davis, 2000, S. 197).

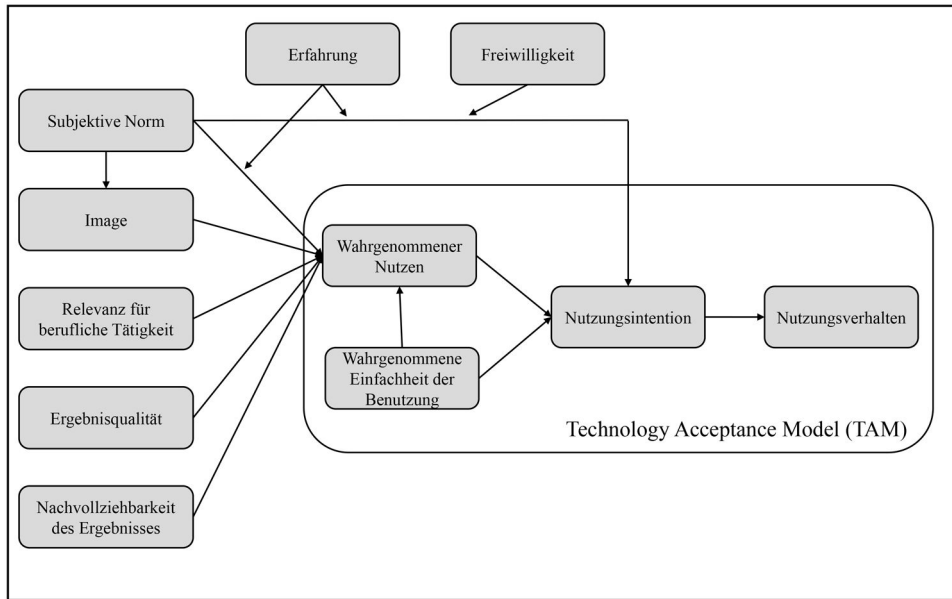


Abbildung 7: Technology Acceptance Model 2 (TAM2). Adaptiert aus Venkatesh und Davis (2000, S. 188)

Die subjektive Norm wird definiert als „die Vorstellung, die die handelnde Person von den Erwartungen anderer relevanter Personen hinsichtlich des fraglichen Verhaltens hat und die Bereitschaft der handelnden Person, diesen Erwartungen zu folgen“ (Bürg & Mandl, 2004, S. 8). Die subjektive Norm lässt sich somit als externer Effekt kategorisieren, welcher auf Grundlage empirischer Prüfung einen direkten Einfluss auf die Handlungsabsicht hat (vgl. Ajzen & Madden, 1986).²⁸ Die *Job Relevance* bildet indes die Wahrnehmung darüber ab, ob das System relevant für den eigenen Aufgabenbereich ist und als Unterstützungssystem fungieren kann (Venkatesh & Davis, 2000, S. 191).

Kritisiert wird der Ansatz insbesondere hinsichtlich der Problematik, dass ein Individuum durchaus positiv eingestellt sein kann, sich aber aus anderen Gründen (z. B. kognitiv) nicht in der Lage sieht, das entsprechende Verhalten auszuüben (vgl. Frey et al., 1993). Daraus lässt sich ableiten, dass einer Modellierung die kognitive Facette im Sinne eines Wissens über die Technologie hinzugefügt werden sollte.

²⁸ Weitere Studienergebnisse (vgl. Venkatesh & Davis, 2000) lassen vermuten, dass dieser Effekt von der Variable „Freiwilligkeit der Nutzung“ moderiert wird.

Klauser (2006, S. 124) kategorisiert im Kontext der Akzeptanz von E-Learning-Anwendungen die Einflussfaktoren auf die individuelle Ausprägung von Akzeptanz in personenbezogene und kontextbezogene Faktoren. Zu den personenbezogenen Merkmalen zählen in diesem Zusammenhang kognitive (z. B. wahrgenommener Nutzen, Relevanz für die beruflichen Aufgaben, Kenntnisse, Fähigkeiten), soziale (z. B. subjektive Norm) sowie affektive (z. B. Motivation) Dispositionen. Die kontextbezogenen Merkmale können unterteilt werden in technische (z. B. einfache Bedienbarkeit, Erlernbarkeit) sowie organisationale (z. B. Freiwilligkeit der Nutzung) Rahmenbedingungen (siehe Abbildung 8). Insbesondere die personenbezogenen Faktoren, die Klauser (2006, S. 124) modelliert, bieten einen Anknüpfungspunkt für die im Rahmen der vorliegenden Dissertation intendierte Modellierung.

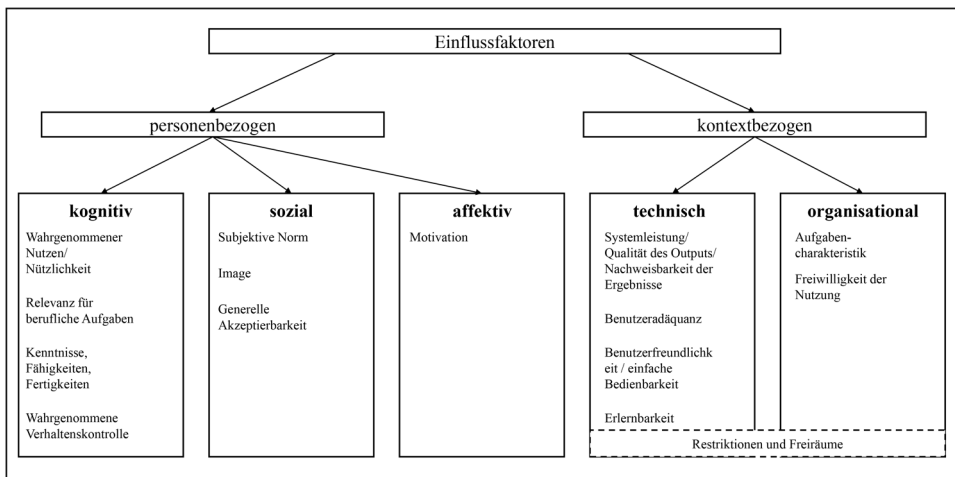


Abbildung 8: Einflussfaktoren auf die Akzeptanz. Übernommen aus Klauser (2006, S. 124)

Im Rahmen der (beruflichen) Bildung bilden die Modelle der Technologieakzeptanzforschung überwiegend die theoretische Grundlage für empirische Studien (vgl. Antonietti et al., 2022; Ifenthaler & Schweinbenz, 2013, 2016; Pletz & Zinn, 2018; Salleh et al., 2021; Scherer et al., 2018; Schmitt et al., 2021; Stobbe, 2021; Wagner, 2016). Die Modelle suggerieren einen gerichteten Zusammenhang zwischen Intention und tatsächlicher Nutzung, der in zahlreichen Regressionsstudien mit empirischer Evidenz unterlegt werden konnte (Dwivedi et al., 2011, S. 164). Diese Einschätzung liefert einen Hinweis dafür, dass bei positiv ausgeprägter Einstellung zur Nutzung von einer tatsächlichen Nutzung als Konsequenz ausgegangen werden kann.

2.1.4 Zwischenfazit und Implikationen für die Untersuchung

Die Diskussion der Bereiche Lehrprofessionalität, digitalisierungsbezogene Kompetenzen sowie Technologieakzeptanz verdeutlicht, dass die kognitiven Dispositionen einen fundamentalen Bestandteil von Kompetenzmodellierungen darstellen, allerdings allein nicht hinreichend sind. Daher wird die kognitive Disposition des Grund-

lagenwissens in der vorliegenden Arbeit um non-kognitive Kompetenzfacetten erweitert, um ein strukturelles Rahmenmodell für KI-bezogene Kompetenz für (angehende) Lehrkräfte im berufsbildenden Bereich abzubilden. In Anlehnung an die diskutierten Modelle werden dabei insbesondere die Facetten (Fach-)Wissen, Einstellungen, Überzeugungen und Motivation in die Modellierung aufgenommen. Das Wissen wird dabei in Anlehnung an das TPACK-Modell in die Facette des TK eingeordnet. Für die Kontextualisierung der Einstellungen erweisen sich die Modelle der Technologieakzeptanzforschung als zweckmäßig, sodass die Einstellungen als Einstellungsakzeptanz in Bezug auf KI modelliert werden können.

Die Perspektive des fachdidaktischen Wissens, die sowohl in Modellen der professionellen Kompetenz als auch in Modellen der digitalisierungsbezogenen Kompetenz inkludiert ist, wird für den Untersuchungskontext zunächst nicht weiter berücksichtigt. Das liegt unter anderem darin begründet, dass die Curricula der kaufmännischen Ausbildungsberufe (Stand 01/2023) bisher keine konkreten Anknüpfungspunkte zu KI aufweisen. Daher wird das Unterrichten von KI-Inhalten zunächst als nachrangig betrachtet und bietet Anknüpfungspunkte für anschließende Forschungsvorhaben.²⁹ Die Dimension der Selbstregulation wird aus testökonomischen Gründen ebenfalls ausgeklammert.

Die kontextbezogenen Einflussfaktoren auf die Akzeptanz von Technologien werden vor dem Hintergrund der Zielsetzung der vorliegenden Arbeit nicht weiter berücksichtigt. Da im Rahmen der Untersuchung kein konkreter Technologieeinsatz erfolgte, besteht keine Basis, auf der die Faktoren erfasst und analysiert werden können. Es werden daher die personenbezogenen Merkmale fokussiert.

Die Kombination von kognitiven und non-kognitiven Kompetenzfacetten wird durch die Einschätzung der SWK (2022, S. 17) gestützt, die unter anderem Überzeugungen und motivationale Dispositionen als zentral für den Einsatz digitaler Medien erachtet (siehe auch Petko 2012a, 2012b). Durch die Einordnung der Untersuchung in die diskutierten Modelle wird durch die theoretische Grundlage der vorliegenden Modellierung der Forderung der SWK Rechnung getragen, die „ein breites Verständnis analog zu generischen Modellen professioneller Kompetenzen (Baumert & Kunter, 2006), [fordert] um professionelle Kompetenzen im Kontext der Digitalisierung modellieren zu können“ (SWK, 2022, S. 120).

Im weiteren Verlauf werden daher nun die kognitiven und non-kognitiven Facetten weiter ausdifferenziert und für den Untersuchungskontext kontextualisiert.

29 Es existieren zahlreiche Studien zur Beziehung zwischen Fachwissen und fachdidaktischem Wissen (einen Überblick liefern Seifried & Wuttke, 2015), die als Anknüpfungspunkt für dementsprechende Modellierungen genutzt werden können. Für die Domäne der Wirtschaftswissenschaften kann auf Kuhn (2014), Kuhn et al. (2014) und Zlatkin-Troitschanskaia et al. (2013a) verwiesen werden.

2.2 Kognitive und non-kognitive Kompetenzfacetten

2.2.1 Wissen

Aus der Betrachtung der Modelle der professionellen Kompetenz (Kapitel 2.1.1) und der digitalisierungsbezogenen Kompetenzen (Kapitel 2.1.2) geht hervor, dass der Facette des Wissens eine fundamentale Bedeutung zugesprochen wird. Das fachspezifische Wissen von Lehrkräften nimmt nachweislich Einfluss auf die Gestaltung von Lehr-Lern-Prozessen (Lindner & Berges, 2020, o. S.). Trotzdem ist zu beobachten, dass der Wissensbegriff in der Fachliteratur häufig negativ konnotiert wird (Happ, 2017, S. 31).³⁰ Es scheint außerdem verwunderlich, dass ein derart zentrales Konstrukt wie das Wissen häufig nicht hinreichend präzise definiert wird (Minnameier, 2005, S. 137; Wuttke, 2005, S. 30–31).

Die Zuweisung einer fundamentalen Rolle (Happ, 2017, S. 32; Minnameier, 2005, S. 139) lässt sich z. B. mit Verweis auf Anderson und Krathwohl (2001) illustrieren, die das Wissen als Grundlage für alle höheren kognitiven Leistungen (z. B. Anwenden, Analysieren) und damit als unverzichtbar bestimmen. Wissen beschreibt die Gesamtheit aller Kenntnisse (Krohne & Hock, 2007, S. 425) und beinhaltet durch kognitive Prozesse verarbeitete Informationen (Schröder, 2001, S. 286). Happ (2017, S. 32) kommt in seiner Arbeit zum volkswirtschaftlichen Grundlagenwissen zum Schluss, dass das Wissen aus einer statischen und einer dynamischen Perspektive betrachtet werden kann (siehe dazu auch Aebli 1978, 1980, 1981; Tergan, 1989; Weber, 1994) und einen immanenten Wahrheitsanspruch innehat (siehe dazu Minnameier, 2005, S. 139, 2013, S. 146).³¹ Minnameier (2005, S. 139) postuliert in Anlehnung an Brendel (1999) drei für den erziehungswissenschaftlichen Kontext anwendbare Bedingungen, mithilfe derer Wissen von einer Meinung abgegrenzt werden kann:

1. Das Individuum ist der Meinung, dass die These stimmt,
2. die These muss wahr sein und
3. das Individuum ist davon überzeugt und hat Gründe dafür, dass die These stimmt.

Auf dieser Grundlage hat sich das Verständnis von Wissen als „wahre, gerechtfertigte Meinung“ (Happ, 2017, S. 34; Minnameier, 2005, S. 140) etabliert. Darüber hinaus lassen sich in diesem Zusammenhang die Bereiche „explizites Wissen“ und „implizites Wissen“ unterscheiden (Minnameier, 2010, S. 65). Das explizite Wissen umfasst nach Minnameier (2010, S. 66) nicht nur einzelne Fakten, sondern „auch kausale oder andere Zusammenhänge oder abstrakte wissenschaftliche Theorien“.

30 Happ (2017, S. 31) bezieht sich dabei zur Illustration auf die Einordnung des Wissens in die unterste Stufe der Taxonomie nach Bloom (1976) und die Einschätzung nach Weinert (1986, S. 102), dass es sich dabei um „geistig nicht verarbeitete, mehr oder minder unverstandene Anhäufungen beliebiger Informationen, die durch verbale Belehrung vermittelt und passiv-mechanisch gelernt“ wurden, handelt.

31 Diese Einschränkung gilt als Kritik an der Einschätzung nach Aebli (1978, S. 9), nach der „Wissen [...] die Basis des Handelns, des Handelns und des Sprechens [ist] und in ihm [...] die Erfahrungen niedergelegt [sind], die der Mensch aus diesen Tätigkeiten gewinnt“. An dieser Stelle kann ergänzend auch auf Minnameier (2005, S. 140) verwiesen werden.

Für die kognitionsbezogene Binnendifferenzierung des Wissensbegriffs existieren überdies verschiedene Ansätze (z. B. Anderson, 1983; Arbinger, 1997; Beck, 1995; Dubs, 1995). Als Orientierungsrahmen kann dabei die Differenzierung in die Bereiche „bereichsbezogenes Wissen, strategisches Wissen und metakognitives Wissen“ (Arbinger, 1997, S. 17) herangezogen werden. Ersteres kann dabei weiter ausdifferenziert werden in die Bereiche „deklaratives Wissen“ (Arbinger, 1997, S. 17) und „prozedurales Wissen“ (Arbinger, 1997, S. 20). Deklaratives Wissen³² beinhaltet nach diesem Begriffsverständnis das Wissen über Fakten in Form von Propositionen, die sich eindeutig als richtig oder falsch klassifizieren lassen (Winther, 2010, S. 54). Die Repräsentation von deklarativem Wissen wird zumeist über Schemata realisiert (Arbinger, 1997, S. 19; Happ, 2017, S. 38; Wuttke, 2005, S. 43). Das prozedurale Wissen wird in der Literatur häufig als „Wissen, wie... oder Handlungswissen“ (Arbinger, 1997, S. 20) charakterisiert und beinhaltet das Wissen über Verfahrensmuster (Happ, 2017, S. 38). Die Repräsentation wird in diesem Fall über Produktionssysteme realisiert (Arbinger, 1997, S. 21; Happ, 2017, S. 38; Wuttke, 2005, S. 43). Das deklarative Wissen folgt nach diesem Verständnis dem „Alles-oder-nichts-Prinzip“, während beim prozeduralen Wissen im Sinne einer graduellen Abstufung das „Mehr-oder-weniger-Prinzip“ Anwendung findet (Arbinger, 1997, S. 22). Der Erwerb des jeweiligen Wissens erfolgt aufgrund seiner Beschaffenheit und Charakteristika verschieden. Während deklaratives Wissen in der Theorie über eine Mitteilung erfolgen kann, bedarf es für den Erwerb prozeduralen Wissens in der Regel Übung (Arbinger, 1997, S. 22). Trotz der zentralen Unterschiede sind die Wissensarten nicht als Antagonisten zu verstehen (Arbinger, 1997, S. 23). Dieser Gedanke findet sich auch in der theoretischen Modellierung nach Beck (1995), in der das deklarative Wissen in die Subfacetten deklarativ-statisch und deklarativ-prozedural binnendifferenziert wird. Nach Dubs (1995, S. 166) handelt es sich dabei um Faktenwissen (deklarativ-statisch) und Strukturwissen (deklarativ-prozedural).³³ Im Bereich des Strukturwissens sind Zusammenhänge aus einer prozeduralen und konditionalen Perspektive bekannt (Wuttke, 2005, S. 35). Das Fehlen dieser Zusammenhänge im Bereich des deklarativ-statischen Wissens ist einer der Gründe dafür, dass dieses als Lernergebnis als nicht hinreichend betrachtet werden kann (Wuttke, 2005, S. 36). Sobald hingegen kohärente Verknüpfungen zwischen den Wissensarten hergestellt und begründet sind, kann von einem „Verstehen“ gesprochen werden (Dubs, 1995, S. 166; Minnameier, 2000, S. 147), das eine interne Verarbeitung und aktive Konstruktion der Strukturen beinhaltet (Happ, 2017, S. 39). Die vorliegende Arbeit folgt dem differenzierten Verständnis von Faktenwissen (deklarativ-statisch) und Strukturwissen (deklarativ-prozedural).

Eine weitere Differenzierung kann über den Aspekt der Domänenspezifität erfolgen (Happ, 2017, S. 40). Dabei wird unter domänenspezifischem Wissen³⁴ „die Sachkenntnis einer Domäne“ (Wuttke, 2005, S. 31) verstanden. Für den allgemeinbildenden

32 Zur Binnendifferenzierung des deklarativen Wissens in semantisches und episodisches Wissen sei auf Tulving (1972) verwiesen.

33 Einen Überblick bietet Wuttke (2005, S. 35).

34 Zur Abgrenzung von domänenspezifischem und domänenübergreifendem Wissen sei auf Renkl (2005) sowie Winther & Achtenhagen (2008) verwiesen.

den Bereich kann zur Binnendifferenzierung auf Arbeiten von Blömeke et al. (2009, 2010, 2011) verwiesen werden, in denen das Wissen des Fachbereichs der Mathematik in die fünf Inhaltsbereiche Arithmetik, Algebra, Funktionen, Geometrie und Statistik unterteilt werden kann. Im berufsbildenden Bereich nimmt Happ (2017) eine Systematisierung für die Volkswirtschaftslehre (VWL) vor. Für den Fachbereich der VWL konkludiert Happ (2017, S. 40) darunter die Kenntnis verschiedener Inhaltsbereiche aus der VWL sowie „das Wissen über bestimmte volkswirtschaftliche Prinzipien“. Vor dem Hintergrund des Gegenstandsbezugs der vorliegenden Arbeit kann als Domäne die Informatik gefasst werden, in der KI ein konkreter Inhaltsbereich ist, der sich in weitere Subbereiche ausdifferenzieren lässt (siehe ausführlicher Kapitel 2.3).

Zusammengefasst wird in der vorliegenden Arbeit als kognitive Leistungskomponente ein domänenspezifisches, deklarativ-statistisches sowie deklarativ-prozedurales Wissen zu KI fokussiert.

2.2.2 Einstellungen

Insbesondere aus der Betrachtung der Modelle aus der Technologieakzeptanzforschung (Kapitel 2.1.3) geht hervor, dass die Einstellungen als erste von zwei Ebenen der Akzeptanz³⁵ modelliert wird. Einstellungen bilden ein Kernkonzept der Sozialpsychologie ab (vgl. Bohner, 2002). Ihnen wird dabei eine verhaltenssteuernde Funktion zugesprochen (vgl. Ajzen, 1991), die dazu führt, dass Individuen ihr Denken und Verhalten anpassen (Bohner, 2002, S. 266, 300). Einstellungen werden definiert als „zusammenfassende Bewertung eines Gegenstands“ (Bohner, 2002, S. 267), die auf Grundlage einer „psychischen Tendenz“ (Eagly & Chaiken, 1998, S. 269) vorgenommen wird. Der fokussierte Gegenstand wird in diesem Zusammenhang als sog. Einstellungsgegenstand definiert, der konkreter oder abstrakter Natur sein kann und bezogen auf seine Beschaffenheit keine Einschränkungen erfährt (Bohner, 2002, S. 267). Das Objekt wird dabei in Anlehnung an Eagly und Chaiken (1998, S. 269) „mit einem gewissen Grad an Zuneigung oder Abneigung bewertet“. Diese Tendenz ist dabei nicht direkt beobachtbar (Bohner, 2002, S. 267) und bildet somit ein latentes Merkmal³⁶ ab.

Es wird angenommen, dass Einstellungen zum Teil auf Erfahrungen beruhen und sich in verschiedenen Formen beobachtbaren Verhaltens äußern (Bohner, 2002, S. 268). Die Einstellungen werden der Annahme nach dreidimensional von kognitiven, affektiven und verhaltensbezogenen Prozessen beeinflusst und initiieren andererseits Reaktionen auf den respektiven Ebenen (Bohner, 2002, S. 268; Siegfried & Wuttke, 2021, S. 124). Klauser (2006, S. 114) konstatiert dazu, dass das Wissen über das Einstellungsobjekt sowie die individuellen Zielsetzungen dabei einen wesentlichen Einfluss auf die Handlungstendenzen nehmen. Die Definition nach Müller-Böling und Müller (1986, S. 26), die Einstellung ist „eine relativ dauerhafte kognitive und affektive Wahrnehmungsorientierung“, lässt vermuten, dass es sich dabei um ein tendenziell stabiles Charakteristikum handelt. Befunde aus Untersuchungen liefern allerdings Hinweise

³⁵ Siehe zur Begriffsklärung Kapitel 2.1.3.

³⁶ Manifeste Merkmale lassen sich direkt beobachten und messen (z. B. über Likert-Skalen). Latente Merkmale hingegen sind nicht direkt beobachtbar und können allenfalls über die Beobachtung und Messung manifester Variablen ermittelt werden (Blasius & Baur, 2022, S. 737).

dafür, dass Einstellungen durch verschiedene Formen von Konditionierung, Persuasionsprozesse oder durch Anreize induzierte Verhaltensänderungen beeinflussbar sind (einen Überblick liefert Bohner, 2002).

Ein Zusammenhang zwischen Einstellungen und Verhalten konnte in Studien auf Grundlage der in Kapitel 2.1.3 diskutierten Modelle mit empirischer Evidenz unterlegt werden (z. B. Ajzen & Fishbein, 1977; Venkatesh & Davis, 2000). Eine positiv gerichtete Einstellung führt demnach eher zu der Ausübung einer aktiven Handlung (z. B. Nutzung eines Systems oder einer Technologie) als eine negativ gerichtete Einstellung (Klauser, 2006, S. 121). In Anlehnung an die Auseinandersetzung mit Modellen der Technologieakzeptanzforschung gilt auch in diesem Zusammenhang die Voraussetzung, dass die Einstellungs- und Verhaltensebene die gleiche Abstraktionsebene innehaben.³⁷ Aus diesem Grundsatz ergeben sich Implikationen für die Operationalisierung der Einstellungen für die vorliegende Untersuchung (siehe ausführlicher Kapitel 3.2.3.2).

2.2.3 Überzeugungen

Überzeugungen werden innerhalb der Dimension der Werthaltungen im Kontext der professionellen Kompetenz von Lehrkräften diskutiert (vgl. Baumert & Kunter, 2006; Berding & Lamping, 2014; Seifried, 2006, 2010). Es gilt als erwiesen, dass Überzeugungen als zentrale non-kognitive Kompetenzfacette einen bedeutsamen Einfluss auf das Lehrerhandeln nehmen (vgl. Krauss, 2011). Das latente Konstrukt der Überzeugungen wird in der Informatik, Psychologie und Pädagogik seit der zweiten Hälfte des 20. Jahrhunderts rege diskutiert (einen Überblick liefert Seifried, 2010). In der Berufs- und Wirtschaftspädagogik kann der Diskurs im Gegensatz dazu noch als vergleichsweise neu bewertet werden (Seifried, 2006, S. 112). Geiser (2022, S. 55) arbeitet heraus, dass in der deutschsprachigen berufs- und wirtschaftspädagogischen Forschung hauptsächlich epistemologische Überzeugungen (z. B. Berding & Lamping, 2014; Berding, 2016; Rebmann & Slopinski, 2018) Gegenstand der Untersuchungen sind, während sich im internationalen Raum analog zu Untersuchungen der allgemeinen Pädagogik und den Erziehungswissenschaften ein eher ausgewogenes Bild zeigt, in dem Zusammenhänge epistemologischer, unterrichtlicher und personenbezogener Überzeugungen zu pädagogischer Professionalität und unterrichtlichem Handeln gleichermaßen untersucht werden. Das Ergebnis eines systematischen Literaturreviews zeigt, dass in 53,8% der Fälle epistemologische Überzeugungen in den 52 analysierten Publikationen der berufs- und wirtschaftspädagogischen Forschung fokussiert werden (Geiser, 2022, S. 56). Der Autor weist jedoch auch darauf hin, dass eine trennscharfe begriffliche Abgrenzung der Konzepte nicht immer zweifelsfrei möglich ist. Insgesamt ist zu beobachten, dass der Begriff der Überzeugungen nicht immer konsistent genutzt wird (vgl. Geiser, 2022; Reusser et al., 2011; Schmechtig et al., 2019; Schmidt et al., 2015; Sembill & Seifried, 2009). Seifried (2006, S. 110) stellt u. a. eine teilweise synonyme Ver-

37 Neben dem sog. Korrespondenzprinzip werden das Aggregationsprinzip und die theoretische Korrespondenz als maßgeblich für empirische Untersuchungen formuliert. Dabei geht es stets um die Untersuchung der Einstellungs- und der Verhaltensebene (Bohner, 2002, S. 301–302). Da in der vorliegenden Untersuchung kein konkretes Verhalten der Probandinnen und Probanden untersucht wird, werden die weiteren Prinzipien in der Studie nicht weiter adressiert.

wendung mit Begriffen wie Einstellungen³⁸ oder subjektiven Theorien³⁹ heraus. Geiser (2022, S. 57) bestätigt, dass Überzeugungen als schwer zu definieren gelten. Insbesondere die kritische begriffliche Abgrenzung zum Begriff der Einstellungen ist in dieser Arbeit von hoher Relevanz und wird über die Theoriebildung hinaus auch in der Phase der Operationalisierung berücksichtigt, um disjunkte Kategorien für die empirische Untersuchung zu schaffen. Geiser (2022, S. 57–58) adressiert außerdem mit Bezug zu Baumert und Kunter (2006)⁴⁰ die enge Verknüpfung der Überzeugungen zur Facette des Wissens.

Die vorliegende Arbeit folgt dem verbreiteten Verständnis von Überzeugungen als epistemologische Überzeugungen (EÜ) zu Wissen und Wissenserwerb. Die Eingrenzung manifestiert sich aus der Annahme, dass epistemologische Überzeugungen einen Zusammenhang mit anderen Kompetenzfacetten, wie z. B. dem Fachwissen, aufweisen (z. B. Schmidt et al., 2015, S. 110). Im internationalen Diskurs werden dafür Begriffe wie „*personal epistemology*“ (vgl. Hofer, 2010), „*epistemological beliefs*“ (vgl. Schommer, 1990) und weitere verwendet (einen Überblick liefert Berding, 2016, S. 83). Die Anfänge der Forschung zu epistemologischen Überzeugungen sind auf erste Arbeiten von Perry in den 1950er-Jahren zurückzuführen (vgl. Perry, 1970). Hofer und Pintrich (1997) definieren epistemologische Überzeugungen anhand von vier Dimensionen, von denen zwei auf das Wissen (Struktur und Sicherheit) und zwei auf den Wissenserwerb (Quelle und Rechtfertigung) rekurrieren. Epistemologische Überzeugungen werden häufig als mehrdimensionales Modell dargestellt (vgl. Schommer, 1990; Schommer-Aikins, 2004), das auf das von Schommer (1990) aufgrund widersprüchlicher empirischer Ergebnisse entwickelte Überzeugungssystem zurückzuführen ist. Die Modellierung erfolgt mehrdimensional, um der Komplexität des Sachverhalts gerecht zu werden (Schommer, 1990, S. 498). Demnach wird unter epistemologischen Überzeugungen ein System von persönlichen Haltungen zu Wissen und Lernen (bzw. Wissenserwerb) verstanden (vgl. Schommer, 1990), sodass sich dieses System auch in der Modellierung von Hofer und Pintrich (1997) abbilden lässt. Der Aspekt des Wissens wird nach Schommer-Aikins (2004, S. 20) in die folgenden Subdimensionen ausdifferenziert:

- (1) Struktur des Wissens
- (2) Stabilität des Wissens
- (3) Quelle des Wissens

Ein zentraler Bestandteil in den Modellen ist dabei die Stabilität/Sicherheit des Wissens (Schmidt et al., 2015, S. 110). Epistemologische Überzeugungen lassen sich in diesem Zusammenhang nach Schommer (1990) auf einem Kontinuum zwischen „naiven“ und „kritischen“ Überzeugungen einordnen. Für den Aspekt der Sicherheit

38 Siehe dazu Kapitel 2.2.2.

39 Subjektive Theorien werden im Rahmen der vorliegenden Arbeit nicht weiter betrachtet. Für eine Auseinandersetzung und Analyse subjektiver Theorien im Bezug zu Qualifikations- und Kompetenzanforderungen siehe z. B. Moschner (2014). Für die Einordnung epistemologischer Überzeugungen in subjektive Theorien sei auf Berding & Lamping (2014) verwiesen. Der zentrale Unterschied liege darin, „dass subjektive Theorien eine Aussage über die innere Struktur des Systems treffen“ (Berding, 2016, S. 83).

40 Siehe dazu Kapitel 2.1.1.

des Wissens lassen sich diese beiden Extrempunkte als absoluten, allgültigen Wissensbestand auf der einen Seite und einen relativen Wissensbestand, der einer gewissen Dynamik unterliegt, auf der anderen Seite konkretisieren. Die epistemologischen Überzeugungen einer Person gelten als veränderbar (vgl. Perry, 1970), wenngleich sie häufig als recht stabiles Merkmal charakterisiert werden (Kuhn, 2014, S. 16).⁴¹ Es hat sich ein Verständnis etabliert, in dem epistemologische Überzeugungen einen domänenspezifischen Charakter aufweisen (vgl. Berding, 2016, S. 85; Buehl & Alexander, 2001; Hofer, 2006; Schmidt et al., 2015), weshalb sie mit Bezug zum Untersuchungsgegenstand zu modellieren sind.

2.2.4 Motivation

Ein weiteres Kernelement professioneller Kompetenz (Kapitel 2.1.1) wird durch motivationale Orientierungen abgebildet (z. B. Kunter, 2011, S. 534). Es gilt als empirisch gesichert, dass Motivation eine individuelle Determinante des erworbenen Kompetenzniveaus ist (Helmke, 2012, S. 71). Mit der Motivation wird eine gerichtete Steuerung des Verhaltens assoziiert (vgl. Deci, 1992; Deci & Ryan, 1985, 1991, 1993; Deci et al., 1991; Ryan & Deci, 2017). Auf eine konkrete Handlung bezogen bedeutet das, dass ein Individuum einen bestimmten Zustand als Ziel fokussiert und das Verhalten zum Zweck der Erreichung dieses Ziels anpasst.⁴² Dabei kann es sich sowohl um einen Wunsch als auch um eine Absicht handeln (Schiefele & Köller, 2010, S. 336). Ob diese Perspektive kurz- oder langfristig intendiert ist, ist dabei für den Grundzusammenhang zunächst unerheblich (Deci & Ryan, 1993, S. 224). Motivationale Dispositionen können dabei von Selbstwirksamkeitserwartungen (vgl. Bandura, 1977) beeinflusst sein (Deci & Ryan, 1993, S. 224).

In der pädagogischen Psychologie wird das Konzept der Motivation meist mit Bezug zu Lernhandlungen plausibilisiert (vgl. Deci & Ryan, 1993, S. 233; Schiefele et al., 2002; Schiefele & Köller, 2010). In der (Berufs-)Bildungsforschung hat sich dabei insbesondere das Verständnis von Deci und Ryan (1993) etabliert. Die Theorie fußt auf der Annahme, dass „der Mensch die angeborene motivationale Tendenz hat, sich anderen Personen in einem sozialen Milieu verbunden zu fühlen, in diesem Milieu effektiv zu wirken (zu funktionieren) und sich dabei persönlich autonom und initiativ zu erfahren“ (Deci & Ryan, 1993, S. 229).⁴³ Deci und Ryan (1993, S. 225) differenzieren die motivationalen Dispositionen anhand des Grads ihrer Selbstbestimmung weiter aus. Dabei beinhaltet die intrinsische Motivation den höchsten Grad an Selbstbestimmung und das Individuum handelt gänzlich ohne externen Druck (Deci & Ryan, 1993,

41 Daraus erwächst die Annahme, dass mit der im Rahmen der vorliegenden Dissertation durchgeführten Veränderungsmessung im vergleichsweise kurzen Zeitraum eines Semesters keine signifikanten Veränderungen in der Merkmalsausprägung gemessen werden können. Dieser Annahme wird im Rahmen der empirischen Erprobung (Kapitel 4) nachgegangen.

42 Die Motivation wird in der pädagogischen Psychologie dabei als Zustand (*state*) verstanden, der im Gegensatz zu überdauernden Personenmerkmalen (*trait*) veränderbar ist (Schiefele & Köller, 2010, S. 338). Zur Diskussion stabiler Persönlichkeitsmerkmale wird auf Child (1968, S. 83) verwiesen. Für die Differenzierung zwischen dispositionaler und habitueller Motivation sei auf Schiefele (1996, 2009) verwiesen.

43 Dabei rekurren die Autoren auf die drei vermuteten menschlichen Bedürfnisse nach Kompetenz oder Wirksamkeit (*effectance*, vgl. White, 1959), Autonomie oder Selbstbestimmung (vgl. DeCharms, 1968) und soziale Eingebundenheit (*social relatedness*) oder soziale Zugehörigkeit (*affiliation*, vgl. Harlow, 1958).

S. 226), während die externe Motivation durch externale Anreize bedingt wird und „mit instrumenteller Absicht“ durchgeführte Handlungen impliziert (Deci & Ryan, 1993, S. 225). Intrinsisch motiviertes Verhalten setzt in diesem Zusammenhang die Perzeption von Kompetenz und Selbstbestimmung voraus (Schiefele & Köller, 2010, S. 339).

Empirische Evidenz legt nahe, dass die intrinsische Motivation als Konsequenz auf extern geschaffene Anreize abnimmt (Deci & Ryan, 1993, S. 226; Schiefele & Köller, 2010, S. 339).⁴⁴ Trotzdem handelt es sich bei den beiden Dispositionen nicht um Antagonisten (Deci & Ryan, 1993, S. 226). Die Analyse von Befunden empirischer Untersuchungen hat dazu geführt, dass die Dispositionen weiter ausdifferenziert wurden und die Selbstbestimmungstheorie auf Basis dessen überarbeitet wurde (vgl. Deci & Ryan, 1991, 1993). Empirische Studien liefern zwar Hinweise dafür, dass primär die intrinsische Motivation statistisch signifikante Effekte auf Testleistungen initiiert (Schiefele & Köller, 2010, S. 337), allerdings zeigt sich bei einer Vielzahl an Lernhandlungen, dass sie sowohl intrinsisch als auch extrinsisch motiviert sind (Deci & Ryan, 1993; Schiefele & Köller, 2010, S. 337), sodass eine Trennung nicht immer zweckmäßig erscheint. Schiefele und Köller (2010, S. 342) sprechen in diesem Zusammenhang von einer „Integration von Ansätzen der intrinsischen und extrinsischen Motivation“ auf Basis der theoretischen und empirischen Erkenntnisgewinne. Unter bestimmten Voraussetzungen kann auch extrinsisch charakterisierte Motivation internalisiert werden (Schiefele & Köller, 2010, S. 341). Studien aus dem berufsbildenden Bereich unterstützen die Annahme, dass motivationale Orientierungen einen bedeutsamen Einflussfaktor auf Fachkompetenz liefern (Zlatkin-Troitschanskaia et al., 2013b, S. 87).

Schiefele (1996) differenziert sowohl die intrinsische als auch die extrinsische Motivation weiter aus. Nach dieser Konzeptionalisierung kann die intrinsische Motivation einen gegenstandsbezogenen oder einen tätigkeitsbezogenen Charakter aufweisen, während die extrinsische Motivation sich als soziale Motivation, Kompetenzmotivation, Wettbewerbsmotivation, Berufsmotivation oder Leistungsmotivation manifestiert (vgl. Schiefele, 2009).

Für den vorliegenden Untersuchungskontext erweist sich vor dem Hintergrund der wachsenden Relevanz von Anwendungen der KI für zahlreiche Lebens- und Arbeitsbereiche (siehe Kapitel 1.1.1) eine Betrachtung von sowohl intrinsischen als auch extrinsischen motivationalen Dispositionen als sinnvoll. Dabei werden die von Schiefele (2009) postulierten Bezugsgrößen für den Untersuchungsgegenstand kontextualisiert und die Motivation sowohl auf persönlicher als auch auf beruflich orientierter Ebene adressiert.

2.2.5 Zwischenfazit

Die Betrachtung der in Kapitel 2.1.4 als für die Untersuchung relevant postulierten kognitiven und non-kognitiven Kompetenzfacetten liefert Hinweise für die theoretische Modellierung, die als Grundlage für die Operationalisierung des Untersuchungsgegenstandes in Kapitel 3 dient.

44 In Studien wurden verschiedene externe Effekte auf ihren Einfluss auf die Motivation untersucht. Einen Überblick liefern Deci und Ryan (1993, S. 230).

Die zentrale kognitive Facette des Wissens (Kapitel 2.2.1) wird demnach im Rahmen der vorliegenden Arbeit als Strukturwissen (deklarativ-prozedural) und Faktenwissen (deklarativ-statisch) definiert und bildet auch in Teilen das darauf operierende Verstehen ab.

Die Facette der Einstellungen (Kapitel 2.2.2) wird in Bezug zu Modellen der Technologieakzeptanzforschung als Einstellungsakzeptanz modelliert, die persönliche Tendenzen in Bezug auf ein Einstellungsobjekt inkludiert. Das Einstellungsobjekt wird in der vorliegenden Arbeit durch KI und KI-gestützte Systeme abgebildet.

Die Überzeugungen (Kapitel 2.2.3) werden im Rahmen der Untersuchung als epistemologische Überzeugungen zu Wissen und Wissenserwerb definiert. Dabei wird insbesondere die zentrale Dimension der Stabilität/Sicherheit des Wissens fokussiert, deren Kontinuum von einem feststehenden, endgültigen Wissensbestand bis zu einer dynamischen Menge an Wissen reicht, die veränderbar ist.

Die motivationalen Orientierungen (Kapitel 2.2.4) werden analog zu den diskutierten theoretischen Ansätzen als intrinsische und extrinsische motivationale Dispositionen zur Auseinandersetzung mit KI modelliert. Zusätzlich wird die in Kapitel 2.1.3 diskutierte Subfacette der subjektiven Norm unter den motivationalen Orientierungen im Modell inkludiert.

Um das Strukturmodell als theoretische Grundlage für die vorliegende Arbeit abschließend modellieren zu können, bedarf es einer inhaltlichen Klassifizierung der kognitiven Facette (Kapitel 2.3). Das Strukturmodell als Basis für die Operationalisierung des Messgegenstandes ist in Kapitel 2.5 abgebildet.

2.3 Inhaltliche Anforderungen an das Wissen zu Künstlicher Intelligenz von (angehenden) Lehrkräften

2.3.1 Künstliche Intelligenz als Fachgebiet der Informatik

Der Begriff der Künstlichen Intelligenz wurde erstmals 1956 im Rahmen der Dartmouth Conference genutzt (vgl. McCarthy et al., 1956) und hat sich seitdem zu einem immensen Forschungsfeld an der Schnittstelle zwischen Informatik, Neurowissenschaften, Linguistik, Philosophie und weiteren Disziplinen etabliert (Kühl et al., 2022, S. 2237). Die Definition des Begriffs und die Eingrenzung dessen, was darunter subsumiert wird, ist bisher weitgehend inkonsistent (André et al., 2021, S. 7; Seufert et al., 2021, S. 11; Peissner et al., 2019, S. 9). Häufig wird eine Annäherung zur Begriffsdefinition über die einzelnen Bestandteile „künstlich“ und „intelligent“ induziert (Seufert et al., 2021, S. 11). Allerdings ist auch dieser Definitionsansatz herausfordernd, da bereits der Intelligenzbegriff nicht hinreichend geklärt ist und kein Konsens über seine explizite Bedeutung besteht (Buxmann & Schmidt, 2021, S. 6; Dietzmann & Alt, 2020, S. 5170; Paaß & Hecker, 2020, S. 1). In verschiedenen Definitionsansätzen wird Intelligenz häufig global postuliert als „kognitive Fähigkeiten [...], die helfen, den Alltag zu bewältigen und Probleme zu lösen“ (Seufert et al., 2021, S. 11). Görz et al. (2021, S. 6) definieren Intelligenz hingegen als „Erkenntnisvermögen, als Urteilsfähigkeit, als das

Erfassen von Möglichkeiten, aber auch als das Vermögen, Zusammenhänge zu begreifen und Einsichten zu gewinnen“. Es ist insgesamt kritisch zu hinterfragen, inwiefern KI mit menschlicher Intelligenz gleichzusetzen ist (Seufert et al., 2021, S. 12). Ohne den Begriff der Intelligenz abschließend zu klären, definieren Kleesiek et al. (2020, S. 24) KI als „Simulation von intelligentem Verhalten in Computern“. An dieser Stelle wird deutlich, dass nicht von Intelligenz *per se*, sondern von intelligentem Verhalten gesprochen wird. Das Verhalten ist eine beobachtbare, manifeste Größe (Blasius & Baur, 2022, S. 737; Vogt, 2022, S. 12), während die Intelligenz ein latentes Merkmal (Blasius & Baur, 2022, S. 737; Gniewosz, 2015, S. 72) abbildet, weshalb dieser Definitionsansatz tangibler scheint. KI hat sich als Sammelbegriff für eine Vielzahl von Systemen etabliert, die intellektuelle Tätigkeiten bis zu einer bestimmten Ausprägung automatisieren (siehe z. B. Peissner et al., 2019). Ein zentraler Unterschied von KI-Systemen zu anderen Technologien besteht darin, dass sie statt eines finalen Ist-Zustandes eine stetige und dynamische Weiterentwicklung implizieren, die auch Auswirkungen auf die Nutzung und dementsprechend auf die Nutzenden hat (André et al., 2021, S. 7). Dabei arbeiten KI-Systeme auf Grundlage von Algorithmen (Holmes et al., 2019, S. 17).

Im öffentlichen Diskurs kann häufig beobachtet werden, dass KI auf maschinelles Lernen (ML) reduziert wird (Holmes et al., 2019, S. 18; Seufert et al., 2021, S. 12; SBF1, 2019) oder die Begriffe zumindest inkonsistent und nicht trennscharf verwendet werden (Kühl et al., 2022, S. 2235). ML lässt sich als Teil von KI einordnen (Wuttke, 2021, S. 56). Aus Abbildung 9 kann die Beziehung von KI, ML und DL entnommen werden.

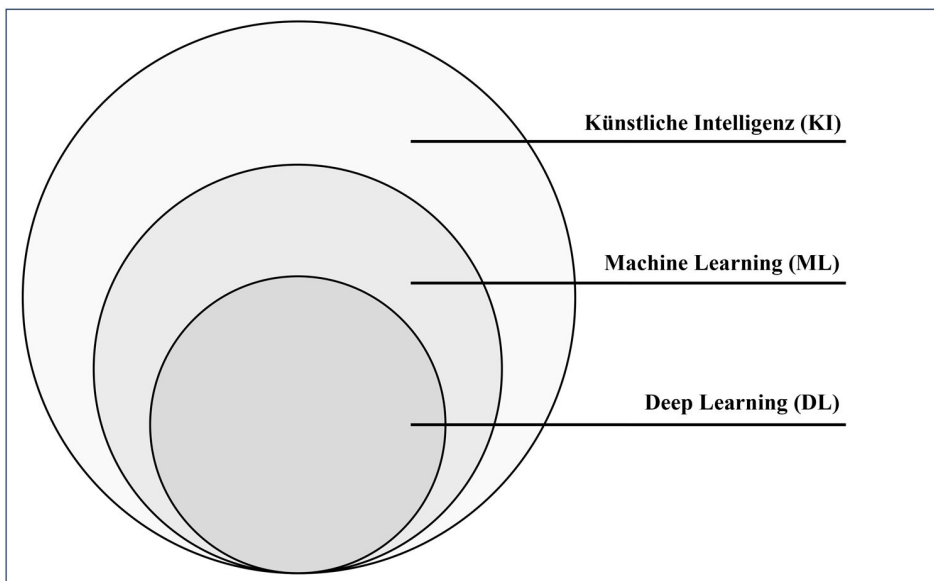


Abbildung 9: Einordnung der Begriffe Künstliche Intelligenz, Machine Learning und Deep Learning. Adaptiert aus Wuttke (2021, S. 56)

Im Rahmen von ML werden auf Basis von Datensätzen Vorhersagen ermöglicht (Holmes et al., 2019, S. 18). Dabei kann ML in die drei primären Typen überwachtes Lernen (*supervised learning*), unüberwachtes Lernen (*unsupervised learning*) und verstärkendes Lernen (*reinforcement learning*) differenziert werden (Buxmann & Schmidt, 2021, S. 11; Holmes et al., 2019, S. 19–20; Kühl et al., 2022, S. 2237). Seit den 2010er-Jahren sind besonders die Entwicklungen im Bereich des ML und des sog. *Deep Learnings* (DL) populär (Ertel, 2016, S. 13; Thiebes et al., 2021, S. 448). Aktuell stehen dabei primär künstliche neuronale Netze (kNN) im Fokus des Diskurses (Dietzmann & Alt, 2020, S. 517)⁴⁵; Holmes et al., 2019, S. 20), auf denen DL basiert (Buxmann & Schmidt, 2021, S. 14). Die Funktionsweise von kNN kann nur schwer nachvollzogen werden (zur sog. *Black-Box-Problematik* siehe Eschenbach, 2021; Guidotti et al., 2018; Lee & Chen, 2008; Zednik, 2021), was dazu führt, dass der Einsatz häufig mit wahrgenommenen Unsicherheiten behaftet ist (Holmes et al., 2019, S. 22). Die Integration von KI-Themen und die Erarbeitung einer KI-Strategie stellt für zahlreiche Unternehmen eine Herausforderung dar (Seufert et al., 2021, S. 13).

Das Ziel der Integration ist es stets, die individuellen Vorteile der menschlichen und der künstlichen Intelligenz gewinnbringend zu verknüpfen (Kühl, 2022, S. 2235). Nicht zuletzt aufgrund der bei zahlreichen potenziellen Nutzenden identifizierten Vorbehalte aufgrund mangelnder Kenntnis oder fehlender Kontrolle (z. B. Overdiek & Petersen, 2022) hat sich der Forschungszweig der *trustworthy AI* etabliert (vgl. Thiebes et al., 2021).⁴⁶ In Anlehnung an Arbeiten der Vertrauensforschung (vgl. Mayer et al., 1995)⁴⁷ werden in diesem Zusammenhang Modelle entwickelt, mit denen das Vertrauen in KI-Systeme operationalisiert werden kann. Aktuell werden überdies vermehrt *Bias- und Diskriminierungsproblematiken* diskutiert (Thiebes et al., 2021, S. 447). Es hat sich gezeigt, dass zahlreiche auf kNN basierende KI-Anwendungen u. a. rassistisch, misogyn oder sexistisch agieren (Pargmann & Berding, 2022, S. 221) und damit gesellschaftlich und pädagogisch als nicht akzeptabel gelten (vgl. Klugmann, 2021). Diese Problematiken ergeben sich durch die im Trainingsprozess implizit oder explizit angelernten Stereotypen und Verhaltensmuster, die durch die initiale Einschätzung der Trainingsdaten durch den Menschen erfolgen (vgl. Klugmann, 2021). Das führt dazu, dass vom Menschen internalisierte Diskriminierungsmuster auf die Datensätze übertragen werden und sich im Lernprozess manifestieren. Es existieren Ansätze, um diese Problematik in der Ursache tiefergehend zu analysieren und zu beheben (Pargmann & Berding, 2022, S. 221).

45 Dietzmann und Alt (2020, S. 517) nehmen in Bezug zu Cattells Distinktion des Intelligenzbegriffs in *fluid intelligence* und *crystallized intelligence* (vgl. Cattell, 1971) eine Differenzierung der Ansätze vor. Konkret bedeutet das, dass ML die *fluid intelligence* (die Fähigkeit, Lösungsstrukturen zu entwickeln und auf neue Situationen zu übertragen) und kNN die *crystallized intelligence* (kognitive Fähigkeiten auf Grundlage von Lernprozessen) widerspiegeln (Dietzmann & Alt, 2020, S. 517). Die Autoren finden überdies heraus, dass Testpersonen tendenziell eher ein der *crystallized intelligence* zuzuordnendes Verständnis von Intelligenz besitzen. Damit können möglicherweise Verständnisschwierigkeiten in Bezug zu KI erklärt werden (Dietzmann & Alt, 2020, S. 5178).

46 In den letzten Jahren haben sich als Reaktion auf die Herausforderungen, die durch den Einsatz von KI wahrgenommen werden, zahlreiche vergleichbare Konzepte zur *trustworthy AI* etabliert (siehe z. B. *beneficial AI*, Future of Life Institute, 2017; *responsible AI*, Wiens et al., 2019; *ethical AI*, Floridi et al., 2018). Einen Überblick bieten Thiebes et al., 2021, S. 452.

47 Siehe zum Vertrauen in digitalisierten Arbeitsumgebungen auch Moschner und Schlicht (2018).

Im berufsbildenden Bereich wird KI häufig über einen anwendungsorientierten Zugang mittels der Stufen der Autonomie definiert (siehe z. B. Becker et al., 2021; Wilbers, 2021). Dabei reicht das Spektrum von Stufe 0 „Keine Autonomie, Mensch hat volle Kontrolle ohne Assistenz“ bis Stufe 5 „Autonomer Betrieb in allen Bereichen, auch in Kooperationen und in sich ändernden Systemgrenzen, Mensch kann abwesend sein“ (Becker et al., 2021, S. 41). Die Auswirkungen der einzelnen Stufen auf die berufliche Tätigkeit des Menschen sind differenziert zu betrachten (Becker et al., 2021, S. 40). Insgesamt stellt die Entwicklung und Implementation von KI-Systemen eine „interdisziplinäre Gestaltungsaufgabe“ dar (Plattform Lernende Systeme, 2019, S. 11). Die Frage nach veränderten Kompetenzanforderungen und die Konsequenzen, die sich daraus für die berufliche Aus- und Weiterbildung ergeben, sind eng damit verknüpft (Ertl & Seifried, 2021, S. 154).

2.3.2 Inhaltliche Strukturierung des Fachgebiets Künstliche Intelligenz

2.3.2.1 Vorgehensweise

Für die inhaltliche Strukturierung des Fachgebiets zur Identifizierung der relevanten Inhalte für das zu entwickelnde Strukturmodell (Kapitel 2.5) erfolgt eine Differenzierung nach fachinhaltlichen Bereichen. Dabei werden mit Bezug zur Zielgruppe insbesondere nicht-technische Inhalte fokussiert (Meß, 2022, S. 15). Zur Identifizierung der relevanten Inhalte eines Fachbereichs bietet sich eine Analyse von Lehrwerken⁴⁸ an (Happ, 2017, S. 55). Darüber hinaus werden die Befunde aus der Literaturanalyse mit den Ergebnissen von durchgeführten Experteninterviews (vgl. Meß, 2022) abgeglichen. Verknüpft mit den erarbeiteten konzeptionellen Grundlagen (Kapitel 2.1 und Kapitel 2.2) wird auf dieser Basis ein Strukturmodell entwickelt, das die für die Untersuchung des Grundlagenwissens zu Künstlicher Intelligenz der anvisierten Zielgruppe (angehende Lehrkräfte im berufsbildenden Bereich) relevanten Inhaltsbereiche inkludiert (siehe Kapitel 2.5). Wie bereits herausgearbeitet, mangelt es bisher an Instrumenten zur systematischen Erfassung des Grundlagenwissens zu KI für (angehende) Lehrkräfte im berufsbildenden Bereich. Das zu erarbeitende Strukturmodell soll als Grundlage für die Operationalisierung⁴⁹ des Messgegenstandes (Kapitel 3) dienen.

2.3.2.2 Auswahl und Analyse relevanter Grundlagenwerke

Für die inhaltliche Strukturierung eines Fachgebiets können insbesondere Lehrbücher herangezogen werden (Happ, 2017, S. 55). In diesem Kontext kann zwischen Bestands- und Defizitanalysen sowie zwischen einer horizontalen und einer vertikalen Perspektive unterschieden werden (Ernst, 2012, S. 42).⁵⁰ In der vorliegenden Arbeit wird eine Bestandsanalyse vorgenommen. Dabei soll es insbesondere darum gehen, welche In-

48 Zur Methode der Lehrbuchanalyse kann auf Adelberg (1984), Fuchs et al. (2014) und Knecht et al. (2014) verwiesen werden.

49 Ergebnisse aus Lehrbuchanalysen werden häufig als Grundlage zur Operationalisierung herangezogen (Härtig et al., 2010, S. 216).

50 Andere Autoren stellen darüber hinaus weitere Differenzierungsprinzipien auf (siehe dazu Krumm, 1973; Reetz & Witt, 1974).

haltsbereiche für ein Rahmenmodell zum Grundlagenwissen berücksichtigt werden müssen. Die Auswahl erfolgt auf Grundlage quantifizierbarer Kennwerte (Meß, 2022, S. 15–17). Bezogen auf die eingenommene Perspektive wird dieser Arbeit eine horizontale Analyse zugrunde gelegt. Bei einer vertikalen Analyse liegt der Fokus auf der Betrachtung von Lehrwerken über einen längeren Zeitraum mit dem Ziel, potenzielle Entwicklungen festzustellen (Ernst, 2012, S. 42). Bei einer horizontalen Analyse wird primär der gegenwärtige Stand im Querschnitt betrachtet. Es wird also der Frage nachgegangen, welche Bereiche zu einem festgelegten Zeitpunkt in einer Vielzahl von Lehrwerken behandelt werden (Happ, 2017, S. 56). Ernst (2012, S. 27) betont, dass die Auswahl der analysierten Werke nachvollziehbar zu begründen ist. Wall (2002, S. 77) arbeitet für den Auswahl- und Begründungsprozess die Richtlinien *Verbreitungsgrad* sowie *Standardcharakter* heraus.

Diese Modellierung im Rahmen der vorliegenden Dissertation erfolgt mit einem stetigen Bezug zur betrachteten Zielgruppe. Für die Zielgruppe wird auf Grundlage der fehlenden curricularen Anknüpfungspunkte ein defizitäres Vorwissen angenommen (vgl. Pargmann & Berding, 2022). Weit verbreitete Grundlagenwerke zu KI (z. B. Russel & Norvig, 2016) formulieren ein technisches Wissen als zwingende Voraussetzung. Mathematikkenntnisse aus der Oberstufe werden dabei als nur bedingt ausreichend angesehen. Ein höheres technisches Wissen kann für die Zielgruppe nicht angenommen werden, weshalb in der begründeten Auswahl der Lehrwerke explizit darauf geachtet wurde, dass „keine Mathematik- oder Programmierkenntnisse auf akademischem Niveau zum Verständnis der Inhalte erforderlich sind“ (Meß, 2022, S. 15). Aufgrund der wachsenden Bedeutung von Online-Lehr-Lern-Medien für die betrachtete Zielgruppe und mit dem Ziel eine breitere Basis abzudecken, wurden in Anlehnung an die didaktische Explikation (Michaeli et al., 2020, S. 73) neben klassischen Print-Lehrbüchern auch E-Learning-Angebote in die Analyse mit aufgenommen, sofern sie wissenschaftlich fundiert sind und eine nicht-technische Einführung in die KI bieten (Meß, 2022, S. 15). Eine Übersicht der analysierten Ressourcen kann Tabelle 1 entnommen werden.

Tabelle 1: Übersicht der analysierten Grundlagenwerke (Meß, 2022, S. 15–16)

Titel	Autorinnen und Autoren, Jahr	Art des Mediums
Einführung in die KI	Elements of AI, 2022	OER
Einführung in die KI	KI-Campus, 2020	OER
Künstliche Intelligenz	Buxmann & Schmidt, 2021	Lehrbuch
Künstliche Intelligenz	Reinhart, Mayer & Greiner, 2021	Lehrbuch
Digitalisierung und Künstliche Intelligenz	Deckert & Meyer, 2020	Lehrbuch
Artificial Intelligence basics, a non-technical introduction	Taulli, 2019	Lehrbuch
Künstliche Intelligenz	Wittpahl, 2019	Lehrbuch

Als Ergebnis der quantifizierten Dokumentenanalyse⁵¹ (Meß, 2022, S. 17–18) gehen die vier Hauptkategorien *Grundlagen der KI*, *Methoden der KI*, *Anwendungen der KI* sowie *Auswirkungen der KI* als relevante Inhaltsbereiche hervor. Dabei stehen die vier Kategorien in einer Beziehung⁵² zueinander (siehe Abbildung 10). Der Bereich der Grundlagen bildet dabei den Bezugs- und Ausgangspunkt für die anderen Subkategorien. Meß (2022, S. 18) begründet die Einschätzung damit, dass diese Ebene in den analysierten Werken grundsätzlich immer die Einführung darstellt und ein stetiger Rückbezug auf die Inhalte dieses Bereichs erfolgt.⁵³

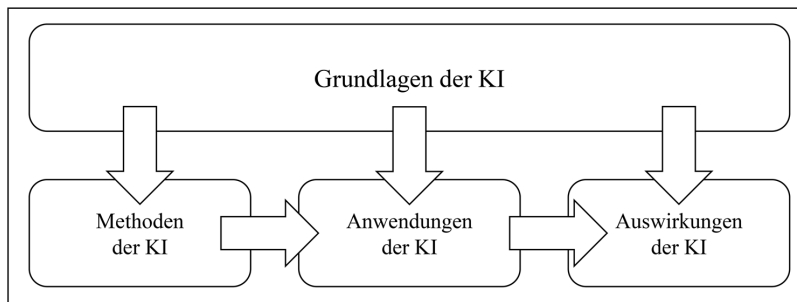


Abbildung 10: Inhaltliche Subdomänen der KI. Adaptiert aus Meß (2022, S. 19)

2.3.2.3 Experteninterviews zur Identifikation relevanter Inhaltsbereiche

Eine weitere Perspektive zur Identifikation relevanter Inhaltsbereiche wird durch Experteninterviews eingenommen. Für die Beurteilung der inhaltlichen Relevanz des Inhalts wurde eine Befragung mit vier Expertinnen und Experten durchgeführt (vgl. Meß, 2022). Ziel der Experteninterviews ist es, Hinweise zu finden, inwieweit die identifizierten Inhaltsbereiche aus den Lehrbüchern auch aus Sicht der Expertinnen und Experten als relevant eingestuft werden können (Meß, 2022, S. 32). Die Interviews zielen damit auf den Wissens- und Erfahrungshorizont der Expertinnen und Experten ab (Bogner et al., 2014, S. 33–34) und sind zu diesem Zweck explorativ angelegt (Meß, 2022, S. 31). Die Auswahl der Expertinnen und Experten erfolgte dabei in Anlehnung an die Kriterien nach Gläser und Laudel (2010, S. 117).⁵⁴ Nach der Durchführung und Transkription der Interviews wurden die generierten qualitativen Forschungsdaten in Anlehnung an Mayring (2015) inhaltsanalytisch ausgewertet (Meß, 2022, S. 37–39).

51 An dieser Stelle muss darauf verwiesen werden, dass im theoretischen Teil der Dissertation auch empirische Elemente (u. a. Dokumentenanalyse und Experteninterviews) zu finden sind. Das soll keine Vermischung mit dem eigentlichen empirischen Teil der Arbeit in Kapitel 4 sein. Die qualitativen Analysen im theoretischen Teil dienen der Ausdifferenzierung der Inhaltsbereiche (inhaltliche Validierung im Sinne der AERA et al., 2014) und dienen daher der theoretischen Modellierung. Die empirischen Validierungsschritte im späteren empirischen Teil der Arbeit münden in eine quantitative Studie an verschiedenen Hochschulen.

52 Zur genaueren Ausdifferenzierung der Beziehungen zwischen den Subkategorien siehe Meß, 2022, S. 18–19.

53 Die Masterarbeit von Meß (2022) wurde am Institut für Wirtschaftspädagogik der Wirtschaftswissenschaftlichen Fakultät der Universität Leipzig verfasst. Die Arbeit wurde eng vom Institut betreut (z. B. Diskussion und Auswahl der relevanten Lehrbücher und Akquise der Expertinnen und Experten). Die Autorin der vorliegenden Dissertation war in die Betreuung federführend eingebunden.

54 Für eine dezidierte Betrachtung der einzelnen Experteninterviews sei auf Meß (2022) verwiesen.

Die Befunde aus der Inhaltsanalyse legen nahe, dass das Modell der inhaltlichen Subdomänen der KI (siehe Abbildung 10) in seiner aus der Literatur abgeleiteten Form bestätigt werden kann (Meß, 2022, S. 41). Das betrifft sowohl die Auswahl der Inhaltsbereiche als auch die Beziehung dieser untereinander.

2.3.3 Zwischenfazit

Aus den Analyseschritten zur inhaltlichen Strukturierung des Fachgebiets haben sich demnach folgende Inhaltsbereiche als relevant ergeben: Grundlagen der KI, Methoden der KI, Anwendungen der KI und Auswirkungen der KI. Diese einzelnen Kategorien lassen sich auf Grundlage der Dokumentenanalyse (Buxmann & Schmidt, 2021; Deckert & Meyer, 2020; Elements of AI, 2022; KI-Campus, 2020; Reinhart et al., 2021; Taulli, 2019; Wittpahl, 2019) und der Ergebnisse der Experteninterviews in konkrete Inhaltselemente ausdifferenzieren, die im Folgenden vorgestellt werden.

Grundlagen der KI

Zu den Grundlagen der KI gehören auf Basis der durchgeführten Analysen z. B. Inhalte zur geschichtlichen Entwicklung des Fachgebiets. Dazu zählen neben der ersten Erwähnung des Begriffs „Künstliche Intelligenz“ im Rahmen der Dartmouth Conference im Jahr 1956 (Buxmann & Schmidt, 2021, S. 4; Taulli, 2019, S. 6) deklarierte Meilensteine der Entwicklung, wie z. B. der *Turing Test*⁵⁵ (vgl. KI-Campus, 2020) und der Sieg des algorithmischen Systems *Deep Blue*⁵⁶ gegen den damaligen Schachweltmeister (Buxmann & Schmidt, 2021, S. 6; Taulli, 2019, S. 13; Wittpahl, 2019, S. 7). Darüber hinaus fallen in diesen Bereich die Zusammenhänge von Künstlicher Intelligenz, Machine Learning und Deep Learning (Taulli, 2019, S. 16; KI-Campus, 2020). Auch die Differenzierung zwischen schwacher (*weak/narrow*) und starker (*strong/general*) KI (Buxmann & Schmidt, 2021, S. 7; Wittpahl, 2019, S. 21) wird zu diesem Bereich gezählt. Diese Typisierung soll eine Beurteilung der Aufgabenerledigung im Vergleich zu der Qualität ermöglichen, die durch einen Menschen erzielt wird. Bei der vergleichenden Betrachtung von menschlich und künstlich generierten Outputs kam es zur Definition des *Moravec'schen Gesetzes* (vgl. KI-Campus, 2020; Taulli, 2019, S. 141), das besagt: „Was für einen Menschen schwer ist, ist für einen Computer leicht und umgekehrt“. Diese Erkenntnis wird ebenfalls zum Bereich der Grundlagen der KI gezählt. Zudem gehört die Definition des Begriffs Algorithmus zum Bereich der Grundlagen.

Methoden der KI

Die Methoden der KI basieren auf statistischen Verfahren (Buxmann & Schmidt, 2021, S. 108). Um auf der Grundlage von Daten Prognosen zu generieren, werden z. B. Regressionsmodelle eingesetzt (vgl. KI-Campus, 2020; Wittpahl, 2019, S. 26). Im Kontext

⁵⁵ Der *Turing Test* rekuriert auf ein von Alan Turing durchgeführtes Experiment (vgl. Turing, 1950). Im Rahmen des Tests agiert eine Person mit einer anderen Person und einem Computer, ohne dabei zu wissen, welcher Interaktionspartner human und welcher maschinell ist. Am Ende muss die Testperson eine Einschätzung abgeben. Sofern sie dabei den Computer für die Person hält, wird die Maschine als „intelligent“ bezeichnet (siehe Taulli, 2019, S. 2).

⁵⁶ Im Jahr 1996 schlug der Computer *Deep Blue* von IBM den damaligen Schachweltmeister Garri Kasparow (Wittpahl, 2019, S. 7).

des ML, das ebenfalls als Methode der KI klassifiziert werden kann, werden die Lernverfahren überwachtes Lernen (*supervised learning*), unüberwachtes Lernen (*unsupervised learning*) und verstärkendes Lernen (*reinforcement learning*) differenziert (Buxmann & Schmidt, 2021, S. 11; Taulli, 2019, S. 50; Wittpahl, 2019, S. 25–29). Die Entwicklung von maschinellern Lernen wird maßgeblich durch Big Data, d. h. das Vorhandensein und Verarbeiten großer Datensätze, begünstigt (Wittpahl, 2019, S. 100). Aktuell von zentraler Bedeutung in der KI-Forschung sind künstliche neuronale Netze (kNN). Bei mehreren Schichten im Netzwerk wird in diesem Kontext von *Deep Learning* (DL) gesprochen (vgl. KI-Campus, 2020; Wittpahl, 2019, S. 36). Das Ziel von maschinellern Lernen besteht grundsätzlich darin, Vorhersagen basierend auf Daten zu treffen (Buxmann & Schmidt, 2021, S. 9; Wittpahl, 2019, S. 26). Je nachdem, wie sicher dies gelingt – d. h. wie hoch die Vorhersagegenauigkeit des Algorithmus ist –, desto besser ist das System zu bewerten. Die Vorhersagegenauigkeit stellt damit ein zentrales Gütekriterium dar (vgl. KI-Campus, 2020).

Anwendungen der KI

In dieser Kategorie werden spezifische Anwendungen und deren Fähigkeiten subsumiert. Darunter fallen z. B. Sprachmodelle, die auf Basis von Computerlinguistik (*Natural Language Processing* (NLP)) arbeiten (Buxmann & Schmidt, 2021, S. 57; Taulli, 2019, S. 104). Als ein sehr bekanntes Sprachmodell hat sich der *Generative Pre-Trained Transformer* (GPT) entwickelt, der mit ungelabelten Daten trainiert wurde (Taulli, 2019, S. 114). Neben der Computerlinguistik ist die Computer Audition ein zentrales Anwendungsfeld (Buxmann & Schmidt, 2021, S. 9). Eine weitere Fähigkeit von KI-gestützten Systemen besteht in der Generalisierung. Diese sagt aus, dass das System ein Zielobjekt auf bisher unbekannten Bildern zuverlässig erkennt (Buxmann & Schmidt, 2021, S. 21). Dabei basiert der Lernprozess immer auf einem sog. Trainingsdatensatz (Buxmann & Schmidt, 2021, S. 11). Die Fähigkeiten von KI können u. a. in wahrnehmende, analytische, generierende und motorische Fähigkeiten differenziert werden (vgl. KI-Campus, 2020). Bei einer Software, die in Echtzeit Staus und Verkehrsbehinderungen erkennt und daraufhin die Navigation anpasst, werden analytische Fähigkeiten des Bereichs Planung und Optimierung (Buxmann & Schmidt, 2021, S. 95) eingesetzt.

Auswirkungen der KI

Die inhaltliche Analyse hat ergeben, dass es sich in der Kategorie *Auswirkungen von KI* hauptsächlich um wahrgenommene Herausforderungen aus dem Einsatz von KI-gestützten Systemen handelt. Darunter fallen z. B. die bereits in Kapitel 1 und Kapitel 2.3.1 adressierten *bias*-Problematiken und wahrgenommene Unsicherheiten, die mit dem Einsatz von KI assoziiert sind (siehe ausführlich dazu Meß, 2022, S. 17). Auch die Befunde aus den Experteninterviews bestätigen, dass es dabei primär um die notwendige Sensibilisierung zum kritischen Hinterfragen von KI-generierten Outputs

geht.⁵⁷ Diese Aspekte können nicht unter den für die vorliegende Dissertation eingegrenzten Wissensbegriff (siehe Kapitel 2.2.1) gefasst werden, sondern sie tangieren tendenziell non-kognitive Facetten der Einstellungen oder Überzeugungen. Aus diesem Grund wird die Kategorie der Auswirkungen der KI für die Modellierung der kognitiven Leistungsdisposition zunächst nicht weiter betrachtet.

2.4 Forschungsstand zur Ausprägung KI-bezogener Kompetenzfacetten bei (angehenden) Lehrkräften

2.4.1 Konzeptionelle Ansätze zu KI-Kompetenzen

Um die Potenziale von KI-gestützten Systemen voll ausschöpfen zu können, sind Qualifikationsprozesse notwendig, die die Kompetenzentwicklung unterstützen (André et al., 2021, S. 7). Wie in Kapitel 2.1.1 dargelegt, nehmen Lehrer:innen im berufsbildenden Bereich eine zentrale Rolle in der Kompetenzentwicklung der (zukünftigen) Fach- und Führungskräfte ein und müssen entsprechend auf diese Gestaltungsaufgabe vorbereitet werden. Dabei stellt sich die Frage, welche Kompetenzen die (angehenden) Lehrkräfte in diesem Zusammenhang benötigen und wie deren Ausprägung erfasst werden kann. In der Forschungsliteratur werden für Kompetenzen im Bereich KI häufig die Begriffe der KI-Kompetenz (vgl. André et al., 2021) und der *AI Literacy* verwendet (vgl. Ng et al., 2021a, 2021b). Ng et al. (2021b, S. 2) zeigen in ihrem systematischen Literaturreview einen rasanten Anstieg an Publikationen, die zwischen 2014 und 2021 unter dem Schlagwort *AI Literacy* erschienen⁵⁸ sind. Während zwischen 2016 und 2019 lediglich 11 relevante Artikel zu diesem Thema identifiziert wurden, sind es in den Jahren 2020 und 2021 insgesamt 19 relevante Publikationen (Ng et al., 2021b, S. 3). Ein Großteil der Publikationen bezieht sich dabei auf die Schüler:innen in Grundschulen oder weiterführenden Schule (93,4%). Lediglich zwei Paper (6,7%) können der Lehrkräftebildung und damit annähernd dem Bezugsgegenstand der vorliegenden Arbeit zugeordnet werden (Ng et al., 2021b, S. 5). Da die meisten Lehrkräfte im allgemeinbildenden Schulbereich, ausgenommen die Informatiklehrkräfte, ebenso keine Lerngelegenheiten zu Anwendungen der KI wahrnehmen wie die berufsbildenden Lehrkräfte, werden in dieser Dissertation auch Studien zu Lehrkräften aus dem allgemeinbildenden Bereich herangezogen. Aus dem Review (vgl. Ng et al., 2021b) gehen vier Aspekte hervor, die trotz z. T. divergierender Begriffsverwendung in den analysierten Studien unter dem Begriff der *AI Literacy* vereint werden können (Tabelle 2).

57 Die Einschätzung wird u. a. gestützt durch die folgenden Aussagen aus den Experteninterviews: „Es soll ein Bewusstsein für die Gefahren und Herausforderungen der KI geschaffen werden“ sowie „Dabei ist es wichtig, dass die Gesellschaft für die Existenz von Deep Fakes sensibilisiert wird und ein Bewusstsein darüber entwickelt, Geschehenes kritisch zu hinterfragen“. Für eine ausführliche Auseinandersetzung mit den Befunden der Interviewstudie wird auf Meß (2022) verwiesen.

58 Das Literaturreview stützt sich dabei auf die Angaben auf Google Scholar (Ng et al., 2021, S. 3).

Tabelle 2: AI Literacy-Aspekte, adaptiert aus Ng et al. (2021a, S. 505)

Literacy-Aspekt	Definition
Wissen und Verstehen	Grundlegende Funktionen von KI kennen und wissen, wie KI-Systeme genutzt werden
Anwenden	Wissen zu KI und KI-Systemen in verschiedenen Situationen anwenden
Evaluierten und Erstellen	Höherrangige Denkleistungen, Evaluierten und Erstellen von KI-Anwendungsszenarien
Ethische Aspekte	Menschenzentrierte Überlegungen (Transparenz, Ethik, Sicherheit, Fairness)

Nahezu alle analysierten Paper beinhalten in ihrer Definition von *AI Literacy* die erste Stufe „*Know & understand AI*“. Darunter werden sowohl inhaltliche Grundlagen und Funktionsweisen von KI-Anwendungen als auch die damit verbundenen Einstellungen zusammengefasst (Ng et al., 2021a, S. 505; Ng et al., 2021b, S. 4). Darüber hinaus wird die Ebene der Anwendung in allen analysierten Studien adressiert. Aus der Beschreibung dieser Stufe kann allerdings vermutet werden, dass darin zumindest zum Teil das Wissen über die Anwendung („*to know how to apply*“, S. 505) sowie das Verständnis über die Auswirkungen der Nutzung („*to understand AI applications and how it can affect our lives, as well as knowing the ethical issues regarding AI technologies*“ (S. 505)) adressiert werden und nicht ausschließlich die tatsächliche Anwendung im Fokus der Betrachtung steht. Die höheren Stufen „*evaluate & create*“ und „*AI ethics*“ werden nur in 63 % bzw. 50 % der analysierten Werke als Bestandteil der *AI Literacy* definiert (Ng et al., 2021a, S. 505; Ng et al., 2021b, S. 4). Aus der Systematisierung von KI-Kompetenzen nach André et al. (2021, S. 11) geht vergleichbar dazu hervor, dass die Aspekte „Anwendung von Fach- und Grundlagenwissen“ und „Umgang mit KI-Systemen“ die Basis darstellen, während Aspekte wie „Gestaltung von Arbeitsprozessen“ darauf aufbauen. Im Gegensatz zu Ng et al. (2021a, 2021b) konstatieren André et al. (2021, S. 11, 18) die zwingende Kontextualisierung von KI-Kompetenz in Bezug auf konkrete Aufgaben und Anwendungen.

Ng et al. (2021b) nutzen für die Definition von *AI Literacy* den Ansatz nach Long und Magerko (2020), nach denen es sich dabei um eine Sammlung von Kompetenzen handelt, die für die Nutzung und kritische Reflexion von KI in verschiedenen Kontexten (z. B. Kommunikation, berufliche und private Nutzung) benötigt werden. In Anlehnung an die Taxonomie nach Bloom (1976)⁵⁹ werden die in Abbildung 11 zusammengefassten Stufen modelliert.

59 An dieser Einordnung ist kritisch anzumerken, dass diese auf Bloom (1976) und nicht die revidierte Version von Anderson und Krathwohl (2001) Bezug nimmt.

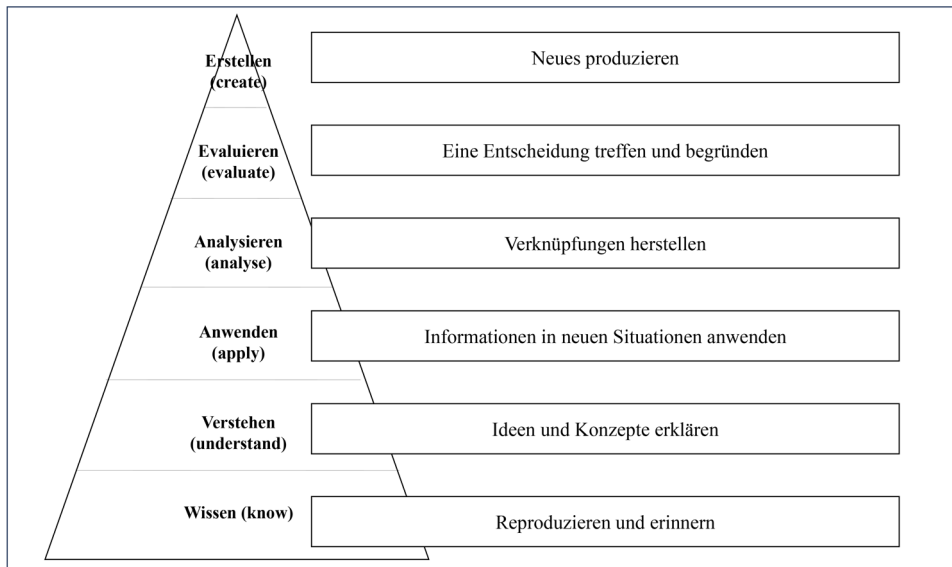


Abbildung 11: Einordnung der AI Literacy in die Taxonomie nach Bloom (1976). Adaptiert aus Ng et al. (2021b, S. 5)

Ng et al. (2021, S. 5) synthetisieren das Konzept der *AI Literacy* außerdem mit dem *TPACK*-Modell (vgl. Mishra & Koehler, 2006, siehe ausführlicher Kapitel 2.1.2). Daraus geht hervor, dass im Bereich *TK* z. B. Kenntnisse über Hard- und Software enthalten sind. Der Bereich *PK* umfasst didaktische Aspekte, wie kollaboratives Lernen und konstruktivistische Ansätze. Im Bereich *CK* geht es um die Inhaltsebene. Darunter fassen die Autorinnen und Autoren auch ethisch-moralische Fragestellungen, die mit dem Einsatz von KI-Technologien assoziiert sind. Sie identifizieren darüber hinaus die „*AI syllabus*“ (vgl. Russel & Norvig, 2009) sowie die „*Five big ideas about AI*“ (vgl. Touretzky et al., 2019) als inkludierten Gegenstandsbereich des *CK*.

In Anlehnung an die in Kapitel 2.3 diskutierten Einschränkungen, die sich aus den formalen Lerngelegenheiten und dem dadurch verbundenen antizipierten defizitären (Vor-)Wissen der Zielgruppe manifestieren, sind diese Inhalte für das vorliegende Rahmenmodell zunächst nachrangig zu betrachten. Die Klassifikation in Anlehnung an das *TPACK*-Modell und die Taxonomie nach Bloom (bzw. Anderson & Krathwohl) können allerdings unter Berücksichtigung der in Kapitel 2.1.2 und Kapitel 2.3 herausgearbeiteten konzeptionellen und inhaltlichen Überlegungen genutzt werden, um den Untersuchungsgegenstand weiter zu kontextualisieren. Daraus lässt sich ableiten, dass das Wissen zu KI in Modellen der *AI Literacy* nicht eindimensional betrachtet wird, sondern in verschiedene Dimensionen und Facetten binnendifferenziert werden kann.

2.4.2 Empirische Studien zu KI-bezogenen Kompetenzen von (angehenden) Lehrkräften

In den letzten Jahren ist ein deutlicher Anstieg an Studien zu vermerken, die sich mit KI-bezogenen Kompetenzen von (angehenden) Lehrkräften beschäftigen (für deutschsprachige Studien z. B. Roppertz, 2021, für internationale Studien z. B. Ng et al., 2023). Vor dem Hintergrund der dargelegten zunehmenden Relevanz von KI-gestützten Systemen in nahezu allen Lebens- und Arbeitsbereichen (siehe Kapitel 1) ist dies nicht verwunderlich. Bei der Analyse der empirischen Studien lassen sich dabei allerdings große Unterschiede feststellen, die sich sowohl auf die konzeptionellen Fundierungen, das angelegte Forschungsdesign und die eingesetzten Instrumente als auch die generierten Befunde beziehen. Erste Studien mit Lehrkräften liefern initiale Aussagen zur Ausprägung des Grundlagenwissens und der non-kognitiven Dispositionen zu Anwendungen der Künstlichen Intelligenz bei Lehrkräften (vgl. Linder & Berges, 2020; Lindner & Romeike, 2019). Allerdings handelt es sich dabei hauptsächlich um Studien aus dem allgemeinbildenden Schulbereich, weshalb diese nur mit Einschränkung auf den Untersuchungskontext der vorliegenden Arbeit übertragbar sind. Für den berufsbildenden Bereich existieren bisher kaum vergleichbare Studien (vgl. Schmidt & Happ, 2022a), sodass die Recherche auf alle Schularten ausgeweitet wurde. Auch angehende Lehrkräfte (vgl. Haseski, 2019) und Bildungspersonal im Hochschulbereich (vgl. Hussain, 2020) werden aufgrund der raren Studienlage in die Analyse einbezogen.

Zahlreiche Studien legen in ihrer Analyse dabei den Fokus nicht explizit auf kognitive oder non-kognitive Facetten, sondern betrachten zum Teil mehrere Kompetenzfacetten in einem integrativen Ansatz. Aus Tabelle 3 können die Studien entnommen werden, die den Fokus (ggf. neben anderen) auf kognitive Dispositionen gesetzt haben. Studien, die darüber hinaus auch andere (non-kognitive) Kompetenzfacetten in der Untersuchung inkludiert haben, enthalten entsprechende Informationen in der rechten Spalte („ergänzende Informationen“).

Tabelle 3: Forschungsstand zum KI-Wissen von (angehenden) Lehrkräften

Autorinnen/ Autoren, Jahr	Forschungs- paradigma	Stichprobe	Land	Ergänzende Informationen
Chiu & Chai, 2020	Qualitativ	24	China	Auch Einstellungen erfasst
Chiu, 2021	Qualitativ	24	China	Mittelschule; auch Einstellungen erfasst
Chounta et al., 2021	Quantitativ	131	Estland	K-12 ⁶⁰ , auch Einstellungen erfasst
Dorotea et al., 2021	Quantitativ	174	Portugal	Informatiklehrkräfte

60 Unter K-12 wird im englischen Sprachgebrauch die Primar- und Sekundarstufe bis zur 12. Klasse subsumiert.

(Fortsetzung Tabelle 3)

Autorinnen/ Autoren, Jahr	Forschungs- paradigma	Stichprobe	Land	Ergänzende Informationen
Gocen & Aydemir, 2020	Qualitativ	19	Türkei	Auch Einstellungen erfasst
Haseski, 2019	Qualitativ	94	Türkei	Angehende Lehrkräfte, auch Einstellungen erfasst
Hofstetter & Massmann, 2021	Mixed-Methods	150	Deutschland	Sekundarstufe I & II
Kaufmann, 2020	Qualitativ	162	Schweiz	(Angehende) Lehrkräfte, auch Einstellungen erfasst
Lindner & Berges, 2020	Qualitativ	23	Deutschland	Gymnasium, Informatiklehrkräfte, auch Einstellungen erfasst
Lindner & Romeike, 2019	Quantitativ	37	Deutschland	Sekundarstufe, Informatiklehrkräfte, auch Einstellungen erfasst
Pfeiffer et al., 2021	Mixed-Methods	60	Keine Angabe	Primar-, Sekundar-, Tertiärstufe, auch Einstellungen erfasst
Pu et al., 2021	Quantitativ	60	China	Angehende Lehrkräfte
Sanusi et al., 2021	Qualitativ	12	Nigeria, Ghana, Tansania, Kenia, Südafrika und Namibia	Highschool, Informatiklehrkräfte, auch Einstellungen erfasst
Serholt et al., 2014	Qualitativ	8	England, Portugal, Schottland, Schweden	Geografielehrkräfte, auch Einstellungen erfasst
Wollowski et al., 2016	Qualitativ	37	Keine Angabe	Tertiärstufe, auch Einstellungen erfasst
Zhong et al., 2020a	Quantitativ	1266	China	Grund- und Sekundarstufe, auch Einstellungen erfasst
Zhong et al., 2020b	Quantitativ	k/A	China	Informatiklehrkräfte, Grund- und Sekundarstufe, auch Einstellungen erfasst

Aus der Analyse geht hervor, dass einige Studien zwar angeben, das Wissen als kognitive Facette zu untersuchen, bei genauerer Betrachtung der eingesetzten Instrumente zeigt sich allerdings, dass die Zuordnung nicht trennscharf ist. Pu et al. (2021, S. 208) setzen zur Erfassung der kognitiven Dispositionen den „*Practical Knowledge Acquisition*“-Test nach Jie (2016) ein. Die Ausrichtung des genannten Beispielitems „*I can learn AI education knowledge*“ (Pu et al., 2021, S. 208) liefert einen Hinweis dafür, dass in Anlehnung an das in Kapitel 2.2.3 präzisierte Begriffsverständnis für die vorliegende Arbeit der eingesetzte Test tendenziell auf epistemologische Überzeugungen zu Wissen und insbesondere Wissenserwerb abzielt. Aufgrund von potenziellen Un-

schärfen in der Zuordnung und der Feststellung, dass in zahlreichen Studien mehrere Facetten adressiert werden, erfolgt die systematische Auswertung entlang aller identifizierten Studien. Aus Tabelle 4 gehen jene Studien hervor, die nach eigener Angabe primär non-kognitive Facetten in Bezug zu KI analysiert haben.

Tabelle 4: Forschungsstand zu non-kognitiven Kompetenzfacetten von (angehenden) Lehrkräften zu KI

Autorinnen/Autoren, Jahr	Forschungsparadigma	Stichprobe	Land	Ergänzende Informationen
Alhashmi et al., 2021	Qualitativ	5	VAE	
Attwood et al., 2020	Mixed-Methods	41	USA	Angehende Lehrkräfte verschiedener allg. Fächer
Bii et al., 2018	Mixed-Methods	10	Kenia	Informatik- und Physiklehrkräfte
Bunting et al., 2021	Qualitativ	11	Schweden	Grundschule
Chocarra et al., 2021	Quantitativ	225	k/A	
Chuah & Kabilan, 2021	Quantitativ	142	Malaysia	Sekundarstufe
Grimes & Warschauer, 2010	Mixed-Methods	k/A	USA	
Hussain, 2020	Quantitativ	196	Pakistan	Hochschullehrende
Kennedy et al., 2016	Quantitativ	35	United Kingdom	
Kim & Lee, 2015	Quantitativ	140	Korea	Grundschule
Lee et al., 2008	Quantitativ	367	Korea	Grund- und Mittelschule
Louie et al., 2021	Qualitativ	8	k/A	
Reich-Stiebert & Eysell, 2016	Quantitativ	59	Deutschland	Grund-, Sekundar-, Berufsschule
Roppertz, 2021	Mixed-Methods	48	Deutschland	Berufsschule
Rosanda & Isenic, 2021	Quantitativ	121	k/A	Angehende Lehrkräfte
Vazhayil et al., 2019	Mixed-Methods	34	Indien	Informatiklehrkräfte, Mittel- und Oberstufe
Wilson et al., 2021	Quantitativ	135	USA	Grundschule

Aus der Übersicht der erfassten Studien (Tabelle 3 und Tabelle 4) lassen sich erste Erkenntnisse über die Ausprägung der kognitiven und non-kognitiven Kompetenzfacetten bei (angehenden) Lehrkräften ableiten. Der Großteil der Studien adressiert den allgemeinbildenden Bereich. Nur zwei der analysierten Studien fokussieren explizit Lehrkräfte im berufsbildenden Bereich (vgl. Roppertz, 2021; Reich-Stiebert & Eysell,

2016). Das zeigt, dass der berufsbildende Bereich in diesem Forschungsfeld bisher eher nachrangig betrachtet wurde.

Aus den in den Studien generierten Befunden geht hervor, dass das Wissen der (angehenden) Lehrkräfte zu KI defizitär ist (vgl. Hofstetter & Massmann, 2021; Lindner & Berges, 2020; Lindner & Romeike, 2019; Nenner et al., 2021; Pfeiffer et al., 2021; Serholt et al., 2014; Zhong et al., 2020a; Zhong et al., 2020b) und maßgeblich durch aktuelle Hypethemen und deren mediale Repräsentation geprägt wird (vgl. Linder & Berges, 2020; Lindner & Romeike, 2019).

Daraus kann ein implizit notwendiger und explizit benannter Weiterbildungsbedarf des Bildungspersonals abgeleitet werden (vgl. Bii et al., 2018; Chiu & Chai, 2020; Chiu, 2021; Chounta et al., 2021; Haseski, 2019; Lindner & Romeike, 2019; Nenner et al., 2021; Pu et al., 2021; Roppertz, 2021; Sanusi et al., 2021; Zhong et al., 2020b).

Die (angehenden) Lehrkräfte, die im Rahmen der analysierten Studien befragt wurden, erkennen mehrheitlich das Potenzial von KI sowohl für ihre eigene berufliche Praxis (vgl. Attwood et al., 2020; Bii et al., 2018; Hofstetter & Massmann, 2021; Roppertz, 2021; Zhong et al., 2020a) als auch für das Privat- und (zukünftige) Arbeitsleben ihrer Schüler:innen (vgl. Hofstetter & Massmann, 2021; Hussain, 2020; Lindner & Romeike, 2019; Roppertz, 2021; Sanusi et al., 2021). Dies wird zum Teil als Erklärungsansatz dafür herangezogen, dass die Probandinnen und Probanden eine Bereitschaft äußern, sich (intensiver) mit dem Themenkomplex auseinanderzusetzen und die neuen Technologien zu nutzen (vgl. Chuah & Kabilan, 2021; Roppertz, 2021; Vazhayil et al., 2019). Dabei wird der Individualisierung von Lehr-Lern-Prozessen durch Anwendungen der KI großes Potenzial zugesprochen (vgl. Alhashmi et al., 2021; Bunting et al., 2021; Chounta et al., 2021; Gocen & Aydemir, 2020; Hussain, 2020; Reich-Stiebert & Eyssel, 2016; Serholt et al., 2014; Wilson et al., 2021). Darüber hinaus werden die KI-gestützte Unterstützung bei Hausaufgaben (vgl. Chuah & Kabilan, 2021) sowie die Verbesserung des Monitorings (vgl. Reich-Stiebert & Eyssel, 2016; Serholt et al., 2014) als Chancen von KI hervorgehoben.

Trotz der erkannten Potenziale sehen zahlreiche Lehrkräfte den Einsatz von KI im Bildungsbereich kritisch (vgl. Alhasmi et al., 2021; Chiu & Chai, 2020; Grimes & Warschauer, 2010; Haseski, 2019; Kaufmann, 2020). Die Kritik bezieht sich dabei u. a. auf die Angst der Lehrkräfte, dass KI sie in ihrer beruflichen Tätigkeit ersetzen könnte (vgl. Gocen & Aydemir, 2020; Serholt et al., 2014) oder dass sie bei der Implementation der Systeme die Kontrolle über die Algorithmen verlieren könnten (vgl. Gocen & Aydemir, 2020; Kaufmann, 2020; Lee et al., 2008; Lindner & Berges, 2020). Ein zentraler Kritikpunkt besteht darüber hinaus in dem wahrgenommenen oder befürchteten Verlust von zwischenmenschlicher Interaktion in Bildungsprozessen (vgl. Alhashmi et al., 2021; Chounta et al., 2021; Kennedy et al., 2016; Kim & Lee, 2015; Reich-Stiebert & Eyssel, 2016; Roppertz, 2021; Rosanda & Istenic, 2021; Serholt et al., 2014). Diese Befürchtungen zeigen sich in der Literatur u. a. durch das Narrativ des sozial isolierten Lernenden, der von einem Roboter unterrichtet wird (Pargmann & Berding, 2022, S. 227). Zusammengefasst zeigen sich darin insbesondere durch Unsicherheiten geprägte kritische Haltungen.

Neben der notwendigen Aus- und Weiterbildung werden auch die institutionellen und infrastrukturellen Voraussetzungen kritisch betrachtet. So ist die technische Ausstattung in den Schulen in vielen Fällen unzureichend, um die Potenziale von KI-gestützten Technologien ausschöpfen zu können (vgl. Bii et al., 2018; Vazhayil et al., 2019; Zhong et al., 2020b).⁶¹ Außerdem wird der Mehraufwand der Lehrpersonen aktuell als Barriere wahrgenommen (vgl. Kennedy et al., 2016; Reich-Stiebert & Eyssel, 2016). Schließlich sind es auch ethisch-moralische Fragestellungen, die mit dem Einsatz von KI-gestützten Technologien assoziiert werden (vgl. Chounta et al., 2021; Hofstetter & Massmann, 2021; Lindner & Berges, 2020; Serholt et al., 2014).

Aus messtheoretischer Sicht lässt sich beobachten, dass ein Großteil der Studien einem qualitativen Forschungsparadigma folgt (siehe Tabelle 3 und Tabelle 4) oder quantitativ die Selbsteinschätzung der Probandinnen und Probanden mittels Likert-Skalen erfasst (z. B. Lindner & Romeike, 2019; Pu et al., 2021).⁶²

Das zeigt deutlich, dass es insgesamt noch an systematischen und standardisierten Verfahren mangelt, die das Grundlagenwissen testbasiert erfassen. Dieser Punkt wird im Rahmen der Operationalisierung (Kapitel 3) weiter verfolgt.

Neben der aggregierten Einschätzung gehen aus den Studien außerdem aufschlussreiche Einzelbefunde hervor. So stellen Reich-Stiebert und Eyssel (2016) vergleichend fest, dass die Lehrkräfte im Grundschulbereich kritischer gegenüber KI sind als Lehrkräfte der weiterführenden und berufsbildenden Schulen. Einige Studien lassen darüber hinaus vermuten, dass das individuelle Interesse an Technologien einen Einfluss auf die Einstellungen nimmt (vgl. Reich-Stiebert & Eyssel, 2016; Pfeiffer et al., 2021).

Roppertz (2021) analysiert explizit die Gruppe der Lehrkräfte im berufsbildenden Bereich und untersucht mittels eines standardisierten Online-Fragebogens deren Einstellungen. Dabei werden die Perspektiven „Wandel der Arbeitswelt“, „Einflüsse auf die Berufsschulpraxis“ und „Qualifizierung des pädagogischen Personals“ eingenommen (Roppertz, 2021, S. 9–10). Insbesondere die dritte Dimension ist dabei für den vorliegenden Untersuchungskontext relevant. Es wird adressiert, welche digitalen Kompetenzen die Lehrkräfte im Zuge der aktuellen Entwicklungen für besonders relevant halten und wie sie ihre eigene Bereitschaft beurteilen würden, an einer entsprechenden Weiterbildung teilzunehmen (Roppertz, 2021, S. 10). Eine Einschätzung über die Ausprägung der Kompetenzfacetten beim beruflichen Bildungspersonal geht aus der Analyse allerdings nicht hervor.

2.4.3 Zwischenfazit und Implikationen für die Untersuchung

Aus den Studien geht ein defizitäres Wissen der (angehenden) Lehrkräfte zu KI hervor. Da es sich bei den Probandinnen und Probanden z. T. um Lehrkräfte aus der In-

61 Dieser ausschließlich in internationalen Studien explizit formulierte Befund lässt sich mit Bezug zu den Ergebnissen der ICIL-Studie zumindest in Teilen auf die Rahmenbedingungen in Deutschland transferieren. Aus der Studie geht hervor, dass die technische Ausstattung von allgemeinbildenden Schulen weit unter dem europäischen Vergleich zurückbleibt. Gerholz et al. (2022b) konstatieren, dass dieser Befund auch auf die Situation in den berufsbildenden Schulen übertragen werden kann.

62 Dieses Vorgehen ist insbesondere aufgrund zahlreicher potenziell „verzerrender“ Effekte zu kritisieren (siehe dazu Nickolaus, 2010, S. 485).

formatik handelt, denen auf Grundlage der inhaltlichen Nähe ein tendenziell stärker ausgeprägtes Wissen zum Themenkomplex unterstellt werden kann, wird in der vorliegenden Untersuchung ebenfalls von einem defizitären Wissen der angehenden Lehrkräfte ausgegangen.

Im Rahmen der Analyse zeichnen sich darüber hinaus positive Einstellungstendenzen der (angehenden) Lehrkräfte ab, die sich insbesondere in der Wahrnehmung der Potenziale KI-gestützter Systeme für die eigene berufliche Praxis und für die Schüler:innen manifestieren. Dagegen stehen Befunde, aus denen eine kritische Haltung der Lehrkräfte abgeleitet werden kann. Das Erkennen von Potenzialen kann in Anlehnung an die Auseinandersetzung mit den Modellen der Technologieakzeptanzforschung (Kapitel 2.1.3) und dem Einstellungsbegriff (Kapitel 2.2.2) in das für die vorliegende Arbeit festgelegte Verständnis von Einstellungen als non-kognitive Facette gefasst werden. Daher wird für die vorliegende Untersuchung von tendenziell positiven Einstellungen der Probandinnen und Probanden ausgegangen. Vor dem Hintergrund des in Kapitel 2.2.3 hergeleiteten Verständnisses von epistemologischen Überzeugungen und den aggregierten Befunden aus der Literaturanalyse werden für die Zielgruppe tendenziell kritische Überzeugungen gegenüber KI angenommen.

Aus den Studien geht außerdem hervor, dass die (angehenden) Lehrkräfte eine Bereitschaft zeigen, sich mit dem Themenkomplex auseinanderzusetzen. Worin diese Bereitschaft begründet liegt und ob sie sich dementsprechend eher als extrinsisch oder intrinsisch motiviert (in Anlehnung an die in Kapitel 2.2.4 vorgenommenen Begriffspräzision) klassifizieren lässt, bleibt zunächst offen.

Zusammengefasst ergeben sich aus der Analyse des Forschungsstandes folgende Annahmen, die im Rahmen der empirischen Untersuchung geprüft werden (siehe Tabelle 5).

Tabelle 5: Abgeleitete Annahmen für die empirische Überprüfung

Facette	Bezeichnung	Annahme
Wissen	AW1	Die angehenden Lehrkräfte im berufsbildenden Bereich weisen ein defizitäres Wissen auf.
Einstellungen	AE1	Die angehenden Lehrkräfte erkennen das Potenzial von KI-gestützten Technologien.
Überzeugungen	AÜ1	Die angehenden Lehrkräfte nehmen eine kritische Haltung gegenüber KI ein.
Motivation	AM1	Die Lehrkräfte zeigen eine Bereitschaft, sich mit dem Themenkomplex auseinanderzusetzen.

2.5 Strukturmodell KI-bezogener Kompetenzfacetten von (angehenden) Lehrkräften im berufsbildenden Bereich

Ein zentrales Element der vorliegenden Arbeit ist eine theoretische Modellierung des (Grundlagen-)Wissens zu Künstlicher Intelligenz als fundamentale kognitive Facette einer KI-bezogenen Kompetenz von (angehenden) Lehrkräften im berufsbildenden Bereich. Wie bereits in Kapitel 1.3 dargelegt, mangelt es bisher an einer grundlegenden Konzeptualisierung des Untersuchungsgegenstandes für die anvisierte Zielgruppe, weshalb noch keine standardisierten Verfahren zur Messung vorliegen. Es wurde bereits herausgearbeitet, dass jüngst Ansätze entwickelt wurden, um das Wissen zu KI bei Lehrkräften (primär im allgemeinbildenden Bereich) zu erfassen. Jedoch sind die Befunde für den vorliegenden Untersuchungskontext nur bedingt übertragbar.

Auf Grundlage der Einordnung des Untersuchungsgegenstandes in die Modelle der professionellen Kompetenz (Kapitel 2.1.1), der digitalisierungsbezogenen Kompetenzen (Kapitel 2.1.2) und der Technologieakzeptanzforschung (Kapitel 2.1.3) konnten die für die Untersuchung relevanten Kompetenzfacetten identifiziert werden, die in Kapitel 2.2 jeweils begrifflich spezifiziert wurden. Durch die Verknüpfung dieser Befunde mit den Ergebnissen der inhaltlichen Strukturierung des Fachbereichs (Kapitel 2.3) und der Analyse des Forschungsstandes zu KI-bezogenen Kompetenzen (Kapitel 2.4) wird das in Abbildung 12 zusammengefasste Strukturmodell erarbeitet, das als Grundlage für die Operationalisierung des Untersuchungsgegenstandes (Kapitel 3) fungiert.

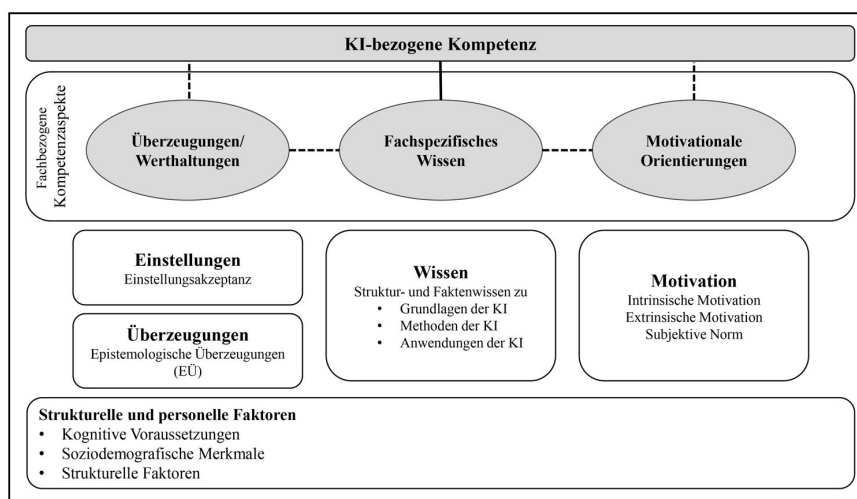


Abbildung 12: Theoretisches Rahmenmodell als Grundlage für die Operationalisierung des Messgegenstandes

Aus dem Modell geht die Differenzierung der KI-bezogenen Kompetenz in die Bereiche Überzeugungen/Werthaltungen, Fachspezifisches Wissen sowie Motivationale

Orientierungen hervor. Erstere werden in Einstellungen im Sinne einer Einstellungsakzeptanz sowie Überzeugungen als epistemologische Überzeugungen zu Wissen und Wissenserwerb spezifiziert. Das fachspezifische Wissen wird als Struktur- und Faktenwissen zu den identifizierten Inhaltsbereichen Grundlagen der KI, Methoden der KI und Anwendungen der KI kontextualisiert. Die motivationalen Orientierungen beinhalten neben intrinsischen und extrinsischen motivationalen Dispositionen ebenfalls die Subfacette der subjektiven Norm.

3 Operationalisierung KI-bezogener Kompetenzfacetten

3.1 Messgegenstand und Validierungsaspekte

3.1.1 Bestimmung des Messgegenstandes

Auf der Grundlage der theoretischen Modellierung (Kapitel 2) ergeben sich theoretische Annahmen für die Operationalisierung des interessierenden Merkmals. Diese Annahmen beziehen sich auf eine quantitative oder qualitative Ausrichtung und granularer auch auf die Abstufungen, mit denen das Personenmerkmal gemessen werden kann (Rost, 2004, S. 32). Aus dem in Kapitel 2.5 entwickelten Strukturmodell⁶³ gehen die einzelnen für die Untersuchung relevanten Kompetenzfacetten hervor, die im Rahmen der Operationalisierung messbar gemacht werden. Dafür ist eine eindeutig abgrenzbare Definition der einzelnen Konstrukte und ihrer für die Messung relevanten Charakteristika essenziell (AERA et al., 2014, S. 11). Im Rahmen der vorliegenden Arbeit handelt es sich um die kognitive Facette des Wissens (siehe Kapitel 2.2.1) und um die non-kognitiven Facetten der Einstellungen (Kapitel 2.2.2), Überzeugungen (Kapitel 2.2.3) und Motivation (Kapitel 2.2.4). Das Grundlagenwissen als kognitive Facette steht dabei im Fokus der Betrachtung der vorliegenden Arbeit. Das ist u. a. mit der zentralen Bedeutung dieser Facette (siehe ausführlicher Kapitel 2.1.1 und Kapitel 2.2.2) zu begründen. Darüber hinaus existieren, wie bereits herausgearbeitet, für die Erfassung des Grundlagenwissens zu KI bei der Zielgruppe keine standardisierten Instrumente, auf die im Rahmen der vorliegenden Untersuchung zumindest in Teilen zurückgegriffen werden kann. Ausgehend von der theoretischen Modellierung wird ein neues Testinstrument entwickelt und erprobt. Bei der Operationalisierung der non-kognitiven Kompetenzfacetten kann auf existierende Messinstrumente zurückgegriffen werden, die adaptiert werden (Kapitel 3.2.3.2). Dieses ist möglich, da zur Erfassung von Einstellungen (vgl. Davis, 1989; Venkatesh & Davis, 2000), Überzeugungen (vgl. Schiefele et al., 2002) und motivationalen Orientierungen (vgl. Ryan & Deci, 2017) bereits eine Vielzahl erprobter Messinstrumente aus der Pädagogischen Psychologie vorliegen (zur Testadaption von Testinstrumenten siehe z. B. Förster et al., 2017b). Der Vorteil der Testadaption kann zum einen darin gesehen werden, dass aus methodischer Sicht gute Item- und Skalenkennwerte bei dem Rückgriff auf erprobte Skalen erzielt werden sollten. Zum anderen kann es aus Sicht des Aufwands als zeitlich sparsamere Variante umschrieben werden. Der erstgenannte Punkt überwiegt bei der Entscheidung zur Testadaption jedoch.

⁶³ Die modellbasierte Testentwicklung hat sich auch in anderen Domänen etabliert. Für den naturwissenschaftlichen Bereich kann auf Tepner et al. (2012) verwiesen werden.

Aus der kognitionsbezogenen Beschreibung des für die Arbeit definierten Begriffs des Grundlagenwissens wurde der Fokus auf das Struktur- und Faktenwissen gelegt, das sich als deklarativ-prozedural bzw. deklarativ-statisch klassifizieren lässt (siehe ausführlicher in Kapitel 2.2.1). In Anlehnung an die Einordnung in die Taxonomie nach Anderson und Krathwohl (2001) wird damit der kognitive Prozess des Erinnerns bzw. darauf aufbauender Verstehensprozesse intendiert. Das impliziert, dass die Bewältigung von Aufgaben in diesem Zusammenhang anhand eines Richtigkeitsmaßstabs (Krohne & Hock, 2007, S. 368) eingeschätzt werden kann. Es können Testaufgaben konstruiert werden, die die adressierten kognitiven Prozesse intendieren und anschließend mit „korrekt“ oder „falsch“ bewertet werden können.

Aus der inhaltsbezogenen Strukturierung des Fachgebiets gehen die Inhaltsbereiche Grundlagen der KI, Methoden der KI und Anwendungen der KI als relevant für die Untersuchung hervor (siehe Kapitel 2.3). Diese Inhaltsbereiche müssen analog zur kognitionsbezogenen Differenzierung in Testaufgaben abgebildet werden.

Folglich kann für den definierten Messgegenstand festgehalten werden, dass Testpersonen Grundlagenwissen zu KI unterstellt werden kann, wenn sie Testaufgaben zu den identifizierten Inhaltsbereichen Grundlagen der KI, Methoden der KI und Anwendungen der KI durch die kognitiven Prozesse des Erinnerns oder darauf operierenden Verstehens bewältigen können. Aus dieser Einschränkung und der in Kapitel 2.2.2 vorgenommenen Binnendifferenzierung des Wissensbegriffs geht hervor, dass es sich bei dem Konstrukt des Grundlagenwissens um ein eher breit formuliertes Konstrukt handelt, welches inhalts- und kognitionsbezogen weiter ausdifferenziert werden muss, um es zu operationalisieren. Zur Systematisierung des Messgegenstandes als Grundlage für die Operationalisierung eignet sich daher die Darstellung in Form einer Matrix (Kuhn, 2014, S. 116).

Tabelle 6: Matrix zur Systematisierung des Grundlagenwissens (Darstellung in Anlehnung an Kuhn, 2014, S. 116)

Inhaltliche Strukturierung	Kognitionsbezogene Differenzierung	
	Deklarativ-statisch	Deklarativ-prozedural
Grundlagen der KI		
Methoden der KI		
Anwendungen der KI		

3.1.2 Aspekte der Validierung

3.1.2.1 Festlegung der Validierungsaspekte und Grundannahmen

Dem definierten Messgegenstand liegen eine Reihe an Annahmen zugrunde, die sich insbesondere aus der theoretischen Modellierung (Kapitel 2) ergeben. Um zu prüfen, inwieweit die formulierten Annahmen im Rahmen der Operationalisierung und empirischen Erprobung des entwickelten Instruments tragen, werden im Folgenden die Aspekte konkretisiert, die im Rahmen der Validitätsbeurteilung näher betrachtet werden.

Im Zuge der Validitätsbeurteilung wird geprüft, „ob die Personen auf die Items so antworten, wie man es aufgrund der Theorie über die zu messende Personeneigenschaft erwarten würde“ (Rost, 2004, S. 35). Aus diesem Grund kann eine umfassende Prüfung der Validität nur auf Grundlage empirischer Befunde durchgeführt werden (Krohne & Hock, 2007, S. 26), es bedarf aber gleichermaßen auch einer Analyse konzeptioneller Annahmen (Krohne & Hock, 2007, S. 65), weshalb die Beurteilung der Validität über alle Phasen der vorliegenden Arbeit adressiert wird. Die Einschätzung der Validität ist dabei zwingend in Bezug zum Zweck der Testung vorzunehmen (Seufert et al., 2020a, S. 2).

Lange Zeit wurde von einem validen Test gesprochen, sofern dieser „misst, was er messen soll“ (Bortz & Döring, 2006, S. 200; Moosbrugger & Kelava, 2012, S. 30; Krohne & Hock, 2007, S. 25; Rost, 2004, S. 34). Dabei wurde Validität primär als Eigenschaft eines Tests verstanden. In der aktuellen Forschungsliteratur wird das Kriterium der Validität differenzierter betrachtet. Insbesondere bei der Überprüfung von Kompetenzen wird das alte Verständnis von Validität als zu kurz greifend kritisiert (z. B. Schaper, 2014, S. 22). Daher hat sich ein Verständnis etabliert, das Validität nicht mehr als Eigenschaft eines Tests ansieht, sondern die Schlussfolgerungen aus Testergebnissen betrifft (Moosbrugger & Kelava, 2020, S. 30). Dadurch rückt die Frage nach der Belastbarkeit der Testwertinterpretationen in den Fokus der Betrachtung (Hartig et al., 2020, S. 531). Es wird sich also der Frage gewidmet, inwiefern auf Grundlage der Testwertergebnisse valide Aussagen abgeleitet werden können (Meinhardt, 2018, S. 12). Für die Erfassung der Validität wird in der Forschungsliteratur häufig in die Inhaltsvalidität, die Kriteriumsvalidität und die Konstruktvalidität differenziert (Bortz & Döring, 2006, S. 200; Bühner, 2011, S. 61; Jensen et al., 2015, S. 12; Krohne & Hock, 2007, S. 65). Wurde in diesem Kontext früher noch von „Arten der Validität“ gesprochen, besteht heute weitestgehend Konsens darüber, dass diese Evidenzen integraler Bestandteil eines zusammenfassenden Gütekriteriums sind (Hartig et al., 2020, S. 531–532). Die Validität gilt dabei insgesamt als das wichtigste Gütekriterium (Krohne & Hock, 2007, S. 26; Rost, 2004, S. 33). Trotz der integrativen Betrachtung sind die Ausprägungen nach wie vor auch einzeln relevant. Im Gegensatz zu anderen Gütekriterien, wie z. B. der Reliabilität, existiert für die Beurteilung der Validität kein einzelner etablierter Kennwert (wie z. B. für die Schätzung der Reliabilität über die interne Konsistenz), auf dessen Grundlage die Einschätzung erfolgen kann (Krohne & Hock, 2007, S. 65). Vielmehr sind es bei der Validität viele einzelne konzeptionelle Überlegungen und empirische Evidenzen, die Rückschlüsse auf die valide Interpretation der Testwerte zulassen und in der Gesamtschau zu betrachten sind.

In der empirischen Sozialforschung haben sich zur Sicherstellung der Validität insbesondere die „*Standards for Educational and Psychological Testing*“ der American Educational Research Association (AERA), American Psychological Association (APA) und des National Council on Measurement in Education (NCME) etabliert (AERA et al., 2014; siehe z. B. Kuhn, 2014; Happ, 2019, S. 28). In diesen Standards sowie in der deutschsprachigen Forschungsliteratur häufig verwendeten deutschen Adaption wird die Validität definiert als „das Ausmaß, in dem empirische Befunde und theoretische

Argumente die Interpretation von Testwerten für die beabsichtigte Verwendung von Tests unterstützen“ (Hartig et al., 2020, S. 530). Insgesamt umfassen die Standards der AERA et al. (2014) die Aspekte „test content“ (AERA et al., 2014, S. 14), „response processes“ (AERA et al., 2014, S. 15), „internal structure“ (AERA et al., 2014, S. 16), „relations to other variables“ (AERA et al., 2014, S. 16) und „consequences of testing“ (AERA et al., 2014, S. 19). Im Rahmen der vorliegenden Dissertation werden zunächst die Aspekte „test content“, „internal structure“ und „relations to other variables“ fokussiert. Der Aspekt „consequences of testing“ wird aufgrund des für die Arbeit vorrangigen Ziels der kollektiven Betrachtung zunächst ausgeklammert.⁶⁴ Der Aspekt „response processes“ bietet Anknüpfungspunkte für weiterführende qualitative Forschungsvorhaben, die in Kapitel 5 weiter präzisiert werden.

Um im Sinne dieser Standards eine fundierte Einschätzung darüber treffen zu können, inwieweit vom erreichten Testscore auf das theoretisch modellierte Konstrukt geschlossen werden kann, ist eine Berücksichtigung dieser Aspekte über die Phasen der Modellierung, der Testentwicklung sowie der empirischen Erprobung essenziell (Kuhn, 2014, S. 116). Tabelle 7 bildet die einzelnen Validierungsschritte anhand der genannten Aspekte ab, die in den folgenden Kapiteln fortlaufend präzisiert werden.

Tabelle 7: Validierungsaspekte in Anlehnung an die Standards der AERA et al. (2014)

Validierungsaspekte und Annahmen	
„test content“	1) Items als repräsentative Menge des Messgegenstandes
	2) Inhaltliche Relevanz der Items
	3) Angemessenheit der Items
„internal structure“	4) Dimensionalität des Konstrukts
„relations to other variables“	5) Ausprägung des Grundlagenwissens in Abhängigkeit zur Studienrichtung im Master
	6) Ausprägung des Grundlagenwissens in Abhängigkeit zum Studienfortschritt

3.1.2.2 Testinhalt

Der Frage nach der Beziehung zwischen dem theoretisch modellierten Konstrukt und dem Testinhalt kommt nach den Standards der AERA et al. (2014, S. 11) eine hohe Bedeutung zu (Hartig et al., 2007, S. 140; Kane, 2013, S. 18).⁶⁵ Dabei wird im Rahmen der Überprüfung der Inhaltsvalidität der Frage nachgegangen, inwieweit die Items den modellierten Untersuchungsgegenstand abbilden (Krohne & Hock, 2007, S. 67). Dafür ist die Betrachtung der entwickelten Items vor dem Hintergrund des theoretisch modellierten Konstrukts essenziell (Krohne & Hock, 2007, S. 65). Um dem Kriterium zu

⁶⁴ In Anlehnung an Kuhn (2014, S. 116) werden mit der Testung primär grundlagenorientierte Zwecke verfolgt und nicht die Verwertung individueller Testwertbefunde (z. B. zur Schätzung der Eignung für den Lehrberuf).

⁶⁵ In deutschsprachigen Werken zu Forschungsmethoden wird dieser Aspekt häufig als „Inhaltsvalidität“ betitelt (Bortz & Döring, 2006, S. 200; Krohne & Hock, 2007, S. 65) und beinhaltet zentral die Aspekte der Relevanz und Repräsentativität der Items (Krohne & Hock, 2007, S. 65).

genügen, sollen die Items den Inhalt „möglichst umfassend, angemessen gewichtet und auf dem adäquaten Schwierigkeitsniveau operationalisiert“ abbilden (Krohne & Hock, 2007, S. 67). Damit wird deutlich, dass es sich im Rahmen der Inhaltsvalidität primär um testinterne Informationen handelt. Mit Bezug zum Zielkonstrukt der vorliegenden Arbeit soll daher geprüft werden, inwieweit die Items aus dem entwickelten Messinstrument geeignet sind, um das Grundlagenwissen zu KI bei angehenden Lehrkräften im berufsbildenden Bereich zu erfassen. Für die Beurteilung der Repräsentativität und Relevanz der Items zur Abbildung des interessierenden Merkmals werden dabei primär konzeptionelle Überlegungen und theoretische Annahmen auf Basis der Modellierung in Kapitel 2 herangezogen.⁶⁶

Insbesondere bei Tests, die in Bezug auf ein Lehrziel konstruiert und eingesetzt werden, ist die Frage nach der Relevanz und Repräsentativität zentral (Krohne & Hock, 2007, S. 541). Krohne und Hock (2007, S. 541) benennen das Erstellen einer „Verhaltens-Inhalts-Matrix“ als nützlich zur Beurteilung der Repräsentativität. Dabei handelt es sich um eine Kreuztabelle, die sowohl die Inhaltsbereiche als auch die adressierten Anforderungsebenen inkludiert. In diese Matrix sind die konstruierten Items eindeutig einzuordnen. Für diese Zuordnung scheint eine Orientierung an Lehrzieltaxonomien sinnvoll, wie sie im Rahmen der theoretischen Modellierung (Kapitel 2) vorgenommen wurde. Eine einfache Abstufung stellt in diesem Zusammenhang die Differenzierung zwischen kognitiven, affektiven und psychomotorischen Lehrzielen dar (Krohne & Hock, 2007, S. 541). In der vorliegenden Untersuchung zum Grundlagenwissen zu KI werden für das vorliegende Testinstrument nur kognitive Dispositionen tangiert. Für die Einordnung der entwickelten Testbestandteile wird demnach die in Kapitel 3.1 aufgestellte Matrix zur Systematisierung entlang der Inhaltsbereiche und Anforderungsbezüge vorgenommen. Ausgehend von dieser Einordnung kann daher die Annahme formuliert werden:

A1: Die Testitems stellen eine inhalts- und kognitionsbezogene repräsentative Menge des in der Arbeit zugrunde gelegten Messgegenstandes des Grundlagenwissens zu KI dar.

Für die Beurteilung der inhaltlichen Relevanz der einzelnen Items wird insbesondere auf die Befunde der in Kapitel 2.3 dargelegten Analyse von Grundlagenwerken und Befragung von Expertinnen und Experten zurückgegriffen. Daraus ergibt sich die in Kapitel 2.3.3 vorgestellte Differenzierung der Inhaltsbereiche in Grundlagen der KI, Methoden der KI sowie Anwendungen der KI. Diese Systematisierung bildet die inhaltliche Grundlage für die Itementwicklung. Um das Konstrukt möglichst vollständig abzubilden, sollen die Inhaltsbereiche quantitativ vergleichbar in den zu entwickelnden Test aufgenommen werden. Daraus ergibt sich die zweite für die Beurteilung der Testvalidität relevante Annahme:

A2: Die Testitems sind inhaltlich relevant, d. h. sie bilden relevante fachliche Aspekte zu Grundlagen der Künstlichen Intelligenz ab.

66 Siehe dazu auch Hartig et al. (2020) sowie Krohne und Hock (2007, S. 66).

Assoziiert mit der Frage der inhaltlichen Relevanz ist die Frage nach der Angemessenheit der Items (AERA et al., 2014, S. 14; Krohne & Hock, 2007, S. 66). Diesem Kriterium kann z. B. durch die Orientierung an beruflich relevanten Anforderungsbezügen Rechnung getragen werden (Kuhn, 2014, S. 118⁶⁷). In diesem Zusammenhang ist davon auszugehen, dass die Angemessenheit bei Probandinnen und Probanden höher zu bewerten ist, die aufgrund ihrer beruflichen Tätigkeiten regelmäßig mit den im Testinstrument abgebildeten Anforderungen konfrontiert werden. Die Beurteilung dieser Anforderungen gestaltet sich für die Zielgruppe der vorliegenden Arbeit mit Bezug zum Zielkonstrukt des Grundlagenwissens zu KI durchaus schwierig. Das liegt zum einen an der in Kapitel 1 herausgearbeiteten, dynamisch und schnell zunehmenden Relevanz des Themenbereichs KI sowohl für die eigene (zukünftige) berufliche Praxis als Lehrkraft als auch für die adäquate handlungsorientierte Ausbildung der Schüler:innen für eine digitalisierte (Arbeits-)Welt. Darüber hinaus existieren keine gesicherten Befunde über konkrete Anforderungen an das Grundlagenwissen von (angehenden) Lehrkräften im berufsbildenden Bereich. Außerdem befinden sich die Probandinnen und Probanden zum Zeitpunkt der Testung in der ersten Phase der Lehrkräfteausbildung an der Universität und können daher in der Regel nur auf wenig unterrichtliche Berufspraxis zurückgreifen. Auf dieser Grundlage wurde der Wissensbegriff in Kapitel 2.2.1 dahingehend spezifiziert, dass er in diesem Zusammenhang ein essenzielles Fundament für alle weiteren kognitiven Leistungen bildet. Es kann aus diesen Gründen nicht davon ausgegangen werden, dass die (angehenden) Lehrkräfte bereits mit den Anforderungen, die in der Testung abgebildet werden, konfrontiert waren. Demnach ist die Einschätzung der Angemessenheit an dieser Stelle zunächst mit Unsicherheiten behaftet und muss im Rahmen weiterer Forschungsvorhaben erneut adressiert werden (siehe dazu Kapitel 5). Grundsätzlich ergibt sich daraus die dritte für die Beurteilung relevante Annahme:

A3: Die Testitems sind insgesamt für die Gruppe angehender Lehrkräfte im berufsbildenden Bereich in der ersten Phase der Lehrkräfteausbildung angemessen.

3.1.2.3 Interne Struktur

Die Analyse der internen Struktur⁶⁸ des Tests liefert Hinweise, inwieweit die theoretisch postulierten Annahmen zum Konstrukt angemessen im entwickelten Testinstrument abgebildet sind (AERA et al., 2014, S. 16). Nur bei einer hohen Übereinstimmung kann davon ausgegangen werden, dass die Testwerte auf Grundlage des modellierten Konstrukts interpretiert und erklärt werden können (AERA et al., 2014, S. 16). Das ist dann gegeben, wenn für das interessierende Merkmal eine homogene Gruppe an Items vorliegt (AERA et al., 2014, S. 16; Kuhn, 2014, S. 120). In der Analyse der internen Struktur kann außerdem identifiziert werden, ob und inwieweit bestimmte Items für unterschiedliche Probandengruppen unterschiedlich funktionieren (AERA et al., 2014, S. 16).

⁶⁷ In der Arbeit von Kuhn (2014) zum fachdidaktischen Wissen von Lehrkräften im berufsbildenden Bereich wird dies u. a. durch die Orientierung an konkreten beruflichen Situationen gewährleistet.

⁶⁸ Auch als „faktorielle Validität“ bezeichnet (Hartig et al., 2020, S. 531).

Die theoretische Modellierung des Messgegenstandes liefert bereits erste Hinweise dafür, dass es sich um ein tendenziell breit gefasstes Konstrukt handelt. Bisher liegen allerdings keine empirischen Befunde vor, um diese Annahme zu prüfen. Aus diesem Grund soll für den weiteren Verlauf zunächst von einem eindimensionalen Konstrukt ausgegangen werden. Aufgrund der dargelegten Relationen der für das Konstrukt als relevant identifizierten Inhaltsbereiche (siehe Kapitel 2.3) ist eine trennscharfe Abgrenzung in einzelne (Sub-)Konstrukte nicht zweckmäßig. Das liegt u. a. an der postulierten Beziehung, in der die Grundlagen der KI als Voraussetzung für die Bereiche Methoden der KI und Anwendungen der KI gelten. Hinsichtlich der kognitionsbezogenen Differenzierung (siehe Kapitel 2.2.2) kann darauf verwiesen werden, dass sowohl das Struktur- als auch das Faktenwissen als deklaratives Wissen zu klassifizieren sind, weshalb auch an dieser Stelle nicht von einer mehrdimensionalen Betrachtung ausgegangen wird. Um die Dimensionalität im Rahmen der internen Strukturprüfung zu analysieren, kann auf Verfahren der Faktorenanalysen zurückgegriffen werden. Dabei ist zu berücksichtigen, ob bereits theoretisch fundierte Erkenntnisse zu trennbaren Konstruktfacetten vorliegen (konfirmatorische Faktorenanalyse, siehe z. B. Bühner, 2011, S. 380) oder ein strukturentdeckendes Verfahren zu bevorzugen ist (explorative Faktorenanalyse, siehe z. B. Bühner, 2011, S. 296). Zunächst gilt die Annahme:

A4: Das Grundlagenwissen zu KI, wie es in der Arbeit inhalts- und kognitionsbezogen modelliert wurde, lässt sich als eindimensionales Konstrukt abbilden.

3.1.2.4 Beziehung zu anderen Merkmalen

Für den Validierungsaspekt der Beziehung zu anderen Merkmalen bieten sich Analysen im Rahmen einer konvergenten oder diskriminanten Validierung an (AERA et al., 2014, S. 16).⁶⁹ Die konvergente Validierung adressiert dabei die Korrelation der erzielten Messwerte mit den erzielten Messwerten unter Anwendung eines Instruments, mit dem das gleiche interessierende Konstrukt gemessen wird (siehe auch Rost, 2004, S. 34). Unter Annahme der konvergenten Validierung soll diese Korrelation hoch ausfallen (AERA et al., 2014, S. 17). Da es allerdings, wie in Kapitel 1 hergeleitet, bisher an systematischen Ansätzen zur theoretischen Modellierung und empirischen Erfassung des Grundlagenwissens der vorliegenden Zielgruppe mangelt, kann der Ansatz der konvergenten Validierung zunächst nicht weiter verfolgt werden.

Diskriminante Evidenz kann hingegen z. B. bei der Betrachtung der Beziehung zwischen Testergebnissen und Messwerten, die theoretisch unterlegt unterschiedliche Konstrukte erfassen, erzielt werden (AERA et al., 2014, S. 17). Für die diskriminante Validierung kann außerdem auf die „Methode der bekannten Gruppen“ (Kuhn, 2014, S. 123; Schnell et al., 2008, S. 156, s. dazu auch AERA et al., 2014, S. 16) zurückgegriffen

⁶⁹ In Lehrbüchern zu Forschungsmethoden werden die Aspekte der konvergenten und diskriminanten Validierung in der Regel unter dem Begriff der Konstruktvalidität subsumiert (Bortz & Döring, 2006, S. 200; Krohne & Hock, 2007, S. 68–70; Schnell et al., 2008, S. 155–157). Aufgrund der Limitationen der Inhalts- und Kriteriumsvalidität wird der Konstruktvalidität häufig besondere Bedeutung zuteil (Bortz & Döring, 2006, S. 201). Diese Messgröße gilt dann als erfüllt, wenn die aus dem Zielkonstrukt ableitbaren Hypothesen mithilfe der erzielten Testwerte bestätigt werden können (Bortz & Döring, 2006, S. 201).

werden. In diesem Zusammenhang wird davon ausgegangen, dass die theoretisch postulierten Unterschiede in der Merkmalsausprägung verschiedener Gruppen empirisch evident sind (AERA et al., 2014, S. 16). Die in der vorliegenden Arbeit betrachteten Gruppen können dabei nach der Art ihrer Ausbildung hinsichtlich des interessierenden Merkmals differenziert werden. Für die diskriminante Validierung im Rahmen der Untersuchung des Grundlagenwissens der KI von angehenden Lehrkräften im berufsbildenden Bereich wird die Gruppe der Studierenden im Fachbereich Wirtschaftsinformatik als geeignet angesehen. Aufgrund der formalen Lerngelegenheiten im Rahmen des Studiums wird dieser Gruppe eine stärkere Merkmalsausprägung unterstellt als der Gruppe der angehenden Lehrkräfte. Der Gruppenvergleich ist insbesondere mit der Zielsubgruppe der Studierenden der Wirtschaftspädagogik tragfähig, da in diesem Vergleich die übereinstimmende Komponente des Kernfachs Wirtschaft vorliegt. Eine Analyse der Curricula der Studienmodelle legt nahe, dass die Kontrastgruppe deutlich häufiger formale Lerngelegenheiten mit Bezug zu KI hat und daher auch über ein stärker ausgeprägtes Grundlagenwissen zu Künstlicher Intelligenz verfügen sollte als die Zielgruppe. Die Annahme lautet daher:

A5: Die Kontrastgruppe der Studierenden des Fachbereichs Wirtschaftsinformatik weist eine stärkere Ausprägung des Grundlagenwissens zu KI auf als die Zielgruppe.

Eine weitere Kontrastgruppe bilden in diesem Zusammenhang Studierende der Wirtschaftspädagogik im Bachelorstudium. Für den Vergleich der Zielgruppe mit dieser Kontrastgruppe werden keine signifikanten Unterschiede in der Merkmalsausprägung erwartet. Zwar kann unter der Entwicklungsperspektive (Kuhn, 2014, S. 125) davon ausgegangen werden, dass mit zunehmendem Studienfortschritt und den damit verbundenen Erfahrungen und Übungen eine positive Entwicklung des Professionswissens (siehe Kapitel 2.1.1) verbunden ist, allerdings existieren für die in der theoretischen Modellierung erfasste Kompetenzfacette des Grundlagenwissens zu KI keine formalen Lerngelegenheiten im konsekutiven Studienmodell.⁷⁰ Wenngleich davon ausgegangen werden kann, dass die beiden Gruppen eine unterschiedliche Merkmalsausprägung im Sinne anderer Kompetenzfacetten (z. B. der Überzeugungen, siehe dazu Schmidt et al., 2015) abbilden, wird diese Annahme für das Grundlagenwissen zu KI als nicht tragfähig erachtet. Daraus ergibt sich die folgende Annahme:

A6: Die Merkmalsausprägung des Grundlagenwissens zu KI ist bei den Bachelorstudierenden der Wirtschaftspädagogik gleich ausgeprägt wie bei der Zielgruppe.

3.1.2.5 Zwischenfazit

Aus den Ausführungen zur Präzisierung des Messgegenstandes und zur Auswahl und Kontextualisierung der Aspekte für die Validitätsbeurteilung gehen die in Tabelle 8 zusammengefassten Validierungsschritte hervor. Dabei werden die Aspekte Testinhalt (Kapitel 2.1.2.2), Interne Struktur (Kapitel 2.1.2.3) sowie Beziehung zu anderen Merk-

⁷⁰ Die Inhalte des Studienmodells können aus der Studienordnung des Bachelorstudiengangs Wirtschaftspädagogik an der Universität Leipzig entnommen werden (vgl. WiFa UL, 2015).

malen (Kapitel 2.1.2.3) in Anlehnung an die Standards der AERA et al. (2014) über die verschiedenen Phasen der Arbeit hinweg berücksichtigt.

Die hergeleiteten Annahmen werden unter Berücksichtigung der diskutierten Standards (vgl. AERA et al., 2014) als zweckmäßig erachtet, um von den erzielten Testwerten auf das theoretisch modellierte Konstrukt des Grundlagenwissens zu KI und die damit verbundene Merkmalsausprägung zu schließen. In Ergänzung zu der Item- und Testkonstruktion (Kapitel 3.2) und der empirischen Überprüfung (Kapitel 4) wird so entlang der einzelnen Befunde ein Gesamturteil über die Validität ermöglicht.

Tabelle 8: Validierungsschritte nach den internationalen Teststandards der AERA et al. (2014). Darstellung in Anlehnung an Kuhn (2014, S. 129)

	Validierungsaspekte und Annahmen	Methode	Kapitel
„test content“	1) Items als repräsentative Menge des Messgegenstandes	Systematisierung des Messgegenstandes (Matrix)	Kapitel 2.3
	2) Inhaltliche Relevanz der Items		
	3) Angemessenheit der Items	Testkonstruktion	Kapitel 3.2
„internal structure“	4) Eindimensionalität des Konstrukts	Faktorenanalyse (N = 153)	Kapitel 4.4
„relations to other variables“	5) Ausprägung des Grundlagenwissens in Abhängigkeit zur Studienrichtung im Master	Methode der bekannten Gruppen	Kapitel 4.5
	6) Ausprägung des Grundlagenwissens in Abhängigkeit zum Studienfortschritt	Methode der bekannten Gruppen	

3.2 Item- und Testkonstruktion

3.2.1 Festlegung allgemeiner Testmerkmale

Tests werden in der pädagogischen Psychologie häufig nach Cronbach (1990, S. 32) als „Instrumente, die der systematischen Beobachtung und Beschreibung von Erleben und Verhalten mit Hilfe von Skalen (numerische Beschreibung) oder Kategorien (klassifizierende Beschreibung) dienen“, definiert.⁷¹ Die Entwicklung eines Testinstruments bedarf einer trennscharfen Bestimmung seiner Charakteristika (vgl. Bühner, 2011; Moosbrugger & Kelava, 2020; Rost, 2004). Auf Basis des bereits spezifizierten Messgegenstandes (Kapitel 3.1) und der fokussierten Test- und Kontrastgruppen wird im Folgenden insbesondere auf die Art und den Umfang des Tests und seiner Durchführung eingegangen. Darunter fallen neben der Testdauer das Testverfahren (Paper-Pencil vs. computerbasiert) und der Aufbau der Items. Der einzusetzende Test besteht aus einzelnen Items, die wiederum jeweils aus einem Itemstamm und einem Antwortformat zusammengesetzt sind (Krohne & Hock, 2007, S. 37). Damit wird das Ziel verfolgt, ein Verhalten in einer standardisierten Situation (durch Itemstamm) in einem

71 Siehe auch Krohne & Hock (2007, S. 23).

vergleichbaren Format (Antwortformat) zu erfassen (Rost, 2004, S. 55). Ziel ist es, das zu erfassende Konstrukt durch die formulierten Items bestmöglich zu repräsentieren (Krohne & Hock, 2007, S. 37, siehe auch Kapitel 3.1). Die Erfassung des Messgegenstandes soll in der vorliegenden Untersuchung systematisch erfolgen. Dem Verständnis von Krohne und Hock (2007, S. 23) folgend meint systematisch, „dass alle Personen, die den Test absolvieren, mit den gleichen Anforderungen konfrontiert werden, also z. B. dieselben Aufgaben bearbeiten“. Ob sich dieser Anspruch auf das Aufgabenformat oder andere Aspekte des Tests bezieht, lässt diese Definition offen. Ein weiterer Grundsatz für den Test besteht darin, dass die gegebenen Antworten zur weiteren Verarbeitung in einen numerischen Wert übertragen werden, um quantitative Aussagen zu ermöglichen (Krohne & Hock, 2007, S. 23).

Unter Berücksichtigung der klassischen Testtheorie (KTT) werden zur Erfassung der einzelnen Merkmale jeweils mehrere Items eingesetzt (Bortz & Döring, 2006, S. 194). Über die Größe der Itemstichprobe, d. h. die Anzahl der eingesetzten Items, lässt sich kein allgemeingültiges Urteil fällen. Rost (2004, S. 57) konstatiert, dass mit einer größeren Itemanzahl eine höhere Messgenauigkeit erwartet werden kann. Gleichzeitig betont er, dass mit einer höheren Anzahl auch negative Effekte wie Ermüdung oder Konzentrationseinbußen der Probandinnen und Probanden einhergehen können. Wie bereits in den einführenden Kapiteln dieser Arbeit dargelegt, besteht die Zielgruppe der Untersuchung aus Studierenden der Fachrichtung Berufs- und Wirtschaftspädagogik (vorzugsweise gegen Ende des Masterstudiums und unmittelbar vor dem Berufseinstieg stehend), die als angehende Lehrkräfte im berufsbildenden Bereich klassifiziert werden können. Dabei fokussiert die vorliegende Untersuchung insbesondere die kollektiven Ergebnisse und nicht die individuellen Entwicklungen der Leistungscharakteristika im Längsschnitt.⁷² Bei einer kollektiven Betrachtung auf Gruppenebene sollten die Itemschwierigkeiten idealerweise breit streuen, um in allen Ausprägungsbereichen gut differenzieren zu können (Bortz & Döring, 2006, S. 219). Der Aspekt der Itemschwierigkeiten wird auf Grundlage der erhobenen Rohdaten in Kapitel 4.3.3 näher betrachtet.

Gemäß der Definition des Messgegenstandes als kognitive Disposition im Sinne eines Grundlagenwissens zu Künstlicher Intelligenz, das innerhalb der Zielgruppe in seiner maximalen Ausprägung erfasst werden soll, handelt es sich bei dem zu entwickelnden Instrument um einen sogenannten *Leistungstest*. Nach Rost (2004, S. 43) schränkt dies die Manipulation der Testwerte bereits vor der Durchführung ein, da diese lediglich „nach unten“ durch Faktoren wie fehlende Motivation oder Testmüdigkeit beeinflusst werden können. Verzerrungen „nach oben“ seien nur durch zufälliges

72 Standardisierte Testinstrumente, die mit der Intention eingesetzt werden, Prognosen auf individueller Ebene vorzunehmen, unterliegen anderen Ansprüchen an Validität (siehe dazu Brückner et al., 2013). Neben dem Kriterium der Validität werden die Kriterien der Reliabilität und der Objektivität als Hauptgütekriterien bezeichnet. Seltener wird auch das Kriterium des Nutzens mit in die Bewertung von Testinstrumenten einbezogen (Krohne & Hock, 2007, S. 26). Dabei werden unter dem Nutzen eines Tests die auf Grundlage der Befunde abgeleiteten „Beurteilungen, Prognosen und darauf aufbauenden Empfehlungen und Entscheidungen“ subsumiert. Darüber hinaus werden diese Gesichtspunkte mit den durch die Testung anfallenden Kosten ins Verhältnis gesetzt, was dazu führen kann, dass ein Test, dessen Validität unzureichend ist, besser bewertet wird.

Raten⁷³ möglich. Im Rahmen der Leistungstests kann der vorliegende Test als sog. Niveautest (*power test*) klassifiziert werden (Krohne & Hock, 2007, S. 369).

Da die Erhebung im Rahmen regulärer universitärer Lehrveranstaltungen erfolgt, wird eine Testzeit von unter 45 Minuten angestrebt.⁷⁴ Dies impliziert Einschränkungen in der maximalen Itemanzahl, die vor dem Hintergrund der in Kapitel 3.1 getroffenen Annahmen (insbesondere zur Repräsentativität des Messgegenstandes) berücksichtigt werden muss.

Aus testökonomischen Gründen und durch den Einfluss der Covid-19-Pandemie zum Zeitpunkt der Datenerhebung wurde der Test computerbasiert umgesetzt. Dafür wurde die Plattform Limesurvey (vgl. Limesurvey, 2011) verwendet, die über den Server der Universität Leipzig betreut wird und eine DSGVO-konforme Verarbeitung der Daten sicherstellt.

Das Antwortformat ist in Abhängigkeit an die intendierten kognitiven Prozesse zu wählen (Kuhn, 2014, S. 139). Bei der Formulierung von Antwortformaten kann zwischen freien (oder offenen) und gebundenen (geschlossenen) Antwortformaten unterschieden werden (Rost, 2004, S. 59). In einem freien Antwortformat erfasste Elemente können bei der Auswertung in ein vorgefertigtes Kategoriensystem eingeordnet werden (Rost, 2004, S. 59). Bei einem gebundenen Antwortformat wählt die Testperson ihre Antwort aus vorgegebenen Antwortalternativen (auch Verhaltensalternativen) aus (Rost, 2004, S. 59). In Abhängigkeit des für die vorliegende Arbeit präzisierten Messgegenstandes kann vor dem Hintergrund der intendierten kognitiven Prozesse ein geschlossenes Format als adäquat eingeordnet werden. Geschlossene Formate in Form von Multiple-Choice (MC)-Verfahren werden im akademischen Bereich sehr häufig angewendet (Happ et al., 2016; Walstad & Wagner, 2016). Insbesondere aus testökonomischen Gründen bietet sich diese Umsetzung an, da eine objektive Auswertung mit vergleichsweise geringem Aufwand ermöglicht wird (Rost, 2004, S. 61). Allerdings führt die Wahl dieses Formats zu einer möglicherweise zeitintensiveren Formulierung von geeigneten Antwortalternativen, da ungeeignete Bestandteile häufig zu Einbußen der Validität führen (Rost, 2004, S. 61). Zudem sind disjunkte Antwortalternativen ein zentraler Anspruch (siehe Kapitel 3.2.2). Für den Messgegenstand kann die Annahme formuliert werden, dass eindeutig korrekte und eindeutig falsche Antworten existieren. Vor diesem Hintergrund werden im vorliegenden Testinstrument Items nach dem „*correct-answer*“⁷⁵-Prinzip entwickelt (Carlson, 1990, S. 161). Um Veränderungen der Merkmalsausprägung mit empirischer Evidenz zu unterlegen, wird der entwickelte Test in einem Pretest-Posttest-Design eingesetzt (siehe ausführlicher dazu Kapitel 4).

Zusammenfassend lässt sich festhalten, dass zur Erfassung des Grundlagenwissens zu Künstlicher Intelligenz bei (angehenden) Lehrkräften im berufsbildenden Bereich (Zielgruppe) ein Test konstruiert werden soll, dessen vollständige Bearbeitung inkl. der An- und Abmoderation nicht länger als 45 Minuten dauert. Um die Anforde-

73 Zur Berechnung von Ratewahrscheinlichkeiten sei auf Walstad et al. (2018) verwiesen.

74 In Anlehnung an Happ (2017, S. 75) sollte eine Testzeit von 45 Minuten nicht überschritten werden.

75 Bei anderen Messgegenständen wird aufgrund mangelnder Trennschärfe von richtigen und falschen Antwortalternativen ein „*best-answer*“-Format bevorzugt (siehe z. B. Kuhn, 2014, S. 134).

rungen an eine repräsentative Abbildung des Messgegenstandes zu erfüllen, erfolgt die Orientierung an der inhaltlichen und kognitionsbezogenen Strukturierung. Für die Durchführung wird die Online-Plattform Limesurvey als zweckmäßig betrachtet. Zur Erfassung des Grundlagenwissens wird ein geschlossenes Aufgabenformat im „correct-answer“-Format genutzt.

Im Folgenden soll nun auf die konkrete Konstruktion der Wissensitems eingegangen werden, wobei die hier diskutierten Aspekte berücksichtigt werden.

3.2.2 Itemformulierung und Itemrevision

Ausgehend von der Festlegung der allgemeinen Testmerkmale ergeben sich Anforderungen an die Formulierung der einzelnen Items. Die Itemformulierung und -entwicklung fand im Frühjahr 2021 statt. Der erste Pool bestand aus 21 Items im Single-Choice (SC)-Format mit jeweils vier Antwortmöglichkeiten, von denen eine korrekt ist. Dabei ist darauf zu achten, dass ein Item disjunkt formuliert wird und jeweils nur einen Aspekt adressiert (Rost, 2004, S. 55). Die Formulierung des Itemstamms in Form einer Frage kann entweder direkt oder indirekt erfolgen. Direkt bedeutet eine unmittelbare Erfragung eines Merkmals (z. B. das Interesse an KI), während über eine indirekte Formulierung Indikatoren erfragt werden, über die indirekt auf das zu erfassende Merkmal geschlossen werden kann (Rost, 2004, S. 58). Bezugnehmend auf das Beispiel „Interesse an KI“ könnte das z. B. über die Häufigkeit der Nutzung einzelner Systeme oder der Auseinandersetzung mit diesen erfolgen. Darüber hinaus wird zwischen einer personalisierten oder depersonalisierten Formulierung der Itemstämme unterschieden (Rost, 2004, S. 58). Um den Fokus auf die individuellen Dispositionen zu verdeutlichen, wurde im vorliegenden Fragebogen ausschließlich mit personalisierten Itemstämmen gearbeitet. Dadurch wird die persönliche Einschätzung zu Künstlicher Intelligenz adressiert.

Die Formulierung der Items kann bereits Einfluss auf die Itemschwierigkeit (siehe Kapitel 4.3.3) nehmen. Diese Besonderheit kann genutzt werden, um die Schwierigkeit des Tests in der Konstruktionsphase aktiv zu beeinflussen (Rost, 2004, S. 59).

3.2.3 Erste Testversion und Rahmenbedingungen der Testung

3.2.3.1 Testversion mit 21 Items zum Grundlagenwissen zu Künstlicher Intelligenz

Dem eingesetzten Test mit 21 Items zum Grundlagenwissen sind Hinweise zur Bearbeitung vorangestellt. Diese richten sich an das Aufgabenformat (geschlossene Items im Single-Choice-Format) und die Art der Bearbeitung (durch Anklicken der als korrekt eingeschätzten Antwortalternative). Weitere Hinweise werden den Probandinnen und Probanden nicht gegeben. Tabelle 9 enthält eine Übersicht der 21 Items, der auch die Zuordnung zu den einzelnen Inhaltsbereichen (siehe Kapitel 2.3) entnommen werden kann. Im Anhang A kann das gesamte Instrument eingesehen werden.

Tabelle 9: Zuordnung der Wissensitems zu den Inhaltsbereichen⁷⁶

Item-Nr.	Grundlagen der KI	Methoden der KI	Anwendungen der KI
W1			x
W2		x	
W3		x	
W4	x		
W5	x		
W6			x
W7			x
W8	x		
W9	x		
W10	x		
W11		x	
W12		x	
W13		x	
W14			x
W15	x		
W16			x
W17		x	
W18			x
W19			x
W20	x		
W21			x
Anzahl	7	6	8

Aus der Darstellung geht hervor, dass die Inhaltsbereiche nahezu gleichverteilt im Instrument abgebildet werden. So werden die Grundlagen der KI mit sieben Items, die Methoden der KI mit sechs Items und die Anwendungen der KI mit insgesamt acht Items erfasst. Die in den Items abgebildeten inhaltlichen Elemente gehen aus den Untersuchungen im Rahmen der Dokumentenanalyse und Experteninterviews (siehe Kapitel 2.3) hervor. Unter Berücksichtigung des in Kapitel 2.2.1 für die Arbeit definierten

⁷⁶ Der Buchstabe W in der Bezeichnung der Items steht für „Wissen“. So ist über die Itemcodes eine eindeutige Zuordnung der Items zu den erfassten Facetten möglich.

Begriffs des Wissens im Sinne eines Struktur- und Faktenwissens, das deklarativ-prozedural bzw. deklarativ-statisch klassifiziert ist, kann eine Einordnung der einzelnen Items in die aufgestellte Matrix (siehe Kap 3.1) vorgenommen werden. Aus der Darstellung (Tabelle 10) geht hervor, dass der Bereich Grundlagen der KI primär über deklarativ-statisches Faktenwissen im entwickelten Instrument abgebildet wird. Mit Bezug zu konkreten Methoden und Anwendungen der KI verlagert sich das Verhältnis von deklarativ-statischen und deklarativ-prozedural fokussierten Items tendenziell.

Tabelle 10: Einordnung der Items in die Matrix zur Systematisierung des Messgegenstandes

Inhaltliche Strukturierung	Kognitionsbezogene Differenzierung	
	Deklarativ-statisch	Deklarativ-prozedural
Grundlagen der KI	W4, W8, W9, W10, W15, W20	W5
Methoden der KI	W11, W12, W17	W2, W3, W13
Anwendungen der KI	W7, W14, W19	W1, W6, W16, W18, W21

Das ist vor dem Hintergrund der theoretischen Modellierung (Kapitel 2.3) mit den inkludierten Inhaltselementen der einzelnen Bereiche zu erklären. Die im Inhaltsbereich Grundlagen der KI enthaltenen Inhaltselemente beziehen sich laut Klassifizierung z. B. auf Meilensteine der KI-Entwicklung und grundlegende Termini. Im Rahmen der Methoden und Anwendungen der KI werden darüber hinaus Funktionsweisen und Zusammenhänge thematisiert, die einen weniger statischen Charakter aufweisen und eher strukturelle Elemente fokussieren.

3.2.3.2 Non-kognitive Kompetenzfacetten

Auf Grundlage der in Kapitel 2.5 erfolgten inhaltlichen Einordnung des Untersuchungsgegenstandes in Modelle der professionellen Kompetenz, digitalisierungsbezogene Kompetenzen und die Technologieakzeptanzforschung (Kapitel 2.1) wurden die im Modell inkludierten kognitiven und non-kognitiven Kompetenzfacetten im Rahmen der vorliegenden Studie fortlaufend operationalisiert. Ausgehend von der theoretischen Modellierung in Kapitel 2 wurden dafür etablierte Skalen aus der Kompetenz- und Technologieakzeptanzforschung (vgl. Davis, 1989; Venkatesh & Davis, 2000; Venkatesh & Bala, 2008) auf den Untersuchungskontext adaptiert. Die Gesamtheit der adaptierten Konstrukte kann aus Tabelle 11 entnommen werden. Die einzelnen Adaptionsschritte werden im Folgenden entlang der operationalisierten Kompetenzfacetten adressiert.

Tabelle 11: Übersicht der adaptierten Konstrukte zur Erfassung der non-kognitiven Kompetenzfacetten

Einstellungen		
Konstrukt		Quelle
<i>Perceived Usefulness</i>	Technology Acceptance Model (TAM)	Davis, 1989, S. 340
<i>Perceived Ease of Use</i>	Technology Acceptance Model (TAM)	Davis, 1989, S. 340
<i>Job Relevance</i>	Technology Acceptance Model 2 (TAM2)	Venkatesh & Davis, 2000, S. 201
<i>Attitude Towards Behavior</i>	Theorie der überlegten Handlung (TRA)	Fishbein & Ajzen, 1975
<i>Computer Playfulness</i>	Technology Acceptance Model 3 (TAM3)	Venkatesh & Bala, 2008, S. 279
Überzeugungen		
Objektivität des Wissens	SMILE-Projekt	Schiefele et al., 2002, S. 73
Motivation		
Gegenstandsbezogene intrinsische Motivation	SMILE-Projekt	Schiefele et al., 2002, S. 41
Berufsbezogene extrinsische Motivation	SMILE-Projekt	Schiefele et al., 2002, S. 47
Subjektive Norm	Theorie der überlegten Handlung (TRA)	Fishbein & Ajzen, 1975

Für die Erfassung der Einstellungen gegenüber KI wurden empirisch erprobte Skalen aus der Technologieakzeptanzforschung (vgl. Davis, 1989; Fishbein & Ajzen, 1975; Venkatesh & Bala, 2008; Venkatesh, 2000; Venkatesh & Davis, 2000; Venkatesh et al., 2003) für die vorliegende Untersuchung kontextualisiert. Im Einzelnen handelt es sich dabei in Anlehnung an die theoretische Modellierung (siehe ausführlicher Kapitel 2.1.3) um die Konstrukte *Perceived Usefulness*, *Perceived Ease of Use*, *Job Relevance*, *Attitude Towards Behaviour* sowie *Computer Playfulness*. Für die Adaption auf den Untersuchungsgegenstand mussten die Originalitems jeweils aus der englischen in die deutsche Sprache übersetzt werden. Die Konstrukte *Perceived Usefulness*, *Perceived Ease of Use*, *Job Relevance* und *Attitude Towards Behavior* zielen in ihrer ursprünglichen Form dabei auf ein konkretes Programm bzw. Produkt ab (Davis, 1989, S. 340; Venkatesh & Davis, 2000, S. 201). Da in der vorliegenden Untersuchung nicht ein konkretes KI-basiertes System adressiert wird, sondern die Anwendung von KI-Technologien generell betrachtet werden soll, wurde die Formulierung im Konjunktiv auch nach der Übersetzung beibehalten. Damit wird dem in Kapitel 2.1.3 herausgearbeiteten Kriterium Rechnung getragen, dass sowohl die Beschreibung der Einstellungs- als auch die der Nutzungsebene die gleiche Abstraktions- bzw. Konkretisierungsstufe innehaben. Darüber hinaus muss mit Bezug zur Zielgruppe explizit von der zukünftigen beruflichen Praxis gesprochen werden, da die in der Untersuchung erfassten Probandinnen und Probanden sich zum Zeitpunkt der Datenerhebung noch in der (universitären) Ausbildung befinden. Die adaptierten Items können aus der Tabelle 12 entnommen werden. In den Studien von Davis (1989), Venkatesh und Davis (2000) sowie Venkatesh und Bala (2008) wurden die einzelnen Items mit einer siebenstufigen Likert-Skala (vgl.

Likert, 1932) versehen. In der vorliegenden Untersuchung findet eine sechsstufige Likert-Skala Anwendung, um die Tendenz zur neutralen Mitte zu limitieren.⁷⁷ Ein hoher Wert impliziert dabei eine hohe Einstellung, während niedrige Werte (nahe 0) als ablehnende Haltung gewertet werden können.

In Anlehnung an die theoretische Modellierung (Kapitel 2.1.3) kann davon ausgegangen werden, dass eine positiv ausgeprägte Einstellung ein verlässlicher Indikator für eine Nutzung der adressierten Technologien darstellt. Daher werden in diesem Zusammenhang hohe Werte angestrebt.

Tabelle 12: Operationalisierung der Einstellungen zu Künstlicher Intelligenz

Konstrukt	Definition	Items
<i>Perceived Usefulness</i>	„the degree to which a person believes that using a particular system would enhance his or her job performance“ (Davis, 1989, S. 320)	E1. Anwendungen der Künstlichen Intelligenz würden mir helfen, Aufgaben in meinem zukünftigen Beruf schneller zu erledigen. E2. Anwendungen der Künstlichen Intelligenz würden meine Job-Performance verbessern. E3. Anwendungen der Künstlichen Intelligenz würden meine Produktivität in meinem zukünftigen Beruf steigern. E4. Anwendungen der Künstlichen Intelligenz würden mir meinen zukünftigen Beruf erleichtern. E5. Ich fände Anwendungen der Künstlichen Intelligenz nützlich in meinem zukünftigen Beruf.
<i>Perceived Ease of Use</i>	“the degree to which a person believes that using a particular system would be free of effort“ (Davis, 1989, S. 320)	E6. Die Bedienung von Anwendungen der Künstlichen Intelligenz zu erlernen, wäre einfach für mich. E7. Mir würde es leichtfallen, Anwendungen der Künstlichen Intelligenz das ausführen zu lassen, wofür sie gedacht sind. E8. Für mich wäre die Interaktion mit Anwendungen der Künstlichen Intelligenz klar und verständlich. E9. Ich fände Anwendungen der Künstlichen Intelligenz flexibel zu bedienen. E10. Es wäre leicht für mich, geschickt im Umgang mit Anwendungen der Künstlichen Intelligenz zu werden. E11. Ich fände Anwendungen der Künstlichen Intelligenz einfach in der Benutzung.
<i>Job Relevance</i>	“[...] an individual's perception regarding the degree to which the target system is applicable to his or her job“ (Venkatesh & Davis, 2000, S. 191)	E12. In meinem zukünftigen Beruf sind Anwendungen der Künstlichen Intelligenz wichtig. E13. In meinem zukünftigen Beruf sind Anwendungen der Künstlichen Intelligenz relevant.

⁷⁷ Mit der Anzahl der Stufen einer Ratingskala steigt zwar die mögliche Differenzierung, allerdings kann es auch dazu führen, dass bei zu vielen Stufen die Differenzierungskapazität der Probandinnen und Probanden erschöpft ist. Darüber hinaus wird das Vorhandensein einer neutralen Position durch eine ungerade Anzahl von Ratingstufen kontrovers diskutiert (Albers et al., 2009, S. 73).

(Fortsetzung Tabelle 12)

Konstrukt	Definition	Items
<i>Computer Playfulness</i>	“the degree of cognitive spontaneity in microcomputer interactions” (Venkatesh, 2000, S. 348)	E14. Wie würden Sie sich selbst bei der Benutzung von Computern beschreiben? ...spontan ...kreativ ...nicht erfinderisch ...spielerisch ...einfallslos
<i>Attitude Towards Behavior</i>	“an individual’s positive or negative feelings about performing the target behavior” (Venkatesh et al., 2003, S. 456)	E15. Anwendungen der Künstlichen Intelligenz zu nutzen, ist eine gute/schlechte Idee. E16. Anwendungen der Künstlichen Intelligenz zu nutzen, ist eine idiotische/weise Idee. E17. Mir gefällt die Idee, Anwendungen der Künstlichen Intelligenz zu nutzen (nicht). E18. Anwendungen der Künstlichen Intelligenz zu nutzen ist angenehm/unangenehm.

Die Überzeugungen werden in der vorliegenden Studie als epistemologische Überzeugungen (vgl. Berding & Lamping, 2014) zu Wissen und Wissenserwerb definiert (siehe Kapitel 2.2.3). Für die Operationalisierung dieser Facette wurde das Konstrukt „Objektivität des Wissens“ (Schiefele et al., 2002, S. 73) auf den Untersuchungsgegenstand adaptiert (siehe Tabelle 13). Die Items wurden in Anlehnung an das Testinstrument aus dem SMILE-Projekt (vgl. Schiefele et al., 2002) formuliert. Dabei wurden die Formulierungen inhaltlich an das Grundlagenwissen zu KI und damit den Gegenstandsbezug des Fachgebiets KI (siehe ausführlicher in Kapitel 2.3) kontextualisiert. Die Auswahl der Items erfolgte dabei in Anlehnung an Studien mit einer vergleichbaren Zielgruppe (vgl. Schmidt et al., 2015).

Tabelle 13: Operationalisierung der Überzeugungen zu Künstlicher Intelligenz

Konstrukt	Adaptierte Items
Objektivität des Wissens (Schiefele et al., 2002, S. 73)	Ü1. Es gibt viele wissenschaftliche Erkenntnisse im Fachgebiet Künstliche Intelligenz, die immer gültig sein werden. Ü2. Im Fachgebiet Künstliche Intelligenz gibt es einen festen Kern von Wissen, der von niemandem infrage gestellt wird. Ü3. Der größte Teil des Wissens im Fachgebiet Künstliche Intelligenz wird auch in Zukunft nicht überholt sein. Ü4. Die Kernaussagen im Fachgebiet Künstliche Intelligenz sind in hohem Maße objektiv. Ü5. Es ist im Fachgebiet Künstliche Intelligenz schwierig, eine eindeutige Position zu beziehen. Ü6. Alle Expertinnen und Experten im Fachgebiet Künstliche Intelligenz kommen wahrscheinlich zu den gleichen Antworten auf fachliche Fragen. Ü7. Es ist im Fachgebiet Künstliche Intelligenz oft notwendig, alternative Standpunkte zu erwägen. Ü8. Wenn man etwas in einem Lehrbuch des Fachbereichs Künstliche Intelligenz liest, dann kann man auch sicher sein, dass es wahr ist.

Die Messung erfolgt hier, wie auch bei den Items zur Erfassung der Einstellungen, über eine sechsstufige Likert-Skala (0–5). Ein hoher Wert ist als Indiz dafür zu werten, dass das Wissen zu Künstlicher Intelligenz als besonders objektiv und stabil wahrgenommen wird. Ein niedriger Wert (nahe 0) hingegen spricht für eine dynamische Wahrnehmung des Untersuchungsgegenstandes. Zusammengefasst bedeutet das, je niedriger der Wert ist, desto kritischer sind die Überzeugungen der Studierenden zu bewerten.⁷⁸ Kritische Überzeugungen sind dabei in diesem Kontext ein favorisiertes Ergebnis (vgl. Schmidt et al., 2015). Studienergebnisse auf Grundlage von Mehrebenenmodellierungen liefern Hinweise dafür, dass die Überzeugungen mit fortschreitender Studienerfahrung kritischer werden (vgl. Schmidt et al., 2015). Das bedeutet, dass die Inhaltsbereiche zunehmend hinterfragt werden. Die Dimension des Fachwissens hat nach Studienergebnissen ebenfalls einen Effekt auf kritischere Überzeugungen (vgl. Ryan, 1984; Schommer, 1993; Hofer, 2000).

Die motivationalen Dispositionen wurden in der vorliegenden Untersuchung analog zur theoretischen Modellierung in Kapitel 2.2.4 untergliedert in die intrinsische und extrinsische Motivation. Dafür wurden die Konstrukte „gegenstandsbezogene intrinsische Motivation“ (Schiefele et al., 2002, S. 41; Schiefele & Köller, 2010, S. 336) sowie „berufsbezogene extrinsische Motivation“ (Schiefele et al., 2002, S. 47; Schiefele & Köller, 2010, S. 337) auf den Untersuchungskontext angepasst. Dabei werden beide Facetten mit Bezug zum Messgegenstand kontextualisiert. Als weitere motivationale Disposition wird die Subjektive Norm in der vorliegenden Untersuchung in Anlehnung an die TRA (vgl. Fishbein & Ajzen, 1975) operationalisiert. Die Erfassung der motivationalen Dispositionen erfolgt im entwickelten Instrument ebenfalls über eine 6-stufige Likert-Skala. Dabei implizieren hohe Werte eine positiv ausgeprägte Motivation mit Bezug zum jeweiligen Konstrukt.

Tabelle 14: Operationalisierung der motivationalen Dispositionen

Konstrukt	Adaptierte Items
Gegenstandsbezogene intrinsische Motivation	M1. Ich beschäftige mich mit Künstlicher Intelligenz, weil mir die Arbeit mit den Inhalten Spaß macht. M2. Ich beschäftige mich mit Künstlicher Intelligenz, weil die Inhalte meinen persönlichen Neigungen entsprechen. M3. Ich beschäftige mich mit Künstlicher Intelligenz, weil ich die Inhalte für sehr bedeutsam halte. M4. Ich beschäftige mich mit Künstlicher Intelligenz, weil ich großes Interesse an den Inhalten habe.

⁷⁸ Dabei ist zu beachten, dass die Items Ü5 und Ü7 im Rahmen des Codierungs- und Auswertungsprozesses umgepolt werden.

(Fortsetzung Tabelle 14)

Konstrukt	Adaptierte Items
Berufsbezogene extrinsische Motivation	M5. Ich beschäftige mich mit Künstlicher Intelligenz, um später gute Berufschancen zu haben. M6. Ich beschäftige mich mit Künstlicher Intelligenz, um später ein finanziell abgesichertes Leben führen zu können. M7. Ich beschäftige mich mit Künstlicher Intelligenz, um später einen gut bezahlten Beruf ausüben zu können. M8. Ich beschäftige mich mit Künstlicher Intelligenz, um meine Einstellungschancen zu erhöhen.
Subjektive Norm	M9. Menschen, die mein Verhalten beeinflussen, denken, dass ich Anwendungen der Künstlichen Intelligenz nutzen sollte. M10. Menschen, die mir wichtig sind, denken, dass ich Anwendungen der Künstlichen Intelligenz nutzen sollte.

3.2.3.3 Ergänzende Testteile

Die Erhebung der im Strukturmodell abgebildeten kognitiven und non-kognitiven Kompetenzfacetten zu KI wurden durch die folgenden Elemente ergänzt:

- (1) Persönliches Anschreiben
- (2) Erstellung eines individuellen Codes
- (3) Soziobiografische Angaben
- (4) Erfahrungen und Interesse
- (5) Testmotivation
- (6) Selbsteinschätzung

Das auf die jeweilige Subgruppe und den jeweiligen Standort angepasste Anschreiben (Teil 1) verfolgt primär den Zweck der Motivation der Probandinnen und Probanden, an der Befragung teilzunehmen (Kuhn, 2014, S. 149). Diesen Zweck soll auch der Hinweis auf die freiwillige Teilnahme sowie die Wahrung der Anonymität erfüllen. Im Anschreiben werden die Probandinnen und Probanden über die Einbindung der Befragung in ein Forschungsprojekt der Universität Leipzig sowie über die Dauer der Befragung informiert. Die Abfrage des 7-stelligen Codes soll eine Zuordnung der Fragebögen zu verschiedenen Messzeitpunkten erlauben, ohne dass Rückschlüsse auf die Person gezogen werden können (vgl. Happ, 2017; Zlatkin-Troitschanskaia, 2013a). Darüber hinaus ist durch die Generierung des personenbezogenen Codes (im Instrument im Anhang A1) eine personalisierte Rückmeldung der Ergebnisse an die Probandinnen und Probanden möglich (vgl. Buske et al., 2011).

Im Teil der soziobiografischen Angaben (Teil 3) werden Informationen erhoben, von denen Effekte auf die Subscores erwartet werden können. Durch den Mangel an Studien zu KI-bezogenen Kompetenzfacetten in der Zielgruppe wurde sich dabei an anderen Studien im Rahmen der Kompetenzforschung bei (angehenden) Lehrkräften im berufsbildenden Bereich orientiert (z. B. im Projekt ILLEV; Kuhn, 2014; Kuhn et al., 2014; oder COACTIV, Brunner et al., 2006). Diese beinhalten neben der Abfrage des Alters und des Geschlechts Fragen zu der bisherigen schulischen, beruflichen und universitären Bildung sowie zu den jeweils erreichten Abschlüssen und den erzielten

Noten. Aufgrund der fachlichen Nähe ist in diesem Kontext besonders interessant, ob die Probandinnen und Probanden im Rahmen ihrer schulischen Vorbildung die Fächer Informatik oder Mathematik als Leistungskurs belegt haben. Darüber hinaus werden in diesem Teil die Vorerfahrungen mit Anwendungen der KI erfragt, und zwar sowohl im beruflichen als auch im privaten Kontext. Studien lassen vermuten, dass die Erfahrungen mit Anwendungen der KI das Verständnis darüber fördern (Dietzmann & Alt, 2020, S. 5174). Die privaten Erfahrungen mit KI-gestützten Systemen wurden darüber hinaus differenziert in Bezug auf Art und Umfang der Erfahrungen erhoben. Dafür wurde den Probandinnen und Probanden eine Reihe von alltäglichen⁷⁹ KI-gestützten Systemen (Sprachassistenten, Chatbots im Kundenservice, Personalisierte Werbung, Film- und Serienvorschläge bei Streamingdiensten oder Kaufvorschläge in Onlineshops) genannt und sie wurden gebeten, auf einer Skala von „nie“ bis „mehrmals täglich“ die Frequenz ihres eigenen Nutzungsverhaltens zu konkretisieren.

Darüber hinaus sind in diesem Teil selbst entwickelte Items zur Erfassung des persönlichen Interesses zu den Inhaltsbereichen Künstliche Intelligenz, Maschinelles Lernen und Big Data inkludiert.

Insbesondere bei Testverfahren mit geringem Erfolgsrisiko gewinnen motivationale und volitionale Dispositionen für die Beurteilung von Testkennwerten an Bedeutung (Förster et al., 2017a, S. 388). Bei diesen sog. *low stake tests* (vgl. Kane, 2006b; Thelk et al., 2009; Wise & DeMars, 2005) kann die Testmotivation einen Effekt auf den erreichten Score haben, der in einigen Studien sogar mit statistischer Signifikanz festgestellt werden konnte (Happ & Förster, 2018, S. 2). Aus diesem Grund wurde in der durchgeführten Untersuchung eine Skala zur Testmotivation mittels Selbsteinschätzung (vgl. Happ & Förster, 2018) eingesetzt. Die Testmotivation wird in diesem Zusammenhang definiert als die Bereitschaft, Mühe in die Bearbeitung von Testaufgaben zu investieren (Happ & Förster, 2018, S. 3).

Neben der standardisierten, testbasierten Erfassung des Grundlagenwissens ist nach jedem Wissensitem eine Selbsteinschätzung der Probandinnen und Probanden vorzunehmen, mit wie viel Sicherheit sie das Item bearbeitet haben. In der Sozialwissenschaft hat sich dafür insbesondere ein vierstufiges Antwortformat (bis 25 %, bis 50 %, bis 75 %, bis 100 %) etabliert (vgl. Kampmeyer, 2014; Kampmeyer et al., 2015). Mit diesem Teil werden primär explorative Ziele verfolgt. Die Angaben sollen im Rahmen zukünftiger Forschungsvorhaben weitere Erklärungen der Testergebnisse ermöglichen.⁸⁰

3.2.3.4 Rahmenbedingungen der Untersuchung

Die Untersuchung wurde im Sommersemester 2021 in einem Pflichtmodul des Masterstudiengangs Wirtschaftspädagogik an der Universität Leipzig (UL) pilotiert. Die

⁷⁹ Die Einschätzung und Auswahl erfolgt in Anlehnung an Michaeli et al. (2020).

⁸⁰ Es sei darauf hingewiesen, dass in dieser Monografie nicht auf die Analysen und Ergebnisse aller in der Studie erfassten Variablen eingegangen wird. Diese werden aber im Rahmen weiterer Forschungsarbeiten vertiefend untersucht und anschließend in geeigneter Form publiziert. Zum Zusammenhang von Antwortrichtigkeit und Antwortsicherheit sei auf Gigerenzer et al., 1991; Juslin, 1994; Juslin & Olson, 1997; Kröner & Biermann, 2007; Stankov, 2000; Stankov & Lee, 2008; Fritzsche, 2012; Kampmeyer, 2014; Kampmeyer et al., 2015 verwiesen.

Testung wurde dafür in ein Seminarkonzept eingebettet, das im Rahmen der Pre-Post-Untersuchung als Treatment fungiert (vgl. Schmidt & Happ, 2022b).

Aufgrund der Covid-19-Pandemie und der damit verbundenen Hygienevorschriften fand die universitäre Lehre der UL zu diesem Zeitpunkt ausschließlich online statt, weshalb das entwickelte Lehr-Lern-Konzept zunächst gänzlich von Präsenzanteilen ab-sah. Nicht zuletzt durch die Covid-19-Pandemie ist der Bedarf an Online-Lehr-Lern-umgebungen sowie hybriden Formaten im Sinne eines Blended-Learning-Ansatzes gestiegen (vgl. Chan et al., 2021; Farnell, 2021; Fretheim, 2021; OECD, 2021; Thatcher, 2020; Yu & Zhang, 2021) und insbesondere Videokonferenzen werden häufig genutzt (vgl. Niegemann, 2020). Für die erfolgreiche Konzeption und Erprobung von digital gestützten Lehr-Lern-Umgebungen sollte der Fokus auf der didaktisch-methodischen Ausgestaltung anstatt auf der Auswahl und dem Einsatz der Technologie liegen. Schneider (2018, S. 200) spricht an dieser Stelle von einer häufig auftretenden Diskrepanz zwischen technologischer Machbarkeit und didaktischer Sinnhaftigkeit. Kerres (2013, S. 130) empfiehlt den sinnbezogenen und adressatengerechten Einsatz digitaler Tools in der Lehre.

Aufgrund der Distanz zwischen den Akteurinnen und Akteuren in Online- oder hybriden Formaten kann eine Interaktion erschwert sein. Das Ermöglichen und Unterstützen von Interaktion zwischen Schülerinnen und Schülern untereinander einerseits und Schülerinnen und Schülern mit den Lehrkräften andererseits stellt daher einen weiteren Gestaltungsgrundsatz dar (Bond et al., 2020, S. 13). Konträr zu den zahlreichen Individualisierungspotenzialen, die Online- und hybriden Lehr-Lern-Formaten zugesprochen werden, steht die Forderung nach strukturgebenden Elementen. Schneider (2018, S. 212) empfiehlt in diesem Zusammenhang eine Balance zwischen zu vielen Freiheiten und zu engen Taktungen. Beides würde nach aktuellem Forschungsstand zu hohen Abbruchraten führen und somit den Erfolg von E-Learning-Maßnahmen behindern.

Das Lehr-Lern-Konzept⁸¹ in der vorliegenden Untersuchung besteht aus einer Kombination aus synchronen Seminarveranstaltungen und asynchronen Selbstlernphasen, in denen die Studierenden ihren Lernprozess selbstständig planen und durchführen können. Durch die Integration der Selbstlernphasen zeigen sich die Vorteile von Blended-Learning-Ansätzen, die ein individualisierteres Lernen ermöglichen (vgl. Schneider, 2018). Ergänzt wird dieses Angebot um die Möglichkeit der individuellen Konsultationstermine mit den Lehrenden, um adressatengerechte Hilfestellungen bei Barrieren zu leisten und den Lernerfolg in den Selbstlernphasen nicht zu gefährden. Dabei fanden die synchronen Einheiten und die individuellen Konsultationen unter Nutzung des Videokonferenzsystems Zoom statt. Für die Bereitstellung von Lehrmaterialien wurde das Learning Management System Moodle verwendet. Die digitale Kollaboration wurde mit einem Online-Whiteboard (MIRO oder Conceptboard) umgesetzt. Im Rahmen des angebotenen Moduls bearbeiteten die Studierenden das Online-Lern-

81 Das entwickelte Konzept wurde im Rahmen der Online-Jahrestagung der Sektion Berufs- und Wirtschaftspädagogik im September 2021 von der Autorin der vorliegenden Dissertation (gemeinsam mit Prof. Dr. Roland Happ) präsentiert. Insgesamt acht weitere Standorte bekundeten daraufhin Interesse, sodass ein umfangreicher Transfer stattfand, auf dessen Grundlage die empirische Studie durchgeführt wurde. Darüber hinaus wird das Konzept fortlaufend weiterentwickelt.

angebot „Einführung in die KI“ auf dem KI-Campus (vgl. KI-Campus, 2020). Da die Curricula der berufs- und wirtschaftspädagogischen Studienmodelle an der UL aktuell keine Schnittstellen für die informationstechnologischen Inhalte bieten, wurden die *Online Educational Resources* (OER) unter einer wirtschaftsdidaktischen Fragestellung bearbeitet. Ausgehend von einem theoriebasierten Evaluationsbogen (siehe ausführlicher Schmidt & Happ, 2022b) bewerteten die Studierenden die Umsetzung der OER. Dabei wurden sowohl mediendidaktische Kriterien, das Design und die Gestaltung, die Aufgabenstellungen und Lernerfolgskontrollen sowie die Möglichkeit von Feedback und Hilfestellungen im Kurs kriteriengeleitet evaluiert (siehe Näheres zum eingesetzten Lehr-Lern-Konzept in Schmidt & Happ, 2022b). Mit Bezug zum interessierenden Konstrukt der Testung und vor dem Hintergrund der Präzisierung des Messgegenstandes wurde im Rahmen des Lehr-Lern-Konzepts sichergestellt, dass alle im Test inkludierten Inhaltsbereiche Teil des Treatments sind.

Neben der Anpassung der Programme an die Entwicklungen in der Wirtschaft und Gesellschaft sowie der curricularen Verankerung ist die regelmäßige kritische Evaluation eine Gelingensbedingung für nachhaltige Bildungsangebote (Kamsker & Riebenbauer, 2022, S. 25). Für die theoretisch fundierte Evaluation des eingesetzten Lehr-Lern-Konzepts wurde eine dreigeteilte Beurteilung durch die Studierenden vorgenommen. Dabei wurden in Anlehnung an Schlicht (2012) die Bereiche „Qualität des Lehrens“, „Qualität des Lernens“ und „Qualität der Bedingungen des Lehrens und Lernens“ adressiert (Näheres dazu siehe Schmidt & Happ, 2022b).

3.3 Objektivität des Tests

Das Kriterium der Objektivität adressiert die Vergleichbarkeit der Testergebnisse verschiedener Probandinnen und Probanden (Krohne & Hock, 2007, S. 24) und gilt als Voraussetzung für die Reliabilität und Validität (Rost, 2004, S. 33, 39). Rost (2004, S. 356) weist darauf hin, dass eine hinreichende Objektivität bereits vor der Testdurchführung zu gewährleisten ist, weshalb die Beurteilung der Objektivität in der vorliegenden Arbeit bereits vor der Betrachtung der empirischen Erprobung (Kapitel 4) vorgenommen wird. In der Forschungsliteratur wird das Kriterium der Objektivität in die Aspekte der Durchführungsobjektivität, der Auswertungsobjektivität und der Interpretationsobjektivität ausdifferenziert (Bortz & Döring, 2006, S. 195; Krohne & Hock, 2007, S. 25). Eine hohe Durchführungsobjektivität beinhaltet eine fixierte Testprozedur (Krohne & Hock, 2007, S. 25) und kann z. B. durch standardisierte Instruktionen erreicht werden (Bortz & Döring, 2006, S. 195). Dies kann im vorliegenden Test als erfüllt angesehen werden, da die Testinstruktion schriftlich und somit für alle Teilnehmenden exakt identisch stattfindet. Die Auswertungsobjektivität verlangt eine vom Auswertenden unbeeinflusste Vergabe von Testpunkten für die Antworten im Test (Bortz & Döring, 2006, S. 195). Die Auswertungsobjektivität wird durch die Vorgabe von Antwortoptionen erhöht (Bortz & Döring, 2006, S. 195). In der vorliegenden Arbeit ist sie damit als erfüllt zu betrachten, da ausschließlich mit geschlossenen Items gearbeitet wurde, die im Rahmen der Auswertung in festgelegte numerische Werte überführt

wurden. Die Auswertung von Multiple-Choice-Testverfahren gilt aufgrund der vorformulierten Antwortoptionen als besonders objektiv (vgl. Förster et al., 2017a). Die Interpretationsobjektivität betrifft die Schlüsse, die auf Basis der Testergebnisse abgeleitet werden (Krohne & Hock, 2007, S. 25). In diesem Zusammenhang muss gewährleistet sein, dass verschiedene Personen auf Grundlage der Testwerte die gleichen Interpretationen ableiten. Vor dem Hintergrund der numerischen Auswertung kann dieses Kriterium als erfüllt angesehen werden.

Insgesamt ist das Kriterium der Objektivität als „recht unproblematisch“ (Bortz & Döring, 2006, S. 195) anzusehen, wenn mit standardisierten Testverfahren gearbeitet wird. Auf Grundlage der Ausführungen zu Art und Umfang des in der vorliegenden Untersuchung eingesetzten Tests wird das Kriterium der Objektivität als hinreichend geprüft und somit als erfüllt bewertet.

3.4 Dokumentation der Items zur Erfassung der KI-bezogenen Kompetenzfacetten

Aus der Entwicklung und Adaption der einzelnen Testteile ergibt sich in Anlehnung an das in Kapitel 2.5 erstellte strukturelle Rahmenmodell für die Untersuchung der kognitiven und non-kognitiven Kompetenzfacetten das in Abbildung 13 abgebildete Messmodell. Eine Übersicht in Form des Gesamtfragebogens ist der vorliegenden Dissertation als Anhang (Anhang A) beigelegt.

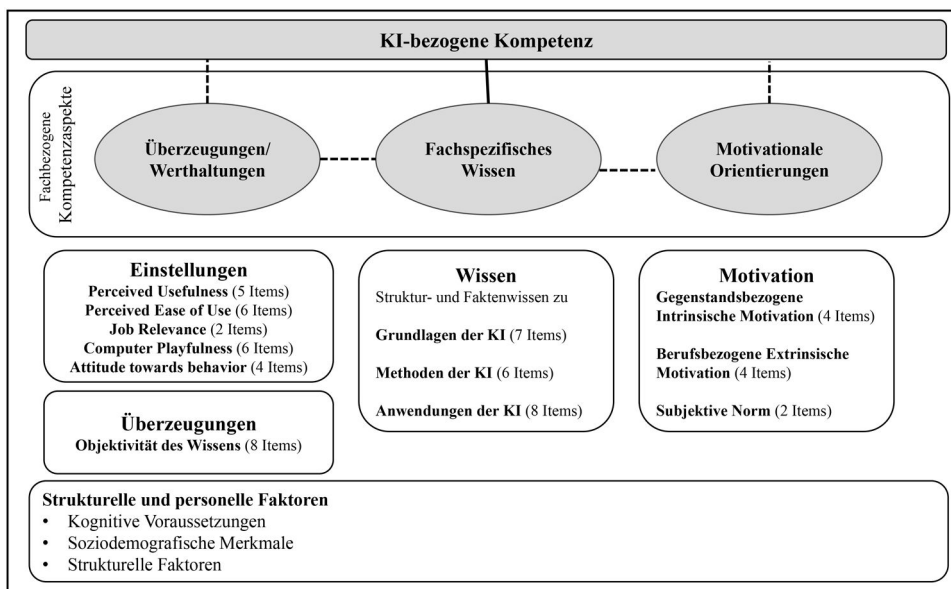


Abbildung 13: Messmodell der kognitiven und non-kognitiven Kompetenzfacetten

4 Empirische Erprobung des Testinstruments

4.1 Design

Die Datenerhebungen, um das auf Basis der theoretischen Modellierung entwickelte Instrument quantitativ zu prüfen und erste empirische Befunde zur Ausprägung der kognitiven und non-kognitiven KI-bezogenen Kompetenzfacetten der Zielgruppe zu generieren, wurden im Zeitraum zwischen April 2021 und Oktober 2022⁸² durchgeführt. Dazu wurden Erhebungen in der Zielgruppe (Masterstudierende der berufs- und wirtschaftspädagogischen Studiengänge) sowie bei den Kontrastgruppen (siehe Kapitel 3.2.1.4) durchgeführt. Die durchschnittliche Bearbeitungsdauer des online umgesetzten Fragebogens (siehe Messmodell in Kapitel 3.4 oder ausführlicher im Anhang A) betrug in der Zielgruppe 30 Minuten zum ersten Messzeitpunkt (t1) und 35 Minuten zum zweiten Messzeitpunkt (t2). Die längere Testzeit zum zweiten Messzeitpunkt ist damit zu erklären, dass in dieser Testversion weitere Items zur Evaluation⁸³ der Lehr-Lern-Umgebung (des Treatments) hinzugefügt wurden. Damit wurde die in Kapitel 3.2.1 maximale Testzeit von 45 Minuten eingehalten.

Die Erhebungen in der Zielgruppe⁸⁴ wurden jeweils innerhalb einer regulären universitären Lehrveranstaltung durchgeführt, um eine möglichst hohe Teilnahmequote sicherzustellen.⁸⁵ Bezogen auf die Zugänglichkeit lässt sich diese Verortung als halboffenes Format einstufen (Döring & Bortz, 2016, S. 337). Das Lehr-Lern-Konzept (siehe Kap. 3.2.3.4 und ausführlicher Schmidt & Happ, 2022b), in das die Pre-Post-Erhebung⁸⁶ eingebettet ist, ist in seiner Konzeption für eine Lernendengruppe von maximal 25 Studierenden ausgerichtet. Damit eine Stichprobe generiert werden kann, auf deren Grundlage belastbare statistische Analysen berechnet werden können⁸⁷, wurde das Lehr-Lern-Konzept (Treatment) inkl. der Datenerhebung im Pretest-Posttest-

82 Die Datenerhebung fand jeweils in der Vorlesungszeit des Sommersemesters 2021, des Wintersemesters 2021/2022, des Frühjahrsemesters 2022 (in der Schweiz) sowie des Sommersemesters 2022 statt. Ein einzelner Pre-Post-Zyklus umfasst dabei jeweils eine Messung zu Beginn des Semesters und eine am Ende des Semesters.

83 Die Evaluation erfolgt in Anlehnung an Schlicht (2012) in den Bereichen „Qualität des Lernens“, „Qualität des Lehrens“ und „Qualität der Bedingungen des Lehrens und Lernens“. Das Instrument wurde dabei in Anlehnung an Klausner et al. (2007), Kramer (2002), Prenzel et al. (1996) und Schlicht (2012) entwickelt. Auf Grundlage der Ergebnisse soll das Seminarekonzept fortlaufend optimiert werden (siehe dazu Schmidt & Happ, 2022b). Darüber hinaus können so ggf. weitere Erklärungsansätze für einzelne Befunde generiert werden. Die Analyse möglicher Effekte ist nicht Teil dieser Monografie und wird in weiteren Forschungsvorhaben adressiert.

84 In Anlehnung an Schnell et al. (2008, S. 267) kann in diesem Fall von einem Auswahlverfahren gesprochen werden.

85 Der Besuch von universitären Lehrveranstaltungen basiert zu einem großen Teil auf Freiwilligkeit, was dazu führen kann, dass Studierende nicht anwesend sind (Mayer & Kriz, 2010, S. 249). Es lässt sich allerdings beobachten, dass insbesondere in den ersten Semesterwochen eine vergleichsweise hohe Teilnahmequote zu verzeichnen ist (Kerres et al., 2012, S. 68). Aus diesem Grund wird die Datenerhebung zu t1 in der ersten (oder alternativ der zweiten) Veranstaltungswoche durchgeführt.

86 Für Studien mit einem ähnlichen Forschungsdesign (Pretest-Posttest-Studie mit Evaluation des Treatments) sei auf Durchholz (2023), Fechter (2015), Krüger & Kaulvers (2019), Nagel (2019) und Stork (2013) verwiesen.

87 Die Genauigkeit von Parameterschätzungen nimmt mit zunehmendem Stichprobenumfang zu (Bortz, 2005, S. 125; Bortz & Döring, 2006, S. 419), weshalb größere Stichproben zu bevorzugen sind.

Design (Döring & Bortz, 2016, S. 738) an insgesamt neun Hochschulstandorten⁸⁸ der Zielgruppe eingesetzt.

Um das Treatment (siehe Kapitel 3.2.3.4) vergleichbar zu machen und somit die einzeln erfassten Subgruppen zu einer Zielgruppe zusammenfassen zu können, erfolgte ein enger Austausch mit den jeweiligen Dozierenden. Insgesamt fanden die folgenden Erhebungsszenarien Anwendung:

1. Pre-Post-Erhebung bei Masterstudierenden der berufs- und wirtschaftspädagogischen Studiengänge, die zwischen den Erhebungszeitpunkten (t1 und t2) ein inhaltlich auf den Wissenstest abgestimmtes Lehr-Lern-Konzept (Treatment) durchliefen ($N = 81$).⁸⁹
2. Pre-Post-Erhebung bei Masterstudierenden der berufs- und wirtschaftspädagogischen Studiengänge, die zwischen t1 und t2 ein anderes Treatment erhielten, das ebenfalls einen inhaltlichen Bezug zu KI aufweist ($N = 6$).⁹⁰
3. Pre-Post-Erhebung bei Bachelorstudierenden der Wirtschaftspädagogik, die zwischen t1 und t2 kein Treatment mit KI-Bezug durchliefen ($N = 19$).
4. Pre-Erhebung bei Masterstudierenden der Wirtschaftsinformatik ($N = 27$).

Zielgruppe

Die Datenerhebung erfolgte ausschließlich im Rahmen regulärer universitärer Lehrveranstaltungen⁹¹ der insgesamt neun teilnehmenden Hochschulstandorte. Um die Motivation zur Teilnahme bei den Probandinnen und Probanden zu steigern, wurde den Studierenden eine individuelle Rückmeldung zu den Testergebnissen und damit die Aussicht auf eine Reflexion der persönlichen Entwicklung im Längsschnitt gegeben.⁹² Die einzelnen Erhebungszeiträume können aus Tabelle 15 entnommen werden.⁹³

88 Die Untersuchung wurde an den folgenden Standorten durchgeführt: Universität Leipzig, Johannes Gutenberg-Universität Mainz, PH Schwäbisch Gmünd, PH Freiburg, TU Kaiserslautern, Georg-August-Universität Göttingen, Universität Stuttgart, Universität Hamburg, PH Zürich.

89 Berichtet wird hier die Stichprobe bestehend aus den Probandinnen und Probanden, die sowohl an t1 als auch an t2 teilgenommen haben und bei denen die beiden Datensätze über den Code und/oder die Kontrollvariablen eindeutig zugeordnet werden konnten.

90 Im Rahmen der vorliegenden Arbeit werden keine expliziten Treatment-Effekte untersucht. Das ist u. a. auch mit den stark unterschiedlich großen Subgruppen zu begründen.

91 Die Erhebungen sind fest in das Seminarkonzept eingeplant. Studierende, die zu den betreffenden Terminen nicht am Seminar teilgenommen haben, wurden per Mail oder Systemnachricht im jeweils genutzten LMS zur Teilnahme aufgefordert. Die Teilnahme ist trotz der Einbettung in die Lehrveranstaltung vollständig freiwillig und eine Teilnahme oder Nicht-Teilnahme hatte für die Studierenden keine Auswirkungen auf den weiteren Verlauf des Moduls.

92 Um Testeffekte zu minimieren, erfolgte die Rückmeldung erst nach dem zweiten Messzeitpunkt. Zur positiven Wirkung von individuellen Rückmeldungen kann z. B. auf Brehm (1993), Groves et al. (2004) und Saßenroth (2013) verwiesen werden. In der Umsetzung wurde jeweils im Rahmen der letzten Veranstaltung im Semester eine gemeinsame Reflexion über die Pre-Post-Daten durchgeführt.

93 Zu den Limitationen dieses Designs siehe Kapitel 5.2.

Tabelle 15: Einzelerhebungen in der Zielgruppe

Standort	Semester	N	
		T1	T2
Universität Leipzig	SoSe 2021	20	17
	SoSe 2022	15	15
TU Kaiserslautern	WiSe 21/22	10	8
PH Freiburg	WiSe 21/22	9	8
Universität Stuttgart	WiSe 21/22	10	1
PH Zürich	FrüSe 2022	18	18
Universität Mainz	SoSe 2022	18	14
Universität Göttingen	SoSe 2022	11	6
PH Schwäbisch Gmünd	SoSe 2022	28	8
Universität Hamburg	SoSe 2022	14	2
Gesamt		153	97

Die Zielgruppe besteht zum ersten Messzeitpunkt (t1) aus $N = 153$ Probandinnen und Probanden, von denen sich 85 (55,6 %) mit dem weiblichen Geschlecht identifizieren. Im Durchschnitt sind die Studierenden 28 Jahre alt. Die gesamte Altersspanne erstreckt sich von 21 bis 54 Jahre ($MW = 28,02$; $SD = 5,79$). Alle Probandinnen und Probanden sind eingeschriebene Studierende in einem Masterstudiengang oder einem vergleichbaren Bildungsangebot.⁹⁴ Das durchschnittliche Alter der Studierenden ist vor diesem Hintergrund als erwartungskonform zu beurteilen.⁹⁵ Zum zweiten Messzeitpunkt (t2) konnten $N = 97$ Probandinnen und Probanden⁹⁶ aus der Zielgruppe erneut befragt werden.

Die Zielgruppe der angehenden Lehrkräfte im berufsbildenden Bereich ist in unterschiedlichen Studiengängen immatrikuliert. Das ist u. a. mit dem weit gefächerten Studienangebot zum beruflichen Lehramt (Frommberger & Lange, 2021, S. 524) zu begründen. Vor dem Hintergrund der Varietät der berufs- und wirtschaftspädagogischen Studiengänge kann die erfasste Zielgruppe in einzelne Subgruppen weiter ausdifferenziert werden. Aus Tabelle 16 kann die Verteilung der einzelnen in der Erhebung erfassten Studiengänge entnommen werden. Daraus wird ersichtlich, dass der Großteil (55,6 %) der erfassten Probandinnen und Probanden das Fach Wirtschaftspädago-

94 Die Probandinnen und Probanden der PH Zürich studieren im Lehrdiplom Berufsmaturität, welches nach der Einordnung im Europäischen Qualifikationsrahmen (EQR) (siehe Schweizer EQR-Zuordnungsbericht in der Version vom 17.12.2015) analog zum Masterstudium in Deutschland auf Niveau 7 zu verorten ist.

95 In Ausbildungszweigen zum berufsbildenden Lehramt ist ein vergleichsweise hohes Alter nicht ungewöhnlich, da häufig Personen mit Berufserfahrung diesen Bildungsweg einschlagen.

96 Für den Panelausfall (*panel drop-out*) können eine Reihe an Erklärungsansätzen herangezogen werden (einen Überblick liefert Happ, 2017, S. 138–140).

gik (insgesamt 28,1 % im M. Sc. WiPäd und 27,5 % im M. Ed. WiPäd)⁹⁷ studiert. Darüber hinaus wurden Studierende in den Studiengängen M. Sc. und M. A. Berufspädagogik sowie im Studiengang M. A. Pflegepädagogik befragt. An der PH Zürich wurde die Studie in einer Seminargruppe im Lehrdiplom Berufsmaturität durchgeführt. In der Gruppe „Sonstiges“ sind jene Probandinnen und Probanden subsumiert, die keine konkrete Angabe zu ihrem Studiengang machten⁹⁸, sowie Studiengänge, die weniger als fünfmal genannt wurden.⁹⁹

In keinem der in Tabelle 16 erfassten Studienmodelle ist davon auszugehen, dass Inhalte zu KI curricular verankert sind.¹⁰⁰ Die unterschiedlichen Modelle werden daher zu der Zielgruppe der angehenden Lehrkräfte im berufsbildenden Bereich zusammengefasst. Wie in Kapitel 1 herausgearbeitet, erstreckt sich das Spektrum der Berufs- und Wirtschaftspädagogik (vgl. Sektion BWP, 2014) über die Bereiche gewerblich-technisch (in diesem Fall Berufspädagogik, Technikpädagogik, Ingenieurpädagogik), kaufmännisch-verwaltend (in diesem Fall Wirtschaftspädagogik) und Gesundheit und Pflege (in diesem Fall Pflegepädagogik), die aus diesem Grund alle in der Untersuchung zu berücksichtigen sind.

Tabelle 16: Übersicht der erfassten Studiengänge (Subgruppen der Zielgruppe)

Studiengang	Häufigkeit absolut	Häufigkeit in %	Alter	Anteil weiblich
M. Sc. Wirtschaftspädagogik	43	28,1	25,63	65 %
M. A./M. Sc. Berufspädagogik	13	8,5	25,62	85 %
M. A. Pflegepädagogik	19	12,4	32,26	69 %
M. Ed. Wirtschaftspädagogik	42	27,5	26,90	36 %
Lehrdiplom Berufsmaturität	18	11,8	35,39	39 %
Sonstiges	18	11,8	26,22	50 %

Von den $N = 153$ erfassten Probandinnen und Probanden streben nach eigener Angabe 95 eine berufliche Tätigkeit im Schuldienst an. Das entspricht 62,1 % der Stichprobe. 39 Probandinnen und Probanden (25,5 %) fokussieren eine Karriere im Unternehmen und 9 Probandinnen und Probanden (5,9 %) antworteten mit „Sonstiges“. 6,5 % (10 Probandinnen und Probanden) machten hierzu keine Angabe. Einige der in der Untersuchung inkludierten Studienmodelle beinhalten neben dem Kernfach der beruflichen

97 In Deutschland wird der Masterstudiengang Wirtschaftspädagogik sowohl mit dem angestrebten Abschluss Master of Science (z. B. Universität Leipzig, TU Dresden, Goethe-Universität Frankfurt, Universität Mannheim) als auch mit dem angestrebten Abschluss Master of Education (z. B. Johannes Gutenberg-Universität Mainz, Friedrich-Schiller-Universität Jena, Georg-August-Universität Göttingen) angeboten.

98 In der Online-Maske gab es die Antwortoption „Sonstiger Studiengang“, die mit einem Freifeld zur Angabe des Studienfachs versehen war. Wurde dieses Feld freigelassen oder konnte der Studiengang keiner der anderen Gruppen zugeordnet werden, wurde die Angabe zur Gruppe „Sonstiges“ gezählt.

99 Darunter befinden sich die Studiengänge Technikpädagogik und Ingenieurpädagogik.

100 Zur grundsätzlichen Ausrichtung der berufs- und wirtschaftspädagogischen Studiengänge sei auf das Basiscurriculum für das universitäre Studienfach Berufs- und Wirtschaftspädagogik (vgl. Sektion BWP, 2014) verwiesen.

Fachrichtung¹⁰¹ ein gymnasiales Zweitfach. In Kapitel 3 wurde bereits präzisiert, dass insbesondere die Schulfächer Mathematik und Informatik eine inhaltliche Nähe zum Untersuchungsgegenstand aufweisen. Aus diesem Grund können durch die Teilnahme an Lerngelegenheiten auf akademischem Niveau in diesen Fächern Effekte auf die Testleistung erwartet werden. Drei Probandinnen und Probanden gaben an, das Zweitfach Mathematik zu studieren. Eine Person studiert das Zweitfach Informatik. Damit beträgt der Anteil der Probandinnen und Probanden mit Mathematik oder Informatik als studiertes Zweitfach lediglich 2,7 %.

Kontrastgruppe

Für die diskriminante Validierung (Schermerle-Engel & Schweizer, 2003; siehe auch Kapitel 3.1.2.4) wurde der Fragebogen zusätzlich zu den Erhebungen in den Zielsubgruppen in zwei Kontrastgruppen eingesetzt ($N = 57$). Diese bestehen wie in Kapitel 3.1.2.4 präzisiert zum einen aus Masterstudierenden der Wirtschaftsinformatik und zum anderen aus Bachelorstudierenden der Wirtschaftspädagogik. Die Datenerhebung fand in diesen Fällen ebenfalls im Rahmen regulärer universitärer Lehrveranstaltungen statt. Die Einzelerhebungen in der Kontrastgruppe der Studierenden der Wirtschaftsinformatik sind in Tabelle 17 abgebildet.

Tabelle 17: Einzelerhebungen in der Kontrastgruppe „Studierende der Wirtschaftsinformatik“

Standort	Semester	N
Universität Leipzig	SoSe 2022	5
Wirtschaftsuniversität Breslau	WiSe 21/22	22

Für die Datenerhebung an der Universität Breslau wurde der vorliegende Fragebogen in die englische Sprache übersetzt.¹⁰² Da die Unterrichtssprache in der Lehrveranstaltung, in der der Fragebogen eingesetzt wurde, ebenfalls englisch ist, sind kaum Effekte zu erwarten, die auf eine potenzielle Sprachbarriere zurückzuführen sind. Die Gesamtheit der Einzelerhebungen über die Ziel- und Kontrastgruppen ist in Tabelle 18 zusammengefasst.

101 In der Stichprobe explizit angegeben werden die beruflichen Fachrichtungen Wirtschaft und Verwaltung, Gesundheit und Pflege sowie Holztechnik.

102 Bei der Übersetzung von Testinstrumenten gibt es einige Einschränkungen zu berücksichtigen. Einen Überblick liefern Förster et al. (2017b, S. 124–126).

Tabelle 18: Zusammenfassung von Ziel- und Kontrastgruppe(n)

Zielgruppe N = 153	N	Alter	Anteil weiblich
Studierende der Berufs- und Wirtschaftspädagogik im Master ¹⁰³	153	28,02	55,6 %
Kontrastgruppe N = 57	N	Alter	Anteil weiblich
Studierende der Wirtschaftspädagogik im Bachelor	30	22,60	60 %
Studierende der Wirtschaftsinformatik	27	23,62	34,5 %

4.2 Datenaufbereitung und Umgang mit fehlenden Werten

Die erhobenen Daten wurden jeweils als csv-Datei aus der Online-Plattform Limesurvey (vgl. Limesurvey, 2011) exportiert und in IBM SPSS, Version 27 (vgl. Brosius, 2018) überführt. Dabei wurden die Angaben der Likert-Skalen von verbalen Antworten in numerische Angaben umgewandelt. Die Codierung der Antworten kann aus Tabelle 19 entnommen werden. Die Antworten der Single-Choice-Items zum Grundlagenwissen wurden dichotom codiert (richtig und falsch). Das bedeutet, dass für die Auswahl einer der drei falschen Antwortmöglichkeiten eine 0 und für die Wahl der korrekten Antwortmöglichkeit eine 1 im Datensatz vermerkt wurde (Bühner, 2011, S. 222). Fehlende Werte wurden im Datensatz mit der Einwahl 99 vermerkt, da dies als ausgeschlossene intendierte Angabe bewertet werden kann. Das Tool Limesurvey wurde mit der Einstellung versehen, dass alle Items, die nicht infolge von Trichterfragen sichtbar wurden, als Pflichtangabe abzugeben sind. So kommt es nur zu fehlenden Werten aufgrund der Trichterselektion, wenn der Test abgebrochen oder wenn die Bearbeitung gar nicht begonnen wurde.

Tabelle 19: Codierung der Likert-Skala

Verbale Aussage in Limesurvey	Numerischer Wert in SPSS
„ich stimme der Aussage nachdrücklich zu“	5
„ich stimme der Aussage zu“	4
„ich stimme der Aussage eher zu“	3
„ich lehne die Aussage eher ab“	2
„ich lehne die Aussage ab“	1
„ich lehne die Aussage nachdrücklich ab“	0

103 Für die Betrachtung der Zielgruppe im Vergleich zu den Kontrastgruppen wurden die Studierenden der Berufs- und Wirtschaftspädagogik im Master unabhängig von der studierten beruflichen Fachrichtung (Wirtschaft und Verwaltung, Technik, Gesundheit und Pflege) und weiteren potenziell differierenden Charakteristika zu einer Gruppe zusammengefasst. Die einzelnen Subgruppen sind aus Tabelle 16 zu entnehmen.

4.3 Beurteilung des Wissenstests anhand von Kriterien der klassischen Testtheorie

4.3.1 Beurteilungsverfahren und Annahmen

Zur Beurteilung, inwieweit ein beobachtbares Merkmal ein latentes Merkmal messen kann und inwieweit die theoretisch modellierten Annahmen auf Grundlage der empirischen Daten verifiziert oder falsifiziert werden können, existieren verschiedene Verfahren der Testtheorie (Döring & Bortz, 2016, S. 461). Eine prominente, an einem naturwissenschaftlichen Messmodell entwickelte Theorie zur Beurteilung der psychometrischen Charakteristika eines Tests stellt dabei die Klassische Testtheorie (KTT, *classical test theory* oder *true score theory*) dar (Döring & Bortz, 2016, S. 461). Zentral in der KTT ist die Annahme eines Messfehlers (E) zusätzlich zum wahren Testwert (T) (Bortz & Döring, 2006, S. 194; Döring & Bortz, 2016, S. 461; Krohne & Hock, 2007, S. 51; Moosbrugger & Kelava, 2020, S. 254). Für die Eigenschaften des Messfehlers E stellen Bortz und Döring (2006, S. 194) insgesamt fünf Axiome auf. Für die vorliegende Untersuchung besonders relevant ist die Annahme, dass das Testergebnis eine additive Zusammensetzung aus dem wahren Wert T und dem Messfehler E abbildet. Der Messfehler würde sich auf Grundlage dieser Axiome erst bei wiederholter Testdurchführung ausmitteln. Darüber hinaus ist die Höhe des Messfehlers unabhängig von der Ausprägung des zu messenden Merkmals und es liegt außerdem keine Korrelation mit Persönlichkeitsmerkmalen vor. Fehlereffekte können bspw. durch Müdigkeit, Unkonzentriertheit oder andere externe Effekte auftreten (Bortz & Döring, 2006, S. 194). Im Rahmen der KTT wird angenommen, dass alle eingesetzten Items dasselbe Konstrukt messen und die Relation zwischen manifesten und latenten Merkmalen linear ist (Moosbrugger & Kelava, 2020, S. 255). Neben der KTT sind die Probabilistische Testtheorie (PTT) und die Item-Response-Theorie (IRT) weit verbreitet (Döring & Bortz, 2016, S. 461). Die IRT-Skalierung muss auf Grundlage der für eine Kalibrierungsstichprobe zur robusten Parameterschätzung benötigten Stichprobengröße ausgeschlossen werden, da diese je nach Modell ein N von > 150 (vgl. Lincare, 1994), > 350 (vgl. Embretson & Reise, 2000), > 500 (Reise & Yu, 1990) und explizit bei dichotomen Items, wie sie in dieser Studie vorliegen, z. T. ein N von > 1.000 (vgl. Cella & Chang, 2000) voraussetzen. Daher erfolgt zunächst eine Einschätzung auf Grundlage der KTT. Die IRT-Skalierung wird im Rahmen anschließender Forschungsvorhaben verfolgt (siehe Kapitel 5).

Zum Zweck der Beurteilung der psychometrischen Eigenschaften des entwickelten Tests anhand der Kriterien der KTT wird im Folgenden eine Analyse auf Testebene (Kapitel 4.3.2), auf Itemebene (Kapitel 4.3.3) und auf Antwortebene (Kapitel 4.3.4) vorgenommen, bevor abschließend die Reliabilität der Messung (Kapitel 4.3.5) geschätzt wird. Die einzelnen Analyseverfahren beinhalten dabei die Einschätzung der Rohwertverteilung, der Schwierigkeit, Varianz und Trennschärfen der Items sowie eine Analyse der Antwortalternativen der MC-Items (Bortz & Döring, 2006, S. 217; Krohne & Hock, 2007, S. 45; Lienert & Raatz, 1998, S. 101–104; Moosbrugger & Kelava, 2020, S. 160). Die Schätzung der Reliabilität ist zur Sicherung der Validität als grundlegend

zu betrachten (Rost, 2004, S. 33). Diese Beurteilung ist im Rahmen der Prüfung entlang der KTT-Kriterien mit der Einschränkung verbunden, dass die Eindimensionalität der Skala zunächst ungeprüft angenommen werden muss (Schermmelleh-Engel & Werner, 2007, S. 124–128). Hohe Trennschärfen der Items können auf eine Homogenität hindeuten (Moosbrugger & Kelava, 2020, S. 156), weshalb deren Berechnung als erstes Indiz herangezogen wird. Die Homogenität der Items ist ebenfalls für die Aussagekraft der Verteilungsanalyse essenziell. In diesen Fällen zeigen sich grundlegende Probleme der Analyse entlang der KTT-Kriterien, die sich u. a. in einer mangelnden Überprüfbarkeit der Modelle äußern (Kuhn, 2014, S. 184, ausführlicher in Moosbrugger & Kelava, 2020, S. 253).

4.3.2 Verteilungsanalyse

Die Analyse der Häufigkeitsverteilung von Rohwerten ist zentral in der Erprobungsphase von Testinstrumenten (Krohne & Hock, 2007, S. 49) und verfolgt das Ziel, einen ersten Überblick zum Antwortverhalten der Probandinnen und Probanden zu generieren (Bortz & Döring, 2006, S. 218). Dabei ist insbesondere interessant, wie die Streuung der Werte zu beurteilen ist und ob die Verteilung der Werte einer Normalverteilung entspricht (Bortz & Döring, 2006, S. 218). Von einer Normalverteilung wird gesprochen, wenn die Verteilung einen glockenförmigen Verlauf zeigt, unimodal und symmetrisch ist und ca. 2/3 der Gesamtfläche zwischen den Wendepunkten liegt, ab denen sich die Verteilung asymptotisch der X-Achse nähert (Bortz, 2005, S. 42, S. 74). Zunächst empfiehlt es sich, durch die Berechnung des Modalwerts, des Medians und des arithmetischen Mittels die zentrale Tendenz der Testwertverteilung zu bestimmen (Moosbrugger & Kelava, 2020, S. 160). Durch den Vergleich der drei Kennzahlen ist eine erste Einschätzung über die Symmetrie der Verteilung möglich. Die ermittelten Messwerte zur Bestimmung der zentralen Tendenz der Messungen sind aus Tabelle 20 zu entnehmen. Auf Grundlage der Kennzahlen kann in der Untersuchung des Grundlagenwissens der Zielgruppe zu beiden Messzeitpunkten zunächst von einer symmetrischen Verteilung ausgegangen werden, wobei die Einschätzung zu t1 aufgrund der Nähe der einzelnen Kennwerte als sicherer zu bewerten ist als für den Messzeitpunkt t2.

Tabelle 20: Bestimmung der zentralen Tendenz der Messung in der Zielgruppe

Kennzahl	Wert t1	Wert t2
Modalwert	9	10
Median	9	11
Arithmetisches Mittel	8,95	11,09

Durch die dichotome Codierung der Itemantworten (Kapitel 4.2) können im Wissenstest maximal 21 Punkte erreicht werden. Der vorliegende Test wurde für die Zielgruppe der angehenden Lehrkräfte im berufsbildenden Bereich konstruiert (siehe Prä-

zisierung des Messgegenstandes in Kapitel 3.1). Daher wird diese in der Analyse der Rohwerte bevorzugt fokussiert (Abbildung 14).

Die gesamte Zielgruppe ($N = 153$) erreichte im Mittel 8,95 Punkte von möglichen 21 Punkten, bei einer Standardabweichung von 2,35 Punkten ($SE = 0,190$). Insgesamt erstrecken sich die Testwerte auf einen Bereich von 4 (Minimum) bis 14 (Maximum). Die durchschnittliche Testleistung liefert einen ersten Hinweis dafür, dass der Test für die vorliegende Zielgruppe als tendenziell zu schwierig eingeschätzt werden kann. Diese Einschätzung erwächst auch aus der Berechnung der Schiefe ($0,110$), die eine leicht rechtsschiefe Verteilung zeigt. Die darüber hinaus eher flache und breitgipflige Verteilung (Kurtosis = $-0,671$) deutet dabei auch auf tendenziell heterogene Untergruppen hin, bei denen die Mittelwerte nicht an einer Stelle gehäuft sind (siehe ausführlicher zur Beurteilung von Testwertverteilungsformen in Krohne & Hock, 2007, S. 50). Diese Annahmen sind im weiteren Verlauf zu prüfen.

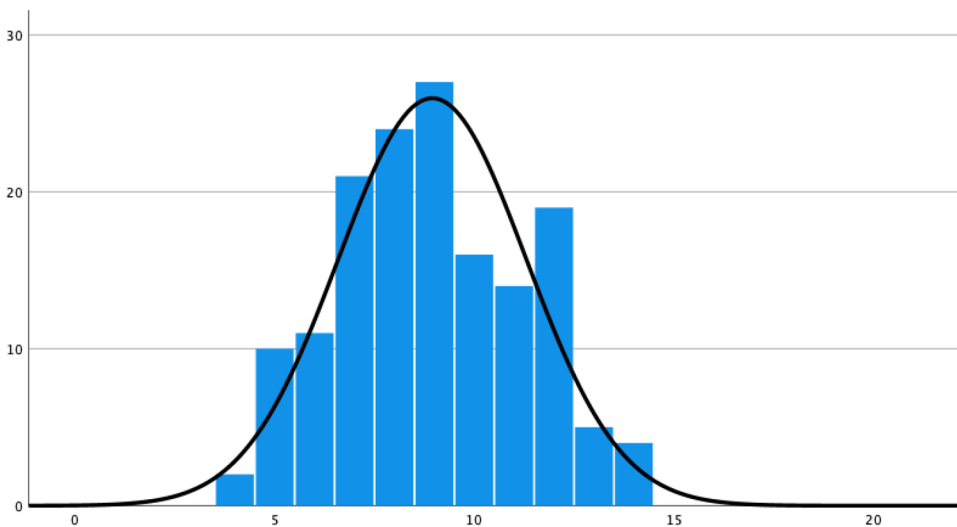


Abbildung 14: Verteilung der Summenscores der Zielgruppe zum Messzeitpunkt t1 ($N = 153$)

Der statistische Test auf Normalverteilung deutet an, dass die Annahme auf Normalverteilung für die vorliegende Zielgruppe zu verwerfen ist (Kolmogorov-Smirnov: Statistik = $0,112$; $df = 153$; $p = ,000$). Die Ergebnisse des Shapiro-Wilk-Tests zeigen allerdings eine deutliche Annäherung an die angenommene Normalverteilung (Statistik = $0,97$; $df = 153$; $p = ,002$), die sich auch im Normalverteilungsdiagramm (Q-Q-Diagramm) grafisch abbilden lässt (Abbildung 15). Darin wird deutlich, dass die erhobenen Daten insgesamt nur leicht von der erwarteten Geraden abweichen. Die Werte der Schiefe ($0,110$) und Kurtosis ($-0,671$) unterstützen dabei die Annahme, dass in diesem Fall approximativ von einer Normalverteilung ausgegangen werden kann (siehe z. B. Urban & Mayerl, 2011, S. 196).

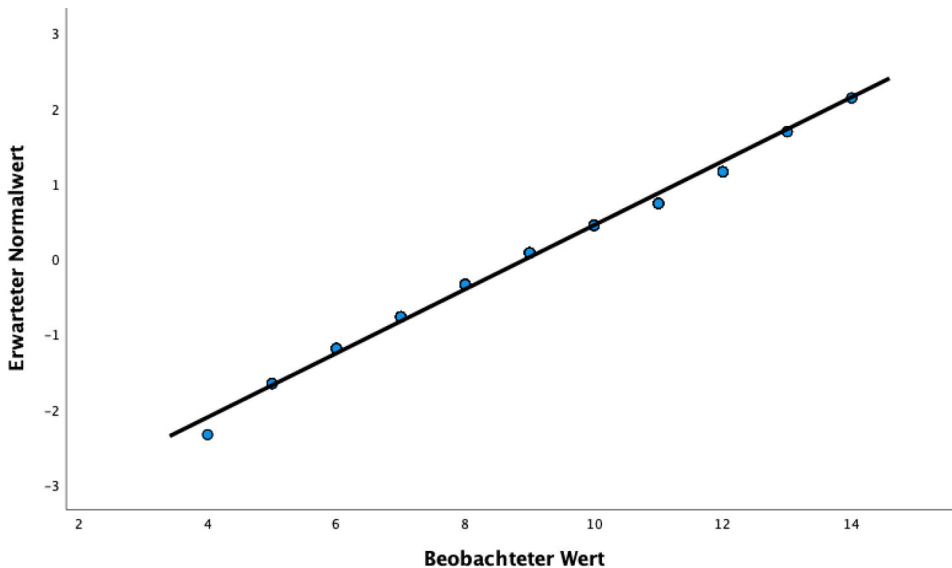


Abbildung 15: Normalverteilungsdiagramm (Q-Q-Diagramm) zum Summenscore der Zielgruppe zum Messzeitpunkt t1

Die Abweichungen von der Normalverteilung in der gesamten Zielgruppe können durch die differenzierte Betrachtung der einzelnen Subgruppen weiter erklärt werden. Aus dieser Analyse ergibt sich, dass die Studierenden im M. Sc. WiPäd den höchsten mittleren Summenscore erreichen ($N = 43$; $MW = 9,63$; $SD = 2,193$). Marginal niedriger sind die Werte der M. Ed. WiPäd ($N = 42$; $MW = 9,12$; $SD = 2,144$) und der sonstigen Studiengänge ($N = 18$; $MW = 9,22$; $SD = 2,756$). Unter dem ermittelten Durchschnitt der gesamten Zielgruppe liegen somit die Subgruppen BPäd ($N = 13$; $MW = 8,92$; $SD = 2,29$), Lehrdiplom Berufsmaturität ($N = 18$; $MW = 8,50$; $SD = 2,662$) sowie M. A. Pflegepädagogik ($N = 19$; $MW = 7,21$; $SD = 1,718$). Gemäß der Prüfung der Normalverteilungsannahme für die einzelnen Subgruppen (Tabelle 21, Histogramme der Subgruppen im Anhang C) muss die Annahme für die Subgruppen M. Ed. WiPäd und M. A. Pflegepädagogik verworfen werden. Für die anderen Subgruppen kann die Annahme der Normalverteilung zunächst beibehalten werden. Für die Kontrastgruppen kann ausgehend von der Statistik ebenfalls von einer Normalverteilung ausgegangen werden.

Tabelle 21: Prüfung der Verteilungsannahmen bei den Subgruppen zu t1

Kolmogorov-Smirnov				Shapiro-Wilk				
Subgruppen	Statistik	df	Signifikanz	Statistik	df	Signifikanz	Schiefe	Kurtosis
Ziel-Subgruppen								
M. Sc. WiPäd	0,124	43	0,093	0,918	18	0,120	−0,010	−0,595
M. Ed. WiPäd	0,223	42	0,000	0,896	42	0,001	0,711	−0,512
M. A. PPäd	0,256	19	0,002	0,902	19	0,053	−0,473	−0,922
M. Sc./M. A. BPäd	0,206	13	0,136	0,942	13	0,479	−0,086	−0,149
Lehrdiplom Berufsmaturität	0,130	18	0,200	0,925	18	0,158	−0,084	−1,205
Sonstige	0,157	18	0,200	0,918	18	0,120	−0,222	−1,331
Kontrast-Subgruppen								
B. Sc. WiPäd	0,100	29	0,200	0,957	29	0,271	−0,525	1,152
M. Sc. WINF	0,136	27	0,200	0,965	27	0,488	−0,135	−0,806

Die Verteilungsanalysen der Rohwerte der Subzielgruppen und der Kontrastgruppen liefern erste Hinweise dafür, dass die theoretischen Annahmen (Kapitel 2) mit dem entwickelten Testinstrument empirisch bestätigt werden können. Insbesondere bezieht sich diese Einschätzung auf die Vermutung des defizitären Wissens innerhalb der Zielgruppe und des besseren Abschneidens der Studierenden der Wirtschaftsinformatik (siehe für ausführlichere Analysen Kapitel 4.5).

Da das eingesetzte Instrument für die Zielgruppe im Pre-Post-Design entwickelt und eingesetzt wurde, schließt sich die Einschätzung auf Basis von Rohwertverteilungen zum zweiten Messzeitpunkt (t2) an. Als Stichprobe fungieren die Probandinnen und Probanden, die im Pre-Post-Design erfasst werden konnten und zwischen den beiden Messzeitpunkten am Treatment (siehe Kapitel 3.2.3.4) teilgenommen haben. Diese Gruppe besteht aus $N = 81$ Probandinnen und Probanden.

Die zum zweiten Messzeitpunkt erfassten Probandinnen und Probanden ($N = 81$), die am Treatment teilgenommen haben, erreichten im Mittel 11,09 Punkte von möglichen 21 Punkten, bei einer Standardabweichung von 3,458 Punkten ($SE = 0,384$). Insgesamt erstrecken sich die Testwerte auf einen Bereich von 2 (Minimum) bis 18 (Maximum). Die Berechnungen der Schiefe (-0,289) und Kurtosis (-0,147) weisen in diesem Zusammenhang auf eine im Gegensatz zur Rohwerteverteilung in t1 tendenziell links-schiefe Verteilung hin. Ähnlich wie zum Messzeitpunkt 1 kann durch den negativen Wert der Kurtosis von einer eher flachen, breitipfligen Verteilung ausgegangen werden. Die Verteilung kann aus Abbildung 16 entnommen werden.

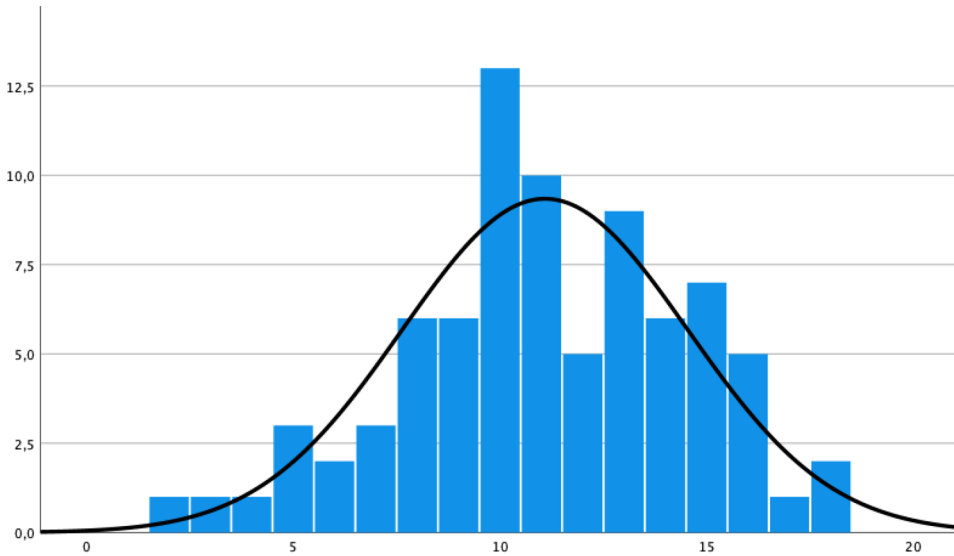


Abbildung 16: Verteilung der Summenscores der Zielgruppe zum Messzeitpunkt t2 (N = 81)

Auf Grundlage der Statistiken zum Test auf Normalverteilung kann im vorliegenden Datensatz eine Normalverteilung angenommen werden (Kolmogorov-Smirnov: Statistik = 0,093; $df = 81$; $p = ,082$ & Shapiro-Wilk: Statistik = 0,981; $df = 81$; $p = ,263$). Die Darstellung im Normalverteilungsdiagramm (Q-Q-Diagramm) unterstützt diese Annahme (Abbildung 17).

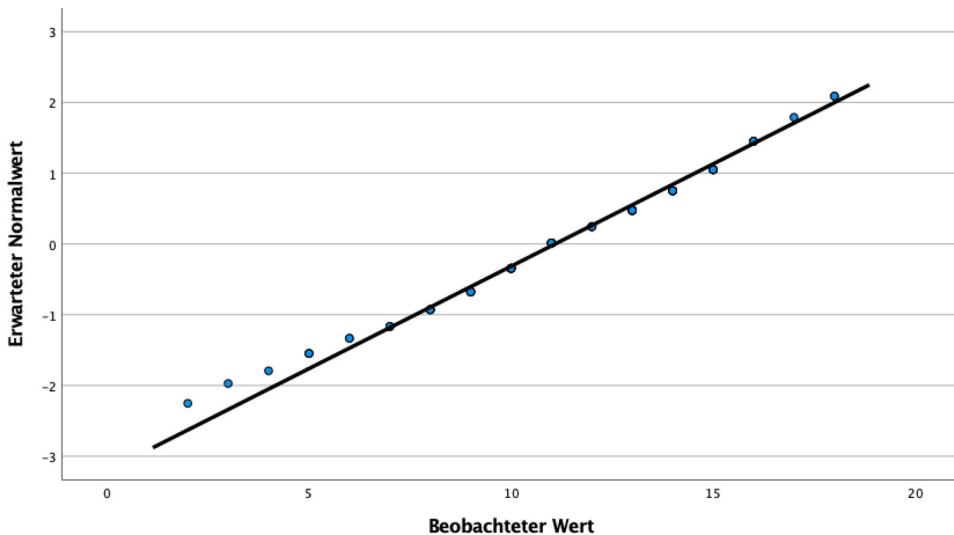


Abbildung 17: Normalverteilungsdiagramm (Q-Q Diagramm) der Zielgruppe zum Messzeitpunkt t2

Die differenzierte Betrachtung der einzelnen Ziel-Subgruppen bestätigt die Annahme einer Normalverteilung auch auf granularer Ebene. Darüber hinaus kann für die Kontrastgruppe der Bachelorstudierenden im Fach Wirtschaftspädagogik zum zweiten Messzeitpunkt ($N = 19$) ebenfalls von einer approximativen Normalverteilung ausgegangen werden (Tabelle 22).

Tabelle 22: Prüfung der Verteilungsannahmen bei den Subgruppen zu t2

Kolmogorov-Smirnov				Shapiro-Wilk				
Subgruppen	Statistik	df	Signifikanz	Statistik	df	Signifikanz	Schiefe	Kurtosis
Ziel-Subgruppen								
M. Sc. WiPäd	0,089	37	0,200	0,977	37	0,625	−0,229	0,480
M. Ed. WiPäd	0,163	19	0,197	0,935	19	0,218	−0,910	1,778
M. Sc./M. A. BPäd	0,305	4	–	0,789	4	0,084	−0,060	−5,652
Lehrdiplom Berufsmaturität	0,179	16	0,180	0,909	16	0,112	−0,093	−1,194
Sonstige	0,341	5	0,058	0,787	5	0,063	−0,597	−3,089
Kontrast-Subgruppen								
B. Sc. WiPäd	0,155	19	0,200	0,954	19	0,467	0,417	0,471

4.3.3 Itemkennwerte

Die Anwendung und Untersuchung von Testgütekriterien und die Analyse von Itemkennwerten sind bei der Neukonstruktion von Testinstrumenten von besonderer Bedeutung. Auch bei der Adaption von bereits hinreichend empirisch geprüften Skalen auf einen anderen Untersuchungsgegenstand ist es empfehlenswert, die Testgüte bei der eigenen Durchführung zu prüfen (Bortz & Döring, 2006, S. 193). In der Analyse auf Itemebene werden insbesondere die Itemschwierigkeiten und die Trennschärfen näher betrachtet.

Die Schwierigkeit bei dichotomen Items wird berechnet, indem die Anzahl der richtigen Lösungen durch die Anzahl der Gesamtantworten geteilt wird (Bortz & Döring, 2006, S. 218; Bühner, 2011, S. 223; Krohne & Hock, 2007, S. 46). Die relative Häufigkeit der korrekten Antworten bildet damit den psychometrischen Schwierigkeitsindex P ab (Bühner, 2011, S. 223). Ein Schwierigkeitsindex von 0,5 bedeutet, dass das betreffende Item von 50 % der Probandinnen und Probanden korrekt gelöst wurde. Ein höherer Wert impliziert, dass das Item von vielen Probandinnen und Probanden

gelöst wurde (Bühner, 2011, S. 223).¹⁰⁴ Einen Überblick für die Itemschwierigkeiten zu t1 liefert Tabelle 23.

Tabelle 23: Schwierigkeitsindex (*P*) der Wissensitems in der Zielgruppe zu Messzeitpunkt t1

Item	<i>P</i>	Varianz
W1	0,52	0,251
W2	0,39	0,238
W3	0,25	0,191
W4	0,22	0,174
W5	0,76	0,185
W6	0,20	0,159
W7	0,09	0,084
W8	0,37	0,235
W9	0,82	0,146
W10	0,34	0,226
W11	0,37	0,235
W12	0,72	0,203
W13	0,59	0,243
W14	0,22	0,174
W15	0,38	0,237
W16	0,48	0,251
W17	0,46	0,250
W18	0,61	0,240
W19	0,13	0,114
W20	0,47	0,251
W21	0,54	0,250

Für eine gute Differenzierung wird es als ratsam erachtet, die Itemschwierigkeiten gleichmäßig zu verteilen, um in allen Bereichen der Merkmalsausprägung ausreichend differenzieren zu können (Krohne & Hock, 2007, S. 47). Welches Maß in einem Schwierigkeitsindex empfohlen wird, ist durchaus unterschiedlich. Krohne und Hock

¹⁰⁴ Der Schwierigkeitsindex *P* sieht sich in der Literatur häufig mit der Kritik konfrontiert, dass es sich bei der Kennzahl der Interpretation nach eher um einen „Leichtigkeitsindex“ handelt. Höhere Zahlen implizieren höhere Lösungshäufigkeiten und damit tendenziell leichtere Items. Der Vorschlag, den Index dementsprechend umzubenennen, hat sich trotzdem nicht durchgesetzt (Bühner, 2011, S. 223; Krohne & Hock, 2007, S. 50).

(2007, S. 47) empfehlen einen Schwierigkeitsindex von 0,2 bis 0,8. Moosbrugger und Kelava (2007, S. 85) hingegen weiten den Bereich auf 0,05 bis 0,95 aus und formulieren somit kaum Restriktionen bezogen auf die Itemschwierigkeit. Förster et al. (2017a, S. 383) verweisen in der kritischen Auseinandersetzung mit Itemschwierigkeiten auf einen akzeptablen Bereich von 0,2 bis 0,8. Aus Tabelle 23 geht hervor, dass zum Messzeitpunkt t1 die Items W7 (0,09) und W19 (0,13) unterhalb und Item W9 (0,82) oberhalb des empfohlenen Schwierigkeitsbereichs von 0,2 bis 0,8 liegen. Mit Bezug zur von Moosbrugger und Kelava (2007, S. 85) aufgestellten Spannweite liegt keines der Items außerhalb des akzeptablen Bereichs. Die grafische Darstellung in Abbildung 18 verdeutlicht die Verteilung der Itemschwierigkeiten im vorliegenden Test.



Abbildung 18: Lösungshäufigkeiten der Zielgruppe zu Messzeitpunkt t1

Aus der Abbildung 18 wird deutlich, dass zum Messzeitpunkt t1 zwei Items (W7 und W19) als schwer (Lösungshäufigkeit von 0,00 bis 0,19), neun Items (W2, W3, W4, W6, W8, W10, W11, W14, W15) als eher schwer (Lösungshäufigkeit von 0,20 bis 0,39), sechs Items (W1, W13, W16, W17, W20, W21) als moderat (Lösungshäufigkeit von 0,40 bis 0,59), drei Items (W5, W12, W18) als eher leicht (Lösungshäufigkeit von 0,60 bis 0,79) und ein Item (W9) als leicht (Lösungshäufigkeit von 0,80 bis 1) kategorisiert werden können. Dabei häuft sich die durchschnittliche Itemschwierigkeit erwartungskonform im mittleren Bereich (siehe dazu Krohne & Hock, 2007, S. 47). Insgesamt weisen mit 14 Items zwei Drittel des Itempools eine Lösungshäufigkeit von 50 % oder weniger auf. Insgesamt weist das entwickelte Instrument bei einem Schwierigkeitsbereich von 0,09 bis 0,82 Messspielräume in beide Richtungen auf. Somit kann davon ausgegangen werden, dass in dieser Version keine Messfehler aufgrund von Decken- (*ceiling*) oder Bodeneffekten (*floor*) zu erwarten sind (siehe dazu Happ, 2017, S. 152; Zlatkin-Troitschanskaia et al., 2013a, S. 78). Das ist insbesondere vor dem Hintergrund positiv

zu bewerten, dass im Rahmen der Professionalisierung von Lehrkräften Instrumente benötigt werden, die Potenziale für Verbesserungen beinhalten (Seufert et al., 2020a, S. 2).

Aus Tabelle 24 geht der Schwierigkeitsindex (P) der Items für die Zielgruppe zum zweiten Messzeitpunkt (t_2) hervor.

Tabelle 24: Schwierigkeitsindex (P) der Wissensitems in der Zielgruppe zu Messzeitpunkt t_2

Item	P	Varianz
W1	0,68	0,221
W2	0,43	0,248
W3	0,46	0,251
W4	0,51	0,253
W5	0,84	0,136
W6	0,28	0,206
W7	0,06	0,059
W8	0,51	0,253
W9	0,89	0,100
W10	0,38	0,239
W11	0,51	0,253
W12	0,65	0,229
W13	0,62	0,239
W14	0,20	0,160
W15	0,63	0,236
W16	0,62	0,239
W17	0,72	0,206
W18	0,80	0,160
W19	0,11	0,100
W20	0,60	0,242
W21	0,59	0,244

Unter Berücksichtigung der zuvor diskutierten Richtwerte für den zu präferierenden Bereich der Itemschwierigkeiten von 0,2 bis 0,8 (Förster et al., 2017a, S. 383; Krohne & Hock, 2007, S. 47) bzw. 0,05 bis 0,95 (Moosbrugger & Kelava, 2007, S. 85) wird deutlich, dass kein Item außerhalb des Bereichs von 0,05 bis 0,95 und insgesamt vier Items (W5, W7, W9, W19) außerhalb des Bereichs von 0,2 bis 0,8 liegen. In Abbildung 19 wird die Verteilung der Itemschwierigkeiten zu t_2 visualisiert.

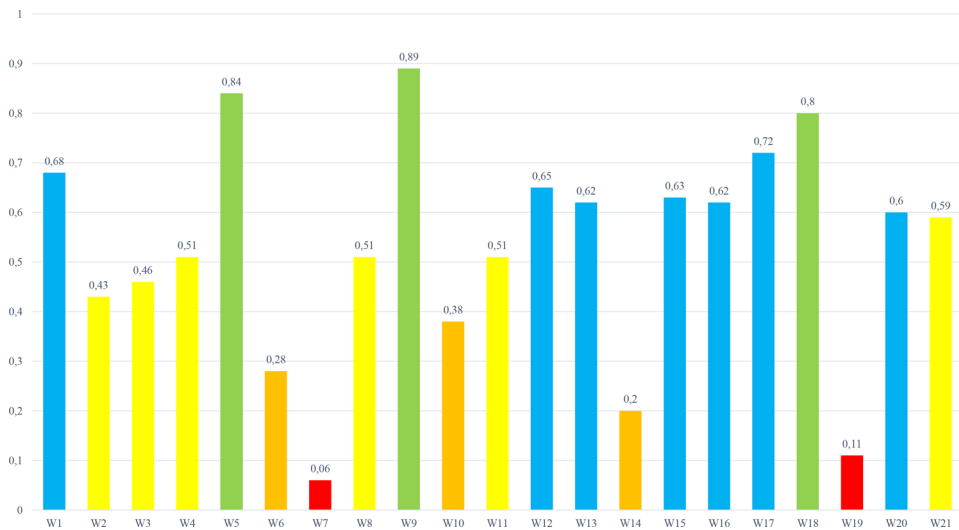


Abbildung 19: Lösungshäufigkeiten in der Zielgruppe zu Messzeitpunkt t2

Aus der Darstellung geht hervor, dass der Anteil an schweren und eher schweren Items im Gegensatz zum Messzeitpunkt t1 zurückgegangen ist. Die erwartete Häufung im mittleren Schwierigkeitsbereich (Krohne & Hock, 2007, S. 47) ist zu t2 noch deutlicher sichtbar als zu t1. Insgesamt können auf dieser Grundlage zwei Items als schwer, drei Items als eher schwer, sechs Items als moderat, sieben Items als eher leicht und drei Items als leicht klassifiziert werden. Damit weisen zum zweiten Messzeitpunkt sieben Items eine relative Lösungshäufigkeit von 50 % oder weniger auf. Im Vergleich zum ersten Messzeitpunkt ist diese Quote um 50 % gesunken. Zum zweiten Messpunkt sind erneut keine Verzerrungen aufgrund von Decken- oder Bodeneffekten sichtbar (siehe dazu Happ, 2017, S. 152).

In der Bewertung und der Auswahl der Itemschwierigkeiten¹⁰⁵ innerhalb eines Tests ist die Zielgruppe zu berücksichtigen. Die Zielgruppe Studierende der Wirtschaftspädagogik weist erwartungsgemäß wenig bis keine Vorkenntnisse auf. Diese Einschätzung erwächst aus der Analyse der formalen Lerngelegenheiten, die der Messung vorangehen (Curricula der Studiengänge, Berufsausbildungen, Schulbildung).

Aus den in Tabelle 23 (für t1) und Tabelle 24 (für t2) berichteten Varianzen lässt sich entnehmen, inwieweit die einzelnen Items zwischen den Probandinnen und Probanden streuen. Nach Rost (2004, S. 56) liefern die Items mit der größten Variation der Itemantworten die meiste Information zur Messung eines Persönlichkeitsmerkmals. Im vorliegenden Fall, in dem die Antworten auf die Wissensitems dichotom kodiert

¹⁰⁵ Die Berechnung von Itemschwierigkeiten basierend auf den Rohdaten ist nur bedingt aussagefähig. Die Einschränkung ist z. T. damit zu begründen, dass aus diesen Daten keine Aussagen zu den angewandten Lösungsstrategien ableitbar sind (Förster et al., 2017a, S. 385). Studien machen deutlich, dass die richtige Lösung neben der Anwendung klassischer vorwärts- oder rückwärtsgerichteter Strategien auch durch die Anwendung von Heuristiken, sog. *testwiseness*, Intuition oder Zufallsraten zurückzuführen ist (vgl. Abele, 2016). Für die Berechnung von Rateeffekten in Multiple-Choice-Assessments sei auf Walstad et al. (2018) verwiesen.

werden, bedeutet das, dass die Varianz maximal ist, wenn das Item eine relative Lösungshäufigkeit von 50 % aufweist. Die maximale Varianz liegt in diesem Fall bei 0,25 (Rost, 2004, S. 56). Aus Tabelle 23 lässt sich entnehmen, dass dies für Messzeitpunkt t1 auf die Items W1, W16, W17, W20 und W21 und für Messzeitpunkt t2 (siehe Tabelle 23) auf die Items W3, W4, W8 und W11 zutrifft. Davon ausgehend sind Items in den Randbereichen der Schwierigkeitsbewertung eher weniger gut geeignet, um zwischen den Probandinnen und Probanden zu differenzieren. Im vorliegenden Test weist erwartungskonform das schwierigste Item (W7) die niedrigste Varianz (0,084 in t1 und 0,059 in t2) auf. Insgesamt weisen zu t1 acht Items und zu t2 sechs Items eine Varianz von $< 0,2$ auf, was dazu führt, dass diese Items als eher ungeeignet zu bewerten sind (Krohne & Hock, 2007, S. 47).

Eine erste Einschätzung darüber, inwieweit die Antwort auf einzelne Items Rückschlüsse auf das Gesamtergebnis im Test ermöglicht, kann über die Berechnung des korrigierten Trennschärfekoeffizienten generiert werden (Bortz & Döring, 2006, S. 219). Dabei ist die Trennschärfe (*item discrimination power*) definiert als „Korrelation der Beantwortung dieses Items mit dem Gesamtwert“ (Bortz & Döring, 2006, S. 219). Der zu berechnende Koeffizient zur Beurteilung der Trennschärfe ist dabei abhängig vom eingesetzten Skalenniveau (vgl. Bortz, 2005). Je höher die Item-Test-Korrelation ausgeprägt ist, desto besser ist das Item geeignet, um zwischen Probandinnen und Probanden mit hohem und Probandinnen und Probanden mit niedrigem Testscore zu differenzieren (Bortz & Döring, 2006, S. 219; Krohne & Hock, 2007, S. 48). Ab welchem Wert Trennschärfekoeffizienten als substanziell und zufriedenstellend bewertet werden, ist nicht konsistent geklärt. Förster et al. (2017a, S. 382) postulieren den Wert von 0,2 als Grenzwert. Krohne und Hock (2007, S. 48) legen einen Richtwert von 0,3 fest. Items, die geringere Trennschärfen aufweisen, sollten aus dem Test entfernt werden. Als „gute Items“ sind jene mit einer Trennschärfe ab 0,50 oder 0,60 zu bewerten (Krohne & Hock, 2007, S. 48). Mit der Entfernung einzelner Items aus dem Test geht allerdings die Gefahr einher, dass das Instrument an inhaltlicher Breite verliert und damit Einbußen im Rahmen der Konstruktrepräsentativität gemacht werden müssen (Förster et al., 2017a, S. 382). Moosbrugger und Kelava (2020, S. 155) empfehlen daher, lediglich Items mit einer Trennschärfe nahe 0 oder im negativen Bereich aus dem Test zu entfernen. Bei einer Trennschärfe nahe 0 kann das Item demnach als ungeeignet bewertet werden, um zwischen Probandinnen und Probanden mit hohem Testwert und Probandinnen und Probanden mit niedrigem Testwert zu differenzieren. Bei Trennschärfen im negativen Bereich ist davon auszugehen, dass insbesondere Probandinnen und Probanden mit niedrigen Testwerten das Item korrekt beantworten, was zu Verzerrungen bei der Testwertbewertung führen kann.

Tabelle 25: Korrigierte Trennschärfen der Fachitems

Korrigierte Trennschärfen		
Item	Zielgruppe t1	Zielgruppe t2
W1	−0,079	0,262
W2	0,163	0,397
W3	−0,007	0,232
W4	0,037	0,242
W5	0,157	0,360
W6	0,032	−0,028
W7	−0,106	0,134
W8	0,205	0,282
W9	0,178	0,292
W10	0,139	0,413
W11	0,025	0,266
W12	0,083	0,165
W13	0,156	0,433
W14	0,050	−0,038
W15	0,084	0,287
W16	−0,053	0,276
W17	0,121	0,254
W18	0,038	0,229
W19	−0,158	−0,055
W20	0,085	0,455
W21	−0,055	0,200

Aus Tabelle 25 sind die korrigierten Trennschärfekoeffizienten der Wissensitems zu entnehmen. Auf Grundlage des berechneten Koeffizienten und der aufgeführten Bezugsgröße von 0,2 (Förster et al., 2017a, S. 382) zeigt sich, dass zum ersten Messzeitpunkt nur ein Item (W8) eine ausreichend hohe Trennschärfe aufweist. Alle anderen Items wären demnach als ungeeignet zu bewerten und aus dem Test zu entfernen. Da auf Basis eines Items nicht mehr von einem Test gesprochen werden kann, werden zunächst nur jene Items kritischer betrachtet, die eine negative Trennschärfe aufweisen. Das betrifft zum Messzeitpunkt t1 insgesamt sechs Items (W1, W3, W7, W16, W19, W21). Mit Bezug zur theoretischen Modellierung (Kapitel 2.3) und der Einordnung der entwickelten Items in die Inhaltsbereiche (Kapitel 3.2.3.1) würde ein Ausschluss dieser Items bedeuten, dass aus dem Bereich „Anwendungen der KI“ fünf von

acht Items zu entfernen sind. Es ist allerdings fraglich, inwieweit das Subkonstrukt auf Basis von drei Items angemessen im Test repräsentiert werden kann.

Bei der Analyse der korrigierten Trennschärfen zu t2 wird deutlich, dass insgesamt 16 Items unter Berücksichtigung des Grenzwerts von 0,2 (Förster et al., 2017a, S. 382) über eine ausreichend hohe Trennschärfe verfügen. Im Gegensatz zur Befundlage zu t1 weisen zum Messzeitpunkt t2 nur drei Items eine negative Trennschärfe auf und sind damit als ungeeignet einzuordnen. Die Diskrepanz der Itemkennwerte zwischen t1 und t2 liefert einen ersten Hinweis für ein eher unsystematisches Antwortverhalten der Zielgruppe zu t1. Diese Annahme wird im Rahmen der Reliabilitätsschätzung (Kapitel 4.3.5) weiter verfolgt.

4.3.4 Analyse der Antwortalternativen

Nachdem durch die Betrachtung der Rohwertverteilung eine Analyse auf Testebene und durch die Betrachtung der Itemkennwerte eine Analyse auf Itemebene durchgeführt wurden, folgt nun die Analyse auf Antwortebene. Wie bereits in Kapitel 3.2 diskutiert, werden mit der Konstruktion von geschlossenen Aufgaben im Antwort-Wahl-Verfahren (MC/SC) zusätzliche Herausforderungen assoziiert. Alle 21 Wissensitems sind im Single Choice-Format (SC) mit jeweils vier Antwortmöglichkeiten entwickelt worden (siehe Kapitel 3.2). Um mögliche Distraktoren zu identifizieren, die die Interpretation der Testwerte erschweren, wird eine Analyse der Antwortalternativen durchgeführt. Unter dem Begriff der Distraktoren sind Antwortoptionen subsumiert, die als Alternativantworten im Test enthalten sind und im Idealfall in etwa gleich häufig ausgewählt werden (Lienert & Raatz, 1998, S. 101). Distraktoren sollten eine negative Trennschärfe besitzen und demnach häufiger von Probandinnen und Probanden mit einer schlechten Testleistung ausgewählt werden als von Probandinnen und Probanden mit einer guten Testleistung (Lienert & Raatz, 1998, S. 103).

Für die Analyse auf Distraktoren werden in Tabelle 26 die absoluten und prozentualen Häufigkeiten der einzelnen Items berichtet. Darüber hinaus enthält die Tabelle die Angabe der mittleren Testleistung der Probandinnen und Probanden, die die betreffende Antwortalternative während der Testdurchführung ausgewählt haben (Darstellung in Anlehnung an Kuhn, 2014, S. 191).

Tabelle 26: Distraktorenanalyse zu t1 und t2

Item	Antwort- option	Codierung (Score)	T1 (N = 153)			T2 (N = 81)		
			Häufigkeit (absolut)	Häufigkeit (%)	Mittlere Testleistung	Häufigkeit (absolut)	Häufigkeit (%)	Mittlere Testleistung
W1	A	1	76	49,7	9,17	55	67,9	12,00
	B	0	26	17,0	9,65	6	7,4	6,50
	C	0	15	9,8	6,93	6	7,4	11,67
	D	0	36	23,5	8,81	14	17,3	9,21
W2	A	0	17	11,1	9,06	9	11,1	8,78
	B	0	54	35,3	8,28	27	33,3	9,41
	C	1	59	38,6	10,02	35	43,2	13,11
	D	0	23	15	7,70	10	12,3	10,60
W3	A	0	24	15,7	7,83	6	7,4	8,17
	B	1	39	25,5	9,67	37	45,7	12,46
	C	0	7	4,6	7,14	4	4,9	7,75
	D	0	83	54,2	9,08	34	42	10,50
W4	A	0	55	35,9	8,45	15	18,5	9,07
	B	1	34	22,2	9,88	41	50,6	12,37
	C	0	15	9,8	8,93	5	6,2	10,20
	D	0	49	32	8,86	20	24,7	10,20
W5	A	1	113	73,9	9,34	68	84,0	11,76
	B	0	5	3,3	9,40	1	1,2	10,00
	C	0	13	8,5	8,23	4	4,9	9,75
	D	0	22	14,4	7,27	8	9,9	6,13
W6	A	1	30	19,6	9,90	23	28,4	11,65
	B	0	67	43,8	8,67	23	28,4	11,43
	C	0	31	20,3	8,26	20	24,7	9,95
	D	0	25	16,3	9,40	15	18,5	11,20
W7	A	0	104	68,0	8,97	65	80,2	10,88
	B	1	18	11,8	9,44	5	6,2	13,80
	C	0	18	11,8	8,67	3	3,7	10,33
	D	0	13	8,5	8,46	8	9,9	11,38
W8	A	0	15	9,8	8,00	8	9,9	7,87
	B	1	57	37,3	10,16	41	50,6	12,49
	C	0	61	39,9	8,38	28	34,6	10,39
	D	0	20	13,1	7,95	4	4,9	8,00

(Fortsetzung Tabelle 26)

Item	Antwort- option	Codierung (Score)	T1 (N = 153)			T2 (N = 81)		
			Häufigkeit (absolut)	Häufigkeit (%)	Mittlere Testleistung	Häufigkeit (absolut)	Häufigkeit (%)	Mittlere Testleistung
W9	A	0	14	9,2	7,14	5	6,2	7,00
	B	0	4	2,6	7,25	2	2,5	6,00
	C	1	126	82,4	9,31	72	88,9	11,54
	D	0	9	5,9	7,44	2	2,5	10,00
W10	A	1	52	34,0	10,04	31	38,3	13,39
	B	0	4	2,6	6,50	3	3,7	11,00
	C	0	42	27,5	8,48	23	28,4	9,17
	D	0	55	35,9	8,45	24	29,6	9,96
W11	A	0	16	10,5	9,88	8	9,9	11,38
	B	0	49	32,0	8,51	24	29,6	9,50
	C	0	31	20,3	7,87	8	9,9	8,63
	D	1	57	37,3	9,65	41	50,6	12,44
W12	A	1	110	71,9	9,35	53	65,4	11,83
	B	0	5	3,3	7,80	3	3,7	9,33
	C	0	26	17,0	7,81	19	23,5	10,00
	D	0	12	7,8	8,25	6	7,4	8,83
W13	A	0	24	15,7	8,29	14	17,3	8,93
	B	0	30	19,6	8,03	12	14,8	8,33
	C	0	11	7,2	7,55	5	6,2	9,00
	D	1	88	57,5	9,61	50	61,7	12,56
W14	A	0	41	26,8	8,41	19	23,5	10,37
	B	0	44	28,8	9,05	31	38,3	11,97
	C	0	34	22,2	8,47	15	18,5	9,60
	D	1	34	22,2	9,94	16	19,8	11,62
W15	A	1	58	37,9	9,81	51	63,0	12,18
	B	0	37	24,2	8,62	14	17,3	9,71
	C	0	29	19,0	8,62	8	9,9	9,63
	D	0	29	19,0	7,97	8	9,9	8,00
W16	A	0	31	20,3	8,39	7	8,6	8,86
	B	0	11	7,2	9,09	6	7,4	8,67
	C	1	73	47,7	9,34	50	61,7	12,18
	D	0	38	24,8	8,61	18	22,2	9,72

(Fortsetzung Tabelle 26)

Item	Antwort- option	Codierung (Score)	T1 (N = 153)			T2 (N = 81)		
			Häufigkeit (absolut)	Häufigkeit (%)	Mittlere Testleistung	Häufigkeit (absolut)	Häufigkeit (%)	Mittlere Testleistung
W17	A	0	27	17,6	8,26	11	13,6	8,91
	B	0	23	15,0	8,65	6	7,4	7,83
	C	0	32	20,9	7,88	6	7,4	10,50
	D	1	71	46,4	9,79	58	71,6	11,90
W18	A	1	92	60,1	9,41	65	80,2	11,66
	B	0	36	23,5	8,56	11	13,6	7,73
	C	0	15	9,8	8,13	3	3,7	10,67
	D	0	10	6,5	7,30	2	2,5	11,50
W19	A	0	91	59,5	9,03	50	61,7	11,00
	B	0	12	7,8	8,83	10	12,3	11,90
	C	1	20	13,1	8,85	9	11,1	11,44
	D	0	30	19,6	8,80	12	14,8	10,50
W20	A	0	19	12,4	8,47	10	12,3	7,20
	B	0	50	32,7	8,30	17	21,0	9,47
	C	1	72	47,1	9,68	49	60,5	12,65
	D	0	12	7,8	8,00	5	6,2	9,00
W21	A	1	83	54,2	9,29	48	59,3	12,04
	B	0	12	7,8	8,42	8	9,9	8,63
	C	0	51	33,3	8,63	21	25,9	10,90
	D	0	7	4,6	8,14	4	4,9	5,50

Aus der Analyse der Distraktoren geht hervor, dass zu Messzeitpunkt t1 insgesamt fünf Items (W3, W8, W10, W14) jeweils einen Distraktor aufweisen, der häufiger ausgewählt wurde als die korrekte Antwortalternative. Das Item W7, welches auf Grundlage der Rohwertverteilung als einfachstes Item gilt, hat mit der Antwortoption A einen eindeutigen Distraktor (68 %) und zusätzlich eine weitere falsche Antwortoption, die genauso häufig ausgewählt wurde wie die richtige Antwort (je 11,8 %). Die Items W4, W6 und W19 weisen jeweils zwei Distraktoren auf, die häufiger ausgewählt wurden als die korrekten Antworten auf die Items.

Zum zweiten Messzeitpunkt kann bei Item W7 ebenfalls ein Distraktor (Antwortoption A) identifiziert werden, der häufiger ausgewählt wurde als die korrekte Antwort. In Item W14 können zu t2 zwei Distraktoren festgestellt werden und bei Item W19 werden zum zweiten Messzeitpunkt alle falschen Antwortoptionen häufiger ausgewählt als die korrekte Antwort.

Insgesamt lässt sich beobachten, dass zu beiden Messzeitpunkten zahlreiche Items nicht dem Anspruch genügen, dass die Alternativantworten weniger häufig aus-

gewählt werden als die korrekte Antwort, und dass die Wahrscheinlichkeit, mit der die Alternativantworten ausgewählt werden, ungefähr gleich ist. Diese Befunde weisen auf einen notwendigen Revisionsprozess hin, der im Rahmen von Kapitel 5 weiter präzisiert wird.

4.3.5 Reliabilität

Wie bereits erwähnt stellt das Kriterium der Reliabilität neben der Objektivität und der Validität ein zentrales Gütekriterium bei der Beurteilung des entwickelten Testinstruments dar (Bühner, 2011, S. 58). Durch die Beurteilung der Objektivität (siehe dazu Kapitel 3.3) wurde die Basis für eine ausreichende Reliabilität der Testwerte geschaffen (Rost, 2004, S. 33). Für die in der vorliegenden Arbeit fokussierte Beurteilung der Validität ist die Schätzung der Reliabilität essenziell. Nach Rost (2004, S. 33) kann mit einer geringen Reliabilität kaum eine hohe Validität einhergehen. Krohne und Hock (2007, S. 25) bezeichnen analog die Reliabilität als „notwendige, aber keine hinreichende Bedingung für die Validität“.

Die Reliabilität einer Testung beschreibt den Grad an Präzision, mit dem das zu erfassende Merkmal gemessen werden konnte (Bortz & Döring, 2006, S. 196; Krohne & Hock, 2007, S. 25). Unter Einhaltung der Kriterien der KTT ist ein Test als reliabel zu bewerten, wenn er unter gleichen Bedingungen reproduzierbare Ergebnisse ermöglicht (Bortz & Döring, 2006, S. 196). Eine Reliabilitätsprüfung ist deshalb sinnvoll, weil in einer Testung nicht alle Aufgaben, die zur vollständigen Erfassung eines Merkmals nötig wären, gestellt werden können. Aus diesem Grund wird von einer sogenannten Verhaltensstichprobe gesprochen (Krohne & Hock, 2007, S. 25). Aufgrund dieser Limitation ist eine gänzlich fehlerfreie Testung nahezu ausgeschlossen.

Moosbrugger und Kelava (2020, S. 281) definieren die Reliabilität als „Anteil der wahren Varianz [$Var(Y)$] an der Gesamtvarianz [$Var(T)$]“. Für die Schätzung der Reliabilität existiert eine Reihe verschiedener Verfahren (Krohne & Hock, 2007, S. 51). Dabei ist besonders die Betrachtung des zu erwartenden (zufälligen und unkorrelierten) Messfehlers (E) relevant. Je kleiner der Messfehler einer Testung ist, desto besser ist seine Reliabilität zu beurteilen (Moosbrugger & Kelava, 2020, S. 282). Der Reliabilitätskoeffizient ($Rel(Y)$) wird demnach berechnet als

$$Rel(Y) = \frac{Var(T)}{Var(Y)} = \frac{Var(T)}{Var(T) + Var(E)}$$

und kann einen Wert zwischen 0 (keine Reliabilität) und 1 (höchste Reliabilität) annehmen, wobei der Wert 1 implizieren würde, dass kein Messfehler existent ist und somit die wahre Varianz und die Testwertvarianz kongruent sind (Moosbrugger & Kelava, 2020, S. 282).

Um die Reliabilität ausgehend von der KTT schätzen zu können, sind mehrere Messungen eines Merkmals zwingend (Moosbrugger & Kelava, 2020, S. 281). Diesem Kriterium kann entweder durch mehrere Items in einem Test, den mehrfachen Einsatz des Tests zu unterschiedlichen Zeitpunkten (Retest-Reliabilität) oder auch mehrere Messungen mit verschiedenen Testversionen (Paralleltest-Reliabilität) entsprochen werden (Krohne & Hock, 2007, S. 57–58; Moosbrugger & Kelava, 2020, S. 322).

Die Parelleltest-Methode kann aufgrund ihrer anspruchsvollen Konstruktion und Durchführung (siehe dazu Schnell et al., 2008, S. 152) als für die vorliegende Untersuchung nicht realisierbar bewertet werden. Eine erneute Testdurchführung bei denselben Probandinnen und Probanden wird grundsätzlich als machbar eingestuft. Es stellt sich in diesem Zusammenhang allerdings z. B. die Frage nach einem geeigneten Zeitintervall, um Erinnerungseffekte zu vermeiden. Darüber hinaus ist die Untersuchung im Pretest-Posttest-Design mit Treatment angelegt, sodass die Ausprägung des interessierenden Konstrukts (Grundlagenwissen der KI) innerhalb dieser Zeitspanne (gewünschten) Veränderungen unterliegt. Das bedeutet, dass selbst bei gleichbleibenden Messfehlerwerten eine Verzerrung der Reliabilitätsschätzung erwartet werden kann. Aus diesen Gründen wird die Option der Testung durch mehrere Items präferiert, auf deren Grundlage die Beurteilung erfolgen kann. Durch die sogenannte interne Konsistenz ist eine recht stabile Schätzung der Reliabilität möglich (Bortz & Döring, 2006, S. 198; Moosbrugger & Kelava, 2020, S. 201). Die interne Konsistenz eines Tests wird hauptsächlich mit dem Alphakoeffizienten (vgl. Cronbach, 1951) gemessen. Das Verfahren ist sowohl für dichotome als auch für polytome Items geeignet (Bortz & Döring, 2006, S. 198) und somit auf den gesamten Test anwendbar.¹⁰⁶ Alpha (α) wird dabei auf Grundlage der Anzahl der Items (n), der Varianz der Items (σ_i^2) sowie der Varianz des Tests (σ_x^2) berechnet als

$$\alpha = \frac{n}{n-1} \left[1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right]$$

und nimmt einen Wert zwischen 0 und 1 an, wobei höhere Werte für eine bessere Reliabilität sprechen (Schnell et al., 2008, S. 153). Als Maßstab für eine akzeptable Reliabilität nach Cronbach's alpha gelten meist Werte ab 0,8 (Krohne & Hock, 2007, S. 65; Moosbrugger & Kelava, 2020, S. 320; Schnell et al., 2008, S. 153). Wenn Testinstrumente lediglich zur Bestimmung von Gruppenmittelwerten eingesetzt werden oder für den Untersuchungszweck noch keine (besseren) Testinstrumente etabliert sind, können auch niedrigere Werte toleriert werden (Krohne & Hock, 2007, S. 65). Dies stellt vor dem Hintergrund des neu entwickelten Instruments zur Betrachtung kollektiver Merkmalsausprägungen auf Gruppenebene in der vorliegenden Untersuchung eine wichtige Erweiterung des Gütekriteriums dar.

Die in Kapitel 4.3.3 analysierten Trennschärfen der einzelnen Items zur Erfassung des Grundlagenwissens zu KI deuten bereits darauf hin, dass die vorausgesetzte τ -Äquivalenz (Krohne & Hock, 2007, S. 97; Moosbrugger & Kelava, 2020, S. 317) der Items nicht angenommen werden kann. Demnach muss sich der Schätzung der Reliabilität nach Cronbach's alpha die Analyse der Faktorladungen anschließen (Kapitel 4.4). Der korrigierte Reliabilitätskoeffizient, der für die Zielgruppe in t1 nach Selektion der Items mit negativen Item-Skala-Korrelationskoeffizienten (insgesamt 8 Items) für eine Skala mit 13 Items berechnet wurde, kann mit einem Wert von 0,437 als

106 Bortz und Döring (2006, S. 198) weisen allerdings darauf hin, dass dieser Wert die Reliabilität insbesondere bei heterogenen bzw. mehrdimensionalen Tests häufig unterschätzt. Da im Rahmen der Testerprobung aktuell ungeprüft von einer eindimensionalen Skala ausgegangen werden muss, ist diese potenzielle Einschränkung bei der Interpretation zu berücksichtigen.

grundsätzlich nicht akzeptabel eingestuft werden. Da es sich hierbei allerdings um ein neu entwickeltes Testinstrument handelt, das auf Grundlage ausschließlich dichotom codierter SC-Items das interessierende Konstrukt zu messen versucht, wird der Test zur Erfassung von Gruppenkennwerten trotzdem als zweckmäßig erachtet. Ein Ausschluss zahlreicher Items würde in diesem Fall mit Einbußen im Rahmen der Konstruktrepräsentativität einhergehen. Darüber hinaus sollte in die Gesamtbetrachtung auch der Wert zum zweiten Messzeitpunkt (t_2) einfließen.

Der bereits in Kapitel 4.3.3 erlangte Eindruck von einem unsystematischen Antwortverhalten der Zielgruppe zu t_1 verstärkt sich bei der vergleichenden Betrachtung der Reliabilitätskoeffizienten des zweiten Messzeitpunkts. Zu t_2 werden drei Items aufgrund negativer Item-Skala-Korrelation (W6, W14, W19) ausgeschlossen, sodass alpha auf Grundlage von 18 Items mit einem Wert von 0,710 geschätzt werden kann und damit deutlich näher an einer akzeptablen Bezugsgröße liegt. Das liefert einen Hinweis für ein systematischeres Antwortverhalten der Zielgruppe zu t_2 . Für die Kontrastgruppe der Wirtschaftsinformatikstudierenden, denen ein höheres fachspezifisches Wissen und damit einhergehend auch ein systematischeres Antwortverhalten unterstellt werden kann (siehe Präzisierung der Kontrastgruppe in Kapitel 3.1.2.4), werden zur Schätzung von alpha zwei Items aufgrund negativer Trennschärfen entfernt (W14, W19). Für eine Skala mit 19 verbliebenen Items kann alpha auf 0,775 geschätzt werden. Das zeigt, dass der Test für einige der betrachteten Gruppen durchaus akzeptable Reliabilitätswerte erzielt, dieses aber nicht für alle Gruppen unterstellt werden kann.

4.4 Validierungsaspekt: Interne Struktur

Im Rahmen der internen Strukturprüfung kann auf Faktorenanalysen zurückgegriffen werden. Durch die Verfahren der Faktorenanalysen werden „Variablen gemäß ihrer korrelativen Beziehung in voneinander unabhängige Gruppen klassifiziert“ (Bortz, 2005, S. 512). Damit werden primär Zwecke der Dimensionsreduktion in Datensätzen intendiert (Bortz, 2005, S. 513), indem anhand einer kleinen Anzahl latenter Variablen (Faktoren) die Zusammenhänge zahlreicher manifester Variablen beschrieben werden (Krohne & Hock, 2007, S. 85).

In Kapitel 3.1.2.3 wurde für den Aspekt der internen Struktur die folgende Annahme aufgestellt:

A4: Das Grundlagenwissen zu KI, wie es in der Arbeit inhalts- und kognitionsbezogen modelliert wurde, lässt sich als eindimensionales Konstrukt abbilden.

Wie bereits in Kapitel 4.3.1 dargelegt, genügt die Datengrundlage der vorliegenden Arbeit nicht den Ansprüchen einer Kalibrierungss Stichprobe zur robusten Parameterschätzung im Rahmen einer IRT-Skalierung. Darüber hinaus mangelt es bisher an theoretisch explizit trennbaren Konstruktfacetten. Aus diesen Gründen wird zur inter-

nen Strukturprüfung ein exploratives Verfahren gewählt.¹⁰⁷ In diesem Zusammenhang wird häufig eine sogenannte Hauptkomponentenanalyse (*principal components analysis (PCA)*) (Bortz, 2005, S. 516) durchgeführt, die sich aufgrund des explorativen Charakters für die Dimensionalitätsprüfung der vorliegenden Daten eignet.

Die PCA wird zunächst für den Datensatz der Zielgruppe zu t1 ($N = 153$) durchgeführt. Die Maße der Stichprobeneignung nach Kaiser-Meyer-Olkin (KMO) (mit 0,511 nur marginal über dem Grenzwert $> 0,5$) sowie der Bartlett-Test auf Sphärizität (Signifikanz = 0,014) liefern erste Hinweise dafür, dass der Datensatz in dieser Form nicht für die Analyse geeignet ist. Die Betrachtung der einzelnen Anti-Image-Korrelationen auf Itemebene bestätigen die Einschätzung. Insgesamt zwölf von 21 Items unterschreiten den Grenzwert von $> 0,5$ und sind somit für die Hauptkomponentenanalyse ungeeignet. Die verbleibenden neun Items erfüllen die Stichprobenanforderungen an das Verfahren (KMO = 0,582; Signifikanz nach Bartlett = 0,009). Auf dieser Grundlage können drei Komponenten mit einem Eigenwert > 1 extrahiert werden, die kumuliert 46,49 % der Varianz erklären.

Für den Datensatz der Zielgruppe zu t2 werden zwar mit allen 21 Items die globalen Voraussetzungen der Stichprobeneignung erfüllt (KMO = 0,560; Signifikanz nach Bartlett = 0,001), allerdings müssen erneut sieben Items aufgrund unzureichender Anti-Image-Korrelation aus dem Set entfernt werden. Nach der Selektion werden die Voraussetzungen für die Hauptkomponentenanalyse (KMO = 0,665; Signifikanz nach Bartlett = 0,000) erfüllt und es werden vier Komponenten extrahiert, mit denen 49,96 % der Varianz erklärt werden können.

Auf Basis der Hauptkomponentenanalyse kann die Annahme A4: „Das Grundlagenwissen zu KI, wie es in der Arbeit inhalts- und kognitionsbezogen modelliert wurde, lässt sich als eindimensionales Konstrukt abbilden“ nicht bestätigt werden. Daraus ergeben sich Implikationen für die Weiterentwicklung des theoretischen Konstrukts und die empirische Arbeit, die in Kapitel 5 präzisiert werden.

4.5 Validierungsaspekt: Beziehung zu anderen Merkmalen

In Kapitel 3.1.2.4 wurden zwei für die diskriminante Validierung geeignete Kontrastgruppen identifiziert. Diese werden im Folgenden zunächst bzgl. ihrer Merkmalsausprägung jeweils mit der gesamten Zielgruppe ($N = 153$) verglichen. Aufgrund der Argumentation in Kapitel 3.1.2.4 erfolgt anschließend eine differenzierte Betrachtung der Kontrastgruppe im Vergleich zur Subzielgruppe der Masterstudierenden der Wirtschaftspädagogik ($N = 85$).¹⁰⁸

Die erste Kontrastgruppe besteht aus Masterstudierenden der Wirtschaftsinformatik ($N = 27$).¹⁰⁹ Die für diese Gruppe aufgrund der inhaltlichen Nähe des Studien-

107 In Abgrenzung zur explorativen Faktorenanalyse können durch Verfahren der konfirmatorischen Faktorenanalysen hypothesengeleitete Modellüberprüfungen vorgenommen werden (Krohne & Hock, 2007, S. 91).

108 Dafür wurden die Studierenden aus den M. Sc. und M. Ed.-Studienmodellen zu einer Subgruppe zusammengefasst.

109 Die deskriptiven Daten dieser Gruppe wurden bereits in Kapitel 4.1 berichtet.

fachs zum Untersuchungsgegenstand im Kontext der vorliegenden Arbeit formulierte Annahme lautet:

A5: Die Kontrastgruppe der Studierenden des Fachbereichs Wirtschaftsinformatik weist eine stärkere Ausprägung des Grundlagenwissens zu KI auf als die Zielgruppe.

Zur Überprüfung der Annahme werden die mittleren Testscores der beiden Gruppen miteinander verglichen. Für die Auswahl der Testverfahren sind zunächst einige Voraussetzungen innerhalb der Teststichproben zu prüfen (Bortz, 2005, S. 141). Die vorausgesetzte Normalverteilungsannahme wurde für beide Gruppen bereits in Kapitel 4.3.2 geprüft. Der *Levene-Test* der Varianzgleichheit ergibt, dass für die beiden Gruppen keine Homoskedastizität angenommen werden kann. Aus diesem Grund¹¹⁰ wird auf nichtparametrische Tests zurückgegriffen und die allgemeine Tendenz mittels *Mann-Whitney-Test* testscorebasiert ermittelt.

Tabelle 27: Mittlere Testscores der Zielgruppe (gesamt) und der Kontrastgruppe WINF

	N	MW	Standardabweichung
Zielgruppe (gesamt)	153	8,95	2,35
Kontrastgruppe WINF	27	10,96	3,93

Aus den deskriptiven Befunden (Tabelle 27) ergibt sich, dass die Zielgruppe im Mittel 8,95 von insgesamt 21 möglichen Punkten erreicht. Die Kontrastgruppe der WINF-Studierenden erreicht im Mittel 10,96 Punkte. Auf Grundlage der rangbasierten Schätzung des *Mann-Whitney-Tests* kann ein signifikanter Unterschied ($p = 0,007$) zwischen den mittleren Testscores der beiden Gruppen festgestellt werden. Die Effektstärke ist mit 0,201 als mittelmäßig zu bewerten.

Tabelle 28: Mittlere Testscores der Zielsubgruppe (WiPäd) und der Kontrastgruppe WINF

	N	MW	Standardabweichung
Zielsubgruppe WiPäd	85	9,38	2,171
Kontrastgruppe WINF	27	10,96	3,93

Für die Analyse des Mittelwertvergleichs zwischen der Kontrastgruppe der Masterstudierenden WINF mit den Masterstudierenden der Wirtschaftspädagogik sind erneut die Voraussetzungen für parametrische Testverfahren zu prüfen. Auf Grundlage der Normalverteilungsprüfung (Kapitel 4.3.2) und der Ergebnisse des *Levene-Tests* auf Homoskedastizität muss auch an dieser Stelle auf nichtparametrische Testverfahren zurückgegriffen werden. Aus Tabelle 28 gehen die mittleren Testscores der beiden Grup-

¹¹⁰ Obwohl der t-Test für unabhängige Stichproben recht robust auf Verletzungen der Voraussetzungen reagiert, muss bei deutlich unterschiedlich großen Stichproben sowie Varianzungleichheit von einer erhöhten Fehlerquote ausgegangen werden (Bortz, 2005, S. 141).

pen hervor. Die rangbasierte Differenz der beiden Gruppen auf Grundlage des *Mann-Whitney-Tests* ist nicht signifikant ($p = 0,033$).

Auf Grundlage der Tests kann die Annahme A5 bestätigt werden. Die Unterschiede im mittleren Testscore zum Grundlagenwissen der KI zeigen, dass das Wissen der Kontrastgruppe (Studierende der WINF) stärker ausgeprägt ist als das Grundlagenwissen der Zielgruppe (angehende Lehrkräfte im berufsbildenden Bereich) und der Zielsubgruppe (Masterstudierende der Wirtschaftspädagogik). Der Unterschied im mittleren Testscore zwischen der Kontrastgruppe und der gesamten Zielgruppe ist dabei signifikant ($p = 0,007$). Auch im Vergleich zur Subgruppe der Masterstudierenden im Fach Wirtschaftspädagogik erreicht die Kontrastgruppe im Mittel 1,58 Punkte mehr, die Teststatistik ist allerdings nicht signifikant ($p = 0,033$).

Die zweite in Kapitel 3.1.2.4 definierte Kontrastgruppe für die diskriminante Validierung besteht aus Bachelorstudierenden des Fachs Wirtschaftspädagogik. Die zu prüfenden Annahme in diesem Zusammenhang lautet:

A6: Die Merkmalsausprägung des Grundlagenwissens zu KI ist bei den Bachelorstudierenden der Wirtschaftspädagogik gleich ausgeprägt wie bei der Zielgruppe.

Zur Prüfung der Annahme werden zunächst die Voraussetzungen für vergleichende Testverfahren überprüft. Obwohl die Stichproben unterschiedlich groß sind (siehe Tabelle 28), kann aufgrund der Ergebnisse des *Levene-Tests* auf Varianzgleichheit ein parametrisches Testverfahren gewählt werden.¹¹¹ Die mittleren Testscores der beiden Gruppen sind aus Tabelle 29 zu entnehmen.

Tabelle 29: Mittlere Testscores der Zielgruppe (gesamt) und der Kontrastgruppe Bachelor WiPäd

	<i>N</i>	<i>MW</i>	Standardabweichung
Zielgruppe (gesamt)	153	8,95	2,35
Kontrastgruppe (B. Sc. WiPäd)	29	8,66	2,30

Die deskriptiven Befunde weisen bereits darauf hin, dass zwischen den beiden Gruppen keine signifikanten Mittelwertdifferenzen existieren. Der t-Test für unabhängige Stichproben unterstützt diese Annahme.

Für eine differenziertere Betrachtung der potenziellen Gruppenunterschiede folgt die vergleichende Analyse der Kontrastgruppe mit der Zielsubgruppe der Masterstudierenden im Fach Wirtschaftspädagogik. Die Voraussetzungen für ein parametrisches Testverfahren sind erfüllt und die deskriptiven Befunde in Tabelle 30 zusammengefasst.

¹¹¹ Ausschlaggebend für diese Einschätzung ist die Robustheit der parametrischen Verfahren bei ungleichen Stichprobengrößen unter der Voraussetzung der Varianzgleichheit (Bortz, 2005, S. 141).

Tabelle 30: Mittlere Testscores der Zielsubgruppe (WiPäd) und der Kontrastgruppe Bachelor WiPäd

	<i>N</i>	<i>MW</i>	Standardabweichung
Zielsubgruppe WiPäd	85	9,38	2,171
Kontrastgruppe (B. Sc. WiPäd)	29	8,66	2,30

Obwohl die Mittelwertdifferenz der Zielsubgruppe und der Kontrastgruppe größer ist als die Differenz der vergleichenden Analyse mit der Gesamtzielgruppe, kann kein signifikanter Befund eruiert werden.

Auf Grundlage der Analysen kann die Annahme A6: „Die Merkmalsausprägung des Grundlagenwissens zu KI ist bei den Bachelorstudierenden der Wirtschaftspädagogik gleich ausgeprägt wie bei der Zielgruppe“ damit bestätigt werden.

4.6 Analyse der quantitativen Daten

4.6.1 Betrachtung im Querschnitt

Die aus Kapitel 4.3 hervorgehende Bewertung des Testinstruments anhand der Kriterien der klassischen Testtheorie liefert Hinweise dafür, dass die mit dem Instrument generierten Befunde mit Einschränkungen zu interpretieren sind. In Anlehnung an die zentrale Ziel- und Fragestellung der vorliegenden Dissertation sowie insbesondere in Bezug zu den aufgestellten Leitfragen 15 und 16 sollen im Folgenden aufgrund der empirischen Befunde Aussagen über die Merkmalsausprägung der kognitiven sowie non-kognitiven Kompetenzfacetten der Zielgruppe generiert werden.

Aus der Analyse der Rohwerte in Kapitel 4.3 konnten bereits Erkenntnisse über die Ausprägung des Grundlagenwissens zu KI innerhalb der Zielgruppe generiert werden. Die deskriptiven Befunde zum ersten Messzeitpunkt sind in Tabelle 31 noch einmal zusammengefasst.

Tabelle 31: Deskriptive Statistik zum Testscore der Zielgruppe zu t1

	<i>N</i>	Min	Max	<i>MW</i>	Standardabweichung
Grundlagenwissen zu KI	153	4	14	8,95	2,350

Demnach erreicht die Zielgruppe im Mittel 8,95 von 21 möglichen Punkten. Die Ausprägung des Grundlagenwissens streut in der Zielgruppe weit. Das verdeutlichen die in der Tabelle gebildeten Werte der Spannweite (Min = 4; Max = 14) und der Standardabweichung. Dieser Befund ist konsistent zu der Annahme, die sich bereits aus der Analyse des Forschungsstandes (Kapitel 2.4) zu der Ausprägung von KI-bezogenen Kompetenzen bei (angehenden) Lehrkräften ergeben hat. Demnach ist das Grundlagenwissen der Zielgruppe zu Künstlicher Intelligenz als defizitär zu bewerten.

Die non-kognitiven Kompetenzfacetten wurden anhand adaptierter Skalen aus der Kompetenz- und Technologieakzeptanzforschung erfasst (siehe dazu Kapitel 3.2.3.2).

Im Sinne einer besseren Interpretierbarkeit der Daten werden die einzelnen Befunde entlang der modellierten Konstrukte ausgewertet. Die Merkmalsausprägung der Einstellungen ist in Tabelle 32 zusammengefasst.

Tabelle 32: Deskriptive Befunde zu den Einstellungen der Zielgruppe zu t1

	N	Min	Max	MW	Standardabweichung
<i>Perceived Usefulness</i>	153	0,00	5,00	2,986	0,938
<i>Perceived Ease of Use</i>	153	0,00	4,67	2,751	0,766
<i>Job Relevance</i>	153	0,00	5,00	2,781	1,071
<i>Computer Playfulness</i>	153	1,40	4,60	3,123	0,738
<i>Attitude Towards Behavior</i>	153	1,25	5,00	2,742	0,659
Einstellungen gesamt	153	0,68	4,36	2,890	0,596

Daraus geht hervor, dass die Einstellungen insgesamt eine positive Tendenz aufweisen. Diese Einschätzung beruht auf der Einordnung des Mittelwerts (2,890) auf der eingesetzten sechsstufigen Likert-Skala. Die erreichten Ausprägungen können nach der Codierung numerische Werte zwischen 0 und 5 annehmen. Der Mittelwert von 2,89 kann somit als positive Tendenz interpretiert werden.

Die Auswertung entlang der einzelnen Konstrukte ergibt, dass die Lehrkräfte das Potenzial von KI-gestützten Systemen für ihre (zukünftige) berufliche Praxis erkennen (*Perceived Usefulness*). Auch dieser Befund ist konsistent zu der in Kapitel 2.4 aus dem Forschungsstand abgeleiteten Annahme, nach der Lehrkräfte das Potenzial von Künstlicher Intelligenz wahrnehmen. Darüber hinaus geht aus den Daten hervor, dass eine positive Tendenz hinsichtlich der wahrgenommenen einfachen Bedienbarkeit (*Perceived Ease of Use*) erkennbar ist. Die Probandinnen und Probanden gehen demnach davon aus, dass sie den Umgang mit KI-gestützten Systemen erlernen können. Außerdem wird die Relevanz von KI für die eigene berufliche Praxis als tendenziell hoch eingeschätzt (*Job Relevance*). Die Probandinnen und Probanden fühlen sich auf Grundlage der empirischen Befunde wohl im Umgang mit Computern (*Computer Playfulness*) und zeigen eine Tendenz zur positiven Nutzungsintention (*Attitude Towards Behavior*).

Obwohl die Daten grundsätzlich eine positive Tendenz aufweisen, bewegen sie sich insgesamt eher in einer neutralen Mitte ohne starke zustimmende oder ablehnende Haltungen abzubilden.

Die Güte der auf den Untersuchungsgegenstand adaptierten Skalen ist auf Grundlage der Reliabilitätsschätzung nach alpha erwartungskonform positiv zu bewerten. Die Werte von alpha für beide Messzeitpunkte sind Tabelle 33 zu entnehmen.

Tabelle 33: Reliabilität der adaptierten Skalen zur Erfassung der Einstellungen

Konstrukt	Alpha (t1)	Alpha (t2)
<i>Perceived Usefulness</i>	,919	,927
<i>Perceived Ease of Use</i>	,881	,867
<i>Job Relevance</i>	,898	,873
<i>Computer Playfulness</i>	,806	,806
<i>Attitude Towards Behavior</i>	,801	,816

Zur Erfassung der epistemologischen Überzeugungen (Kapitel 2.2.3) wurde eine Skala zur Objektivität des Wissens auf den Untersuchungsgegenstand adaptiert (Kapitel 3.2.3.2). Die deskriptiven Befunde zur Merkmalsausprägung der Zielgruppe zum Messzeitpunkt t1 sind in Tabelle 34 zusammengefasst.

Tabelle 34: Deskriptive Befunde zu den Überzeugungen der Zielgruppe zu t1

	N	Min	Max	MW	Standardabweichung
Objektivität des Wissens	153	0,63	3,38	2,067	0,5585

Aus den Befunden ergibt sich eine tendenziell kritische Haltung der Probandinnen und Probanden zu Künstlicher Intelligenz. Diese Einschätzung erwächst aus dem Mittelwert (2,067), der auf der sechsstufigen Likert-Skala von „ich lehne die Aussage nachdrücklich ab“ bis „ich stimme der Aussage nachdrücklich zu“ eher auf der ablehnenden Seite zu verorten ist. Mit Bezug zum adaptierten Konstrukt bedeutet das, dass die Probandinnen und Probanden das Fachgebiet der KI als eher dynamisch und veränderbar wahrnehmen und die KI-bezogenen Inhalte aus verschiedenen Informationsquellen dementsprechend kritisch hinterfragen. In der Literatur zeigt sich, dass kritische Überzeugungen in der Bildungsforschung (vgl. Schmidt et al., 2015) und speziell mit Bezug zu KI wünschenswert sind (Dietzmann & Alt, 2020, S. 5175). Auch dieser Befund ist vereinbar mit den angenommenen kritischen Haltungen der (angehenden) Lehrkräfte, die sich aus der Analyse des Forschungsstandes ergeben (Kapitel 2.4).

Die Güte ist nach der Schätzung von alpha für die adaptierte Skala als akzeptabel einzuordnen (Tabelle 35).

Tabelle 35: Reliabilität der adaptierten Skala zur Erfassung der epistemologischen Überzeugungen

Konstrukt	Alpha (t1)	Alpha (t2)
Objektivität des Wissens	,801	,806

Die motivationalen Dispositionen wurden im Rahmen der Modellierung in extrinsische und intrinsische Motivation binnendifferenziert (Kapitel 2.2.4). Die Operationalisierung erfolgte unter einem berufsbezogenen (extrinsischen) sowie einem gegen-

standsbezogenen (intrinsischen) Bezug. Ergänzt wurden die motivationalen Dispositionen um die Subfacette der Subjektiven Norm, die insbesondere in empirischen Studien auf Grundlage von Modellen der Technologieakzeptanzforschung eine nachgewiesen handlungsleitende Funktion einnimmt (vgl. Venkatesh & Davis, 2000). Die deskriptiven Befunde zur Merkmalsausprägung sind in Tabelle 36 zusammengefasst.

Tabelle 36: Deskriptive Befunde zu den motivationalen Dispositionen der Zielgruppe zu t1

	N	Min	Max	MW	Standardabweichung
Gegenstandsbezogene intrinsische Motivation	153	0,00	4,75	2,525	0,997
Berufsbezogene extrinsische Motivation	153	0,00	4,75	1,918	1,085
Subjektive Norm	153	0,00	5,00	1,882	1,019

Auf Grundlage der empirischen Befunde im Querschnitt kann geschlussfolgert werden, dass die Probandinnen und Probanden eher gegenstandsbezogen intrinsisch als berufsbezogen extrinsisch motiviert sind. Das bedeutet, dass die Probandinnen und Probanden sich eher mit dem Gegenstand der KI aus persönlichen Neigungen beschäftigen, als dass sie mit der Auseinandersetzung berufsbezogene Ziele verfolgen. Die subjektive Norm, mit der der wahrgenommene Handlungsdruck aufgrund der Einschätzung relevanter Dritter gemeint ist (siehe Kapitel 2.1.3 und Kapitel 3.2.3.2), ist in der vorliegenden Stichprobe tendenziell gering ausgeprägt. Aus der Analyse des Forschungsstandes (Kapitel 2.4) konnte die Annahme abgeleitet werden, dass die (angehenden) Lehrkräfte eine Bereitschaft zeigen, sich mit KI auseinanderzusetzen. Diese Annahme kann auf Grundlage der empirischen Befunde, wenn auch mit Einschränkung, bestätigt werden.

Die Güte der adaptierten Skalen kann auf Grundlage der Werte nach Cronbach's alpha (Kapitel 4.3) als zufriedenstellend eingeschätzt werden (Tabelle 37).

Tabelle 37: Reliabilität der adaptierten Skalen zur Erfassung der Motivation

Konstrukt	Alpha (t1)	Alpha (t2)
Gegenstandsbezogene intrinsische Motivation	,923	,900
Berufsbezogene extrinsische Motivation	,932	,951
Subjektive Norm	,842	,810

4.6.2 Analyse der Daten im Längsschnitt

Nach der Analyse der quantitativen Daten im Querschnitt erfolgt auf Grundlage des Pretest-Posttest-Designs eine Analyse der Befunde im Längsschnitt. Für die Durchfüh-

rung von Veränderungsmessungen existieren in der Forschungsliteratur eine Reihe von Voraussetzungen (einen Überblick liefert Happ, 2017, S. 125–126). Wesentliche Voraussetzungen für die Durchführung einer Veränderungsmessung betreffen dabei die Anzahl an Messzeitpunkten und zentrale Charakteristika des eingesetzten Messinstruments (Happ, 2017, S. 125–126). Für eine Veränderungsmessung müssen nach Klauer (2009, S. 526) Daten zu mindestens zwei Messzeitpunkten vorliegen, wobei mehr Messzeitpunkte besser zu bewerten sind. Das eingesetzte Instrument muss darüber hinaus dem Kriterium der Messinvarianz entsprechen (Happ, 2017, S. 125; Helmreich, 1977, S. 23; Ifenthaler, 2006, S. 60; Klauer, 2009, S. 528). Diesem Kriterium wird in der vorliegenden Studie Rechnung getragen, indem zu beiden Messzeitpunkten das exakt gleiche Instrument eingesetzt wurde.¹¹² Dieses muss nach Petermann (2010, S. 896) und Happ (2017, S. 126) darüber hinaus veränderungssensitiv sein. Die Veränderungssensitivität des in der vorliegenden Arbeit eingesetzten Instruments wird durch den Vergleich von Pre- und Postdaten geprüft. Als Vergleichsgröße werden in Anlehnung an Lienert und Raatz (1998) die jeweils erreichten Summenscores herangezogen, anhand derer der Wachstumsverlauf¹¹³ nachvollzogen werden kann. Die vorliegende Studie verfolgt den Zweck einer kollektiven Gruppenbetrachtung (siehe Kapitel 3.1), weshalb keine individuellen Verläufe analysiert werden. Aus Tabelle 38 kann der mittlere Testscore der Zielgruppe zu beiden Messzeitpunkten entnommen werden.

Tabelle 38: Testscore der Zielgruppe im Pre-Post-Vergleich

	<i>N</i>	<i>MW</i>	Standardabweichung
Summenscore t1	81	9,32	2,366
Summenscore t2	81	11,09	3,458

Aus den Befunden geht eine Mittelwertdifferenz von 1,765 Punkten zwischen t1 und t2 hervor. Der t-Test für verbundene Stichproben ist im Pre-Post-Vergleich für die Mittelwertdifferenz der Zielgruppe signifikant ($p = ,000$; Cohen's $d = 0,54$). Demnach konnten die Probandinnen und Probanden im Laufe eines Semesters ihr Grundlagenwissen signifikant erweitern. Nach der Interpretation des Kennwerts nach Cohen's d liegt ein mittlerer Effekt vor.

Die non-kognitiven Kompetenzfacetten wurden ebenfalls mit dem gleichen Instrument wie zum ersten Messzeitpunkt am Ende des Semesters erneut erfasst (siehe ausführlicher Anhang B). In Tabelle 39 sind die Pre-Post-Daten für die Subkonstrukte der Einstellungen zusammengefasst.

¹¹² Rost (2004, S. 277) formuliert eine Einschränkung für den Fall des Vorliegens von Decken- oder Bodeneffekten und rät dazu, die Items zwischen den Messzeitpunkten zu variieren. Das Vorliegen von Decken- und Bodeneffekten wurde im Rahmen der vorliegenden Arbeit bereits im Rahmen der Verteilungsanalyse (siehe Kapitel 4.3.2) ausgeschlossen, sodass mit einem unveränderten Messinstrument vorgegangen werden kann.

¹¹³ In Anlehnung an Happ (2017, S. 123–124) ist darauf hinzuweisen, dass mit einem Wachstumsverlauf nicht zwangsläufig die positiv gerichtete Veränderung gemeint ist. Vielmehr müsse in Anbetracht potenzieller Faktoren wie Vergessenseffekten eine Betrachtung von sowohl positiv als auch negativ gerichteten Veränderungsverläufen fokussiert werden. Die Analyse von individuellen Wachstumsverläufen ist nicht Gegenstand der vorliegenden Arbeit.

Tabelle 39: Einstellungen der Zielgruppe im Pre-Post-Vergleich

	T1			T2		
	N	MW	Standardabweichung	MW	Standardabweichung	Sig.
<i>Perceived Usefulness</i>	81	3,047	0,883	3,047	0,889	1,000
<i>Perceived Ease of Use</i>	81	2,800	0,701	2,893	0,681	,254
<i>Job Relevance</i>	81	2,963	1,036	2,920	1,005	,725
<i>Computer Playfulness</i>	81	3,067	0,746	3,143	0,741	,255
<i>Attitude Towards Behavior</i>	81	2,858	0,699	2,735	0,497	,099

Daraus geht hervor, dass die marginalen Differenzen zwischen Pre- und Postdaten nicht statistisch signifikant sind. Dabei können die Einstellungen zu Künstlicher Intelligenz im Rahmen der vorliegenden Untersuchung als tendenziell zeitstabil charakterisiert werden.

In Tabelle 40 sind die empirischen Daten im Pre-Post-Vergleich für die epistemologischen Überzeugungen abgebildet.

Tabelle 40: Überzeugungen der Zielgruppe im Pre-Post-Vergleich

	T1			T2		
	N	MW	Standardabweichung	MW	Standardabweichung	Sig.
Objektivität des Wissens	81	2,076	0,561	1,9691	0,577	,082

Auf Grundlage der empirischen Daten konnte keine signifikante Veränderung der Merkmalsausprägung im Pre-Post-Vergleich für die epistemologischen Überzeugungen der Zielgruppe zu KI gemessen werden. Genau wie die Einstellungen, können die epistemologischen Überzeugungen in diesem Kontext eher zeitstabil bewertet werden.

Für die motivationalen Dispositionen sind die empirischen Daten im Pre-Post-Vergleich in Tabelle 41 zusammengefasst.

Tabelle 41: Motivation der Zielgruppe im Pre-Post-Vergleich

	T1			T2		
	N	MW	Standardabweichung	MW	Standardabweichung	Sig.
Gegenstandsbezogene intrinsische Motivation	81	2,491	1,008	2,620	0,932	,238
Berufsbezogene extrinsische Motivation	81	1,870	1,047	1,959	1,192	,352
Subjektive Norm	81	1,846	0,986	1,741	1,000	,329

Der t-Test für verbundene Stichproben liefert für die Mittelwertdifferenzen der motivationalen Dispositionen im Pre-Post-Vergleich keine signifikanten Befunde.

Zusammenfassend kann für die kognitive Facette des Grundlagenwissens ein signifikanter Zuwachs im Pre-Post-Vergleich festgestellt werden. Die non-kognitiven Kompetenzfacetten der Einstellungen, Überzeugungen und Motivation sind hingegen auf Grundlage der Datenbasis als stabile Charakteristika einzuschätzen. Damit bestätigt sich die Annahme aus Kapitel 2.2.3, dass es sich bei den non-kognitiven Facetten zwar um grundsätzlich veränderbare Personenmerkmale handelt, diese Veränderungen aber tendenziell nur über einen längeren Zeitraum hinweg entstehen und gemessen werden können.

5 Das Grundlagenwissen zu Künstlicher Intelligenz bei angehenden Lehrkräften: neue Erkenntnisse und Perspektiven

5.1 Zusammenfassung der zentralen Ergebnisse

Das übergeordnete Ziel der vorliegenden Dissertation besteht darin, das Konstrukt des Grundlagenwissens zu Künstlicher Intelligenz von (angehenden) Lehrkräften im berufsbildenden Bereich theoretisch zu modellieren und empirisch erfassbar zu machen. Dafür gliedert sich das Vorgehen dieser Arbeit in drei zentrale Phasen, die strukturell an Studien zum wirtschaftsdidaktischen Wissen von (angehenden) Lehrkräften im berufsbildenden Bereich orientiert sind (vgl. Kuhn, 2014): die Phase der Modellentwicklung (Kapitel 2), die Phase der Operationalisierung (Kapitel 3) sowie die Phase der empirischen Erprobung des entwickelten Instruments (Kapitel 4).

Im Folgenden werden die zentralen Ergebnisse der drei Phasen mit Bezug zu den aufgestellten Leitfragen aus Kapitel 1 zusammenfassend dargestellt und hinsichtlich ihrer Implikationen für weitergehende Forschungsvorhaben diskutiert.

Phase der theoretischen Modellierung (Leitfragen 1 bis 6)

Leitfrage 1: Wie lässt sich das Grundlagenwissen zu KI in Modelle der Lehrprofessionalität (1.1), der digitalisierungsbezogenen Kompetenzen (1.2) und der Technologieakzeptanzforschung (1.3) einordnen und wie können diese Ansätze für den Untersuchungsgegenstand adaptiert werden?

Da es bisher an einer grundlegenden Systematisierung des Untersuchungsgegenstandes mangelt, sind für diese Phase der Arbeit zunächst Modelle der professionellen Kompetenz von Lehrkräften, Modelle der digitalisierungsbezogenen Kompetenzen sowie Modelle der Technologieakzeptanzforschung hinsichtlich ihrer Anschlussfähigkeit für die theoretische Konzeption analysiert worden (Kapitel 2.1). Es konnte herausgestellt werden, dass Modelle der professionellen Kompetenz (z. B. Baumert & Kunter, 2006; Zlatkin-Troitschanskaia et al., 2013a) einen Anknüpfungspunkt zur grundlegenden Systematisierung des Strukturmodells auf der Basis kognitiver und non-kognitiver Kompetenzfacetten liefern (Kapitel 2.1.1). Im Rahmen der Analyse der Modelle zu digitalisierungsbezogenen Kompetenzen ist insbesondere das *TPACK*-Modell (vgl. Mishra & Koehler, 2006) für den Untersuchungsgegenstand kontextualisiert worden. Das für die Arbeit relevante Grundlagenwissen zu KI kann als inhaltliche Erweiterung in die Facette des *TK* eingeordnet werden (Kapitel 2.1.2). Da die potenzielle Nutzung von Technologien eng mit deren Akzeptanz verknüpft ist (vgl. Davis, 1989; Kollmann, 1998, 2000), sind etablierte Modelle der Technologieakzeptanzforschung hinsichtlich ihrer

für die vorliegende Untersuchung relevanten Einflussgrößen analysiert worden. Als Ergebnis dieser Phase können insbesondere kognitive, motivationale und affektive Dispositionen identifiziert werden (Kapitel 2.2.3, siehe z. B. Klauser, 2006), die in die anschließende theoretische Modellierung aufgenommen werden.

Leitfrage 2: Welcher Wissensbegriff kann der Arbeit zugrunde gelegt werden (kognitionsbezogene Modellierung)?

Das Wissen wird im Rahmen der Modellierung als zentrale, kognitive Leistungsdisposition eingeordnet. Diese Facette gilt dabei als Fundament für alle höheren kognitiven Leistungen (vgl. Anderson & Krathwohl, 2001). Die kognitionsbezogene Differenzierung erfolgt dabei in Anlehnung an Arbinger (1997), Beck (1995), Dubs (1995) und Wuttke (2005) in ein Fakten- und Strukturwissen. Diese Facetten lassen sich als deklarativ-statisch und deklarativ-prozedural klassifizieren (Kapitel 2.2.1). Die damit intendierten kognitiven Prozesse beinhalten das Erinnern und darauf operierende Verstehensprozesse (vgl. Wuttke, 2005).

Leitfrage 3: Welche non-kognitiven Kompetenzfacetten müssen in der Untersuchung berücksichtigt und für den Untersuchungsgegenstand definiert werden?

Als non-kognitive Kompetenzfacetten werden in Anlehnung an die untersuchten Modelle (z. B. Baumert & Kunter, 2006; Zlatkin-Troitschanskaia et al., 2013a) die Facetten Einstellungen (Kapitel 2.2.2), Überzeugungen (Kapitel 2.2.3) und motivationale Dispositionen (Kapitel 2.2.4) zu Künstlicher Intelligenz in die Modellierung aufgenommen. Die Einstellungen zu KI werden im Rahmen der Arbeit als Einstellungsakzeptanz definiert (vgl. Böhner, 2002; Klauser, 2006) und adressieren konkrete, mit einem Gegenstand verbundene, wertende Tendenzen eines Individuums. Im Rahmen der Untersuchung ist damit konkret der Grad der Abneigung oder Zuneigung zur Nutzung von KI-gestützten Systemen gemeint (Kapitel 2.2.2). Vor dem Hintergrund der (zukünftigen) beruflichen Praxis der Zielgruppe wird dabei besonders auf die potenzielle Nutzung von KI zur Bewältigung von Arbeitsaufgaben abgezielt. Die Überzeugungen werden im Kontext der Arbeit als epistemologische Überzeugungen zu Wissen und Wissenserwerb im Kontext der Künstlichen Intelligenz modelliert (Kapitel 2.2.3). Die Betrachtung dieser Facette wird durch den mit empirischer Evidenz unteretzten Einfluss der epistemologischen Überzeugungen auf das Handeln als Lehrkraft plausibilisiert (siehe dazu z. B. Berding & Lamping, 2014).

Die motivationalen Dispositionen im Kontext der vorliegenden Arbeit betreffen insbesondere die Bereitschaft zur Auseinandersetzung mit Künstlicher Intelligenz. In Anlehnung an etablierte Ansätze der Selbstbestimmungstheorie findet eine Binnendifferenzierung in extrinsische und intrinsische Motivation Anwendung (Kapitel 2.2.4).

Leitfrage 4: Welche Inhaltsbereiche sind für das Grundlagenwissen von (angehenden) Lehrkräften zu Künstlicher Intelligenz relevant (inhaltliche Modellierung)?

Neben der kognitionsbezogenen Modellierung ist eine inhaltliche Strukturierung des Fachgebiets der Künstlichen Intelligenz notwendig. Auf der Basis von Befunden aus Lehrbuchanalysen und Experteninterviews (vgl. Meß, 2022) kann eine inhaltliche Modellierung vorgenommen werden (Kapitel 2.3). Daraus geht hervor, dass die Inhaltsbereiche Grundlagen der KI, Methoden der KI und Anwendungen der KI als relevant für die in der vorliegenden Arbeit anvisierte Zielgruppe gelten (Kapitel 2.3, Leitfrage 4). Die Bereiche sind dabei nicht isoliert voneinander zu betrachten, sondern beinhalten inhaltliche Verknüpfungen. So gilt der Bereich der Grundlagen der KI als Bezugs- und Ausgangspunkt für die anderen beiden Subdimensionen.

Leitfrage 5: Wie ist der Forschungsstand zu KI-bezogenen Kompetenzfacetten bei (angehenden) Lehrkräften?

Die Analyse des Forschungsstandes zu KI-bezogenen Kompetenzen bei (angehenden) Lehrkräften bestätigt ein existierendes Forschungsdefizit (Kapitel 2.4). Insbesondere die für die vorliegende Dissertation relevante Gruppe der (angehenden) Lehrkräfte im berufsbildenden Bereich wurde bisher nachrangig betrachtet. Neben diesem quantitativen Mangel können darüber hinaus qualitative Defizite identifiziert werden. Diese betreffen u. a. das weitgehende Fehlen systematischer Ansätze zur Erfassung der Merkmalsausprägung. In den analysierten Studien sind darüber hinaus primär Verfahren der Selbsteinschätzung eingesetzt worden. Diese sind vor dem Hintergrund der nur bedingt möglichen Abgrenzbarkeit zu Selbstwirksamkeitserfahrungen (vgl. Nickolaus, 2010; SWK, 2022) kritisch zu hinterfragen. Auf Grundlage von Selbsteinschätzungen sind in der Regel keine validen Testwertinterpretationen möglich (vgl. Kane, 2013).

Leitfrage 6: Wie lässt sich der Untersuchungsgegenstand in einem theoretischen Modell abbilden?

Das zentrale Ergebnis der theoretischen Modellierung besteht in der Entwicklung des Strukturmodells für die KI-bezogene Kompetenz von (angehenden) Lehrkräften im berufsbildenden Bereich (Kapitel 2.5). Ausgehend von den Befunden der inhaltlichen und kognitionsbezogenen Strukturierung wird ein deklarativ-statisches Faktenwissen und deklarativ-prozedurales Strukturwissen zu den Bereichen Grundlagen der KI, Methoden der KI und Anwendungen der KI als zentrale kognitive Facette modelliert. Die non-kognitiven Facetten werden als epistemologische Überzeugungen zu KI, Einstellungen (im Sinne einer Einstellungsakzeptanz) zu KI und motivationalen Dispositionen als extrinsische und intrinsische Motivation zur Auseinandersetzung mit KI im Modell aufgegriffen.

Phase der Operationalisierung und Testentwicklung (Leitfragen 7 bis 11)

Leitfrage 7: Wie können die Validierungsaspekte der „Standards for Educational and Psychological Testing“ (vgl. AERA et al., 2014) auf die vorliegende Untersuchung angewendet werden?

Die in den Erziehungs- und Sozialwissenschaften hoch angesehenen „Standards for Educational and Psychological Testing“ (vgl. AERA et al., 2014) beinhalten Leitlinien für die Beurteilung des zentralen Testgütekriteriums der Validität. Insbesondere die Aspekte „test content“, „internal structure“ und „relations to other variables“ werden als relevant für die durchgeführte Untersuchung betrachtet (Kapitel 3.1.2). Während die Annahmen zum Testinhalt weitestgehend auf konzeptionellen Überlegungen sowie (qualitativen) Dokumentenanalysen basieren, sind für die Prüfung der internen Struktur und der Beziehung zu anderen Merkmalen quantitative empirische Analysen notwendig. Für die Prüfung der internen Struktur wird dabei auf explorative Verfahren der Dimensionsreduktion zurückgegriffen. Für die Überprüfung der Beziehung zu anderen Merkmalen wird auf die Methode der bekannten Gruppen zurückgegriffen (vgl. Schnell et al., 2008), die im Rahmen einer diskriminanten Validierung präzisiert werden. Dafür werden die Gruppen der Masterstudierenden im Fachbereich Wirtschaftsinformatik sowie die Gruppe der Bachelorstudierenden im Fachbereich Wirtschaftspädagogik miteinbezogen.

Leitfrage 8: In welcher Form lässt sich der Messgegenstand „Grundlagenwissen zu KI von (angehenden) Lehrkräften“ erfassen?

Auf Grundlage der strukturellen Modellierung und der Ergebnisse der Forschungsstandanalyse ist der Messgegenstand inhaltlich und kognitionsbezogen präzisiert worden (Kapitel 3.1). Daraus ergibt sich die Darstellung in Form einer Matrix als Grundlage für die Operationalisierung der kognitiven Leistungsdisposition des Grundlagenwissens zu KI angehender Lehrkräfte im berufsbildenden Bereich. Die intendierten kognitiven Prozesse, die sich aus der kognitionsbezogenen Differenzierung und der Präzisierung des Messgegenstandes ableiten lassen, können als „Erinnern“ und darauf aufbauendem „Verstehen“ eingeordnet werden. Für die Testentwicklung bedeutet das, dass die kognitive Facette in einem geschlossenen Format operationalisiert werden kann (Kapitel 3.2). Das impliziert, dass die Merkmalsausprägung durch die Bewältigung von Testaufgaben, die anhand eines eindeutigen Richtigkeitsmaßstabs bewertbar sind, beurteilt werden kann.

Leitfrage 9: Welche Skalen können für die Erfassung der non-kognitiven Kompetenzfacetten adaptiert werden?

Aus der theoretischen Modellierung ergeben sich bereits Anknüpfungspunkte für die Adaption von Messinstrumenten zur Erfassung der non-kognitiven Facetten. In Anlehnung an Kapitel 2 werden etablierte Skalen aus der Kompetenz- und Technologieakzeptanzforschung für den Untersuchungsgegenstand kontextualisiert (Kapitel 3.2.3.2). Die Einstellungen werden mit Bezug zu etablierten Drittvariablenansätzen (vgl. Fish-

bein & Ajzen, 1985; Davis, 1989; Venkatesh & Davis, 2000) durch die Konstrukte *Perceived Usefulness*, *Perceived Ease of Use*, *Job Relevance*, *Computer Playfulness* und *Attitude Towards Behavior* im Testinstrument abgebildet. Zentral ist damit die Frage assoziiert, inwieweit die Probandinnen und Probanden das Potenzial von KI für ihre (zukünftige) berufliche Praxis bewerten und wie ihre darauf aufbauende Nutzungsintention zu bewerten ist. Die epistemologischen Überzeugungen zu KI werden mit der adaptierten Skala zur Objektivität des Wissens erfasst (vgl. Schiefele et al., 2002). Damit wird der Frage nachgegangen, ob die Probandinnen und Probanden das Fachgebiet der KI als eher statisch oder eher dynamisch wahrnehmen und wie sie die Beständigkeit des Wissens dazu einschätzen. Die motivationalen Dispositionen differenzieren in diesem Zusammenhang die gegenstandsbezogene intrinsische und die berufsbezogene extrinsische Perspektive (vgl. Schiefele et al., 2002). Es soll untersucht werden, ob eher selbstbestimmte Faktoren die Auseinandersetzung mit KI-Inhalten begünstigen oder ob die wahrgenommenen beruflichen Auswirkungen handlungsleitend sind. Die Erfassung der non-kognitiven Kompetenzfacetten erfolgt dabei durch den Einsatz von Likert-Skalen, um eine persönliche Tendenz abbilden zu können.

Leitfrage 10: In welchem organisatorischen Rahmen kann die Testung bei der Zielgruppe durchgeführt werden?

Um die Daten zur Beantwortung der empirisch angelegten Fragestellungen zu erheben, wurde eine Untersuchung an neun Hochschulstandorten (Deutschland und Schweiz)¹¹⁴ geplant und durchgeführt. Da KI-bezogene Inhalte bisher nicht Teil der Curricula sind, wurde ein Lehr-Lern-Konzept entwickelt, in dem die informationstechnologischen Inhalte mit genuin wirtschafts- und berufsdidaktischen Fragestellungen verknüpft sind (Kapitel 3.2.3.4). Das Lehr-Lern-Konzept fungiert im Rahmen der Pre-test-Posttest-Studie als Treatment. Auf dieser Grundlage sind neben Befunden im Querschnitt auch Analysen im Längsschnitt über ein Semester möglich.

Leitfrage 11: Wie ist die Objektivität des Tests zu beurteilen?

Die Itemkonstruktion und -adaption mündet in einen computerbasierten Test (Kapitel 3.2), der aufgrund von Art, Umfang und Durchführungsrahmen als hinreichend objektiv eingeordnet werden kann (Kapitel 3.3). Die Einschätzung erwächst insbesondere aus dem standardisierten Testverfahren. Die Vergleichbarkeit wird daher als sehr hoch eingeschätzt.

¹¹⁴ Die Lehrkräfteausbildung in Deutschland und der Schweiz zeigt im berufsbildenden Bereich starke Parallelen, was sich auch in einer gemeinsamen Sektion Berufs- und Wirtschaftspädagogik für den deutschsprachigen Raum (Deutschland, Schweiz, Österreich) zeigt. Daher wurde der Einbezug der Schweiz (PH Zürich) als zielführend erachtet. Zukünftig wird eine Ausweitung auf Österreich (Karl-Franzens-Universität Graz, Institut für Wirtschaftspädagogik, Prof. Dr. Michaela Stock) angestrebt.

Phase der empirischen Erprobung (Leitfragen 12 bis 17)

Leitfrage 12: Wie kann das entwickelte Instrument anhand der Kriterien der klassischen Testtheorie auf Test- (12.1), Item- (12.2) und Antwortebene (12.3) beurteilt werden?

Die Datenerhebung für die empirische Erprobung des Instruments hat im Zeitraum zwischen April 2021 und Oktober 2022 stattgefunden. Um eine ausreichend große Stichprobe zu akquirieren und damit belastbare statistische Analysen durchführen zu können, ist die Studie an insgesamt neun Hochschulstandorten durchgeführt worden, wodurch eine Reihe an verschiedenen Studienmodellen des Fachs Berufs- und Wirtschaftspädagogik in die Untersuchung einbezogen werden kann (Kapitel 4.1). Im Rahmen der empirischen Erprobung wird zunächst eine Analyse der Rohwertverteilung (Kapitel 4.3.2, Leitfrage 12.1) durchgeführt, um einen ersten Eindruck über die Güte anhand einer Beurteilung entlang der Kriterien der klassischen Testtheorie zu erlangen. Auf Basis der Analyse kann von einer approximativen Normalverteilung ausgegangen werden. Messmethodisch sind für das Pre-Post-Design keine Einschränkungen durch Decken- oder Bodeneffekte zu erwarten, das Instrument kann Veränderungen für die Postmessung abbilden. Darüber hinaus wird das Instrument auf Item- (Kapitel 4.3.3, Leitfrage 12.2) und Antwortebene (Kapitel 4.3.4, Leitfrage 12.3) fortlaufend geprüft. Die Analyse der Schwierigkeiten hat ergeben, dass das Testinstrument nicht in allen Ausprägungsbereichen hinreichend differenziert. Diese Analysen haben bereits erste Hinweise über die Ausprägung des Grundlagenwissens zu KI innerhalb der Zielgruppe sowie Anknüpfungspunkte für Weiterentwicklungs- und Revisionsprozesse geliefert.

Leitfrage 13: Wie ist die Reliabilität des Tests zu beurteilen?

Im Rahmen der Reliabilitätsschätzung mittels Cronbach's alpha (Kapitel 4.3.5) können einige Wissensitems mit negativen Trennschärfen identifiziert werden. Die Beurteilung der Reliabilität mittels Cronbach's alpha ist bei dichotomen Variablen durchaus problematisch zu beurteilen, da die Varianz mit dem maximalen Wert 0,25 limitiert ist. Nach Ausschluss der Items mit unzureichenden Trennschärfen ergibt sich für den Test in der Zielgruppe zum ersten Messzeitpunkt (t1) mit 13 verbleibenden Items ein Reliabilitätskoeffizient von 0,437, was auf inhaltliche und messmethodische Herausforderungen verweist. Zum zweiten Messzeitpunkt kann für die Zielgruppe für 18 (von ursprünglich 21) Items ein alpha von 0,710 geschätzt werden. Das bestätigt die Annahme, dass die Zielgruppe zu t1 aufgrund mangelnden Vorwissens unsystematisch antwortet und daher die Schätzung von alpha verzerrt ist. Für die Kontrastgruppe der Wirtschaftsinformatikstudierenden kann für den Test mit insgesamt 19 Items ein alpha von 0,775 geschätzt werden. Für die weitere Forschung muss daher eine nochmalige Anpassung und eine zielgruppenspezifischere Ausgestaltung des Instruments anvisiert werden. Es ist außerdem zu beachten, dass die Reliabilität immer in Bezug zum Einsatzzweck zu beurteilen ist (siehe Kapitel 4.3.5). Der in der Diagnostik etablierte Richtwert von 0,8 gilt dabei primär für die Analyse von Einzelfalldiagnostiken. Da im vorliegenden Test nur aggregierte Werte für die Zielgruppe im Sinne einer kol-

lektiven Auswertung untersucht werden, kann das Instrument trotz der vergleichsweise niedrigen Reliabilität für diesen Zweck und im Rahmen der ersten Pilotierung eingesetzt werden.

Leitfrage 14: Wie ist die Validität der Testwerte auf Basis des quantitativen Zugangs zu beurteilen?

Im Rahmen der empirischen Erprobung sind die in Kapitel 3 aufgestellten Annahmen für die einzelnen Aspekte der Validitätsbeurteilung überprüft worden. Die Annahme der Eindimensionalität im Rahmen des Validierungsaspekts der internen Struktur kann auf Grundlage der Hauptkomponentenanalyse nicht bestätigt werden (Kapitel 4.4). Die postulierten Beziehungen zu anderen Merkmalen im Rahmen der diskriminanten Validierung mit den Kontrastgruppen der Studierenden der Wirtschaftsinformatik sowie der Bachelorstudierenden der Wirtschaftspädagogik können im Gegensatz dazu aufgrund inferenzstatistischer Verfahren (Mann-Whitney-Test) mit empirischer Evidenz untersetzt werden (Kapitel 4.5).

Leitfrage 15: Wie ist das Grundlagenwissen zu KI in der Zielgruppe ausgeprägt?

Die mit dem entwickelten und an neun Hochschulstandorten eingesetzten Instrument generierten Befunde zur Merkmalsausprägung der kognitiven Kompetenzfacetten zu Künstlicher Intelligenz bei angehenden Lehrkräften im berufsbildenden Bereich lassen auf ein defizitäres Wissen dieser Zielgruppe schließen (Kapitel 4.6, Leitfrage 15). Im Mittel können weniger als die Hälfte der Fragen korrekt beantwortet werden (8,95 von 21 möglichen Punkten). Dabei weisen die Daten bei einer Spannweite von 4 (Min) bis 14 (Max) eine große Streuung auf. Für das Pre-Post-Design können sich demnach sowohl Studierende am unteren als auch oberen Rand der Wissensverteilung zu t1 in der Posttestung ausreichend entwickeln, es liegen demnach keine Decken- und Bodeneffekte vor, die ansonsten messmethodisch zu Problemen führen.

Leitfrage 16: Wie kann die Ausprägung der non-kognitiven Kompetenzfacetten bei der Zielgruppe bewertet werden?

Die Einstellungen der Zielgruppe zu KI sind als tendenziell positiv zu bewerten. Vor dem Hintergrund der adaptierten Skalen und der im Strukturmodell (Kapitel 2.5) und Messmodell (Kapitel 3.4) abgebildeten Konstrukte zur Erfassung der Einstellungen bedeutet das, dass die Probandinnen und Probanden die Relevanz von KI für ihre eigene (zukünftige) berufliche Praxis wahrnehmen (*Job Relevance*). Darüber hinaus erkennen die Probandinnen und Probanden auch das Potenzial von Anwendungen der KI für den eigenen Beruf (*Perceived Usefulness*). Die Überzeugungen, die im Kontext der Arbeit als epistemologische Überzeugungen zu Wissen und Wissenserwerb modelliert und in Anlehnung an eine Skala zur Objektivität des Wissens kontextualisiert sind, sind für die Zielgruppe als eher kritisch zu bewerten ($MW = 2,076$, $SD = 0,0623$). Das bedeutet, dass die angehenden Lehrkräfte den Untersuchungsgegenstand als eher dynamisch wahrnehmen und die Gültigkeit der Erkenntnisse kritisch hinterfragen. In

Anlehnung an Studien mit ähnlichen Zielgruppen sind kritische Überzeugungen ein bevorzugtes Ergebnis (z. B. Schmidt et al., 2015). Für die motivationalen Dispositionen zeigt sich, dass die Zielgruppe eine mittelstark ausgeprägte intrinsische Motivation aufweist ($MW = 2,491$; $SD = 0,112$). Die berufsbezogene extrinsische Motivation liegt im Mittel darunter ($MW = 1,870$; $SD = 0,116$). Das liefert einen Hinweis dafür, dass die Bereitschaft, sich mit dem Inhalt auseinanderzusetzen, tendenziell eher selbstbestimmt intrinsisch motiviert ist. Die Güte der Messungen ist darüber hinaus für die non-kognitiven Kompetenzfacetten, die mit etablierten Skalen auf den Untersuchungsgegenstand adaptiert wurden, als sehr zufriedenstellend zu bewerten. Die Schätzung der internen Konsistenz nach alpha wird für die einzelnen Skalen zu beiden Messzeitpunkten konsistent auf > 0.8 geschätzt (Kapitel 4.6).

Leitfrage 17: Welche Veränderungen lassen sich in der Ausprägung des Grundlagenwissens (17.1) und in der Ausprägung der non-kognitiven (17.2) Kompetenzfacetten im Verlauf eines Semesters mit dem entwickelten Testinstrument erfassen?

Im Rahmen der Analyse des Längsschnitts im Pretest-Posttest-Vergleich kann für die kognitive Facette des Grundlagenwissens innerhalb der Zielgruppe ein signifikanter Zuwachs von durchschnittlich 1,77 Punkten festgestellt werden (Kapitel 4.7, Leitfrage 17.1). Die non-kognitiven Kompetenzfacetten werden auf Grundlage der analysierten Daten als tendenziell zeitstabile Charakteristika bewertet. Für die Einstellungen, Überzeugungen sowie die motivationalen Dispositionen können demnach keine signifikanten Veränderungen im Pre-Post-Vergleich gemessen werden (Leitfrage 17.2).

5.2 Limitationen und anschließende Forschungsdesiderata

Um das übergeordnete Ziel der vorliegenden Dissertation, das Grundlagenwissen als kognitive Komponente und darüber hinaus die relevanten non-kognitiven Facetten zu Künstlicher Intelligenz von (angehenden) Lehrkräften im berufsbildenden Bereich theoretisch zu modellieren und empirisch erfassbar zu machen, zu erreichen, sind entlang der einzelnen Phasen der Arbeit einige (z. T. forschungspragmatisch begründete) Einschränkungen vorgenommen worden, die die Aussagekraft der erzielten Befunde limitieren. Die Limitationen der Arbeit können in theoretische und empirische Grenzen untergliedert werden.

Zu den *theoretischen* Limitationen zählt, dass das Modell aus Kap. 2.5 auf Basis der aktuellen Forschungsliteratur bereits ergänzt werden sollte. Dieses ist damit zu erklären, dass es bislang kein etabliertes Modell zur Integration von kognitiven und non-kognitiven Anforderungen an Lehrkräfte aus dem berufsbildenden Bereich gibt. Aus diesem Grund sind bereits jetzt weitere Aspekte identifiziert worden, die in das Modell Einzug finden sollten, um diese theoretisch und empirisch zu testen. Zu diesen Kompetenzfacetten zählen insbesondere ethisch-moralische Aspekte. Studien liefern Hinweise dafür, dass ethisch-moralische Fragestellungen bedeutsam für die Integration von KI-gestützten Systemen sind (vgl. Jobin et al., 2019). Auch die Befunde aus der

Analyse der Lehrwerke und der Experteninterviews (vgl. Meß, 2022; siehe Kapitel 2.3) liefern bereits Hinweise dafür, dass diese unter dem Bereich „Auswirkungen der KI“ subsumierten Fragestellungen von hoher Bedeutung sind. Auch die von Pargmann und Berding (2022) herausgearbeiteten Befunde, dass KI-Anwendungen häufig wegen diskriminierender, misogynen oder rassistischer Entscheidungsmuster in der Kritik stehen, sprechen der Integration dieser Fragestellungen eine zentrale Rolle zu.

An der Studie sind auch aus *empirischer* Sicht Grenzen kritisch zu würdigen. Die vorliegende Studie zeichnet sich durch ein Pretest-Posttest-Design aus. Um die kognitiven und non-kognitiven Kompetenzfacetten im Quer- als auch Längsschnitt zu erfassen, steht die psychologische Diagnostik vor der Herausforderung, geeignete Kontrollgruppen zu identifizieren, die auch erfasst werden können (Bortz & Döring, 2006, S. 113). Um dieser Limitation zu begegnen, wurden verschiedene Erhebungsszenarien (siehe Kapitel 4.1) genutzt, die u. a. thematisch vergleichbare, aber nicht identische Interventionen (als Treatment) beinhalten. Aufgrund einer unzureichenden Wiedererfassungsquote zum zweiten Messzeitpunkt konnte allerdings keine hinreichend große Stichprobe generiert werden, um interventionsbedingte Effekte auch für die Kontrollgruppen untersuchen zu können. In weiteren Forschungsprojekten sind diese Effekte weiter zu kontrollieren.

Darüber hinaus wurden in der empirischen Studie unter der Zielgruppe der angehenden Lehrkräfte im berufsbildenden Bereich Studierende aus unterschiedlichen Studienmodellen subsumiert (einen Überblick beinhaltet Kapitel 4.1). Für eine bindendifferenzierte Untersuchung der einzelnen Subgruppen ist die akquirierte Stichprobe aktuell noch zu klein. Perspektivisch sollten unterschiedliche Merkmalsausprägungen zwischen den Studiengängen mit Mehrebenenmodellierungen (vgl. Hox et al., 2012) empirisch untersucht werden.

In den Studiengängen der Berufs- und Wirtschaftspädagogik werden die Studierenden grundsätzlich polyvalent ausgebildet (vgl. BWP, 2014). Das bedeutet, dass nicht alle Probandinnen und Probanden eine Tätigkeit im Schuldienst anstreben. Aus den erhobenen Daten konnten keine signifikanten Unterschiede in der mittleren Testleistung zwischen Probandinnen und Probanden mit unterschiedlicher beruflicher Perspektive festgestellt werden. Ob es sich dabei um belastbare und generalisierbare Befunde handelt, muss in der zukünftigen Forschung herausgearbeitet werden.

Eine weitere Limitation besteht in dem Umstand, dass die beiden Testungen als sog. *low stake tests* (vgl. Förster et al., 2017a; Kane, 2006b; Thelk et al., 2009; Wise & DeMars, 2005) gelten. Bei *low stake tests* erwachsen für die Probandinnen und Probanden keine negativen Folgen aus den Ergebnissen der Testung, woraus in einigen Fällen eine geringe Teilnahmemotivation beobachtbar ist. Zwar wurde im Rahmen der vorliegenden Studie eine Skala zur Erfassung der Testmotivation (vgl. Happ & Förster, 2018) eingesetzt, um deren potenzielle Effekte zu erfassen. Allerdings kann dieser Effekt bestenfalls damit allenfalls nur kontrolliert, nicht aber ausgeschlossen werden. Die Befunde zeigen, dass die auf einer 6-stufigen Likert-Skala (Werte 0–5) erfasste Testmotivation der Probandinnen und Probanden zu beiden Testzeitpunkten mittelmäßig ausgeprägt ist ($t_1 = 2,45$; $t_2 = 2,56$). Auf Basis der erhobenen Daten können darüber

hinaus keine signifikanten Unterschiede in der mittleren Testleistung zwischen Probandinnen und Probanden mit einer eher hohen und Probandinnen und Probanden mit einer eher niedrigen Testmotivation festgestellt werden. Dennoch können verzerrende Effekte aus den Rahmenbedingungen der *low stake*-Testsituation nicht ausgeschlossen werden.

Darüber hinaus ist die akquirierte Stichprobe trotz der Implementation der Untersuchung an insgesamt neun Hochschulstandorten vergleichsweise klein. Im Rahmen der empirischen Überprüfung des Testinstruments ergibt sich daraus z. B. die Einschränkung, dass nicht auf eine IRT-Skalierung zurückgegriffen werden kann, da die Anforderungen an eine Kalibrierungsschichprobe zur robusten Parameterschätzung nicht erfüllt werden (siehe Kapitel 4.3). Ausgehend von der kleinen Stichprobe ist außerdem die Frage nach der Generalisierbarkeit der erzielten Befunde kritisch zu betrachten. Daher soll im Rahmen weiterer Forschungsvorhaben eine größere Stichprobe akquiriert werden. So wurde für das Wintersemester 2023/24 vereinbart, dass das Lehrkonzept und die Messinstrumente auch an der Karl-Franzens-Universität Graz in einer Lehrveranstaltung des Instituts für Wirtschaftspädagogik eingesetzt werden. Österreich zählt auch zum deutschsprachigen Raum der Sektion Berufs- und Wirtschaftspädagogik, sodass aus dieser Sichtweise Befunde der Übertragbarkeit des Lehrkonzepts und des Messinstruments erwartbar sind. Außerdem wird die zugrunde liegende Stichprobe aktuell Semester für Semester erweitert. So fanden auch im SoSe 2023 weitere Befragungen an unterschiedlichen Hochschulstandorten statt.

Die Perspektive des Längsschnitts ist bei zwei Messzeitpunkten deutlich eingeschränkt (Klauer, 2009, S. 526). Um belastbare Aussagen über Entwicklungsverläufe zu ermöglichen, sollten aus diesem Grund weitere Messzeitpunkte eingeführt werden. Denkbar wäre z. B. eine Follow-Up-Messung nach Abschluss des betreffenden Semesters. Mit Längsschnittuntersuchungen sind im akademischen Bereich allerdings regelmäßig Limitationen aufgrund von Problemen des Feldzugangs (siehe dazu Happ, 2017, S. 135–137) sowie der Panelmortalität verbunden (siehe dazu Happ, 2017, S. 138–140). Die Limitationen, die durch die Datenerhebung bei Studierenden in der akademischen Bildung verbunden sind, wurden zum Erhebungszeitraum der Daten für die vorliegende Dissertation noch durch Restriktionen durch die Covid-19-Pandemie verstärkt. Die Einbettung in Präsenzlehrveranstaltungen, um die Teilnahmequote zu erhöhen, war unter diesen Bedingungen aufgrund von Hygienevorschriften und Schließungen des Präsenzbetriebs nur sehr eingeschränkt möglich.

Eine weitere Limitation ergibt sich aus dem Einsatz des Tools Limesurvey für die Erhebung der Daten. Durch die Voreinstellung, dass keine Frage zum Grundlagenwissen ausgelassen werden kann, werden die Probandinnen und Probanden möglicherweise zum Raten animiert. Das kann zu verzerrten Testwerten führen.¹¹⁵ Die Rateeffekte können durch die Integration von Selbsteinschätzungsskalen nach jeder Frage allenfalls annähernd kontrolliert, nicht aber ausgeschlossen werden.

115 Zur Berechnung von Rateeffekten kann auf Walstad et al. (2018) verwiesen werden.

5.3 Implikationen

5.3.1 Implikationen für die theoretische Weiterentwicklung des Modells

Aus der Betrachtung der theoretischen Limitationen ergeben sich Implikationen für die Weiterentwicklung des aufgestellten Strukturmodells (Kapitel 2.5). Die inhaltliche Strukturierung des Untersuchungsgegenstandes (Kapitel 2.3) hat ergeben, dass neben den Bereichen Grundlagen der KI, Methoden der KI und Anwendungen der KI auch die Auswirkungen der KI relevant sind. Dabei handelt es sich hauptsächlich um ethisch-moralische Fragestellungen, die mit dem Einsatz KI-gestützter Systeme in verschiedenen Lebens- und Arbeitsbereichen assoziiert werden. Die ethisch-moralischen Fragestellungen können nicht unter dem in Kapitel 2.2 für die Arbeit definierten Wissensbegriff im Sinne eines Fakten- und Strukturwissens subsumiert werden. Sie bilden neben der kognitiven Facette des Wissens und den bereits inkludierten non-kognitiven Facetten der Einstellungen, Überzeugungen und Motivation eine weitere relevante non-kognitive Facette ab, die auf Grundlage der Literaturanalyse und Experteninterviews im Kontext von KI in der beruflichen Bildung betrachtet werden sollte. Da ethisch-moralische Fragestellungen demnach einen hohen Stellenwert im Kontext KI haben, wird das Rahmenmodell fortlaufend um eine respektive Facette erweitert, die die ethisch-moralischen Werthaltungen der Probandinnen und Probanden erfasst. Das betreffende Instrument wird aktuell entwickelt und auf Grundlage erster Befunde umfassend validiert (siehe dazu Guggemos et al., in print). In zukünftigen Untersuchungen wird diese Facette demnach inkludiert sein.

Das Strukturmodell bildet aktuell kognitive und non-kognitive Kompetenzfacetten einer KI-bezogenen Kompetenz ab. Zum aktuellen Zeitpunkt sind noch keine Aussagen über den Modellcharakter sowie über direkte oder indirekte Effekte auf Grundlage empirischer Befunde möglich. In nachfolgenden Analysen sollen demnach potenzielle Effekte auf Grundlage von Strukturgleichungsmodellierungen untersucht werden. Mit der Modellierung im Rahmen der vorliegenden Dissertation wurde auf Grundlage des unzulänglichen Forschungsstandes in diesem Bereich (siehe Kapitel 1) die konzeptionelle Grundlage für die Ausweitung der Modellierung einer KI-bezogenen Kompetenz geschaffen. Zu dieser Weiterentwicklung gehört darüber hinaus auch die Frage nach einer entwicklungsorientierten Perspektive und der Graduierbarkeit im Rahmen der Kompetenzmodellierung sowie der Eignung des Modells als Niveau- und Entwicklungsmodell (vgl. Nickolaus, 2018).

In Anlehnung an die Diskussion des *TPACK*-Modells wäre eine Modellierung weiterer Facetten, wie dem *TCK* (siehe Kapitel 2.1.2, vgl. Mishra & Koehler, 2006) denkbar. In dieser Facette können substanzielle Veränderungen von beruflichen Tätigkeiten, die den Gegenstandsbezug der beruflichen Bildung abbilden (siehe Kapitel 1), aufgrund digitaler (KI-gestützter) Technologien verortet werden. Eine Erweiterung des Modells um diese Perspektive würde überdies nicht nur konzeptionelle Überlegungen zu KI-bedingten Veränderungen in einzelnen Tätigkeitsfeldern verlangen, sondern auch die Erweiterung der kognitionsbezogenen Differenzierung um weitere Arten des (Lehrer-)Wissens. Dies wiederum impliziert Herausforderungen in der Operationa-

lisierung und Erfassung des Untersuchungsgegenstandes, da die Erweiterung über die bisher im Modell intendierten kognitiven Prozesse des „Erinnerns“ und „Verstehens“ hinausgehen würde, was dementsprechend im Rahmen der Test- und Itemkonstruktion zu berücksichtigen wäre.

5.3.2 Implikationen für die Weiterentwicklung des Instruments

Die vorliegende Arbeit umfasst gemäß der „*Standards for Educational and Psychological Testing*“ (vgl. AERA et al., 2014) Untersuchungen zur Beurteilung der Validität im Sinne des Testinhalts, der internen Struktur sowie der Beziehung zu anderen Merkmalen. Die Validierungsaspekte „Antwort- und Denkprozesse“, „Interne Struktur“, sowie „Konsequenzen der Testung“ wurden bisher nicht analysiert. Da die einzelnen Validierungsaspekte auch in Beziehung zueinander zu betrachten sind (Happ, 2019, S. 28), wird eine Untersuchung der verbleibenden Aspekte dringend empfohlen. Für die Untersuchung der Antwort- und Denkprozesse während der Aufgabenbewältigung bieten sich z. B. kognitive Interviews im Sinne von Think-Aloud-Studien an (siehe z. B. Heidel & Happ, 2022; Heidel et al., in review; Kuhn & Brückner, 2013; Leighton, 2017). In diesen Studien werden die Proband:innen gebeten, ihre Denkprozesse während der Bearbeitung von Aufgaben zu verbalisieren und laut auszusprechen. Die inhaltsanalytische Auswertung der transkribierten Interviews erlaubt somit Rückschlüsse auf ggf. wahrgenommene Barrieren oder Distraktoren. Auf Grundlage dieser Befunde ist eine Weiterentwicklung des Instruments möglich.

Ein weiterer qualitativer Zugang ist die Einschätzung der inhaltlichen Relevanz der Testitems auf Grundlage von Experteninterviews (vgl. Beck, 2020). Mit diesem Vorgehen können substanzielle Einschätzungen darüber erlangt werden, inwieweit die abgedeckten Inhaltsbereiche als relevant und die eingesetzten Items in ihrer Formulierung als angemessen angesehen werden (vgl. Kuhn, 2014). Dadurch kann ein iterativer Prozess der Itemrevision und -weiterentwicklung initiiert werden.

Aufgrund der problematischen Schätzung von Cronbach's alpha als Maß der Reliabilität, die z. T. mit der Dichotomie der Items einhergeht, wäre aus messtheoretischer Sicht eine Überprüfung mittels Kennzahlen anzustreben, die sich in der psychologischen Diagnostik für diese Anwendungsfälle etabliert haben. Möglich wäre die Schätzung über Omega (vgl. Béland & Falk, 2022). Die Reliabilitätsschätzung nach Omega setzt eine unidimensionale Menge an Items voraus. Die in der vorliegenden Studie durchgeführten explorativen Faktorenanalysen (siehe Kapitel 4.4) lassen allerdings vermuten, dass diese Voraussetzungen für diese Studie nicht als erfüllt bewertet werden können. Darüber hinaus ist durch eine Generierung einer größeren Stichprobe eine IRT-Skalierung anzustreben (siehe Kapitel 4.1). Insbesondere für die Stichprobe, die geringe Erfahrungen zu Inhalten der KI vermuten lassen, ist eine Überarbeitung des Messinstruments wünschenswert.

5.3.3 Implikationen für die Aus- und Weiterbildung von (beruflichen) Lehrkräften

Aus der vorliegenden Dissertation geht hervor, dass die KI-gestützte digitale Transformation die (angehenden) Lehrkräfte im berufsbildenden Bereich vor eine fachwissenschaftliche und fachdidaktische Herausforderung stellt (siehe Kapitel 1, z. B. Meiners et al., 2022). Der Aus- und Weiterbildung von beruflichen Lehrkräften mangelt es bisher an substanziellen Anknüpfungspunkten für KI-bezogene Inhalte. Im Rahmen der vorliegenden Studie wurde ein Konzept vorgestellt, mit dem Inhalte zu Künstlicher Intelligenz adressatengerecht in die Ausbildung integriert werden können (siehe Kapitel 3.2.3.4 und ausführlicher Schmidt & Happ, 2022b). Die Befunde der Pre-Post-Studie haben gezeigt, dass die angehenden Lehrkräfte ihr Wissen zu KI über die Dauer eines Semesters signifikant erweitern konnten. Das impliziert, dass durch die Verknüpfung von KI-Inhalten mit genuinen Fragestellungen der Studienrichtung ein substanzieller Beitrag zur Förderung von KI-bezogenen Kompetenzfacetten der angehenden Lehrkräfte im berufsbildenden Bereich möglich ist.

Mit der vorliegenden Arbeit wurde die erste Phase der beruflichen Lehrkräftebildung fokussiert. Die Aus- und Weiterbildung von beruflichem Bildungspersonal ist essenziell, um die Qualität der beruflichen Bildung sicherzustellen (Frommberger & Lange, 2020, S. 519). Um die Qualität ganzheitlich zu gewährleisten, sind neben der ersten Phase der Lehrkräftebildung zwingend auch die zweite und dritte Phase zu fokussieren. Dabei müssen die Bildungsangebote jeweils den Spezifika der aktuellen Aus- und Weiterbildungsphase gerecht werden. Erste Ansätze für ein Aus- und Weiterbildungsangebot für berufliche Lehrkräfte in allen drei Phasen werden aktuell im Projekt KIWi-MOOC entwickelt. Am 01. Juni 2023 ist das vom Bundesministerium für Bildung und Forschung (BMBF) geförderte Kompetenzzentrum für digitales und digital gestütztes Unterrichten in den wirtschaftswissenschaftlichen Fächern offiziell gestartet. Das Institut für Wirtschaftspädagogik der Universität Leipzig ist Teil des Kompetenzzentrums, das sich insgesamt aus 14 führenden Universitätsstandorten deutschlandweit zusammensetzt. Unter der Leitung von Prof. Jens Klusmeyer (Universität Kassel) werden durch das Kompetenzzentrum im Rahmen des Verbundprojekts „Wirtschaftspädagogik und Ökonomische Bildung: Lehrkräftebildung und Unterricht digital (WÖRLD)“ evidenzbasierte digitale und hybride Lernumgebungen sowie digital gestützte Lehr- und Lernprozesse im Bereich der Wirtschaftswissenschaften erforscht. Das Ziel ist es, die digitalisierungsbezogene Kompetenzentwicklung von Lehrkräften über alle drei Phasen der Lehrkräftebildung zu fördern. Das Institut für Wirtschaftspädagogik unter der Leitung von Prof. Dr. Happ und die Pädagogische Hochschule Schwäbisch Gmünd, vertreten durch Jun.-Prof. Dr. Josef Guggemos, beteiligen sich mit dem Teilprojekt „KIWi-MOOC - Ein Massive Open Online Course zur Förderung und Erfassung von KI-Kompetenzen in der Domäne Wirtschaft“ mit einem Verbundprojekt an WÖRLD. Das Hauptziel des Teilprojekts besteht darin, durch die Entwicklung innovativer Lehr-Lern-Konzepte die Kompetenzen von Wirtschaftslehrkräften im Bereich der Anwendung Künstlicher Intelligenz zu fördern und die erzielten Kompetenzzuwächse zu evaluieren. Im von Juni 2023 bis Dezember 2025 laufenden Projekt

werden die in dieser Dissertation erzielten Befunde genutzt. Die im Rahmen der vorliegenden Dissertation entwickelten Instrumente werden im Rahmen von summativen und formativen Assessments auf der Online-Plattform weiterentwickelt. Auch ethisch-moralische Aspekte werden in dem Projekt durch die Zusammenarbeit mit Jun.-Prof. Josef Guggemos mit in die Modellierung aufgenommen.

Speziell für den Freistaat Sachsen wird aufbauend auf den Befunden aus der Dissertation ein Folgeprojekt angestrebt, das die Lehrkräftefortbildungen an ansässigen Schulen in den Blick nimmt. Dabei wird ausgehend von den im Rahmen der vorliegenden Dissertation generierten empirischen Befunden zum Vorwissen von (angehenden) Lehrkräften einerseits und der Evaluation des entwickelten Lehr-Lern-Konzept andererseits ein Weiterbildungsangebot für *in-service*-Lehrkräfte im berufsbildenden Bereich entwickelt, erprobt und umfassend evaluiert. Auf Grundlage der in Kapitel 1 diskutierten Relevanz von KI im Bildungsbereich ist davon auszugehen, dass der Bedarf an Weiterbildungsangeboten dieser Art stetig zunehmen wird. Dabei müssen zwingend auch jene Lehrkräfte adressiert werden, die nicht mehr in der (universitären) Ausbildung sind und dementsprechend andere inhaltliche und organisatorische Ansprüche an ein Lehr-Lern-Angebot haben. Im Sinne einer adressatengerechten Aus- und Weiterbildung ist eine präzise Bestimmung der Ausgangsbedingungen zwingend erforderlich.

Mit der Perspektive von zwei Forschungs- und Transferprojekten, die direkt anschließen, sollen die breitenwirksame Nutzbarkeit und die iterative Weiterentwicklung der Modellierung und Testentwicklung zu KI-bezogenen Kompetenzfacetten von (angehenden) Lehrkräften im berufsbildenden Bereich, die im Rahmen der vorliegenden Dissertation entwickelt wurden, gewährleistet werden.

Literaturverzeichnis

- Abele, S. (2016). Umgang mit Komplexität: Eine bedeutsame psychische Voraussetzung des domänenspezifischen Problemlösens? *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 112(1), 37–59.
- Abele, S., Behrendt, S., Weber, W., & Nickolaus, R. (2016). Berufsfachliche Kompetenzen von Kfz-Mechatronikern – Messverfahren, Kompetenzdimensionen und erzielte Leistungen (KOKO Kfz). In K. Beck, M. Landenberger & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der Förderinitiative ASCOT* (S. 171–203). W. Bertelsmann Verlag.
- Achtenhagen, F., & Pätzold, G. (2010). Lehr-Lernforschung und Mikrodidaktik. In R. Nickolaus, G. Pätzold, H. Reinisch & T. Tramm (Hrsg.), *Handbuch Berufs- und Wirtschaftspädagogik* (S. 137–159). Klinkhardt UTB.
- Adelberg, A. H., & Razek, J. (1984). The Cloze Procedure: A Methodology for Determining the Understandability of Accounting Textbooks. *The Accounting Review*, 59(1), 109–122.
- Aebli, H. (1978). Zur Einleitung. In D. A. Norman & D. E. Rumelhart (Hrsg.), *Strukturen des Wissens: Wege der Kognitionsforschung* (S. 9–12). Klett-Cotta.
- Aebli, H. (1980). *Denken: Das Ordnen des Tuns*, Bd. 1: *Kognitive Aspekte der Handlungstheorie*. Klett-Cotta.
- Aebli, H. (1981). *Denken: Das Ordnen des Tuns*, Bd. 2: *Denkprozesse*. Klett-Cotta.
- Ajzen, I. (1985). From cognition to actions. A theory of planned behavior. In J. Kuhl & J. Beckmann (Hrsg.), *Action Control: From Cognition to Behavior* (S. 11–39). Springer.
- Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational behavior and human decision processes*, 50, 179–211.
- Ajzen, I., & Madden, T. J. (1986). Prediction of goal directed behavior: attitude, intentions and perceived behavioral control. *Journal of Experimental Social Psychology*, 22, 453–474.
- Albers, S., Klapper, D., Konradt, U., Walter, A., & Wolf, J. (Hrsg.) (2009). *Methodik der empirischen Forschung* (3., überarbeitete Auflage). Gabler.
- Alhashmi, M., Mubin, O., & Baroud, R. (2021). Examining the Use of Robots as Teacher Assistants in UAE Classrooms: Teacher and Student Perspectives. *Journal of Information Technology Education: Research*, 20, 245–261.
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council of Measurement in Education (NCME) (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Anderson, J. R. (1983). *The Architecture of Cognition*. Harvard University Press.
- Anderson, L. W., & Krathwool, D. R. (Hrsg.) (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Allyn & Bacon (Pearson Education Group).

- André, E., Aurich, J. C., Bauer, W., Bullinger-Hoffmann, A., Heister, M., Huchler, N., Neuburger, R., Peissner, M., Stich, A., Suchy, O., Ramin, P., & Wächter, M. (2021). *Kompetenzentwicklung für KI: Veränderungen, Bedarfe und Handlungsoptionen*. Plattform Lernende Systeme Whitepaper. https://doi.org/10.48669/pls_2021-2
- Antonietti, C., Cattaneo, A., & Amenduni, F. (2022). Can teachers' digital competence influence technology acceptance in vocational education? *Computers in Human Behavior*, 132, 1–9. <https://doi.org/10.1016/j.chb.2022.107266>
- Arbinger, R. (1997). *Psychologie des Problemlösens: Eine anwendungsorientierte Einführung*. Wissenschaftliche Buchgesellschaft.
- Attwood, A., Bruster, B., & Bruster, G. (2020). An Exploratory Study of Preservice Teacher Perception of Virtual Reality and Artificial Intelligence for Classroom Management Instruction. *SRATE Journal*, 29(2), 1–9. <https://eric.ed.gov/?id=EJ1268557>
- Baacke, D. (1996). Medienkompetenz – Begrifflichkeit und sozialer Wandel. In A. von Rein (Hrsg.), *Medienkompetenz als Schlüsselbegriff* (S. 112–124). Deutsches Institut für Erwachsenenbildung (DIE).
- Bach, A. (2016). Nutzung von digitalen Medien an berufsbildenden Schulen – Notwendigkeit, Rahmenbedingungen, Akzeptanz und Wirkungen. In J. Seifried, S. Seeber & B. Ziegler (Hrsg.), *Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2016* (S. 107–123). Verlag Barbara Budrich.
- Baethge-Kinsky, V. (2019). Digitalisierung in der industriellen Produktion und Facharbeit: Gefährdung 4.0? *Mitteilungen aus dem SOFI*, 30(13), 2–5. http://www.sofi-goettingen.de/fileadmin/SOFI-Mitteilungen_30_screen.pdf
- Baethge-Kinsky, V., Baethge, M., & Lischewski, J. (2016). Bedingungen beruflicher Kompetenzentwicklung: institutionelle und individuelle Kontextfaktoren (SiKoFak). In K. Beck, M. Landenberger & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der Förderinitiative ASCOT* (S. 265–299). W. Bertelsmann Verlag.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Beck, K. (1995). Wirtschaftskundliches Wissen bei kaufmännischen Lehrlingen – Zur Diagnose kognitiver Strukturen mit dem WBT. In C. Metzger & H. Seitz (Hrsg.), *Wirtschaftliche Bildung: Träger, Inhalte und Prozesse* (S. 375–395). Verlag des Schweizerischen Kaufmännischen Verbandes.
- Beck, K. (2020). Ensuring content validity of psychological and educational tests – the role of experts. *Frontline Learning Research*, 8(6), 1–37.
- Beck, K., Landenberger, M., & Oser, F. (Hrsg.) (2016). *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der BMBF-Förderinitiative ASCOT*. W. Bertelsmann Verlag.

- Becker, M. (2016). Arbeitsprozesse und Berufsbildung im Kontext von „Handwerk 4.0“. In S. Jascke, U. Schwenger & T. Vollmer (Hrsg.), *Digitale Vernetzung der Facharbeit. Gewerblich-technische Berufsbildung in einer Arbeitswelt des Internets der Dinge* (Bd. 43, S. 71–86). W. Bertelsmann Verlag, Reihe Berufsbildung, Arbeit und Innovation.
- Becker, M., Spöttl, G., & Windelband, L. (2021). Künstliche Intelligenz und Autonomie der Technologien in der gewerblich-technischen Berufsbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik, Beiheft 31: Künstliche Intelligenz in der Berufsbildung*, 31–54.
- Béland, S., & Falk, C. F. (2022). A Comparison of Modern and Popular Approaches to Calculating Reliability for Dichotomously Scored Items. *Applied Psychological Measurement*, 46(4) 321–337. <https://doi.org/10.1177/01466216221084210>
- Bendel, O. (2021). Strukturelle und organisationale Rahmenbedingungen für den Einsatz von Pflegerobotern. *Zeitschrift für Berufs- und Wirtschaftspädagogik, Beiheft 31: Künstliche Intelligenz in der Berufsbildung*, 129–151.
- Berding, F. (2016). Welche Bedeutung haben epistemische Überzeugungen für Lernende in der beruflichen Bildung? *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 112, 80–107.
- Berding, F., Jahncke, H., & Holt, K. (2021). Learning Analytics in der Wirtschaftspädagogik. Eine Simulationsstudie für die Anwendung überwachten maschinellen Lernens für Inhaltsanalysen am Beispiel von Grundvorstellungen und (Selbst-)Reflexionskompetenz. *Zeitschrift für Berufs- und Wirtschaftspädagogik, Beiheft 31: Künstliche Intelligenz in der Berufsbildung*, 237–291.
- Berding, F., & Lamping, C. (2014). Epistemologische Überzeugungen als Bestandteil der professionellen Kompetenz von Lehrkräften und ihre Bedeutung für die Auswahl und Bewertung von Lernaufgaben aus Schulbüchern des Wirtschaftslehreunterrichts: Eine explorative Studie. In K. Rebmann (Hrsg.), *Schriften zur Berufs- und Wirtschaftspädagogik* (Bd. 12). Rainer Hampp.
- Bii, P. K., Too, J. K., & Mukwa, C. W. (2018). Teacher Attitude towards Use of Chatbots in Routine Teaching. *Universal Journal of Educational Research*, 6(7), 1586–1597. <https://doi.org/10.13189/ujer.2018.060719>
- Bindner, K., & Cramer, C. (2021). Digitalisierung in der Fachliteratur zum Lehrer*innenberuf – Eine Bestandsaufnahme und Verhältnisbestimmung mittels critical review. *Herausforderung Lehrer*innenbildung*, (4), 329–343. <https://doi.org/10.11576/hlz-4518>
- Bittner, E. A. C., Oeste-Reiß, S., Ebel, P. A., & Söllner, M. (2019). Mensch-Maschine-Kollaboration: Grundlagen, Gestaltungsherausforderungen und Potenziale für verschiedene Anwendungsdomänen. *HMD*, 56, 34–49.
- Blasius, J., Baur, N. (2022). Multivariate Datenstrukturen. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 727–748). Springer VS. https://doi.org/10.1007/978-3-658-37985-8_45
- Blömeke, S. (2003). Erwerb medienpädagogischer Kompetenz in der Lehrerbildung. Modell der Zielqualifikation, Lernvoraussetzungen der Studierenden und Folgerungen für Struktur und Inhalt des medienpädagogischen Lehramtsstudiums. *Medienpädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 3, 231–244.

- Blömeke, S. (2004). Empirische Befunde zur Wirksamkeit der Lehrerbildung. In S. Blömeke, P. Reinhold, G. Tulodziecki & J. Wildt (Hrsg.), *Handbuch Lehrerbildung* (S. 56–91). Westermann Klinkhardt.
- Blömeke, S. (2007). Messung der professionellen Kompetenz zukünftiger Lehrpersonen: Standards empirischer Lehrerausbildungsforschung, bildungstheoretische Herausforderungen und exemplarische Ergebnisse einer Studie in Deutschland. In C. Kraler & M. Schratz (Hrsg.), *Ausbildungsqualität und Kompetenz im Lehrberuf* (S. 191–208). LIT Verlag.
- Blömeke, S., Kaiser, G., & Döhrmann, M. (2011). Bedingungsfaktoren des fachbezogenen Kompetenzerwerbs von Lehrkräften. Zum Einfluss von Ausbildungs-, Persönlichkeits- und Kompositionsmerkmalen in der Mathematiklehrausbildung für die Sekundarstufe I. In W. Helsper & R. Tippelt (Hrsg.), *Pädagogische Professionalität* (S. 77–103). Beltz. DOI: 10.25656/01:7088
- Blömeke, S., Schwarz, B., Kaiser, G., Seeber, S., & Lehmann, R. (2009). Untersuchungen zum mathematischen und mathematikdidaktischen Wissen angehender GHR- und Gymnasiallehrkräfte. *Journal für Mathematik-Didaktik*, 30, 232–255. <https://doi.org/10.1007/BF03339081>
- Blömeke, S., Suhl, U., Kaiser, G., Felbrich, A. & Schmotz, Ch. (2010). Lerngelegenheiten und Kompetenzerwerb angehender Mathematiklehrkräfte im internationalen Vergleich. *Unterrichtswissenschaft*, 38(1), 29–50.
- Bloom, B. S. (1976). *Taxonomie von Lernzielen im kognitiven Bereich* (5. Auflage). Beltz.
- Bogner, A., Littig, B., & Menz, W. (2014). *Interviews mit Experten: Eine praxisorientierte Einführung*. Springer.
- Bohner, G. (2002) Einstellungen. In W. Stroebe, K. Jonas & M. Hewstone (Hrsg.), *Sozialpsychologie*. (S. 265–315). Springer.
- Böhner, M. (2009). Wirkungen des Vorbereitungsdienstes auf die Professionalität von Lehrkräften. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (S. 439–450). Beltz.
- Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology in higher education: a systematic evidence map. *International Journal of Educational Technology in Higher Education*, 17(2), 1–20.
- Bonin, H., Gregory, T., & Zierahn, U. (2015). *Übertragung der Studie Frey/Osborne (2013) auf Deutschland. Endbericht*. Zentrum für Europäische Wirtschaftsforschung GmbH.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* (6. Auflage). Springer.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler* (4. überarbeitete Auflage.). Springer.
- Botturi, L. (2019). Digital and media literacy in pre-service teacher education. *Nordic Journal of Digital Literacy*, 14, 3–4, 147–163.
- Brehm, J. (1993). *The phantom respondents: Opinion surveys and political representation*. The University of Michigan Press.

- Brendel, E. (1999). Externalistische vs. Internalistische Wissensanalysen. In J. Nida-Rümelin (Hrsg.), *Rationalität, Realismus, Revision – Vorträge des 3. Internationalen Kongresses der Gesellschaft für Analytische Philosophie vom 15. bis 18. September 1997 in München* (S. 188–195). De Gruyter.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie, D Praxisgebiete, 1 Pädagogische Psychologie, Bd. 3 Psychologie des Unterrichts und der Schule* (S. 177–212). Hogrefe.
- Brosius, F. (2018). *SPSS: Umfassendes Handbuch zu Statistik und Datenanalyse* (8. Auflage). Mitp.
- Brückner, S., Happ, R., & Schmidt, S. (2013). Validitätsproblematik bei der Prognose des Studienerfolgs mittels standardisierter Testverfahren in der Studiendomäne Wirtschaftswissenschaften. In O. Zlatkin-Troitschanskaia, R. Nickolaus & K. Beck (Hrsg.), *Kompetenzmodellierung und Kompetenzmessung bei Studierenden der Wirtschaftswissenschaften und der Ingenieurwissenschaften. Lehrerbildung auf dem Prüfstand. Sonderheft* (S. 7–25). Verlag Empirische Pädagogik.
- Brunner, M., Kunter, M., Krauss, S., Klusmann, U., Baumert, J., Blum, W., Neubrand, M., Dubberke, T., Jordan, A., Löwen, K., & Tsai, Y.-M. (2006). Die professionelle Kompetenz von Mathematiklehrkräften: Konzeptualisierung, Erfassung und Bedeutung für den Unterricht. Eine Zwischenbilanz des COACTIV-Projekts. In M. Prenzel & L. Alolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 54–82). Waxmann.
- Buehl, M. M., & Alexander, P. A. (2001). Beliefs About Academic Knowledge. *Educational Psychology Review*, 13(4), 385–418.
- Bühner, M. (2021). *Einführung in die Test- und Fragebogenkonstruktion*. Pearson.
- Bundesregierung (2018). *Strategie Künstliche Intelligenz der Bundesregierung*. https://www.bmbf.de/bmbf/de/forschung/digitale-wirtschaft-und-gesellschaft/kuenstliche-intelligenz/kuenstliche-intelligenz_node.html
- Bunting, L., Segerstad, Y., & Barendregt, W. (2021). Swedish teachers' views on the use of personalised learning technologies for teaching children reading in the English classroom. *International Journal of Child-Computer Interaction*, 27, 1–9. <https://doi.org/10.1016/j.ijcci.2020.100236>
- Bürg, O., & Mandl, H. (2004). *Akzeptanz von E-Learning in Unternehmen. Forschungsbericht Nr. 167*. Ludwigs-Maximilians-Universität München.
- Buske, R., Förster, M., Klink, S., Kuhn, C., Preuß, D., & Zlatkin-Troitschanskaia, O. (2011). Innovativer Lehr-Lernortverbund (ILLEV) in der akademischen Hochschulausbildung. In B. Schwarz, P. Nenniger & R. S. Jäger (Hrsg.), *Erziehungswissenschaftliche Forschung – nachhaltige Bildung. Beiträge zur 5. DGfE-Sektionstagung „Empirische Bildungsforschung“* (S. 27–33). Empirische Pädagogik.
- Busse, J., Geiser, P., Schumann, M., Seeber, S., Weber, S., Hackenberg, T., Zarnow, S., & Hiller, F. (2022). Didaktische Bedeutung der Digitalisierung für die kaufmännische Berufsausbildung. In S. Schumann, S. Seeber & S. Abele (Hrsg.), *Digitale Transformation in der Berufsbildung: Konzepte, Befunde und Herausforderungen* (S. 75–102). wbv Publikation.

- Buxmann, P., & Schmidt, H. (2021). *Künstliche Intelligenz. Mit Algorithmen zum wirtschaftlichen Erfolg*. (2. Aufl.). Springer Gabler.
- Carlson, R. (1990). Assessing teachers' pedagogical content knowledge: Item development issues. *Journal of Personnel Evaluation in Education*, 4, 157–173.
- Carretero, S., Vuorikari, R., & Punie, Y. (2017). *DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use*. Doi: 10.2760/38842
- Cattell, R. B. (1971). *Abilities: their structure, growth and action*. Houghton Mifflin.
- Cella, D., & Chang, C.-H. (2000). A Discussion of Item Response Theory and Its Applications in Health Status Assessment. *Medical Care*, 38(9), 66–72.
- Chai, C. S., Koh, J. H. L., & Tsai, C.-C. (2016). A Review of the Quantitative Measures of Technological Pedagogical Content Knowledge (TPACK). In M. C. Herring, M. J. Koehler & P. Mishra (Hrsg.), *Handbook of technological pedagogical content knowledge (TPACK) for educators* (S. 87–106). Taylor & Francis.
- Chan, R., Bista, K., & Allen, R. (2021). *Online teaching and learning in higher education during COVID-19: International perspectives and experiences*. Routledge.
- Child, I. L. (1968). Personality in culture. In E. F. Borgatta & W. W. Lambert (Hrsg.), *Handbook of personality theory and research* (S. 82–145). Rand McNally.
- Chiu, T. K. F., & Chai, C. (2020). *Sustainable Curriculum Planning for Artificial Intelligence Education: A Self-determination Theory Perspective*. <https://doi.org/10.3390/su12145568>
- Chiu, T. K. F. (2021). A Holistic Approach to the Design of Artificial Intelligence (AI) Education for K-12 Schools. *TechTrends*, 65(5), 796–807. <https://doi.org/10.1007/s11528-021-00637-1>
- Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2022). Systematic literature review on opportunities, challenges and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4, 1–15. <https://doi.org/10.1016/j.caeai.2022.100118>
- Chocarro, R., Cortiñas, M., & Marcos-Matás, G. (2021). 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), 295–313. <https://doi.org/10.1080/03055698.2020.1850426>
- Chomsky, N. (1976). *Aspects of the theory of syntax* (11. Aufl.). The M. I. T. Press.
- Chounta, I. A., Bardone, E., Raudsep, A., & Pedaste, M. (2021). Exploring Teachers' Perceptions of Artificial Intelligence as a Tool to Support their Practice in Estonian K-12 Education. *International Journal of Artificial Intelligence in Education*, 2022(32), 725–755. <https://doi.org/10.1007/s40593-021-00243-5>
- Chuah, K. M., & Kabilan, M. K. (2021). Teachers' Views on The Use of Chatbots to Support English Language Teaching in a Mobile Environment. *International Journal of Emerging Technologies in Learning*, 16(20), 223–237.
- Cochran-Smith, M., & Zeichner, K. (2005). *Studying Teacher Education: The report of the AERA Panel on Research and Teacher Education*. Mahwah.

- Cohen, F., & Hemmerich, F. (2020). *Nutzung digitaler Medien für die pädagogische Arbeit in der Kindertagesbetreuung: Kurzexpertise im Auftrag des Bundesministeriums für Familie, Senioren, Frauen und Jugend*. https://www.forum-transfer.de/fileadmin/uploads/Bibliothek/Kurzexpertise_Digitalisierung_Kindertagesbetreuung.pdf
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). Harper & Row.
- Darling-Hammond, L., & Bransford, J. (2005). *Preparing teachers for a changing world. What teachers should learn and be able to do*. San Francisco.
- Daugherty, P. R., & Wilson, H. J. (2018). *Human + Machine: Reimagining Work in the Age of AI*. Harvard Business Review Press.
- Davenport, T. H., & Kirby, J. (2016). Just how smart are smart machines? *MIT Sloan Management Review* 57(3), 21–25.
- Davies, R. S., & West, R. E. (2014). Technology Integration in Schools. In J. M. Spector, M. D. Merrill, J. Elen & M. J. Bishop (Hrsg.), *Handbook of Research on Educational Communications and Technology* (S. 841–853). Springer. http://doi.org/10.1007/978-1-4614-3185-5_68
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), S. 319–340.
- DeCharms, R. (1968). *Personal causation: The internal affective determinants of behavior*. Academic Press.
- De Witt, C. (2013). Vom E-Learning zum Mobile Learning – wie Smartphones und Tablet PCs Lernen und Arbeit verbinden. In C. de Witt & A. Sieber (Hrsg.), *Mobile Learning, Potenziale, Einsatzszenarien und Perspektiven des Lernens mit mobilen Endgeräten* (S. 13–26). Springer.
- Deci, E. L. (1992). Interest and intrinsic motivation of behavior. In K. A. Renninger, S. Hidi & A. Krapp (Hrsg.), *The role of interest in learning and development* (S. 43–70). Hillsdale.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.
- Deci, E. L., & Ryan, R. M. (1991). A motivational approach to self: Integration in personality. In R. Dienstbier (Hrsg.), *Nebraska symposium on motivation: Perspectives on motivation*, 38 (S. 237–288).
- Deci, E. L., & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39(2), 223–238.
- Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and Education: The Self-Determination Perspective. *Educational Psychologist*, 26(3/4), 325–346.
- Deckert, R., & Meyer, E. (2020). *Digitalisierung und Künstliche Intelligenz. Kooperation von Menschen und Maschinen aktiv gestalten*. Springer Fachmedien Wiesbaden.
- Dengler, K., & Matthes, B. (2018). Substituierbarkeitspotenziale von Berufen. Wenige Berufsbilder halten mit der Digitalisierung Schritt. *IAB Kurzbericht*, 4, 1–11.

- Deutscher, V., Seifried, J., Rausch, A., Thomann, H., & Braunstein, A. (2022). Die LUCA Office Simulation in der Lehrerinnen- und Lehrerbildung – Didaktische Design-Empfehlungen und erforderliche Lernkompetenzen. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach & M. Stock (Hrsg.), *Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation*. (S. 107–122). wbv Publikation.
- Dietzen, A., Tschöpe, T., Monnier, M., & Srbeny, C. (2016). Berufsspezifische Messung sozialer Kompetenzen auf Basis eines Situational-Judgement-Tests bei Medizinischen Fachangestellten (CoSMed). In K. Beck, M. Landenberger & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der Förderinitiative ASCOT* (S. 225–241). W. Bertelsmann Verlag.
- Dietzmann, C., & Alt, R. (2020). Assessing the business impact of Artificial Intelligence. *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 5170–5179. DOI:10.24251/HICSS.2020.635
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5., vollständig überarbeitete, aktualisierte und erweiterte Aufl.). Springer.
- Döring, O., Wittmann, E., Weyland, U., Nauwerth, A., Hartig, J., Kaspar, R., Möllers, M., Rechenbach, S., Simon, J., Worofka, I., & Kraus, K. (2016). Technologiebasierte Messung von beruflichen Kompetenzen für die Pflege älterer Menschen: berufsfachliche Kompetenzen, allgemeine Kompetenzen und Kontextfaktoren (TEMA). In K. Beck, M. Landenberger, & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der Förderinitiative ASCOT* (S. 243–264). W. Bertelsmann Verlag.
- Dorotea, N., Piedade, J., & Pedro, A. (2021). Mapping K-12 Computer Science Teacher Interest, Self-Confidence, and Knowledge about the Use of Educational Robotics to Teach. *Education Science*, 11(8), 1–15.
- Dubs, R. (1995). *Lehrerverhalten: Ein Beitrag zur Interaktion von Lehrenden und Lernenden im Unterricht*. Verlag des Schweizerischen Kaufmännischen Verbandes.
- Dubs, R. (2018). Gedanken zur Zukunft der kaufmännischen Grundbildung. In J. Schlicht & U. Moschner (Hrsg.), *Berufliche Bildung an der Grenze zwischen Wirtschaft und Pädagogik: Reflexionen aus Theorie und Praxis* (S. 37–58). Springer VS.
- Durchholz, M. M. (2023). *Multimodale Intervention bei Kindern mit Lernschwierigkeiten und gleichzeitig auftretenden Verhaltensauffälligkeiten*. [Dissertation]. Ludwig-Maximilians-Universität München. https://edoc.ub.uni-muenchen.de/31422/2/Durchholz_Martina_Maria.pdf
- Dwivedi, Y. K., Nripendra, P. R., Chen, H., & Williams, M. D. (2011). A Meta-Analysis of the Unified Theory of Acceptance and Use of Technology (UTAUT). In M. Nüttgens, A. Gadatsch, K. Kautz, I. Schirmer & N. Blinn (Hrsg.), *Governance and Sustainability in Information Systems: Managing the Transfer and Diffusion of IT* (S. 155–170). Springer.
- Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In D. Gilbert, S. T. Fiske & G. Lindzey (Hrsg.), *Handbook of social psychology* (4. Aufl., S. 269–322). McGraw-Hill.

- Eberle, F., Schumann, S., Kaufmann, E., Jüttler, A., & Ackermann, N. (2016). Modellierung und Messung wirtschaftsbürgerlicher Kompetenz von kaufmännischen Auszubildenden in der Schweiz und in Deutschland. In K. Beck, M. Landenberger, & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der Förderinitiative ASCOT* (S. 93–117). W. Bertelsmann Verlag.
- Eickelmann, B., Bos, W., & Labusch, A. (2019). Die Studie ICILS 2018 im Überblick – Zentrale Ergebnisse und mögliche Entwicklungsperspektiven. In B. Eickelmann, W. Bos, J. Gerick, F. Goldhammer, H. Schaumburg, K. Schwippert, M. Senkbeil & J. Vahrenhold (Hrsg.), *ICILS 2018 #Deutschland. Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking* (S. 7–32). Waxmann. <https://doi.org/10.25656/01:18166>
- Elements of AI (2022). *Einführung in die KI*.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Erlbaum.
- Ertel, W. (2016). *Grundkurs Künstliche Intelligenz. Eine praxisorientierte Einführung* (4. Auflage). Springer Vieweg.
- Ertl, H., & Seifried, J. (2021). Auswirkungen künstlicher Intelligenz auf wichtige Berufsfelder. Ein Kommentar zu Teil A des Beihefts. *Zeitschrift für Berufs- und Wirtschaftspädagogik, Beiheft 31: Künstliche Intelligenz in der Berufsbildung*, 153–160.
- Ernst, F. (2012). *Fachdidaktische Analyse von Lehrbüchern für den Rechnungswesenunterricht in Deutschland und den USA*. Dissertation. Friedrich-Schiller-Universität Jena.
- Eschenbach, W. J. von (2021). Transparency and the Black Box Problem: Why We Do Not Trust AI. *Philosophy & Technology*, 34, 1607–1622. <https://doi.org/10.1007/s13347-021-00477-0>
- Euler, D., & Wilbers, K. (2020). Berufsbildung in digitalen Lernumgebungen. In R. Arnold, A. Lipsmeier & M. Rohs (Hrsg.), *Handbuch Berufsbildung* (S. 427–438). Springer.
- Europäische Kommission (2010). *Eine digitale Agenda für Europa. Mitteilung der Kommission an das europäische Parlament, den Rat, den europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen*. Europäische Kommission. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:de:PDF>
- Faath-Becker, A., & Walker, F. (2020). Development of a Video-based Test Instrument for the Assessment of Professional Competence in the Vocational Teacher Training. In E. Wuttke, J. Seifried & H. Niegemann (Hrsg.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (S. 109–130). Barbara Budrich.
- Farnell, T. (2021). *The impact of COVID-19 on higher education: A review of emerging evidence: analytical report*. Publications Office of the European Union.
- Fechter, R. (2015). *Neue Wege in der Berufsorientierung: Auswirkungen einer handlungsorientierten Lernsequenz im FabLab auf die Lernmotivation, das kreative Problemlösen und die Einstellung zu Technik* [Dissertation]. Technische Universität Dresden. <https://core.ac.uk/download/pdf/236372826.pdf>

- Fischer, S., & Barabasch, A. (2020). Gamification: A Novel Didactical Approach for 21st Century Learning. In E. Wuttke, J. Seifried & H. Niegemann (Hrsg.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (S. 89–106). Barbara Budrich.
- Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention and behavior: An introduction to theory and research. Addison-Wesley.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People – An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Förster, M., Brückner, S., Happ, R., Beck, K., & Zlatkin-Troitschanskaia, O. (2017a). Strukturanalyse eines kognitiven Messinstruments im Multiple-Choice-Format: Das Beispiel des Test of Economic Literacy (TEL4-G). *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 113(3), 366–396.
- Förster, M., Happ, R., & Molerov, D. (2017b). Using the U. S. Test of Financial Literacy in Germany – Adaptation and validation. *Journal of Economic Education*, 48(2), 123–135. <https://doi.org/10.1080/00220485.2017.1285737>
- Förster-Kuschel, J. (2022). Umgang mit Heterogenität – ein Thema in Lehrbüchern zur Vorbereitung auf die AEVO-Prüfung? *Berufsbildung in Wissenschaft und Praxis*, 2022(4), 41–45. BIBB.
- Förster-Kuschel, J., & Fürstenau, B. (2020). Wie empfinden betriebliche Ausbilderinnen und Ausbilder den Umgang mit Heterogenität der Lernenden? *BWP*, 49(1), 48–49.
- Fretheim, A. (2021). Relationship between teaching modality and COVID-19, well-being, and teaching satisfaction (Campus & Corona): A cohort study among students in higher education. *bioRxiv.org*. 1–20. <https://doi.org/10.1101/2021.03.26.21254388>
- Frey, D., Stahlberg, D., & Gollwitzer, P. M. (1993). Einstellung und Verhalten: Die Theorie des überlegten Handelns und die Theorie des geplanten Verhaltens. In D. Frey & M. Irle (Hrsg.), *Theorien der Sozialpsychologie* (Bd. 1, S. 361–398). Huber.
- Frey, C. B., & Osborne, M. A. (2013). *The future of employment: How susceptible are jobs to computerization*. University of Oxford.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerization? *Technological forecasting and social change*, 14, 254–280.
- Fritzsche, E. (2012). *Antwortsicherheit von Grundschulkindern: Ein Maß für das Selbstkonzept?* Dissertation. Friedrich-Alexander-Universität Erlangen-Nürnberg.
- Frommberger, D., & Lange, S. (2020). Professionalisierung des berufsschulischen Bildungspersonals – Status Quo und Reformansätze. In R. Arnold, A. Lipsmeier & M. Rohs (Hrsg.), *Handbuch Berufsbildung* (3. Aufl., S. 519–531). Springer.
- Fuchs, E., Niehaus, I., & Stoletzki, A. (2014). *Das Schulbuch in der Forschung – Analysen und Empfehlungen für die Bildungspraxis*. V&R unipress.
- Future of Life Institute (2017). *Asilomar AI Principles*. <https://futureoflife.org/ai-principles/>
- Geiser, P. (2022). *Lehrerüberzeugungen zur Bedeutung der Digitalisierung. Eine Interviewstudie mit Lehrkräften zur Ausbildung kaufmännischer Fachkräfte*. wbv Publikation.

- Geiser, P., Busse, J., Seeber, S., Schumann, S., Weber, S., Zarnow, S., Hiller, F., Hackenberg, T., & Lange, A. (2021). Kompetenzen in digitalisierten kaufmännischen Arbeitsplatzsituationen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 117(4), 630–657.
- Gentilin, O. (2019). KI in der Schule: Digitale Lehrkonzepte und Anwendungsbeispiele für den Fremdsprachenunterricht. *Information – Wissenschaft & Praxis*, 71(1), 5–16. doi:10.1515/iwp-2019-2056
- Gerholz, K.-H., Schlottmann, P., Slepcevic-Zach, P., & Stock, M. (Hrsg.) (2022a). *Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation*. wbv Publikation.
- Gerholz, K.-H., Schlottmann, P., Slepcevic-Zach, P., & Stock, M. (2022b). Digital Literacy in der beruflichen Lehrerinnen- und Lehrerbildung – Einleitung. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach & M. Stock (Hrsg.), *Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation*. (S. 11–18). wbv Publikation.
- Gerholz, K.-H., Maidanuk, I., & Schlottmann, P. (2022c). Virtual Reality in der (beruflichen) Lehrerinnen- und Lehrerbildung – Hochschuldidaktische Einordnung und empirische Befunde auf Basis eines systematischen Literaturreviews. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach & M. Stock (Hrsg.), *Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation*. (S. 185–198). wbv Publikation.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A brunswikian theory of confidence. *Psychological Review*, 98, 506–528.
- Gläser, J., & Laudel, G. (2010). *Experteninterviews und qualitative Inhaltsanalyse*. Springer.
- Gniewosz, B. (2015). III-3 Testverfahren. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung* (S. 71–81). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-19992-4_5
- Gocen, A., & Aydemir, F. (2020). Artificial Intelligence in Education and Schools. *Research on Education and Media*, 12(1), 13–21. <https://doi.org/10.2478/rem-2020-0003>
- Görz, G., Braun, T., & Schmid, U. (2021). Einleitung. In G. Görz, U. Schmid & T. Braun (Hrsg.), *Handbuch der Künstlichen Intelligenz* (6. Aufl., S. 1–26). Oldenbourg.
- Gössling, B., Hagemeyer, D., & Sloane, P. F. E. (2019). Berufsbildung 4.0 als didaktische Herausforderung: Zum Umgang von Lehrkräften an berufsbildenden Schulen mit digitalisierten Arbeitswelten. *Zeitschrift für Berufs- und Wirtschaftspädagogik* (4), 546–566.
- Graham, C. R. (2011). Theoretical considerations for understanding technological pedagogical content knowledge (TPACK). *Computers and Education*, 57(3), 1953–1960. <https://doi.org/10.1016/j.compedu.2011.04.010>
- Graham, C. R., Borup, J., & Smith, N. B. (2012). Using TPACK as a framework to understand teacher candidates' technology integration decisions. *Journal of Computer Assisted Learning*, 530–546. <https://doi.org/10.1111/j.1365-2729.2011.00472.x>
- Grimes, D., & Warschauer, M. (2010). Utility in a Fallible Tool: A Multi-Site Case Study of Automated Writing Evaluation. *Journal of Technology, Learning, and Assessment*, 8(6).

- Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly*, 68(1), 2–31.
- Gruber, H., Mandl, H., & Renkl, A. (1999). *Was lernen wir in Schule und Hochschule: Träges Wissen?* (Forschungsbericht Nr. 101). München: Ludwig-Maximilians-Universität.
- Guggemos, J., Schmidt, J., & Happ, R. (in print). *A matter of power – prospective teachers' attitudes towards the ethical principles of artificial intelligence use in education.*
- Guggemos, J., & Seufert, S. (2020). Teaching with and teaching about technology – Evidence for professional development of in-service teachers. *Computers in Human Behavior*, 115, 1–11.
- Guggemos, J., Seufert, S., & Sonderegger, S. (2020). Humanoid robots in higher education: Evaluating the acceptance of Pepper in the context of an academic writing course using the UTAUT (Nr. 5). *British Journal of Educational Technology*, 51, 1864–1883.
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, 51(5), 1–42. <https://doi.org/10.1145/3236009>
- Happ, R. (2017). *Die Entwicklung des volkswirtschaftlichen Grundlagenswissens im Studienverlauf – Effekte von Eingangsvoraussetzungen auf den Wissenserwerb.* Verlag Empirische Pädagogik.
- Happ, R. (2019). *Der Einfluss des Migrationshintergrundes auf unterschiedliche Facetten des ökonomischen Wissens – Implikationen für die Wirtschaftsdidaktik.* Kumulative Habilitationsschrift. Johannes Gutenberg-Universität Mainz.
- Happ, R., & Förster, M. (2018). The correlation between vocational school students' test motivation and the performance in a standardized test of economic knowledge: using direct and indirect indicators of test motivation. *Empirical Research in Vocational Education and Training*. <https://doi.org/10.1186/s40461-018-0071-x>
- Happ, R., Zlatkin-Troitschanskaia, O., & Schmidt, S. (2016). An Analysis of Economic Learning among Undergraduates in Introductory Economics Courses in Germany. *The Journal of Economic Education*, 47(4), 300–310. <https://doi.org/10.1080/00220485.2016.1213686>
- Harlow, H. F. (1958). The nature of love. *American Psychologist*, 13, 673–685.
- Hartig, J., Frey, A., & Jude, N. (2020). Validität von Testwertinterpretationen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 529–545).
- Hartig, J., Klieme, E., & Leutner, D. (2008). *Assessment of competencies in educational contexts: State of the art and future prospects.* Hogrefe & Huber.
- Härtig, H., Kauertz, A., Neumann, K., & Fischer, H. E. (2010). Inhaltsvalidität von Tests: Expertenrating versus Schulbuchanalyse. In D. Höttercke (Hrsg.), *Gesellschaft für Didaktik der Chemie und Physik: Entwicklung naturwissenschaftlichen Denkens zwischen Phänomen und Systematik* (S. 161–163). Lit.
- Haseski, H. I. (2019). What Do Turkish Pre-Service Teachers Think About Artificial Intelligence? *International Journal of Computer Science Education in Schools*, 3(2), 1–17. <https://doi.org/10.21585/ijcses.v3i2.55>

- Hasse, A., Cortesi, S. C., Lombana, A., & Gasser, U. (2019). *Youth and artificial intelligence: Where we stand*. Youth and Media, Berkman Klein Centre for Internet & Society. <https://cyber.harvard.edu/publication/2019/youth-and-artificial-intelligence/where-we-stand>
- Hauser, C., & Klotz, U. (2022). Ausgewählte Anwendungen der Künstlichen Intelligenz und deren Auswirkungen auf die beruflichen Tätigkeiten – eine Momentaufnahme. *Anwendungen und Konzepte der Wirtschaftsinformatik*, 16, 57–68.
- Heidel, S., & Happ, R. (2023). Challenges in Understanding Western Economic and Financial Concepts from the Perspective of Young Adults with a Post-Soviet Migration Background in Germany – Findings from a Qualitative Interview Study. *Journal of Risk and Financial Management*, 16(3), 1–21. <https://doi.org/10.3390/jrfm16030165>
- Heidel, S., Schmidt, J., & Moschner, U. (2023). Informelle Lerngelegenheiten für den Erwerb ökonomischen und finanziellen Wissens bei jungen Erwachsenen ohne wirtschaftswissenschaftliche Vorkenntnisse in Deutschland – die (besondere) Rolle der Eltern. In: M. Hommel, C. Aprea & K. Heinrichs (Hrsg.), *bwp@ Profil 8: Netzwerke – Strukturen von Wissen, Akteuren und Prozessen in der beruflichen Bildung. Digitale Festschrift für Bärbel Fürstenau zum 60. Geburtstag* (S. 1–21).
- Helmke, A. (2012). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts*. Klett-Kallmeyer.
- Helmreich, R. (1977). *Strategien zur Auswertung von Längsschnittdaten: Ein Beitrag zur Messung von Veränderung in der empirischen Sozialforschung*. Klett.
- Hew, K. F., Lan, M., Tang, Y., Jia, C., & Lo, C. K. (2019). Where is the theory within the field of educational technology research? *British Journal of Educational Technology*, 50(3), 956–971. <https://doi.org/10.1111/bjet.12770>
- Hoc, J.-M. (2000). From human-machine interaction to human-machine cooperation. *Ergonomics*, 43(7), 833–843.
- Hofer, B. K. (2000). Dimensionality and disciplinary differences in personal epistemology. *Contemporary Educational Psychology*, 25(4), 378–405.
- Hofer, B. K. (2006). Beliefs About Knowledge and Knowing: Integrating Domain Specificity and Domain Generality: A response to Muis, Bendixen and Haerle (2006). *Educational Psychology Review*, 18(1), 67–76.
- Hofer, B. K. (2010). Personal epistemology in Asia: Burgeoning research and future directions. *The Asia-Pacific Education Researcher*, 19(1), 179–184.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories. *Review of educational research*, 67(1), 88–140.
- Hoffmann, C. P., Weber, J., Zepic, R., Greger, V., & Krcmar, H. (2019). Dimensionen digitaler Mündigkeit und politische Beteiligung im Netz. In I. Engelmann, M. Legrand & H. Marzinkowski (Hrsg.), *Politische Partizipation im Medienwandel* (S. 79–99). Berlin. <https://doi.org/10.17174/dcr.v6.4>
- Hofstetter, A., & Massmann, C. (2021). *Digital Skills Gap Report 2020 – Skills we need, Skills we teach*. <https://kohorten.com/2020/05/13/digital-skill-gap-studie/>

- Holmes, W., Anastopoulou, S., Schaumburg, H., & Mavrikis, M. (2018). *Technology-enhanced Personalised Learning: Untangling the Evidence*. Robert Bosch Stiftung. https://www.bosch-stiftung.de/sites/default/files/publications/pdf/2018-08/Study_Technology-enhanced%20Personalised%20Learning.pdf
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign.
- Holtel, S. (2016). Artificial Intelligence Creates a Wicked Problem for the Enterprise. *Proceedia Computer Science*, 99, 171–180.
- Hox, J. J., de Leeuw, E. D., Brinkhuis, M. J. S., & Ooms, J. (2012). Multigroup and multilevel approaches to measurement equivalence. In S. Salzborn, E. Davidov & J. Reinecke (Hrsg.), *Methods, Theories and Empirical Applications in the Social Sciences* (S. 91–96). Springer.
- HSG UniStGallen (14.02.2023). *Wie ChatGPT und Co. das Bildungswesen umkrepeln*. [Video] Youtube. <https://www.youtube.com/watch?v=gxPg9Y861co>
- Huchler, N. (2016). Die Grenzen der Digitalisierung. Neubestimmung der hybriden Handlungsträgerschaft zwischen Mensch und Technik und Implikationen für eine humane Technikgestaltung. *HMD*, 53, 109–123.
- Hussain, I. (2020). Attitude of University Students and Teachers towards Instructional Role of Artificial Intelligence. *International Journal of Distance Education and E-Learning (IJDEEL)*, 5(2), 158–178.
- Ifenthaler, D. (2006). *Diagnose lernabhängiger Veränderung mentaler Modelle: Entwicklung der SMD-Technologie als methodologisches Verfahren zur relationalen, strukturellen und semantischen Analyse individueller Modellkonstruktionen*. FreiDok.
- Ifenthaler, D., & Schweinbenz, V. (2013). The acceptance of Tablet-PCs in classroom instruction: The teachers' perspectives. *Computers in Human Behavior*, 29(3), 523–534. <https://doi.org/http://dx.doi.org/10.1016/j.chb.2012.11.004>
- Ifenthaler, D., & Schweinbenz, V. (2016). Students' acceptance of tablet PCs in the classroom. *Journal of Research on Technology in Education*, 48(4), 306–321. <https://doi.org/http://dx.doi.org/10.1080/15391523.2016.1215172>
- Ifenthaler, D., & Widanapathirana, C. (2014). Development and Validation of a Learning Analytics Framework: Two Case Studies Using Support Vector Machines. *Tech Know Learn*, 19, 221–240. <https://doi.org/10.1007/s10758-014-9226-4>
- Ifenthaler, D., & Yau, J. Y.-K. (2021). Learning Analytics zur Unterstützung von Lernerfolg. Ausgewählte Ergebnisse einer systematischen Übersichtsarbeit. *Zeitschrift für Berufs- und Wirtschaftspädagogik, Beiheft 31: Künstliche Intelligenz in der Berufsbildung*, 215–235.
- Institute for the Future [IFTF] (2011). *Future Work Skills 2020*. <https://www.iftf.org/projects/future-work-skills-2020/>
- Jarrahi, M. H. (2018). Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making. *Business Horizons* 61(4), 577–586.

- Jenßen, L., Dunekacke, S., & Blömeke, S. (2015). Qualitätssicherung in der Kompetenzforschung. Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis. In S. Blömeke & O. Zlatkin-Troitschanskaia (Hrsg.), *Kompetenzen von Studierenden* (S. 11–31). Beltz Juventa.
- Jie, D. (2016). *A study of novice teacher's practical knowledge acquisition*. Dissertation. XiNam University.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes*, 57, 226–246.
- Juslin, P., & Olsson, H. (1997). Thurstonian and brunswikian origins of uncertainty in judgement: A sampling model of confidence in sensory discrimination. *Psychological Review*, 104, 344–366.
- Kampmeyer, D. (2014). *Gut geraten oder sicher gewusst? Eine Querschnittstudie über den Zusammenhang von Antwortrichtigkeit und Antwortsicherheit bei der Beantwortung von Multiple-Choice-Fragen durch Medizinstudierende verschiedener Ausbildungsstadien*. Dissertation. Universität zu Köln.
- Kampmeyer, D., Matthes, J., & Herzig, S. (2015). Lucky guess or knowledge: a cross-sectional study using the Bland and Altman analysis to compare confidence-based testing of pharmacological knowledge in 3rd and 5th year medical students. *Adv. Health Sci. Educ.*, 20, 431–40.
- Kamsker, S., & Riebenbauer, E. (2022). Digitalisierung in der Aus- und Fortbildung von Lehrkräften in der Wirtschaftspädagogik – eine erste Bestandsaufnahme aus Österreich. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach & M. Stock (Hrsg.), *Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation*. (S. 21–34). wbv Publikation.
- Kane, M. T. (1992). An argument-based approach to validation. *Psychological Bulletin*, 112(3), 527–535.
- Kane, M. T. (2006a). Validation. In R. Brennan (Hrsg.), *Educational measurement 4th edition* (S. 17–64). American Council on Education and Praeger.
- Kane, M. T. (2006b). Content-related validity evidence in test development. In T. M. Haladyna & S. M. Downing (Hrsg.), *Handbook of test development* (S. 131–153). Lawrence Erlbaum.
- Kane, M. T. (2012). Validating score interpretations and uses: Messick Lecture, Language Testing Research Colloquium, Cambridge, April 2010. *Language Testing* 29(1), 3–17. <https://doi.org/10.1177/0265532211417210>
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Kaufmann, E. (2020). Algorithm appreciation or aversion? Comparing in-service and pre-service teachers' acceptance of computerized expert models. *Computers and Education: Artificial Intelligence*, 2. <https://doi.org/10.1016/j.caeai.2021.100028>

- Kennedy, J., Lemaignan, S., & Belpaeme, T. (26.–31.8. 2016). *The cautious attitude of teachers towards social robots in schools*. IEEE RO-MAN, New York.
- Kerres, M., Hanft, A., Wilkesmann, U., & Wolff-Bendik, K. (2012). *Studium 2020: Positionen und Perspektiven zum lebenslangen Lernen an Hochschulen*. Waxmann.
- Kerres, M. (2013). *Mediendidaktik: Konzeption und Entwicklung mediengestützter Lernangebote*. (4. überarbeitete und aktualisierte Aufl.). Oldenbourg.
- Kerres, M. (2017). Digitalisierung als Herausforderung für die Medienpädagogik: Bildung in einer digital geprägten Welt. In C. Fischer (Hrsg.), *Pädagogischer Mehrwert? Münstersche Gespräche zur Pädagogik* (Bd. 33, S. 85–104). Waxmann.
- KI-Campus. (2020). *Lernangebote*. <https://ki-campus.org/courses/einfuehrungski2020>
- Kim, S. W., & Lee, Y. (2015). A Survey on Elementary School Teachers' Attitude toward Robot. *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, 1802–1807*.
- Kirschner, P. A., & Bruyckere, P. de (2017). The myths of the digital native and the multitasker. *Teaching and Teacher Education*, 67, 135–142. <https://doi.org/10.1016/j.tate.2017.06.001>
- Klauer, J. (2009). Veränderungsmessung. In H. Holling (Hrsg.), *Enzyklopädie der Psychologie, B Methodologie und Methoden, IV Evaluation, Bd. 1 Grundlagen und statistische Methoden der Evaluationsforschung* (S. 525–563). Hogrefe.
- Klauser, F. (1999). Zur Entwicklung der Professionalität von Wirtschaftslehrerinnen und Wirtschaftslehrern im fortschreitenden Transformationsprozeß der neuen Bundesländer. In T. Tramm, D. Sembill, F. Klauser & E. G. John (Hrsg.), *Professionalisierung kaufmännischer Berufsbildung. Beiträge zur Öffnung der Wirtschaftspädagogik für die Anforderungen des 21. Jahrhunderts* (S. 403–419). Peter Lang.
- Klauser, F. (2006). Wissenschaftliche Grundlagen zur Erforschung der Akzeptanz computer- und netzbasierter Lernangebote. In G. Minnameier & E. Wuttke (Hrsg.), *Berufs- und wirtschaftspädagogische Grundlagenforschung: Lehr-Lern-Prozesse und Kompetenzdiagnostik; Festschrift für Klaus Beck*. (S. 111–128). Peter Lang.
- Klauser, F., Schlicht, J., & Wirth, K. (2007). *Evaluation und Qualitätskontrolle des Pilotlehrgangs Personalstaboffizier/Personaloffizier der Streitkräfte – Abschlussbericht*. (Unveröffentlichte Arbeit). Lehrstuhl für Berufs- und Wirtschaftspädagogik, Universität Leipzig.
- Kleesiek, J., Murray, J. J., Strack, C., Kaissis, G., & Braren, R. (2020). Wie funktioniert maschinelles Lernen? *Der Radiologe*, 60(1), 24–31.
- Klieme, E., & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In M. Prenzel, I. Gogolin & H.-H. Krüher (Hrsg.), *Kompetenzdiagnostik* (S. 11–29). Springer.
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52(6), 876–903.

- Klimczak, P., Kusche, I., Tschöpe, C., & Wolff, M. (2019). Menschliche und maschinelle Entscheidungsrationalt  t: Zur Kontrolle und Akzeptanz K  nstlicher Intelligenz. *Zeitschrift f  r Medienwissenschaft*, 11(2), 39–45. <https://doi.org/10.25969/mediarep/12631>
- Klugmann, C. M. (2021). Black Boxes and Bias in AI Challenge Autonomy. *The American Journal of Bioethics: AJOB*, 21(7), 33–35. <https://doi.org/10.1080/15265161.2021.1926587>
- KMK (2007). *Handreichung f  r die Erarbeitung von Rahmenlehrpl  nen der Kultusministerkonferenz f  r den berufsbezogenen Unterricht in der Berufsschule und ihre Abstimmung mit Ausbildungsordnungen des Bundes f  r anerkannte Ausbildungsberufe*. Sekretariat der Kultusministerkonferenz. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2007/2007_09_01-Handreich-RLpl-Berufsschule.pdf
- KMK (2017). *Bildung in der digitalen Welt. Strategie der Kultusministerkonferenz*. Beschluss der Kultusministerkonferenz vom 08.12.2016. https://www.kmk.org/fileadmin/pdf/PresseUndAktuelles/2018/Digitalstrategie_2017_mit_Weiterbildung.pdf
- Knecht, P., Matthes, E., Sch  tze, S., & Aamotsbakken, B. (2014). *Methodologie und Methoden der Schulbuch- und Lehrmittelforschung/Methodology and methods of research on textbooks and educational media*. Verlag Julius Klinkhardt.
- K  gler, K., Egloffstein, M., & Sch  nberger, B. (2020). Openness in MOOCs for Training and Professional Development – An Exploration of Entry and Participation Barriers. In E. Wuttke, J. Seifried & H. Niegemann (Hrsg.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (S. 205–223). Barbara Budrich.
- Kollmann, T. (1998). Akzeptanz innovativer Nutzungsg  ter und -systeme: Konsequenzen f  r die Einf  hrung von Telekommunikations- und Multimediasystemen. Gabler.
- Kollmann, T. (2000). Die Messung der Akzeptanz bei Telekommunikationssystemen. *Wissenschaftsjournal*, 2, 68–77.
- Korbach, A., & Niegemann, H. M. (2020). Microlearning via Smartphones in VET for Professional Drivers: The Case of Securing Cargo for International Transport. In E. Wuttke, J. Seifried & H. Niegemann (Hrsg.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (S. 183–201). Barbara Budrich.
- Kramer, K. (2002). *Die F  rderung von motivationsunterst  tzendem Unterricht – Ansatzpunkte und Barrieren*. [Dissertation]. Universit  t Kiel.
- Kr  mer, H., Jordanski, G., & Goertz, L. (2017). Medien anwenden und produzieren – Entwicklung von Medienkompetenz in der Berufsbildung. *Wissenschaftliche Diskussionspapiere*. <https://www.bibb.de/dienst/veroeffentlichungen/de/publication/download/8275>
- Krauss, S. (2011). Das Experten-Paradigma in der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 171–191). Waxmann.
- Krohne, H.-W., & Hock, M. (2007). *Psychologische Diagnostik: Grundlagen und Anwendungsfelder*. Kohlhammer.
- Kr  ner, S., & Biermann, A. (2007). The relationship between confidence and self-concept – Towards a model of response confidence. *Intelligence*, 35, 580–590.

- Krüger, M., & Kaulvers, J. (2019). Einstellungsveränderungen von BASportlehramtsstudierenden im Kontext eines Professionalisierungsangebots für den sprachbildenden Sportunterricht. *Zeitschrift für Studium und Lehre in der Sportwissenschaft*, 1(3), 14–25. <https://doi.org/10.25847/zsls.2018.008>
- Krumm, V. (1973). *Wirtschaftslehreunterricht: Analyse von Lehrplänen und Lehrinhalten an kaufmännischen Berufs- und Berufsfachschulen in der Bundesrepublik Deutschland*. Klett.
- Kühl, N., Schemmer, M., Goutier, M., & Satzger, G. (2022). Artificial Intelligence and machine learning. *Electronic Markets*, 32, 2235–2244. <https://doi.org/10.1007/s12525-022-00598-0>
- Kuhn, C. (2014). *Fachdidaktisches Wissen von Lehrkräften im kaufmännisch-verwaltenden Bereich. Modellbasierte Testentwicklung und Validierung*. Verlag Empirische Pädagogik.
- Kuhn, C., & Brückner, S. (2013). Analyse des fachdidaktischen Wissens von (angehenden) Lehrkräften in der kaufmännisch-verwaltenden Bildung mit der Methode des lauten Denkens. *bwp@ Berufs- und Wirtschaftspädagogik online* Nr. 24. https://www.bwpat.de/ausgabe24/kuhn_brueckner_bwpat24.pdf
- Kuhn, C., Happ, R., Zlatkin-Troitschanskaia, O., Beck, K., Förster, M., & Preuße, D. (2014). Kompetenzentwicklung angehender Lehrkräfte im kaufmännisch-verwaltenden Bereich – Erfassung und Zusammenhänge von Fachwissen und fachdidaktischem Wissen. In E. Winther & M. Prenzel (Hrsg.), *Perspektiven der empirischen Berufsbildungsforschung: Kompetenz und Professionalisierung*. *Zeitschrift für Erziehungswissenschaft*, 17(1), 149–167. Springer.
- Kuhn, C., Zlatkin-Troitschanskaia, O., Saas, H., & Brückner, S. (2018). Konstruktion, Implementation und Evaluation eines multimedialen Assessmenttools in der wirtschaftspädagogischen Ausbildung. In J. Schlicht & U. Moschner (Hrsg.), *Berufliche Bildung an der Grenze zwischen Wirtschaft und Pädagogik: Reflexionen aus Theorie und Praxis* (S. 339–355). Springer.
- Kunter, M. (2011). Forschung zur Lehrermotivation. In E. Terhart, H. Bennis & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 527–539). Waxmann.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Waxmann.
- Lee, S. H., & Chen, W. (2008). A comparative study of uncertainty propagation methods for black-box-type problems. *Struct Multidisc Optim*, 37, 239–253. <https://doi.org/10.1007/s00158-008-0234-7>
- Lee, E., Lee, Y., Kye, B., & Ko, B. (2008). Elementary and middle school teachers', students' and parents' perception of robot-aided education in Korea. *EdMedia+ Innovate Learning*, 175–183.
- Leighton, J. (2017). *Using think-aloud interviews and cognitive labs in educational research*. Oxford University Press.
- Lienert, G. A., & Ratz, U. (1998). *Testaufbau und Testanalyse* (6. Auflage). Beltz.
- Likert, R. (1932). A Technique for Measurement of Attitudes. *Archives of Psychology*, 140, 1–55.
- Linacre, J. M. (1994). *Many-Facet Rasch Measurement*. MESA Press.

- Lindner, A., & Berges, M.-P. (2020). *Can you explain AI to me? Teachers' pre-concepts about Artificial Intelligence*. 2020 IEEE Frontiers in Education Conference (FIE).
- Lindner, A., & Romeike, R. (2019). *Teachers' Perspectives on Artificial Intelligence. Conference Paper*. ISSEP 2019: 12th International conference on informatics in schools. Situation, evaluation and perspectives.
- Limesurvey (2011). *LimeSurvey: The Open Source Survey Application*.
- Lipp, S., & Stock, M. (2022). Mit Learning Analytics zu Digital Literacy – konzeptionelle Überlegungen eines digitalen Lernraums zur (Weiter-) Entwicklung von Digital Literacy. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach & M. Stock (Hrsg.), *Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation*. (S. 51–62). wbv Publikation.
- Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *CHI '20: Tagungsband der CHI-Konferenz 2020 zu Human Factors in Computing Systems, April 2020*. <https://doi.org/10.1145/3313831.3376727>
- Louie, B., Björling, E. A., & Kuo, A. C. (2021). The Desire for Social Robots to Support English Language Learners: Exploring Robot Perceptions of Teachers, Parents, and Students. *Frontiers in Education*, 6, 1–15. <https://doi.org/10.3389/feduc.2021.566909>
- Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J. H., Ogata, H., Baltes, J., Guerra, R., Li, P., & Tsai, C.-C. (2020). Challenges and Future Directions of Big Data and Artificial Intelligence in Education. *Frontiers in Psychology*, 11, 1–11. <https://doi.org/10.3389/fpsyg.2020.580820>
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed. An argument for AI in education*. Pearson.
- Lytras, M., Sarirete, A., & Damiani, E. (2020). Technology-enhanced learning research in higher education: A transformative education primer. *Computers in Human Behavior*, 109. <https://doi.org/10.1016/j.chb.2020.106350>
- Martin, A. (2006). An european framework for digital literacy. *Nordic Journal of Education*, 1(2), 1–15.
- Martín-Núñez, J. L., Ar, A. Y., Fernández, R. P., Abbas, A., & Radovanović, D. (2023). Does intrinsic motivation mediate perceived artificial intelligence (AI) learning and computational thinking of students during the COVID-19 pandemic? *Computers and Education: Artificial Intelligence*, 4, 1–8. <https://doi.org/10.1016/j.caeai.2023.100128>
- Mayer, H., & Kriz, W. (Hrsg.), (2010). *Evaluation von eLernprozessen*. Oldenbourg Wissenschaftsverlag.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.5465/amr.1995.9508080335>
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12., überarbeitete Aufl.). Beltz.
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1956). A proposal for the Dartmouth summer research project on artificial intelligence. *Dartmouth Conference*. <https://doi.org/10.1609/aimag.v27i4.1904>

- McClelland, D. C. (1973). Testing for competence rather than testing for “intelligence”. *American Psychologist*, 28(1), 1–14.
- McKinsey Global Institute (2018). *Notes from the frontier modelling the impact of AI on the World Economy*. Discussion Paper.
- Meiners, H., Hartmann, P., Niegemann, H., Seeber, S., Wuttke, E., & Schumann, S. (2022). Digitale Medienkompetenz als Voraussetzung für die Erstellung von Prüfungsaufgaben. In S. Schumann, S. Seeber & S. Abele (Hrsg.), *Digitale Transformation in der Berufsbildung: Konzepte, Befunde und Herausforderungen* (S. 123–144). wbv Publikation.
- Meinhardt, C. (2018). *Entwicklung und Validierung eines Testinstruments zu Selbstwirksamkeitserwartungen von (angehenden) Physiklehrkräften in physikdidaktischen Handlungsfeldern*. Logos Verlag. <https://doi.org/10.30819/4712>
- Meß, M. (2022). *Wissenstests zu Anwendungen der Künstlichen Intelligenz – Validierung von Items eines Fragebogens durch Experteninterviews*. Unveröffentlichte Masterarbeit. Universität Leipzig.
- Michaeli, T., Seegerer, S., & Romeike, R. (2020). *Alexa, was ist eigentlich Künstliche Intelligenz? Einblicke in KI als Gegenstand und Methode fachdidaktischer Forschung*. Freie Universität Berlin.
- Minnameier, G. (2000). *Entwicklung und Lernen: kontinuierlich oder diskontinuierlich?* Waxmann.
- Minnameier, G. (2005). *Wissen und inferentielles Denken*. Peter Lang.
- Minnameier, G. (2010). Aufbau von Wissen und Kompetenzen. In R. Nickolaus, G. Pätzold, H. Reinisch & T. Tramm (Hrsg.), *Handbuch Berufs- und Wirtschaftspädagogik* (S. 65–68). Klinkhardt UTB.
- Minnameier, G. (2013). The Inferential Construction of Knowledge in the Domain of Business and Economics. In K. Beck & O. Zlatkin-Troitschanskaia (Hrsg.), *From Diagnostics to Learning Success – Proceedings in Vocational Education and Training* (S. 141–156). Sense Publisher.
- Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record*, 108(6), 1017–1054.
- Montebello, M. (2018). *AI Injected E-Learning: The Future of Online Education*. *Studies in Computational Intelligence Series*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-67928-0>
- Moosbrugger, H., & Kelava, A. (Hrsg.) (2020). *Testtheorie und Fragebogenkonstruktion*. (3., vollständig neu bearbeitete, erweiterte und aktualisierte Auflage). Springer.
- Moschner, U. (2014). *Subjektive Theorien von Top-Managern zu Qualifikationsanforderungen an kaufmännische Fach- und Führungskräfte in der Energie- und Wasserwirtschaft*. Verlag Dr. Kovac.
- Moschner, U., & Schlicht, J. (2018). Interpersonales Vertrauen – eine Komponente der Komplexitätsreduktion in digitalisierten Arbeitsumgebungen. In J. Schlicht & U. Moschner (Hrsg.), *Berufliche Bildung an der Grenze zwischen Wirtschaft und Pädagogik: Reflexionen aus Theorie und Praxis* (S. 179–198). Springer.

- Müller, S. (2018). Berufsbildung 4.0. Was müssen die Fachkräfte der Zukunft können? In C. Bär, T. Grädler & R. Mayr (Hrsg.), *Digitalisierung im Spannungsfeld von Politik, Wirtschaft, Wissenschaft und Recht* (S. 292–298). Springer Gabler.
- Müller-Böling, D., & Müller, M. (1986). *Akzeptanzfaktoren der Bürokommunikation*. Oldenbourg.
- Nagel, N. (2019). *Entwicklung von Handlungskompetenz in der beruflichen Bildung: Analyse einer internetunterstützten Fortbildung von Berufspädagogen zum Thema „schulische Krisenprävention* [Inaugural-Dissertation]. Pädagogische Hochschule Weingarten. https://hsbwgt.bsz-bw.de/frontdoor/deliver/index/docId/293/file/Dissertation_NadineNagel.pdf
- Nenner, C., Damnik, G., & Bergner, N. (2021). Weil informatorische Bildung nicht erst in der Sekundarstufe beginnen darf: Integration informatorischer Bildung ins Grundschullehramtsstudium. In L. Humbert (Hrsg.), *Informatik – Bildung von Lehrkräften in allen Phasen* (S. 103–112). Gesellschaft für Informatik.
- Neuweg, G. H. (2014). Das Wissen der Wissensvermittler. Problemstellungen, Befunde und Perspektiven der Forschung zum Lehrerwissen. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. überarbeitete und erweiterte Aufl., S. 583–614). Waxmann.
- Ng, D. T. K., Leung, J. K. L., Chu, K. W. S., & Qiao, M. S. (2021a). *AI Literacy: Definition, Teaching, Evaluation and Ethical Issues*. 84th Annual Meeting of the Association for Information Science & Technology, Oct 29th – Nov, 3rd, Salt Lake City.
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Shen, M. Q. (2021b). Conceptualizing AI literacy: An exploratory overview. *Computers and Education: Artificial Intelligence*, 2, 1–11. <https://doi.org/10.1016/j.caeai.2021.100041>
- Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W., & Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Education Tech Research Dev* 71, 137–161.
- Nickolaus, R. (2010). Erklärungsmodelle für die Entwicklung der Fachkompetenz – Anmerkungen zu ihren Geltungsansprüchen und didaktischen Implikationen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 106(4), 481–490.
- Nickolaus, R. (2018). Kompetenzmodellierungen in der beruflichen Bildung – eine Zwischenbilanz. In J. Schlicht & U. Moschner (Hrsg.), *Berufliche Bildung an der Grenze zwischen Wirtschaft und Pädagogik: Reflexionen aus Theorie und Praxis* (S. 255–282). Springer.
- Niederhauser, D. S., & Lindstrom, D. L. (2018). Instructional Technology Integration Models and Frameworks: Diffusion, Competencies, Attitudes, and Dispositions. In J. Voogt (Hrsg.), *Handbook of Information Technology in Primary and Secondary Education* (S. 1–21). Springer. https://doi.org/10.1007/978-3-319-53803-7_23-1
- Niegemann, H. M. (2020). Introduction – A Look Back Ahead. In E. Wuttke, J. Seifried & H. Niegemann (Hrsg.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (S. 15–19). Barbara Budrich.

- Nistor, N., Wagner, M., & Heymann, J. O. (2012). Prädiktoren und Moderatoren der Akzeptanz von Bildungstechnologien. Die Unified Theory of Acceptance and Use of Technology auf dem Prüfstand. *Empirische Pädagogik*, 26(3), 343–371.
- OECD (2021). The state of higher education: One year in to the COVID-19 pandemic. OECD Publishing. <https://doi.org/10.1787/83c41957-en>
- OpenAI (2023). *GPT-4 Technical Report*. <https://doi.org/10.48550/arXiv.2303.08774>
- Oser, F. (2001). Modelle der Wirksamkeit in der Lehrer- und Lehrerinnenausbildung. In F. Oser & J. Oelkers (Hrsg.), *Die Wirksamkeit der Lehrerbildungssysteme* (S. 67–96). Chur.
- Oser, F., & Oelkers, J. (Hrsg.) (2001). *Die Wirksamkeit der Lehrerbildungssysteme*. Chur.
- Overdiek, M., & Petersen, T. (2022). *Was Deutschland über Algorithmen und Künstliche Intelligenz weiß und denkt: Ergebnisse einer repräsentativen Bevölkerungsumfrage: Update 2022*. Bertelsmann Stiftung.
- Paaß, G., & Hecker, D. (2020). *Künstliche Intelligenz. Was steckt hinter der Technologie der Zukunft?* Springer.
- Pargmann, J., & Berding, F. (2022). Künstliche Intelligenz in der beruflichen Bildung – Utopie oder Dystopie? Chancen und Herausforderungen von KI als Bildungstechnologie. In M. Eckelt, T. Ketschau, J. Klassen, J. Schauer, J. Schmees & C. Steib (Hrsg.), *Berufsbildungspolitik: Strukturen – Krise – Perspektiven* (S. 215–232). <https://doi.org/10.3278/9783763972623>
- Partnership for 21st century skills (2016). *Framework for 21st Century Learning*. https://www.marietta.edu/sites/default/files/documents/21st_century_skills_standards_book_2.pdf
- Peissner, M., Kötter, F., & Zaiser, H. (2019). Künstliche Intelligenz – Anwendungsperspektiven für Arbeit und Qualifizierung. *BWP*, 3/2019, BIBB, 9–13.
- Pellegrino, J. W. (2010). *The Design of an Assessment System for the Race to the Top: A Learning Sciences Perspective on Issues of Growth and Measurement*. Princeton: Educational Testing Service.
- Petermann, F. (2010). Veränderungsmessung. In J. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 896–902). Beltz.
- Petko, D. (2012a). Hemmende und förderliche Faktoren des Einsatzes digitaler Medien im Unterricht: Empirische Befunde und forschungsmethodische Probleme. In R. Schulz-Zander, B. Eickelmann, H. Moser, H. Niesyto & P. Grell (Hrsg.), *Jahrbuch Medienpädagogik 9. Qualitätsentwicklung in der Schule und medienpädagogische Professionalisierung* (S. 29–50). VS-Verlag.
- Petko, D. (2012b). Teachers' pedagogical beliefs and their use of digital media in classrooms: Sharpening the focus of the 'will, skill, tool' model and integrating teachers' constructivist orientations. *Computers and Education*, 58(4), 1351–1359. <https://doi.org/10.1016/j.compedu.2011.12.013>
- Petko, D. (2020). *Quo vadis TPACK? Scouting the road ahead. Proceedings of EdMedia + Innovative Learning* (S. 1277–1286). Association for the Advancement of Computing in Education (AACE).
- Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. Holt, Rinehart and Winston.

- Pfeiffer, A., Bezzina, S., Dingli, A., Wernbacher, T., Denk, N., & Fleischhacker, M. (2021). *Adaptive Learning and Assessment: From the Teachers' Perspective. Proceedings of INTED 2021 Conference*. https://www.research-gate.net/publication/350428795_ADAPTIVE_LEARNING_AND_ASSESSMENT_FROM_THE_TEACHERS%27_PERSPECTIVE
- Pfeiffer, S. (2020). Kontext und KI: Zum Potenzial der Beschäftigten für Künstliche Intelligenz und Machine-Learning. *HMD*, 57, 465–479.
- Plattform Lernende Systeme (Hrsg.) (2019). *Arbeit, Qualifizierung und Mensch-Maschine Interaktion* – Whitepaper.
- Pletz, C., & Zinn, B. (2018). Technologieakzeptanz von virtuellen Lern- und Arbeitsumgebungen in technischen Domänen. *Journal of Technical Education*, 6(4), 83–105. <http://www.journal-of-technical-education.de>
- Prensky, M. (2001). Digital Natives, Digital Immigrants Part 2: Do they really think differently? *On the Horizon*, 9(6), 1–6.
- Prenzel, M., Kristen, A., Dengler, P., Ettl, R., & Beer, T. (1996). Selbstbestimmtes motiviertes und interessiertes Lernen in der kaufmännischen Erstausbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 13. Beiheft: Künstliche Intelligenz in der Berufsbildung, 108–127.
- Pu, S., Ahmad, A., Khambari, M., Yap, N., & Ahrari, S. (2021). Improvement of Pre-Service Teachers' Practical Knowledge and Motivation about Artificial Intelligence through a Service-learning-based Module in Guizhou, China: A Quasi-Experimental Study. *Asian Journal of University Education*, 17(3), 203–219. <https://doi.org/10.24191/ajue.v17i3.14499>
- Puentedura, R. R. (2006). *Transformation, Technology and Education*. http://hippasus.com/resources/tte/puentedura_tte.pdf
- Puentedura, R. R. (2012). *The SAMR model: Background and exemplars*. https://hippasus.com/rrpweblog/archives/2012/08/23/SAMR_BackgroundExemplars.pdf
- Rausch, A., Deutscher, V., Seifried, J., Brandt, S., & Winther, E. (2021). Die webbasierte Bürosimulation LUCA – Funktionen, Einsatzmöglichkeiten und Forschungsausblick. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 117, 372–394.
- Rebmann, K., & Slopinski, A. (2018). Zum Diskrepanztheorem der (Berufs-)Bildung für eine nachhaltige Entwicklung. In J. Schlicht & U. Moschner (Hrsg.), *Berufliche Bildung an der Grenze zwischen Wirtschaft und Pädagogik. Reflexionen aus Theorie und Praxis*. (S. 73–90). Springer.
- Rebmann, K., Tenfelde, W., & Schlömer, T. (2011). Berufs- und Wirtschaftspädagogik. *Eine Einführung in Strukturbegriffe* (4. Aufl.). Gabler.
- Reetz, L., & Witt, R. (1974). *Berufsausbildung in der Kritik: Curriculumanalyse Wirtschaftslehre*. Hoffmann & Campe.
- Redecker, C. (2017). *European Framework for the Digital Competence of Educators: DigCompEdu*. Publications Office of the European Union. <https://publications.jrc.ec.europa.eu/repository/handle/JRC107466>
- Reetz, L., & Seyd, W. (1995). Curriculare Strukturen beruflicher Bildung. In R. Arnold & A. Lipsmeier (Hrsg.), *Handbuch der Berufsbildung* (S. 203–219). Leske + Budrich.

- Reich-Stiebert, N., & Eyssel, F. (2016). Robots in the Classroom: What Teachers Think About Teaching and Learning with Education Robots. In A. Agah, J. J. Cabibihan, A. M. Howard, M. A. Salichs & H. He (Hrsg.), *Social Robotics. Lecture Notes in Artificial Intelligence (LNAI)*, 9979. (S. 671–680). Springer.
- Reinhart, J., Mayer, O., & Greiner, C. (2021). Künstliche Intelligenz. Grundlagen, Anwendungsbeispiele und Umsatzstrategien für Unternehmen. Vogel Buchverlag.
- Reise, S. P., & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement*, 27, 133–144.
- Renkl, A. (2005). Fünf Dogmen in der Diskussion zum Lernen und Lehren: Provokante Thesen zu dysfunktionalen Voreingenommenheiten. In S. R. Schilling, J. R. Sparfeldt & C. Pruisken (Hrsg.), *Aktuelle Aspekte pädagogisch-psychologischer Forschung: Detlef H. Rost zum 60. Geburtstag* (S. 11–24). Waxmann.
- Reusser, K., Pauli, C., & Elmer, A. (2011). Berufsbezogene Überzeugungen von Lehrerinnen und Lehrern. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 478–495). Waxmann.
- Riebenbauer, E., Berding, F., & Flick-Holtsch, D. (2022). Fachwissenschaftliche Implikationen für die Aus- und Fortbildung von Lehrkräften zur Digitalisierung im Rechnungswesen. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach & M. Stock (Hrsg.), *Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation*. (S. 199–212). wbv Publikation.
- Roll, J. J., & Ifenthaler, D. (2020). The Impact of Learning Factories on Multidisciplinary Digital Competencies. In E. Wuttke, J. Seifried & H. Niegemann (Hrsg.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (S. 23–38). Barbara Budrich.
- Roppertz, S. (2021). Die Rolle und Bedeutung von Künstlicher Intelligenz in der Berufsausbildung – Implikationen für angehende Berufs- und Wirtschaftspädagog*innen. *bwp@*, 40, 1–23.
- Rosanda, V., & Istenić, A. (2021). A Stranger in the Classroom: Pre-service Teachers' Anxiety and Negative Attitudes Toward Humanoid Social Robots. In M. Rauterberg (Hrsg.), *Culture and Computing. Design Thinking and Cultural Computing* (S. 461–473). Springer.
- Rosch, M., & Frey, D. (1994). Soziale Einstellungen. In D. Frey & S. Greif (Hrsg.), *Sozialpsychologie. Ein Handbuch in Schlüsselbegriffen* (S. 296–305). Beltz.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. vollständig überarbeitete und erweiterte Aufl.). Verlag Hans Huber.
- Russel, S. J., & Norvig, P. (2009). *Artificial Intelligence: A modern approach*. Prentice Hall.
- Ryan, M. P. (1984). Monitoring Text Comprehension: Individual Differences in Epistemological Standards. *Journal of Educational Psychology*, 76(2), 248–258.
- Ryan, R. M. & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. The Guildford Press: New York.
- Saas, H., Kuhn, C., & Zlatkin-Troitschanskaia, O. (2020). Ein videobasiertes Lehr-Lernformat als innovativer hochschuldidaktischer Ansatz in der wirtschaftspädagogischen Lehramtsausbildung. In I. Gogolin, B. Hannover & A. Scheunpflug (Hrsg.), *Evidenzbasierung in der Lehrkräftebildung* (S. 315–340). Springer.

- Salleh, S., Musa, J., Jaidin, J. H., & Shahrill, M. (2021). Development of TVET Teachers' Beliefs about Technology Enriched Instruction through Professional Development Workshops: Application of the Technology Acceptance Model (Nr. 2). *Journal of Technical Education and Training*, 13, 25–33.
- Sánchez-Caballé, A., Gisbert-Cervera, M., & Esteve-Mon, F. (2020). The digital competence of university students: a systematic literature review, *Aloma*, 38(19), 63–74.
- Sanusi, T. I., Sunday Oyelere, S., & Olamide Omidiora, J. (2021). Exploring Teachers Pre-concepts of Teaching Machine Learning in High School: A preliminary Insight from Africa, *Computers and Education Open*. <https://doi.org/10.1016/j.caeo.2021.100072>
- Saßenroth, D. (2013). *The impact of personality on participation decisions in surveys: A contribution to the discussion on unit nonresponse*. Springer VS.
- Schäfer, P., Link, N., & Walker, F. (2020). Assessing the Professional Knowledge of Teachers at Vocational Schools – Using the Example of a Professional Development for Automation and Digitized Production. In E. Wuttke, J. Seifried & H. Niegemann (Hrsg.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (S. 131–154). Barbara Budrich.
- Schaper, N. (2014). Validitätsaspekte von Kompetenzmodellen und -tests für hochschulische Kompetenzdomänen. In F. Musekamp & G. Spöttl (Hrsg.), *Kompetenz im Studium und in der Arbeitswelt. Nationale und internationale Ansätze zur Erfassung von Ingenieurkompetenzen* (S. 21–48). Peter Lang Verlag.
- Scheiter, K. (2021). Lernen und Lehren mit digitalen Medien: Eine Standortbestimmung. *Zeitschrift für Erziehungswissenschaft*, 24(5), 1039–1060.
- Scherer, R., Siddiq, F., & Tondeur, J. (2018). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*.
- Schermelleh-Engel, K., & Werner, C. (2007). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 113–133). Springer.
- Scheuer, D. (2020). *Akzeptanz von Künstlicher Intelligenz: Grundlagen intelligenter KI-Assistenten und deren vertrauensvolle Nutzung*. Springer Vieweg.
- Schiefele, U. (1996). *Motivation und Lernen mit Texten*. Hogrefe.
- Schiefele, U. (2009). Motivation. In F. Wild & J. Möller (Hrsg.), *Einführung in die Pädagogische Psychologie* (S. 151–177). Springer.
- Schiefele, U., & Köller, O. (2010). Intrinsische und extrinsische Motivation. In J. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 336–344). Beltz.
- Schiefele, U., Moschner, B., & Husstegge, R. (2002). *Skalenhandbuch SMILE-Projekt*. Bielefeld: Universität Bielefeld, Abteilung für Psychologie.
- Schlicht, J. (2012). *Kosten-Nutzen-Analyse von beruflicher Weiterbildung. Untersuchungen zur Wirtschaftlichkeit und pädagogischen Wirksamkeit*. Gabler.
- Schlicht, J., Hommel, M., & Fürstenau, B. (2022). Hochschuldidaktik – Digital gestütztes, problembasiertes forschendes Lernen von (künftigen) Lehr- und Führungskräften. *Bildung und Beruf*, (5), 259–264.

- Schlicht, J., Hommel, M., Fürstenau, B., & Klauser, F. (2017). WiWiPäd: Ein komplexes Lehr-Lern-Arrangement für forschendes Lernen in wirtschaftswissenschaftlichen und wirtschaftspädagogischen Studiengängen. *HDS Journal*, (1), 42–48.
- Schlottmann, P., & Gerholz, K.-H. (2022). Digital Literacy für Wirtschaftspädagog:innen – eine konzeptionelle Modellierung für die berufliche Lehrer:innenbildung. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach & M. Stock (Hrsg.), *Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation*. (S. 35–50). wbv Publikation.
- Schlottmann, P., Gerholz, K.-H., & Winther, E. (2021). Digital Literacy für Wirtschaftspädagog*innen – Modellierung des domänenspezifischen Fachwissens in der beruflichen Lehrerbildung. *bwp@ Berufs- und Wirtschaftspädagogik – online*, (40), 1–20.
- Schmechtig, N., Puderbach, R., Schellhammer, S., & Gehrmann, A. (2019). *Einsatz von und Umgang mit digitalen Medien und Inhalten in Unterricht und Schule: Befunde einer Lehrkräftebefragung zu beruflichen Erfahrungen und Überzeugungen von Lehrerinnen und Lehrern in Sachsen 2019*. Dresden: Zentrum für Lehrerbildung, Schul- und Berufsbildungsforschung.
- Schmid, M., Brianza, E., & Petko, D. (2020). Developing a short assessment instrument for Technological Pedagogical Content Knowledge (TPACK.xs) and comparing the factor structure of an integrative and a transformative model. *Computers and Education*, 157, 1–12. <https://doi.org/10.1016/j.compedu.2020.103967>
- Schmidt, J., & Happ, R. (2022a). Fachspezifisches Wissen, Einstellungen und Überzeugungen zu Anwendungen der Künstlichen Intelligenz bei Studierenden der Wirtschaftspädagogik. In K. Kögler, U. Weyland & H. H. Kremer (Hrsg.), *Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2022* (S. 89–102). Verlag Barbara Budrich.
- Schmidt, J., & Happ, R. (2022b). Befunde aus einem Lehr- und Forschungsprojekt zur Integration von Inhalten zu KI-Anwendungen bei angehenden Berufsschullehrkräften. In D.-K. Mah & C. Torner (Hrsg.), *Künstliche Intelligenz mit offenen Lernangeboten an Hochschulen lehren. Erfahrungen und Erkenntnisse aus dem Fellowship-Programm des KI-Campus* (S. 57–72). KI-Campus.
- Schmidt, M., Makhkamova, A., Spilski, J., Berg, M., Pietschmann, M., Exner, J.-P., Rugel, D., & Lachmann, T. (2020). Competence Development with Digital Learning Stations in VET in the Crafts Sector. In E. Wuttke, J. Seifried & H. Niegemann (Hrsg.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (S. 39–61). Barbara Budrich.
- Schmidt, S., Zlatkin-Troitschanskaia, O., Happ, R., & Kuhn, C. (2015). Die Entwicklung der epistemologischen Überzeugungen von angehenden Lehrkräften im kaufmännisch-verwaltenden Bereich – Analyse des Zusammenspiels der Überzeugungen mit fachspezifischem Wissen und intellektueller Leistungsfähigkeit. In A. Rausch, J. Warwas, J. Seifried & E. Wuttke (Hrsg.), *Konzepte und Ergebnisse ausgewählter Forschungsfelder der beruflichen Bildung – Festschrift für Detlef Sembill* (S. 107–130). Schneider Verlag Hohengehren.

- Schmitt, B., Klaffke, H., Sievers, T., Tracht, K., & Petersen, M. (2021). Veränderungen der Kompetenzanforderungen durch Zukunftstechnologien in der industriellen Fertigung. *Zeitschrift für Berufs- und Wirtschaftspädagogik, Beiheft 31: Künstliche Intelligenz in der Berufsbildung*, 103–127.
- Schneider, W. (2018). E-Learning zwischen Euphorie und Ernüchterung. In J. Schlicht & U. Moschner (Hrsg.), *Berufliche Bildung an der Grenze zwischen Wirtschaft und Pädagogik: Reflexionen aus Theorie und Praxis* (S. 199–215). Springer.
- Schnell, R., Hill, P. B., & Esser, E. (2008). *Methoden der empirischen Sozialforschung* (8. Aufl.). Oldenbourg.
- Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, 82(3), 498–504.
- Schommer, M. (1993). Comparisons of beliefs about the nature of knowledge and learning among postsecondary students. *Research in Higher Education*, 34(3), 355–370.
- Schommer-Aikins, M. (2004). Explaining the epistemological belief system: Introducing the embedded systematic model and coordinated research approach. *Educational Psychologist*, 39(1), 19–29.
- Schoop, E., Sonntag, R., Altmann, M., & Sattler, W. (2021). Stell dir vor, es ist „Corona“ – und keiner hat's gemerkt. *Lessons Learned*, 1 (1/2), <https://doi.org/10.25369/ll.v1i1/2.33>
- Schröder, H. (2001). *Didaktisches Wörterbuch* (3. erweiterte und aktualisierte Aufl.). Oldenbourg.
- Schulmeister, R. (2012). Vom Mythos der Digital Natives und der Net Generation. *Berufsbildung in Wissenschaft und Praxis* 41(3), 42–45.
- Schumann, M., & Lange, A. (2019). *Digitalisierung als Game Changer (Arbeitsbericht)*. Georg-August-Universität Göttingen.
- Schumann, S., Seeber, S., & Abele, S. (Hrsg.) (2022). *Digitale Transformation in der Berufsbildung. Konzepte, Befunde und Herausforderungen*. wbv Publikation.
- Sczogiel, S., Malapally, A., Schmitt-Rueth, S., & Williger, B. (2019). *Future Digital Job Skills: Die Zukunft kaufmännischer Berufe*. Industrie- und Handelskammer Nürnberg für Mittelfranken.
- Seeber, S., & Nickolaus, R. (2010). Kompetenzmessung in der beruflichen Bildung. *BWP*, 39, 10–13.
- Seeber, S., Schumann, M., Ketschau, T. J., Rüter, T., & Kleinhans, J. (2016). Modellierung und Messung von Fachkompetenzen Medizinischer Fachangestellter (CosMed). In K. Beck, M. Landenberger & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der Förderinitiative ASCOT* (S. 205–223). W. Bertelsmann Verlag.
- Seifried, J. (2006). Überzeugungen von (angehenden) Handelslehrern. In J. Seifried & J. Abel (Hrsg.), *Empirische Lehrerbildungsforschung. Stand und Perspektiven. Beiträge aus den Symposien „Empirische Lehrerbildungsforschung“ und „Kompetenzentwicklung in der Lehrerbildung“*, gehalten auf der Herbsttagung 2005 der Arbeitsgruppe für Empirische Pädagogische Forschung in Salzburg (S. 109–128). Waxmann.
- Seifried, J. (2009). *Unterricht aus Sicht von Handelslehrern*. Internationaler Verlag der Wissenschaften.

- Seifried, J. (2010). Sichtweisen von Lehrkräften an kaufmännischen Schulen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 106(2), 199–219.
- Seifried, J., & Ertl, H. (2021). Forschungsrichtungen zur künstlichen Intelligenz in der beruflichen Bildung – Ein Kommentar zu Teil B des Beihefts. *Zeitschrift für Berufs- und Wirtschaftspädagogik, Beiheft 31: Künstliche Intelligenz in der Berufsbildung*, 341–347.
- Seifried, J., Rausch, A., Kögler, K., Brandt, S., Eigenmann, R., Schley, T., Siegfried, C., Egloffstein, M., Küster, J., Wuttke, E., Sembill, D., Martens, T., & Wolf, K. D. (2016). Problemlösekompetenz angehender Industriekaufleute – Konzeption des Messinstruments und ausgewählte empirische Befunde (DomPL-IK). In K. Beck, M. Landenberger & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der BMBF-Förderinitiative ASCOT* (S. 119–138). W. Bertelsmann Verlag.
- Seifried, J., & Wuttke, E. (2015). Was wissen und können (angehende) Lehrkräfte an kaufmännischen Schulen? Empirische Befunde zur Modellierung und Messung der professionellen Kompetenz von Lehrkräften. *Empirische Pädagogik*, 29(1), 125–146.
- Seifried, J., & Wuttke, E. (2016). Professionelle Kompetenzen von Lehrkräften – Das Beispiel kognitiver Aktivierung. *bwp@ Profil 4. Digitale Festschrift für Sabine Matthäus*, 1–18.
- Sektion Berufs- und Wirtschaftspädagogik (BWP) der Deutschen Gesellschaft für Erziehungswissenschaft (2014). *Basiscurriculum für das universitäre Studienfach Berufs- und Wirtschaftspädagogik im Rahmen berufs- und wirtschaftspädagogischer Studiengänge*. https://www.dgfe.de/fileadmin/OrdnerRedakteure/Sektionen/Sek07_BerWiP/2014_Basiscurriculum_BWP.pdf
- Sembill, D., & Seifried, J. (2009). Konzeptionen, Funktionen und intentionale Veränderungen von Sichtweisen. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrerprofessionalität: Bedingungen, Genese, Wirkungen und ihre Messung* (S. 345–354). Beltz.
- Serholt, S., Barendregt, W., Leite, I., Hastie, H., Jones, A., Paiva, A., Vasalou, A., & Castellano, G. (2014). Teachers' views on the use of empathic robotic tutors in the classroom. *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 955–960.
- Seufert, S., & Guggemos, J. (2022). Digitale Kompetenzen von Lehrpersonen – Ergebnisse einer empirischen Untersuchung in der Berufsbildung. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach & M. Stock (Hrsg.), *Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation*. (S. 213–226). wbv Publikation.
- Seufert, S., Guggemos, J., & Ifenthaler, D. (2021). Zukunft der Arbeit mit intelligenten Maschinen: Implikationen der Künstlichen Intelligenz für die Berufsbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, (31. Beiheft), 9–27.
- Seufert, S., Guggemos, J., & Sailer, M. (2020a). Technology-related knowledge, skills and attitudes of pre-service and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior*, 115, 1–7. <https://doi.org/10.1016/j.chb.2020.106552>

- Seufert, S., Guggemos, J., & Sonderegger, S. (2020b). Digitale Transformation der Hochschullehre: Augmentationsstrategien für den Einsatz von Data Analytics und Künstlicher Intelligenz. *Zeitschrift für Hochschulentwicklung*, 15(1), 81–101. <https://doi.org/10.3217/zfhe-15-01/05>
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22.
- Siegfried, C., & Hermkes, R. (2020). Tablet PCs in Economics Classes – An Empirical Study on Motivational Experiences and Cognitive Load. In E. Wuttke, J. Seifried & H. Niegemann (Hrsg.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (S. 63–87). Barbara Budrich.
- Siegfried, C., & Wuttke, E. (2021). Development of economic competence of students at general grammar schools through a non-formal opportunity to learn with experts. *Citizenship, Social and Economics Education*, 20(2), 122–142.
- Slepcevic-Zach, P., Stock, M., & Köck, V. (2022). Lehren und Lernen in der COVID-19-Pandemie im Masterstudium Wirtschaftspädagogik an der Universität Graz. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach & M. Stock (Hrsg.), *Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation*. (S. 271–284). wbv Publikation.
- Söllner, M., Janson, A., Rietsche, R., & Thiel, M. (2021). Individualisierung in der beruflichen Bildung durch Hybrid Intelligence. Potentiale und Grenzen. *Zeitschrift für Berufs- und Wirtschaftspädagogik, Beiheft 31: Künstliche Intelligenz in der Berufsbildung*, 163–181.
- Spante, M., Lundin, M., Hashemi, S. S., & Algers, A. (2018). Digital competence and digital literacy in higher education research: Systematic review of concept use. *Cogent Education*, 5(1), 1–21.
- Staatssekretariat für Bildung, Forschung und Innovation (SBFI) (2015). *Schweizer EQR-Zuordnungsbericht vom 17.12.2015*. https://www.sbf.admin.ch/dam/sbf/de/dokumente/schweizer_eqr-zuordnungsbericht.pdf.download.pdf/schweizer_eqr-zuordnungsbericht.pdf
- Staatssekretariat für Bildung, Forschung und Innovation (SBFI) (2019). *Herausforderungen der künstlichen Intelligenz. Bericht der interdepartementalen Arbeitsgruppe „Künstliche Intelligenz“ an den Bundesrat*. SBFI. <https://www.sbf.admin.ch/sbf/de/home/bfi-politik/bfi-2021-2024/transversale-themen/digitalisierung-bfi/kuenstliche-intelligenz.html>
- Ständige Wissenschaftliche Kommission (SWK) (2022). *Digitalisierung im Bildungssystem: Handlungsempfehlungen von der Kita bis zur Hochschule: Gutachten der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz (SWK)*. https://www.kmk.org/fileadmin/Dateien/pdf/KMK/SWK/2022/SWK-2022-Gutachten_Digitalisierung.pdf
- Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. *Intelligence*, 28, 121–143.

- Stankov, L., & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology*, 100, 961–976.
- Steil, J., & Wrede, S. (2019). Maschinelles Lernen und lernende Assistenzsysteme: Neue Tätigkeiten, Rollen und Anforderungen für Beschäftigte? *BWO*, (3/2019), BiBB.
- Steppuhn, D. (2023). Der Einsatz Künstlicher Intelligenz in Schulen. *Bildung und Beruf*, 6 (April 2023), 122–125.
- Stobbe, L. (2021). Akzeptanz Lernender gegenüber Learning Analytics: Untersuchung des Einflusses der individuellen Technikeinstellung (Nr. 6). *IPTB Preprint Journal (Online Working Papers der Professur für Ingenieurpädagogik und Didaktik der technischen Bildung)*, 1–16.
- Stock, M., Slepcevic-Zach, P., & Kopp, M. (2022). Haben oder nicht haben, das ist hier die Frage! Eine empirische Studie zur digitalen Kompetenz von Studienanfängerinnen und Studienanfängern. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach & M. Stock (Hrsg.), *Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation*. (S. 169–183). wbv Publikation.
- Stork, J. H. (2013). Der Einfluss unterschiedlicher Unterrichtsmodelle auf den Erwerb von mathematischen und kaufmännischen Kompetenzen im beruflichen Unterricht. In U. Fasshauer, B. Fürstenau & E. Wuttke (Hrsg.), *Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2013* (S. 11–22). Barbara Budrich. <https://library.oapen.org/handle/20.500.12657/25702>
- Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial Intelligence (AI) Literacy in Early Childhood Education: The Challenges and Opportunities. *Computers and Education: Artificial Intelligence*, 4, 1–14. <https://doi.org/10.1016/j.caeai.2023.100124>
- Taulli, T. (2019). *Artificial intelligence basics: A non-technical introduction*. Apress.
- Teichmann, M., Ulrich, A., Wenz, J., & Gronau, N. (2020). Herausforderungen und Handlungsempfehlungen betrieblicher Weiterbildungspraxis in Zeiten der Digitalisierung. *HMD*, 57, 512–527.
- Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S., Leutner, D., Neuhaus, B. J., Sandmann, A., Sumfleth, E., Thillmann, H., & Wirth, J. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften*, (18), 7–28.
- Tergan, S.-O. (1989). Psychologische Grundlagen der Erfassung individueller Wissensrepräsentationen. Teil I: Grundlagen der Wissensmodellierung. *Sprache und Kognition*, 8(3), 152–165.
- Terhart, E. (Hrsg.) (2000). *Perspektiven der Lehrerbildung in Deutschland. Abschlussbericht der von der Kultusministerkonferenz eingesetzten Kommission*. Weinheim.
- Thatcher, A. (2020). Predicting the impact of COVID-19 on Australian universities. *Journal of risk and financial management*. <https://doi.org/10.3390/jrfm13090188>
- Thelk, A. D., Sundre, D. L., Horst, S. J., & Finney, S. J. (2009). Motivation matters: Using the Student Opinion Scale to make valid inferences about student performance. *J Gen Educ* 58(3), 129–151. <https://doi.org/10.1353/jge.0.0047>
- Thiebes, S., Lins, S., & Sunayev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, 31, 447–464. <https://doi.org/10.1007/s12525-020-0041-4>

- Tóth, T., Viragh, R., Hallova, M., Stuchly, P., & Hennyeyová, K. (2022). Digital Competence of Digital Native Students as Prerequisite for Digital Transformation of Education. *International Journal of Emerging Technologies in Learning (IJET)*, 17(16), 150–166. <https://doi.org/10.3991/ijet.v17i16.31791>
- Touretzky, D., Garner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What should every child know about AI? In *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1), 9795–9799.
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Hrsg.), *Organization of memory*. Academic Press.
- Turing, A. M. (1950). Mind. *Mind*, 59(236), 433–460.
- Urban, D., & Mayerl, J. (2011). *Regressionsanalyse: Theorie, Technik und Anwendung* (4. Aufl.). VS Verlag für Sozialwissenschaften.
- Vazhayil, A., Shetty R., Bhavani, R. R & Akshay, N. (2019). Focusing on Teacher Education to Introduce AI in Schools: Perspectives and Illustrative Findings. *Proceedings – IEEE Tenth International Conference on Technology for Education*, 71–77. <https://doi.org/10.1109/T4E.2019.00021>
- Venkatesh, V. (2000). Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. *Information Systems Research*, 11(4), 342–365.
- Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, 39(2), 273–315. <https://doi.org/10.1111/j.1540-5915.2008.00192.x>
- Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model; Four Longitudinal Field Studies. *Management Science* 46(2), S. 186–204.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3), 425–478.
- Vereinigung der Bayerischen Wirtschaft [VBW] (2018). *Digitale Souveränität und Bildung*. Gutachten. Waxmann. <https://doi.org/10.25656/01:16569>
- Vogt, K. (2022). Einführung in die pädagogische Diagnostik am Beispiel der Messung überfachlicher Kompetenzen. In K. Kramer & B. Hoyer (Hrsg.), „INDIVIDUELL FÖRDERN“ – wissenschaftlicher Hintergrund sowie Ansatzpunkte aus und für die Praxis. *FAU Lehren und Lernen Band 8* (S. 9–18). FAU University Press. DOI: 10.25593/978-3-96147-602-2
- Wagner, M. (2016). *Entwicklung und Überprüfung eines konsolidierten Akzeptanzmodells für Lernmanagementsysteme: Ein Vergleich zwischen Lehrkräften und Studierenden*. Dissertation. Ludwigs-Maximilians-Universität München.
- Wagner-Herrbach, C., Tafner, G., Hüttner, A., & Richter, P. (2022). Was Corona lehrt: Erfahrungen mit Präsenz-, Distanz- und Hybridunterricht und Schlussfolgerungen für die Weiterentwicklung der Studiengänge Wirtschaftspädagogik in Berlin. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach & M. Stock (Hrsg.), *Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation*. (S. 241–254). wbv Publikation.
- Wahlster, W. (2017). Künstliche Intelligenz als Treiber der zweiten Digitalisierungswelle. *IM+io Das Magazin für Innovation, Organisation und Management*, (2), 10–13.

- Walker, F., Link, N., von Wavere, L., Hedrich, M., Geißel, B., & Nickolaus, R. (2016). Berufsfachliche Kompetenzen von Elektronikern für Automatisierungstechnik – Kompetenzdimensionen, Messverfahren und erzielte Leistungen (KOKO EA). In K. Beck, M. Landenberger & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der Förderinitiative ASCOT* (S. 139–169). W. Bertelsmann Verlag.
- Wall, F. (2002). Das Instrumentarium zur Koordination als Abgrenzungsmerkmal des Controlling? In J. Weber & B. Hirsch (Hrsg.), *Controlling als akademische Disziplin: Eine Bestandsaufnahme* (S. 67–90). Deutscher Universitätsverlag.
- Walstad, W. B., Schmidt, S., Zlatkin-Troitschanskaia, O., & Happ, R. (2018). *Pretest-posttest measurement of the economic knowledge of undergraduates – Estimating guessing effects*. Discussion Paper for the AEA Annual Meeting 2018 – Philadelphia.
- Walstad, W. B., & Wagner, J. (2016). The disaggregation of value-added test scores to assess learning outcomes in economic courses. *The Journal of Economic Education*, 47(4), 121–131.
- Weber, S. (1994). *Vorwissen in der betriebswirtschaftlichen Ausbildung: Eine struktur- und inhaltsanalytische Studie*. Gabler.
- Weber, S., Draxler, C., Bley, S., Wiethe-Körprich, M., Weiß, C., & Gürer, C. (2016). Large scale assessments in der kaufmännischen Berufsbildung – Intrapreneurship (CoBaLIT). In K. Beck, M. Landenberger & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der Förderinitiative ASCOT* (S. 75–92). W. Bertelsmann Verlag.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Hrsg.), *Defining and Selecting Key Competencies* (S. 45–65). Hogrefe und Huber.
- Wessels, D. (2023). Was bedeutet ChatGPT für Schulen? *Interview in Bildung und Beruf*, 6, (April 2023), 126–128.
- White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, 66, 297–333.
- Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Dohi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., Thadaney-Israni, S., & Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. *Nature medicine*, 25(9), 1337–1340. <https://doi.org/10.1038/s41591-019-0548-6>
- Wilbers, K. (2012). Entwicklung der Kompetenzen von Lehrkräften berufsbildender Schulen für digitale Medien. *Berufsbildung in Wissenschaft und Praxis*, 41(3), 38–41.
- Wilbers, K. (2019). Kaufmännische Digitalkompetenzen als Ausgangspunkt der digitalen Transformation beruflicher Bildung. In K. Wilbers (Hrsg.), *Digitale Transformation kaufmännischer Bildung. Ausbildung in Industrie und Handel hinterfragt* (S. 11–27). epubli GmbH.
- Wilbers, K. (2021). Kaufmännische Aus- und Weiterbildung in der Industrie im Umbruch. Digitale Transformation im Zuge von Industrie 4.0 und künstlicher Intelligenz. *Zeitschrift für Berufs- und Wirtschaftspädagogik, Beiheft 31: Künstliche Intelligenz in der Berufsbildung*, 55–77.

- Wilson, J., Huang, Y., Palermo, C., Beard, G., & MacArthur, C. A. (2021). Automated Feedback and Automated Scoring in the Elementary Grades: Usage, Attitudes, and Associations with Writing Outcomes in a Districtwide Implementation of MI Write. *International Journal of Artificial Intelligence in Education*, 31(2), 234–276.
- Winkler, F., & Schwarz, H. (2021). IT-Berufe im Wandel. *Zeitschrift für Berufs- und Wirtschaftspädagogik, Beiheft 31: Künstliche Intelligenz in der Berufsbildung*, 77–102.
- Winther, E. (2010). *Kompetenzmessung in der beruflichen Bildung*. W. Bertelsmann Verlag.
- Winther, E. (2019). Enquetekommission I Digitale Transformation der Arbeitswelt. *Stellungnahme 17/1669 vom 08.07.2019 zum Thema „Kompetenzen und ihre Vermittlung in der digitalen Arbeitswelt“*. Landtag Nordrhein-Westfalen. <https://www.landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMST17-1669.pdf>
- Winther, E., & Achtenhagen, F. (2008). Kompetenzstrukturmodell für die kaufmännische Bildung: Adaptierbare Forschungslinien und theoretische Ausgestaltung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 104(4), 512–538.
- Winther, E., Seeber, S., Festner, D., Sangmeister, J., & Liedtke, M. (2016). Large scale assessments in der kaufmännischen Berufsbildung – Das Unternehmensassessment ALUSIM (CoBALIT). In K. Beck, M. Landenberger & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der BMBF-Förderinitiative ASCOT* (S. 55–73). W. Bertelsmann Verlag.
- Wirtschaftswissenschaftliche Fakultät der Universität Leipzig [WiFa UL] (2015). *Studienordnung für den Bachelorstudiengang Wirtschaftspädagogik (Business Education and Management Training) an der Universität Leipzig*. http://db.uni-leipzig.de/bekanntmachung/dokudownload.php?dok_id=4305
- Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stake assessments: problems and potential solutions. *Educ Assess*, 10, 1–17.
- Wittpahl, V. (2019). *Künstliche Intelligenz. Technologie, Anwendung, Gesellschaft*. Springer Verlag. <https://doi.org/10.1007/978-3-662-58042-4>
- Wölfel, M. (2021). Besonderheiten zum Einsatz von immersiven Augmented und Virtual Reality Lernanwendungen. *Zeitschrift für Berufs- und Wirtschaftspädagogik, Beiheft 31: Künstliche Intelligenz in der Berufsbildung*, 319–339.
- Wollowski, M., Selkowitz, R., Brown, L., Goel, A., Luger, G., Marshall, J., Neel, A., Neller, T., & Norvig, P. (2016). A Survey of Current Practice and Teaching of AI, *Proceedings of the Sixth Symposium on Educational Advances in Artificial Intelligence (EAAI-16)* , 4119–4124. www.aaai.org
- Wuttke, E. (2005). *Unterrichtskommunikation und Wissenserwerb*. Peter Lang.
- Wuttke, L. (2021). *Praxisleitfaden für Künstliche Intelligenz in Marketing und Vertrieb: Beispiele, Konzepte und Anwendungsfälle*. Springer.
- Wuttke, E., Seifried, J., & Niegemann, H. (2020). Editorial: Vocational Education and Training in the Age of Digitization: Challenges and Opportunities. In E. Wuttke, J. Seifried & H. Niegemann (Hrsg.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (S. 9–13). Barbara Budrich.

- Yu, S., & Zhang, J. (2021). Technical and Individual Factors Influencing Mobile Learning in China's Higher Education during the Outbreak of Covid-19. *Revista Romaneasca pentru Educatie Multidimensionala*, 13(1), 41–53. <https://doi.org/10.18662/rrem/13.1/358>
- Yueh, H.-P., Chiang, F.-K. (2020). Editorial: AI and robotics in reshaping the dynamics of learning. *British Journal of Education Technology*, 5(51). <https://doi.org/10.1111/bjet.13017>
- Zednik, C. (2021). Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence. *Philosophy & Technology*, 34, 265–288. <https://doi.org/10.1007/s13347-019-00382-7>
- Zhong, Z. R., Gao, S. Y., & Yu, X. Y. (2020). 学科教师人工智能基础及智能教育需求的现状和建议 [Der aktuelle Stand der KI-Grundlagen für Fachlehrer sowie der Bedarf und die Empfehlungen für intelligente Bildung]. *Journal of Beijing Institute of education*, 34(2), 28–34.
- Zhong, Z. R., Yu, X. Y., Qin, K., & Hu, S. J. (2020). 信息技术学科教师开展人工智能教育的现状与需求调研 [Forschung über die aktuelle Situation und den Bedarf von IT-Lehrern in der KI-Ausbildung]. *China Academic Journal Electronic Publishing House*, 18, 101–104.
- Ziegler, B., Frey, A., Seeber, S., Balkenhol, A., & Bernhardt, R. (2016). Adaptive Messung allgemeiner Kompetenzen (mAk-adapt). In K. Beck, M. Landenberger & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der Förderinitiative ASCOT* (S. 33–54). W. Bertelsmann Verlag.
- Zlatkin-Troitschanskaia, O., Förster, M., Brückner, S., Hansen, M., & Happ, R. (2013a). Modellierung und Erfassung der wirtschaftswissenschaftlichen Fachkompetenz bei Studierenden im deutschen Hochschulbereich. *Lehrerbildung auf dem Prüfstand, Sonderheft*, 108–133.
- Zlatkin-Troitschanskaia, O., Happ, R., Förster, M., Preuße, D., Schmidt, S., & Kuhn, C. (2013b). Analyse der Ausprägung und Entwicklung der Fachkompetenz von Studierenden der Wirtschaftswissenschaften und der Wirtschaftspädagogik. In O. Zlatkin-Troitschanskaia, R. Nickolaus & K. Beck (Hrsg.), *Kompetenzmodellierung und Kompetenzmessung bei Studierenden der Wirtschaftswissenschaften und der Ingenieurwissenschaften* (S. 69–92). *Lehrerbildung auf dem Prüfstand (Sonderheft)*. Verlag Empirische Pädagogik.
- Zlatkin-Troitschanskaia, O., & Kuhn, C. (2010). Messung akademisch vermittelter Fertigkeiten und Kenntnisse von Studierenden bzw. Hochschulabsolventen – Analyse zum Forschungsstand. *Arbeitspapiere Wirtschaftspädagogik, Johannes Gutenberg-Universität Mainz*. <http://www.wipaed.uni-mainz.de/lis/382.php>
- Zlatkin-Troitschanskaia, O., & Seidel, J. (2011). Kompetenz und ihre Erfassung – das neue „Theorie-Empirie-Problem“ der empirischen Bildungsforschung? In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen empirischer Bildungsforschung* (S. 218–233). Springer. https://doi.org/10.1007/978-3-531-94025-0_16

Anhang A: Fragebogen im Pretest (Version Universität Leipzig, 2022)

KiWiPäd



UNIVERSITÄT
LEIPZIG

PreTest Fragebogen zum Projekt KiWiPäd

Liebe Studierende,

im Rahmen des Moduls Bildungswissenschaften VI bitten wir Sie, an der folgenden Umfrage teilzunehmen. Bearbeiten Sie alle Elemente und lassen Sie keine Fragen aus, auch wenn Sie denken, nicht die richtige Antwort zu wissen.

Ihre Daten werden selbstverständlich anonym und vertraulich behandelt. Die Erstellung des Codes am Anfang des Fragebogens ermöglicht ein Matching von Antworten verschiedener Erhebungszeitpunkte - darüber hinaus werden keine Rückschlüsse auf Ihre Person gezogen. Die Teilnahme an der Umfrage ist freiwillig.

Der Fragebogen besteht insgesamt aus 6 Teilen. Nehmen Sie sich etwa 30 Minuten Zeit für die Bearbeitung.

Herzlichen Dank für Ihre Teilnahme!

In dieser Umfrage sind 115 Fragen enthalten.

Bitte erstellen Sie einen Code aus:

- 1. Und 2. Buchstaben des Vornamens Ihrer Mutter
- Tagesangabe Ihres eigenen Geburtsdatums – z. B. 05
- 3. Und 4. Buchstaben Ihres Geburtsortes
- 1. Buchstaben des Vornamens Ihres Vaters

Beispiel: >He14ipO< für „Helga“, „14“, „Leipzig“, „Olaf“

--

Teil 1: Persönliche und strukturelle Angaben

P1. Wie alt sind Sie?

P2. Mit welchem Geschlecht identifizieren Sie sich?

- ☐ weiblich
- ☐ männlich
- ☐ divers

S1. In welchem Studiengang sind Sie aktuell immatrikuliert?

- ☐ Master of Science Wirtschaftspädagogik
- ☐ Master of Education Wirtschaftspädagogik
- ☐ Sonstiges:

S1.1 Falls Sie Wirtschaftspädagogik studieren: Welches Zweitfach belegen Sie?

- ☐ Mathematik
- ☐ Informatik
- ☐ BWL
- ☐ Sonstiges:

S1.2 Nach meinem Abschluss sehe ich mich beruflich am ehesten...

- ☐ ...im Schuldienst.
- ☐ ...im Unternehmen.
- ☐ Sonstiges:

S2. Haben Sie vor Ihrem aktuellen Studium bereits einen anderen Studiengang besucht?

- ☐ ja, abgeschlossen
- ☐ ja, nicht abgeschlossen
- ☐ nein

Filter: falls S2 mit „ja, abgeschlossen“ beantwortet wurde.

S2.1 Welchen Abschluss haben Sie durch Ihr vorheriges Studium erworben?

☐ Bachelor

☐ Master

☐ Diplom

☐ Sonstiges:

S2.2 Welchem Fachbereich lässt sich ihr vorheriges Studium zuordnen?

☐ Ingenieurwissenschaften

☐ Informationstechnologie

☐ Naturwissenschaften

☐ Wirtschaftswissenschaften

☐ Sonstiges:

S3. Welche Gesamtnote haben Sie bei Ihrem Bachelorabschluss erreicht?

S4. In welchem Fachsemester studieren Sie in Ihrem aktuellen Studiengang?

S5. In welchem Hochschulsesemester (alle besuchten Studiengänge zusammen) studieren Sie aktuell?

S6. Wie viele Leistungspunkte (Credits) haben Sie in Ihrem aktuellen Studiengang bereits erreicht?

S7. Welche Durchschnittsnote haben Sie bei Ihrer Hochschulzugangsberechtigung (z. B. Abitur) erreicht?

S8. In welchem Jahr haben Sie Ihre Hochschulzugangsberechtigung erworben?

S9. Haben Sie in der Schule einen oder mehrere folgender Schwerpunkte bzw. Leistungskurse besucht? (Mehrfachnennung möglich)

☐ Mathematik

☐ Informatik

☐ Sonstiges:

S10. In welchem Bundesland haben Sie Ihre Hochschulzugangsberechtigung erworben? (als Dropdown)

☐ Baden-Württemberg

☐ Bayern

☐ Berlin

☐ Brandenburg

☐ Bremen

☐ Hamburg

☐ Hessen

☐ Mecklenburg-Vorpommern

☐ Niedersachsen

☐ Nordrhein-Westfalen

☐ Rheinland-Pfalz

☐ Saarland

☐ Schleswig-Holstein

☐ Sachsen-Anhalt

☐ Sachsen

☐ Thüringen

S11. Haben Sie eine Berufsausbildung absolviert?

☐ ja

☐ nein

S12. Wie sehr interessieren Sie sich für die benannten Inhaltsbereiche?

	Gar nicht	Wenig	Etwas	Viel	Sehr viel
Künstliche Intelligenz	☐	☐	☐	☐	☐
Big Data	☐	☐	☐	☐	☐
Machine Learning	☐	☐	☐	☐	☐

S13. Ich habe bereits im beruflichen Kontext Erfahrungen mit Anwendungen der Künstlichen Intelligenz gesammelt.

- ☐ ja, als Anwender:in
- ☐ ja, als Entwickler:in
- ☐ nein

S14. Ich habe bereits im privaten Kontext Erfahrungen mit Anwendungen der Künstlichen Intelligenz gesammelt (z. B. Sprachassistenten, Chatbots, Kaufvorschläge, Film-/Serienempfehlungen).

- ☐ ja, regelmäßig
- ☐ ja, manchmal
- ☐ nein

S15. Wie oft nutzen Sie die folgenden Anwendungen?

	Nie	Monatlich	Wöchentlich	Täglich	Mehrmals Täglich
Sprachassistenten (z. B. Siri, Alexa)	☐	☐	☐	☐	☐
Chatbots (z. B. im Kundenservice)	☐	☐	☐	☐	☐
Personalisierte Werbung	☐	☐	☐	☐	☐
Kaufvorschläge (Online-Shops)	☐	☐	☐	☐	☐
Film-/Serienvorschläge (z. B. Netflix, Amazon Prime Video)	☐	☐	☐	☐	☐

Teil 2: Fachspezifisches Wissen

W1. Ein Chatbot antwortet individuell und eindeutig auf die Nachrichten eines Users. Auch bei ähnlicher Texteingabe generiert der Chatbot unterschiedliche Antworten für unterschiedliche User. Welche *Fähigkeit* hat dieses System?

- ☐ Computer Audition
- ☐ Computerlinguistik
- ☐ Computer Vision
- ☐ Computer Consciousness

W1.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W2. Ein Wettermodell sagt voraus, dass heute in einer Woche eine Temperatur von 27 Grad Celsius erreicht wird. Welches Verfahren wurde für diese Prognose wahrscheinlich angewandt?

- ☐ Klassifikation
- ☐ Clustering
- ☐ Regression
- ☐ Strategieentwicklung

W2.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W3. Ein System soll lernen, ein bestimmtes Objekt (z. B. ein Gänseblümchen) auf beliebigen Bildern zu erkennen. Um der Maschine zu helfen, werden die Trainingsbilder mit einem sog. Label („Gänseblümchen“ oder „Kein Gänseblümchen“) versehen. Die Zielvariable ist also schon bekannt. Wie wird diese Art des maschinellen Lernens genannt?

- ☐ Reinforcement Learning
- ☐ Supervised Learning
- ☐ Unsupervised Learning
- ☐ Label Learning

W3.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W4. In welchem Jahr wurde der Begriff *Künstliche Intelligenz* zuerst verwendet?

- ☐ 1983
☐ 1956
☐ 1991
☐ 1965

W4.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W5. Wie hängen die Begriffe *Künstliche Intelligenz*, *Machine Learning* und *Deep Learning* zusammen?

- ☐ Die Begriffe können synonym verwendet werden.
☐ *Künstliche Intelligenz* ist ein Fachbereich, in dem *Machine Learning* ein Teilgebiet ist. *Deep Learning* ist in diesem Teilgebiet eine Methode.
☐ *Machine Learning* ist der Oberbegriff für *Künstliche Intelligenz* und *Deep Learning*.
☐ Gemeinsam ergeben sie das Konstrukt *Big Data*.

W5.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W6. Ein gutes System kann das Zielobjekt auf bisher unbekannten Bildern erfassen. Es erkennt also zum Beispiel einen Hund auf einem Foto, welches das System zuvor noch nicht gesehen hat. Wie wird diese Fähigkeit genannt?

- ☐ Generalisierung
☐ Abstraktion
☐ Analyse
☐ Diagnostik

W6.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W7. Was ist ein Anwendungsbeispiel für *Computer Audition*?

- ☐ Noise Cancelling
☐ Sprachassistenten
☐ Nachrichtenanalyse
☐ Textgenerierung

W7.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W8. Welcher Test gilt als großer Durchbruch für das Fachgebiet der Künstlichen Intelligenz?

- ☐ Alan Test
☐ Turing Test
☐ Miller Test
☐ Thinking Test

W8.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W9. Wie heißt der IBM Supercomputer, der im Jahr 1997 erstmalig gegen den damaligen Weltmeister eine Schachpartie gewann?

- ☐ HighGreen
☐ BrightRed
☐ DeepBlue
☐ DarkOrange

W9.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W10. Was wird unter einer „spezifischen KI“ (*narrow AI*) verstanden?

- ☐ Ein System, das eine konkrete Aufgabe lösen kann.
- ☐ Ein System, das alle Aufgaben einer Branche lösen kann.
- ☐ Ein System, das mit künstlich generierten Trainingsdaten gelernt hat.
- ☐ Ein System, das nur in bestimmten Domänen (Sprache, Naturwissenschaft usw.) eingesetzt werden kann.

W10.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W11. Ab wie vielen Schichten in einem neuronalen Netz wird von *Deep Learning* gesprochen?

- ☐ mehr als eine
- ☐ mehr als fünf
- ☐ mehr als sieben
- ☐ mehr als drei

W11.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W12. Was ist das zentrale Ziel von Maschinellem Lernen?

- ☐ Vorhersagen basierend auf Daten zu treffen
- ☐ Erklärungen für Ereignisse in der Vergangenheit zu liefern
- ☐ Ursache-Wirkungs-Zusammenhänge für Phänomene aufzudecken
- ☐ Beschreibungen für komplexe Ereignisse abzugeben

W12.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W13. Wie hängen Big Data und Maschinelles Lernen zusammen?

- ☐ Aus Big Data hat sich Maschinelles Lernen entwickelt.
- ☐ Big Data begünstigt die Entwicklung von Maschinellern Lernen.
- ☐ Ohne Maschinelles Lernen gäbe es kein Big Data.
- ☐ Zusammen ergeben sie Deep Learning.

W13.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W14. Der GPT-3 (*Generative Pretrained Transformer 3*) stellt einen Durchbruch in der Computerlinguistik dar. Welche Aussage zu diesem System ist **nicht** zutreffend?

- ☐ Das System basiert auf neuronalen Netzen.
- ☐ Das System sagt die Wahrscheinlichkeit der Existenz eines Satzes vorher.
- ☐ Es ist derzeit das größte Sprachmodell, das je geschaffen wurde.
- ☐ Die Trainingsdaten sind gelabelt.

W14.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W15. Was besagt das Moravec'sche Gesetz?

- ☐ Was für einen Menschen schwer ist, ist für einen Computer leicht, und umgekehrt.
- ☐ Der Mensch wird der Maschine immer überlegen sein.
- ☐ Menschen und Maschinen haben die gleichen Schwierigkeiten beim Lernen.
- ☐ Der Mensch wird die Kontrolle über die Maschine verlieren.

W15.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W16. Der Trainingsprozess einer Maschine kann vereinfacht in 5 Schritten abgebildet werden. Welche Reihenfolge ist korrekt?

- ☐ Zieldefinition, Analyse, Anpassung, Optimierung, Ausprobieren
- ☐ Ausprobieren, Anpassung, Zieldefinition, Analyse, Optimierung
- ☐ Zieldefinition, Ausprobieren, Analyse, Anpassung, Optimierung
- ☐ Analyse, Zieldefinition, Ausprobieren, Anpassung, Optimierung

W16.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W17. Was ist ein Messwert, um die Güte eines *Machine Learning*-Systems zu beschreiben?

- ☐ Regelpräzision
- ☐ Vorhersagegenauigkeit
- ☐ Abstraktionsebene
- ☐ Codestabilität

W17.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W18. Eine Bank interessiert sich für eine Anwendung der Künstlichen Intelligenz, die Unterschriften erkennen und ihre Echtheit bestätigen soll. Welche *Fähigkeiten* werden hierbei benötigt?

- ☐ Wahrnehmende und analytische Fähigkeiten
- ☐ Analytische und generierende Fähigkeiten
- ☐ Wahrnehmende und motorische Fähigkeiten
- ☐ Generierende und motorische Fähigkeiten

W18.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W19. Welches ist **kein** Teilgebiet der Computerlinguistik?

- ☐ Spracherklärung
- ☐ Spracherkennung
- ☐ Spracheinordnung
- ☐ Sprachverständnis

W19.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W20. Wie lässt sich ein Algorithmus allgemein beschreiben?

- ☐ Eine einfache Form der Datenanalyse
- ☐ Eine wiederholte Bewertung verschiedener Inputvariablen
- ☐ Eine eindeutig reproduzierbare Handlungsanweisung
- ☐ Eine Sammlung von weiteren Verhaltensoptionen

W20.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W21. Per Handynavigation können wir Staus bzw. erhöhtes Verkehrsaufkommen in Echtzeit vermeiden. Welche Fähigkeit hat das System?

- ☐ Planung und Optimierung
- ☐ Robotik und Steuerung
- ☐ Computer Vision
- ☐ Computer Audition

W21.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

Teil 3: Testmotivation

T0. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

[illegible]

E0. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

[illegible]

E15. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

[illegible]

E16. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

[illegible]

E17. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

[illegible]

E18. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

[illegible]

Teil 5: Epistemologische Überzeugungen

Ü0. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

[illegible]

Teil 6: Motivation

M0. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

[illegible]

Beruf ausüben zu
können.

Ich beschäftigte mich
mit Künstlicher

Intelligenz, um meine
Einstellungschancen zu
erhöhen.

☐ ☐ ☐ ☐ ☐ ☐ ☐

Menschen, die mein
Verhalten beeinflussen,
denken, dass ich
Anwendungen der
Künstlichen Intelligenz
nutzen sollte.

☐ ☐ ☐ ☐ ☐ ☐ ☐

Menschen, die mir
wichtig sind, denken,
dass ich Anwendungen
der Künstlichen
Intelligenz nutzen
sollte.

☐ ☐ ☐ ☐ ☐ ☐ ☐

Herzlichen Dank für Ihre Teilnahme!

Anhang B: Fragebogen im Posttest (Version Universität Leipzig, 2022)

KiWiPäd



UNIVERSITÄT
LEIPZIG

PostTest Fragebogen zum Projekt KiWiPäd

Liebe Studierende,

im Rahmen des Moduls Bildungswissenschaften VI bitten wir Sie, an der folgenden Umfrage teilzunehmen. Bearbeiten Sie alle Elemente und lassen Sie keine Fragen aus, auch wenn Sie denken, nicht die richtige Antwort zu wissen.

Ihre Daten werden selbstverständlich anonym und vertraulich behandelt. Die Erstellung des Codes am Anfang des Fragebogens ermöglicht ein Matching von Antworten verschiedener Erhebungszeitpunkte - darüber hinaus werden keine Rückschlüsse auf Ihre Person gezogen. Die Teilnahme an der Umfrage ist freiwillig.

Der Fragebogen besteht insgesamt aus 7 Teilen. Nehmen Sie sich etwa 30 Minuten Zeit für die Bearbeitung.

Herzlichen Dank für Ihre Teilnahme!

In dieser Umfrage sind 139 Fragen enthalten.

Bitte erstellen Sie einen Code aus:

- 1. Und 2. Buchstaben des Vornamens Ihrer Mutter
- Tagesangabe Ihres eigenen Geburtsdatums – z. B. 05
- 3. Und 4. Buchstaben Ihres Geburtsortes
- 1. Buchstaben des Vornamens Ihres Vaters

Beispiel: >He14ipO< für „**He**lga“, „**14**“, „**Leip**zig“, „**O**laf“

Teil 1: Persönliche und strukturelle Angaben

P1. Wie alt sind Sie?

P2. Mit welchem Geschlecht identifizieren Sie sich?

- ☐ weiblich
- ☐ männlich
- ☐ divers

S1. In welchem Studiengang sind Sie aktuell immatrikuliert?

- ☐ Master of Science Wirtschaftspädagogik
- ☐ Master of Education Wirtschaftspädagogik
- ☐ Sonstiges:

S1.1 Falls Sie Wirtschaftspädagogik studieren: Welches Zweitfach belegen Sie?

- ☐ Mathematik
- ☐ Informatik
- ☐ BWL
- ☐ Sonstiges:

S1.2 Nach meinem Abschluss sehe ich mich beruflich am ehesten...

- ☐ ...im Schuldienst.
- ☐ ...im Unternehmen.
- ☐ Sonstiges:

S4. In welchem Fachsemester studieren Sie in Ihrem aktuellen Studiengang?

S5. In welchem Hochschulsemester (alle besuchten Studiengänge zusammen) studieren Sie aktuell?

S6. Wie viele Leistungspunkte (Credits) haben Sie in Ihrem aktuellen Studiengang bereits erreicht?

S12. Wie sehr interessieren Sie sich für die benannten Inhaltsbereiche?

	Gar nicht	Wenig	Etwas	Viel	Sehr viel
Künstliche Intelligenz	☐	☐	☐	☐	☐
Big Data	☐	☐	☐	☐	☐
Machine Learning	☐	☐	☐	☐	☐

S13. Ich habe bereits im beruflichen Kontext Erfahrungen mit Anwendungen der Künstlichen Intelligenz gesammelt.

- ☐ ja, als Anwender:in
- ☐ ja, als Entwickler:in
- ☐ nein

S14. Ich habe bereits im privaten Kontext Erfahrungen mit Anwendungen der Künstlichen Intelligenz gesammelt (z. B. Sprachassistenten, Chatbots, Kaufvorschläge, Film-/Serienempfehlungen).

- ☐ ja, regelmäßig
- ☐ ja, manchmal
- ☐ nein

S15. Wie oft nutzen Sie die folgenden Anwendungen?

	Nie	Monatlich	Wöchentlich	Täglich	Mehrmals Täglich
Sprachassistenten (z. B. Siri, Alexa)	☐	☐	☐	☐	☐
Chatbots (z. B. im Kundenservice)	☐	☐	☐	☐	☐
Personalisierte Werbung	☐	☐	☐	☐	☐
Kaufvorschläge (Online-Shops)	☐	☐	☐	☐	☐
Film-/Serienvorschläge (z. B. Netflix, Amazon Prime Video)	☐	☐	☐	☐	☐

Teil 2: Fachspezifisches Wissen

W1. Ein Chatbot antwortet individuell und eindeutig auf die Nachrichten eines Users. Auch bei ähnlicher Texteingabe generiert der Chatbot unterschiedliche Antworten für unterschiedliche User. Welche *Fähigkeit* hat dieses System?

- ☐ Computer Audition
- ☐ Computerlinguistik
- ☐ Computer Vision
- ☐ Computer Consciousness

W1.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W2. Ein Wettermodell sagt voraus, dass heute in einer Woche eine Temperatur von 27 Grad Celsius erreicht wird. Welches Verfahren wurde für diese Prognose wahrscheinlich angewandt?

- ☐ Klassifikation
- ☐ Clustering
- ☐ Regression
- ☐ Strategieentwicklung

W2.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W3. Ein System soll lernen, ein bestimmtes Objekt (z. B. ein Gänseblümchen) auf beliebigen Bildern zu erkennen. Um der Maschine zu helfen, werden die Trainingsbilder mit einem sog. Label („Gänseblümchen“ oder „Kein Gänseblümchen“) versehen. Die Zielvariable ist also schon bekannt. Wie wird diese Art des maschinellen Lernens genannt?

- ☐ Reinforcement Learning
- ☐ Supervised Learning
- ☐ Unsupervised Learning
- ☐ Label Learning

W3.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W4. In welchem Jahr wurde der Begriff *Künstliche Intelligenz* zuerst verwendet?

- ☐ 1983
☐ 1956
☐ 1991
☐ 1965

W4.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W5. Wie hängen die Begriffe *Künstliche Intelligenz*, *Machine Learning* und *Deep Learning* zusammen?

- ☐ Die Begriffe können synonym verwendet werden.
☐ Künstliche Intelligenz ist ein Fachbereich, in dem Machine Learning ein Teilgebiet ist. Deep Learning ist in diesem Teilgebiet eine Methode.
☐ Machine Learning ist der Oberbegriff für Künstliche Intelligenz und Deep Learning.
☐ Gemeinsam ergeben sie das Konstrukt „Big Data“.

W5.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W6. Ein gutes System kann das Zielobjekt auf bisher unbekannten Bildern erfassen. Es erkennt also zum Beispiel einen Hund auf einem Foto, welches das System zuvor noch nicht gesehen hat. Wie wird diese Fähigkeit genannt?

- ☐ Generalisierung
☐ Abstraktion
☐ Analyse
☐ Diagnostik

W6.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W7. Was ist ein Anwendungsbeispiel für *Computer Audition*?

- ☐ Noise Cancelling
☐ Sprachassistenten
☐ Nachrichtenanalyse
☐ Textgenerierung

W7.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W8. Welcher Test gilt als großer Durchbruch für das Fachgebiet der Künstlichen Intelligenz?

- ☐ Alan Test
☐ Turing Test
☐ Miller Test
☐ Thinking Test

W8.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W9. Wie heißt der IBM Supercomputer, der im Jahr 1997 erstmalig gegen den damaligen Weltmeister eine Schachpartie gewann?

- ☐ HighGreen
☐ BrightRed
☐ DeepBlue
☐ DarkOrange

W9.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W10. Was wird unter einer „spezifischen KI“ (*narrow AI*) verstanden?

- ☐ Ein System, das eine konkrete Aufgabe lösen kann.
- ☐ Ein System, das alle Aufgaben einer Branche lösen kann.
- ☐ Ein System, das mit künstlich generierten Trainingsdaten gelernt hat.
- ☐ Ein System, das nur in bestimmten Domänen (Sprache, Naturwissenschaft usw.) eingesetzt werden kann.

W10.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W11. Ab wie vielen Schichten in einem neuronalen Netz wird von *Deep Learning* gesprochen?

- ☐ mehr als eine
- ☐ mehr als fünf
- ☐ mehr als sieben
- ☐ mehr als drei

W11.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W12. Was ist das zentrale Ziel von Maschinellem Lernen?

- ☐ Vorhersagen basierend auf Daten zu treffen
- ☐ Erklärungen für Ereignisse in der Vergangenheit zu liefern
- ☐ Ursache-Wirkungs-Zusammenhänge für Phänomene aufzudecken
- ☐ Beschreibungen für komplexe Ereignisse abzugeben

W12.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W13. Wie hängen Big Data und Maschinelles Lernen zusammen?

- ☐ Aus Big Data hat sich Maschinelles Lernen entwickelt.
- ☐ Big Data begünstigt die Entwicklung von Maschinellern Lernen.
- ☐ Ohne Maschinelles Lernen gäbe es kein Big Data.
- ☐ Zusammen ergeben sie Deep Learning.

W13.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W14. Der GPT-3 (*Generative Pretrained Transformer 3*) stellt einen Durchbruch in der Computerlinguistik dar. Welche Aussage zu diesem System ist **nicht** zutreffend?

- ☐ Das System basiert auf neuronalen Netzen.
- ☐ Das System sagt die Wahrscheinlichkeit der Existenz eines Satzes vorher.
- ☐ Es ist derzeit das größte Sprachmodell, das je geschaffen wurde.
- ☐ Die Trainingsdaten sind gelabelt.

W14.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W15. Was besagt das Moravec'sche Gesetz?

- ☐ Was für einen Menschen schwer ist, ist für einen Computer leicht, und umgekehrt.
- ☐ Der Mensch wird der Maschine immer überlegen sein.
- ☐ Menschen und Maschinen haben die gleichen Schwierigkeiten beim Lernen.
- ☐ Der Mensch wird die Kontrolle über die Maschine verlieren.

W15.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W16. Der Trainingsprozess einer Maschine kann vereinfacht in 5 Schritten abgebildet werden. Welche Reihenfolge ist korrekt?

- ☐ Zieldefinition, Analyse, Anpassung, Optimierung, Ausprobieren
- ☐ Ausprobieren, Anpassung, Zieldefinition, Analyse, Optimierung
- ☐ Zieldefinition, Ausprobieren, Analyse, Anpassung, Optimierung
- ☐ Analyse, Zieldefinition, Ausprobieren, Anpassung, Optimierung

W16.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W17. Was ist ein Messwert, um die Güte eines *Machine Learning*-Systems zu beschreiben?

- ☐ Regelpräzision
- ☐ Vorhersagegenauigkeit
- ☐ Abstraktionsebene
- ☐ Codestabilität

W17.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W18. Eine Bank interessiert sich für eine Anwendung der Künstlichen Intelligenz, die Unterschriften erkennen und ihre Echtheit bestätigen soll. Welche *Fähigkeiten* werden hierbei benötigt?

- ☐ Wahrnehmende und analytische Fähigkeiten
- ☐ Analytische und generierende Fähigkeiten
- ☐ Wahrnehmende und motorische Fähigkeiten
- ☐ Generierende und motorische Fähigkeiten

W18.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W19. Welches ist **kein** Teilgebiet der Computerlinguistik?

- ☐ Spracherklärung
- ☐ Spracherkennung
- ☐ Spracheinordnung
- ☐ Sprachverständnis

W19.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W20. Wie lässt sich ein Algorithmus allgemein beschreiben?

- ☐ Eine einfache Form der Datenanalyse
- ☐ Eine wiederholte Bewertung verschiedener Inputvariablen
- ☐ Eine eindeutig reproduzierbare Handlungsanweisung
- ☐ Eine Sammlung von weiteren Verhaltensoptionen

W20.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

W21. Per Handynavigation können wir Staus bzw. erhöhtes Verkehrsaufkommen in Echtzeit vermeiden. Welche Fähigkeit hat das System?

- ☐ Planung und Optimierung
- ☐ Robotik und Steuerung
- ☐ Computer Vision
- ☐ Computer Audition

W21.1 Zu wie viel Prozent sind Sie davon überzeugt, dass die von Ihnen gegebene Antwort richtig ist?

- ☐ 0 % ☐ 25 % ☐ 75 % ☐ 100 %

Teil 3: Testmotivation

T0. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

[illegible]

Teil 4: Einstellungen

E0. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

[illegible]

[illegible][illegible][illegible][illegible]

Teil 5: Epistemologische Überzeugungen

Ü0. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

[illegible]

Teil 6: Motivation

M0. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

[illegible]

Beruf ausüben zu können.

Ich beschäftigte mich mit Künstlicher Intelligenz, um meine Einstellungschancen zu erhöhen.

☐ ☐ ☐ ☐ ☐ ☐

Menschen, die mein Verhalten beeinflussen, denken, dass ich Anwendungen der Künstlichen Intelligenz nutzen sollte.

☐ ☐ ☐ ☐ ☐ ☐

Menschen, die mir wichtig sind, denken, dass ich Anwendungen der Künstlichen Intelligenz nutzen sollte.

☐ ☐ ☐ ☐ ☐ ☐

Teil 7: Qualität der Lernumgebung

K0. War Ihnen der KI-Campus bereits vor dem Besuch des Moduls bekannt?

- ☐ Ja, ich habe bereits vorher Inhalte des KI-Campus bearbeitet.
- ☐ Ja, ich kannte den KI-Campus schon vorher, habe aber keine Inhalte bearbeitet.
- ☐ Nein, der KI-Campus war mir vor dem Semester nicht bekannt.

K1. Haben Sie die von den Dozierenden vorgegebenen Einheiten auf dem KI-Campus bearbeitet?

- ☐ Ich habe alle vier Einheiten bearbeitet.
- ☐ Ich habe drei Einheiten bearbeitet.
- ☐ Ich habe zwei Einheiten bearbeitet.
- ☐ Ich habe eine Einheit bearbeitet.
- ☐ Ich habe keine der Einheiten bearbeitet.

Trichter, wenn K1 nicht „keine“:

K01.1. Haben Sie sich an die Reihenfolge der einzelnen Inhalte im KI-Campus gehalten?

- ☐ Ja, ich habe die Inhalte in der vorgegebenen Reihenfolge bearbeitet.
- ☐ Nein, ich habe die Reihenfolge verändert.

K01.2. Haben Sie einzelne Inhalte in den Sektionen ausgelassen?

- ☐ Ja, ich habe Inhalte ausgelassen/übersprungen.
- ☐ Nein, ich habe die Sektionen vollständig bearbeitet/bearbeitet.

(Trichter, wenn K01.2 „ja“):

K01.2.1 Welche Inhalte haben Sie übersprungen/ausgelassen (welche Art von Inhalten oder konkrete Sektionen)?

K01.2.2 Aus welchem Grund haben Sie Inhalte übersprungen/ausgelassen?

K01.3. Wie haben Sie die Einheiten des KI-Campus zeitlich überwiegend bearbeitet?

- ☐ Zu einem festen Zeitpunkt (fester Wochentag/feste Uhrzeit)
- ☐ Zu wechselnden Zeitpunkten (zeit- und ortsunabhängig)

K01.4. Wie haben Sie die Einheiten des KI-Campus organisatorisch überwiegend bearbeitet?

- ☐ Einzelne Einheiten ohne Unterbrechungen
- ☐ Einzelne Einheiten mit Unterbrechungen

K2. Wie oft haben Sie im Seminar (synchrone Veranstaltungen über Videokonferenz/in Präsenz) gefehlt?

- ☐ Nie gefehlt
- ☐ 1-2 Mal gefehlt
- ☐ 3-4 Mal gefehlt
- ☐ 5-6 Mal gefehlt

K3. Wie sehr interessieren Sie sich für die in den KI-Campus Einheiten adressierten Themengebiete?

- ☐ Gar nicht
- ☐ Wenig
- ☐ Etwas
- ☐ Viel
- ☐ Sehr viel

K4. Wie oft haben Sie auf den KI-Campus zugegriffen?

- ☐ Mehrmals täglich
- ☐ Täglich
- ☐ Mehrmals pro Woche
- ☐ Einmal pro Woche
- ☐ Seltener
- ☐ Nie

K5. Wie viel Zeit haben Sie insgesamt in den vier Selbstlernphasen auf dem KI-Campus verbracht (Schätzung in Minuten)?

K6. Haben Sie zusätzliche Angebote des KI-Campus (abgesehen von den im Seminar vorgegebenen Einheiten) bearbeitet?

- ☐ Ja
- ☐ Nein

Trichter, wenn K6. „ja“

K061 Welche Angebote des KI-Campus haben Sie bearbeitet?

K7. Haben Sie zusätzliche Online-Lernangebote (nicht KI-Campus) bearbeitet, um mehr über die Thematik zu lernen?

- ☐ Ja
- ☐ Nein

Trichter, wenn K7. „ja“

K071 Welche Online-Angebote haben Sie bearbeitet?

K8. Haben Sie zusätzliche Literatur recherchiert, um die Inhalte zu Anwendungen der KI zu vertiefen?

- ☐ Ja
☐ Nein

K9. Haben Sie mit Ihren Kommiliton:innen außerhalb der Seminarzeiten über die fachlichen Inhalte des KI-Campus (KI Inhalte) diskutiert?

- ☐ Ja
☐ Nein

K10. Haben Sie mit Ihren Kommiliton:innen außerhalb der Seminarzeiten über die fachlichen Inhalte des Kurses (Evaluation von Online-Lernumgebungen) diskutiert?

- ☐ Ja
☐ Nein

K11. Wie viel Zeit (in Minuten) haben Sie für die Vor- und Nachbereitung der Seminare aufgebracht (ohne Nutzungszeit des KI-Campus)?

- ☐ Ich habe keine Zeit in die Vor- und Nachbereitung der Seminare investiert und lediglich die KI-Campus Inhalte bearbeitet.
- ☐ Ich habe zusätzlich zur Nutzungszeit auf dem KI-Campus Zeit in die Vor- und Nachbereitung der Seminare investiert. Meine durchschnittliche Vor- und Nachbereitungszeit beträgt in Minuten:

K12. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

[illegible]

[illegible]

K14. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

[illegible]

wurde in der
Veranstaltung sinnvoll
eingesetzt.

Die Nutzung des
Online-Whiteboards
(Miro/Conceptboard)
hat mir geholfen, die
Inhalte zu
strukturieren.

☐ ☐ ☐ ☐ ☐ ☐

Durch das Online-
Whiteboard
(Miro/Conceptboard)
konnte ich den
Fortschritt während des
Semesters gut
nachvollziehen.

☐ ☐ ☐ ☐ ☐ ☐

Ich möchte auch in
Zukunft mit Online-
Whiteboards
(Miro/Conceptboard)
in Lehrveranstaltungen
arbeiten.

☐ ☐ ☐ ☐ ☐ ☐

Der Einsatz der
Technik erfolgte in
diesem Kurs ohne
Probleme.

☐ ☐ ☐ ☐ ☐ ☐

Das Verhältnis
zwischen synchronen
Veranstaltungen (über
zoom) und
asynchronen
Selbstlernphasen war
angemessen.

☐ ☐ ☐ ☐ ☐ ☐

Die Auswahl der
Inhalte auf dem KI-
Campus habe ich als
passend für das
besuchte Modul
empfunden.

☐ ☐ ☐ ☐ ☐ ☐

Der Umfang der
ausgewählten KI-
Campus Inhalte je
Selbstlernphase war
angemessen.

☐ ☐ ☐ ☐ ☐ ☐

Die bereitgestellten
Materialien
(Evaluationsbogen und
Miro/Conceptboard)

☐ ☐ ☐ ☐ ☐ ☐

haben mich bei der
Bearbeitung der
Aufgaben unterstützt.

Die Aufgaben der Selbstlernphasen (Evaluation einer KI-Campus Einheit anhand einer Kategorie aus dem Evaluationsbogen) waren gut zu bewältigen.

□ □ □ □ □ □

Ich habe mich während der asynchronen Selbstlernphasen allein gelassen gefühlt.

--	--	--	--	--	--

K15. Bitte nehmen Sie zu den folgenden Aussagen Stellung. Antworten Sie spontan!

Beim Lernen/Arbeiten in diesem Kurs...

Ich lehne die Aussage nachdrücklich ab.	Ich lehne die Aussage ab.	Ich lehne die Aussage eher ab.	Ich stimme der Aussage eher zu.	Ich stimme der Aussage zu.	Ich stimme der Aussage nachdrücklich zu.
---	---------------------------	--------------------------------	---------------------------------	----------------------------	--

[illegible]

möglichst richtig/gut zu
machen.

...war mir klar, dass ich das für meinen Beruf können muss.	☐	☐	☐	☐	☐	☐
...wollte ich selbst den Stoff verstehen/beherrschen.	☐	☐	☐	☐	☐	☐
...habe ich mich eingesetzt, weil ich meinen eigenen Zielen ein Stück näherkommen konnte.	☐	☐	☐	☐	☐	☐
...machte das Lernen/Arbeiten richtig Spaß.	☐	☐	☐	☐	☐	☐
...war ich wissbegierig/neugierig.	☐	☐	☐	☐	☐	☐
...verging die Zeit (in den synchronen Veranstaltungen) wie im Flug.	☐	☐	☐	☐	☐	☐
...hat mich der Inhalt so fasziniert, dass ich mich voll eingesetzt habe.	☐	☐	☐	☐	☐	☐
...befasste ich mich mit anregenden Problemen, über die ich mehr erfahren will.	☐	☐	☐	☐	☐	☐
...stieß ich auf anregende Themen, über die ich mit anderen sprechen will.	☐	☐	☐	☐	☐	☐

Herzlichen Dank für Ihre Teilnahme!

Anhang C: Verteilung der Summenscores (Wissenstest) der Subgruppen zu beiden Messzeitpunkten

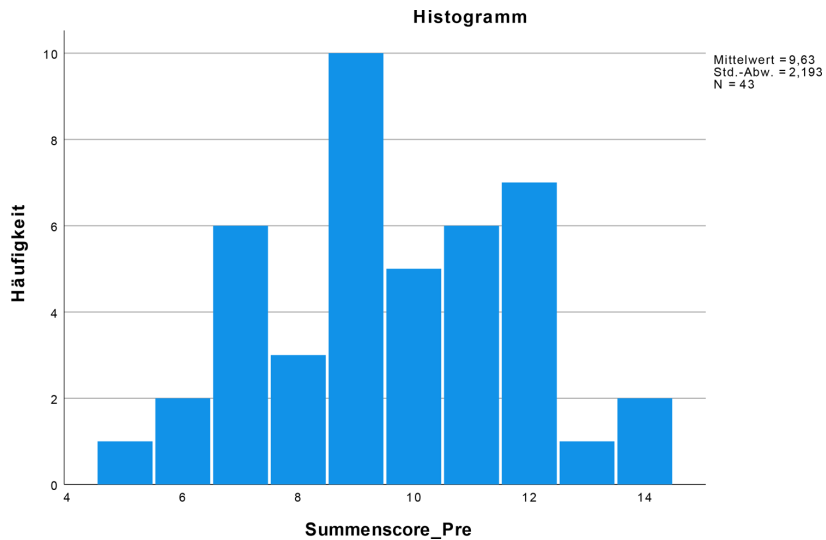


Abbildung C1: Verteilung der Summenscores Subgruppe M. Sc. WiPäd t1

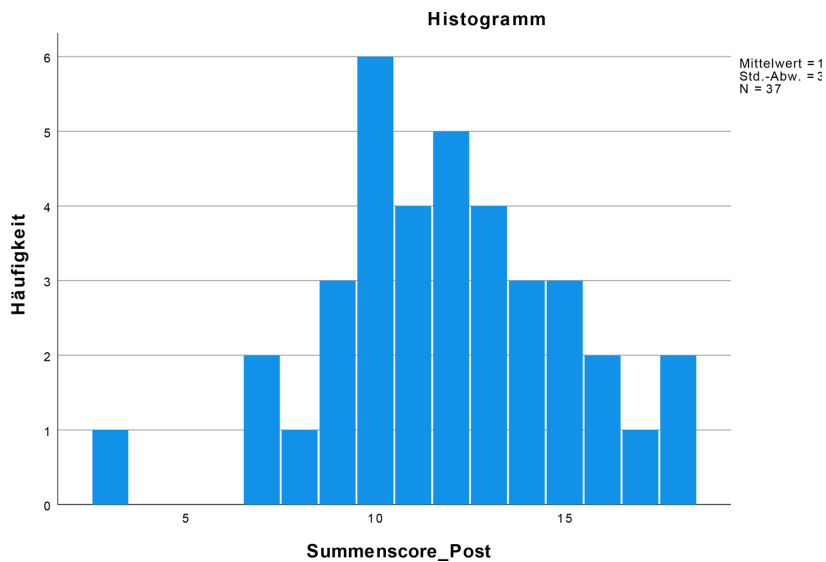


Abbildung C2: Verteilung der Summenscores Subgruppe M. Sc. WiPäd t2

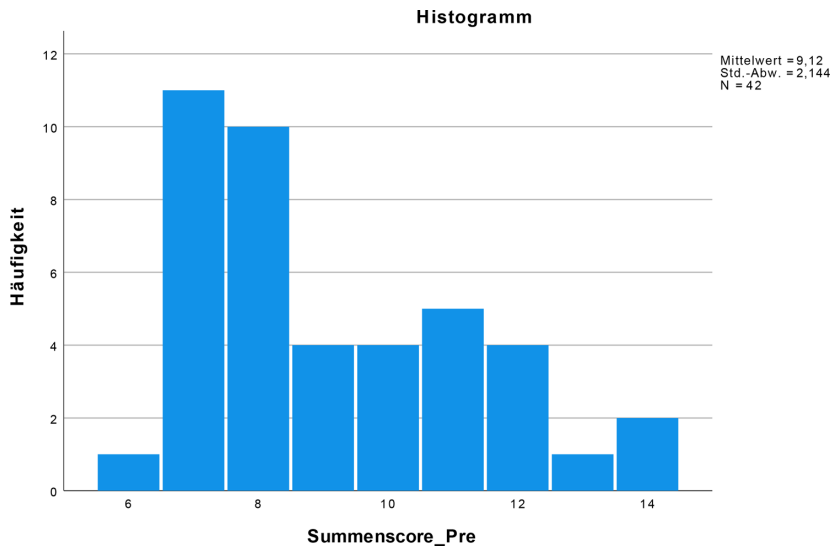


Abbildung C3: Verteilung der Summenscores Subgruppe M. Ed. WiPäd t1

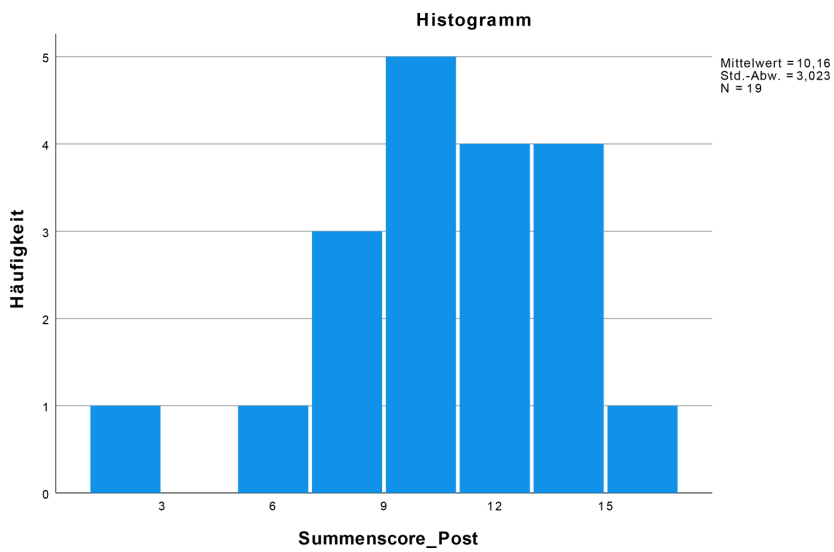


Abbildung C4: Verteilung der Summenscores Subgruppe M. Ed. WiPäd t2

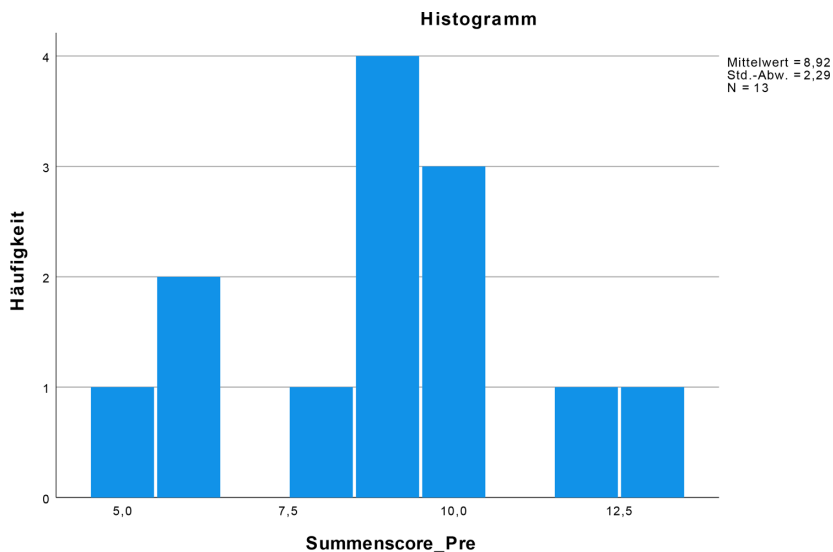


Abbildung C5: Verteilung der Summenscores Subgruppe M. A/M. Sc. Berufspädagogik t1

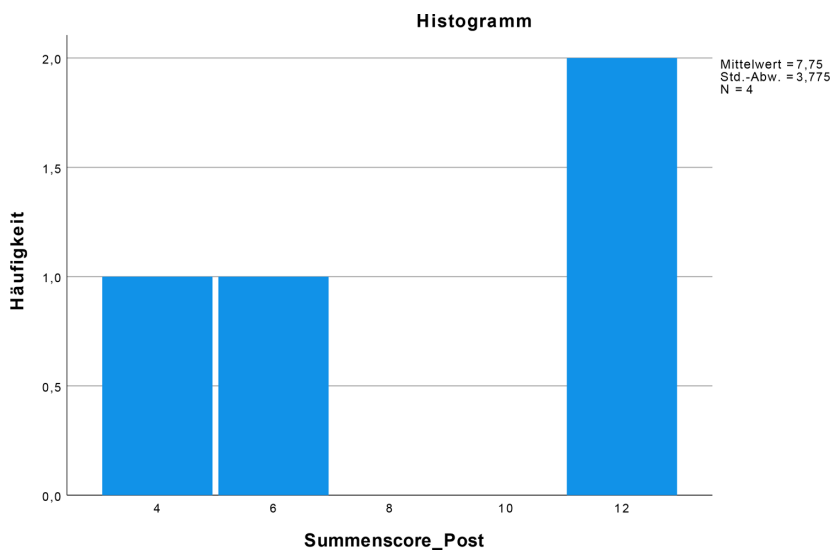


Abbildung C6: Verteilung der Summenscores Subgruppe M. A/M. Sc. Berufspädagogik t2

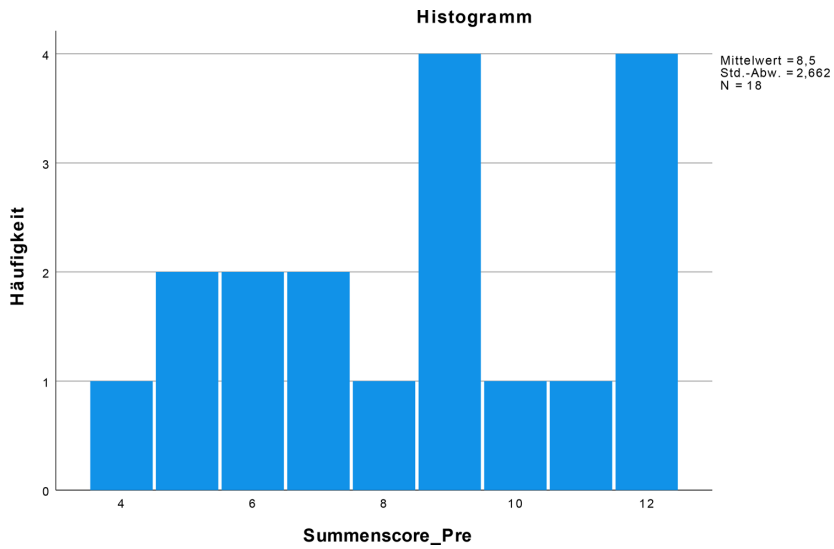


Abbildung C7: Verteilung der Summenscores Subgruppe Lehrdiplom Berufsmaturität t1

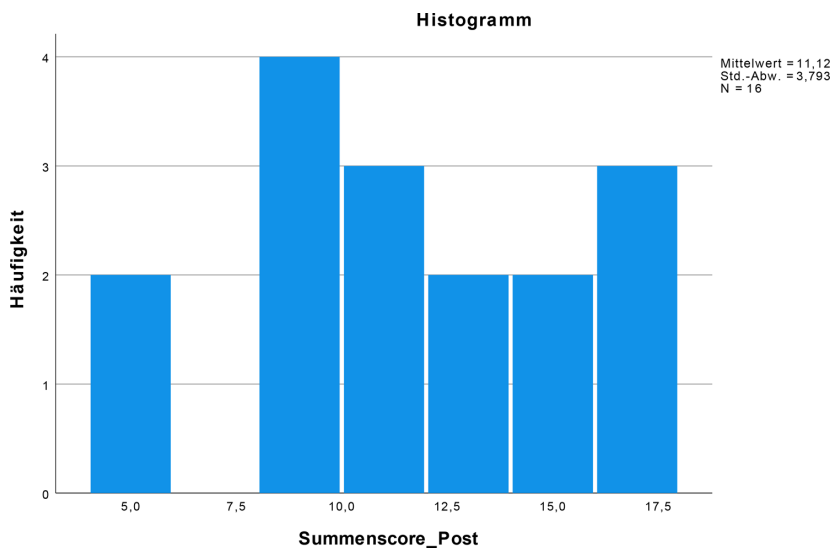


Abbildung C8: Verteilung der Summenscores Subgruppe Lehrdiplom Berufsmaturität t2

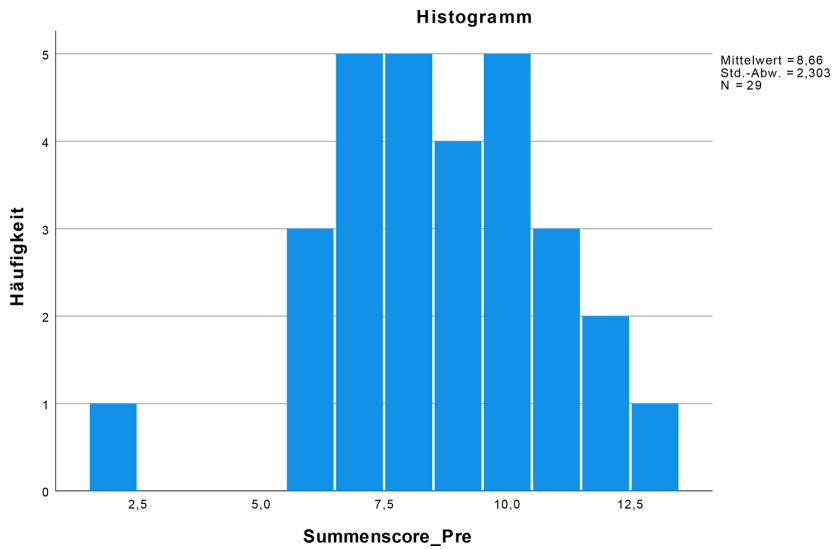


Abbildung C9: Verteilung der Summenscores Kontrastgruppe B. Sc. WiPäd t1

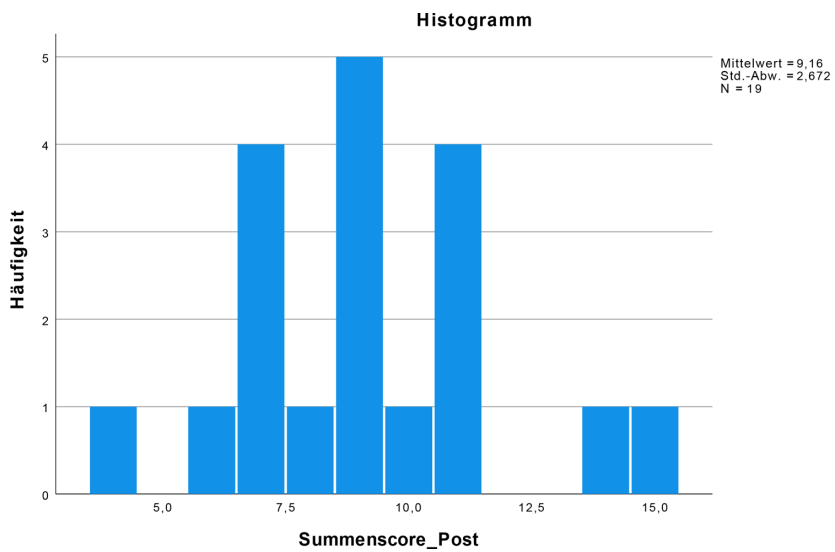


Abbildung C10: Verteilung der Summenscores Kontrastgruppe B. Sc. WiPäd t2

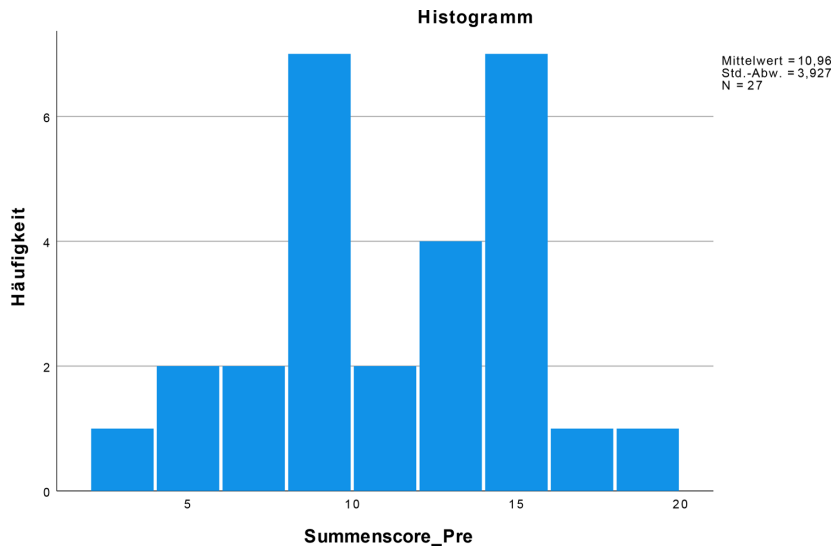


Abbildung C11: Verteilung der Summenscores Kontrastgruppe Wirtschaftsinformatik t1



Berufsbildung, Arbeit und Innovation, 73
2024, 264 S., 49,90 € (D)
ISBN 978-3-7639-7394-1
E-Book im Open Access

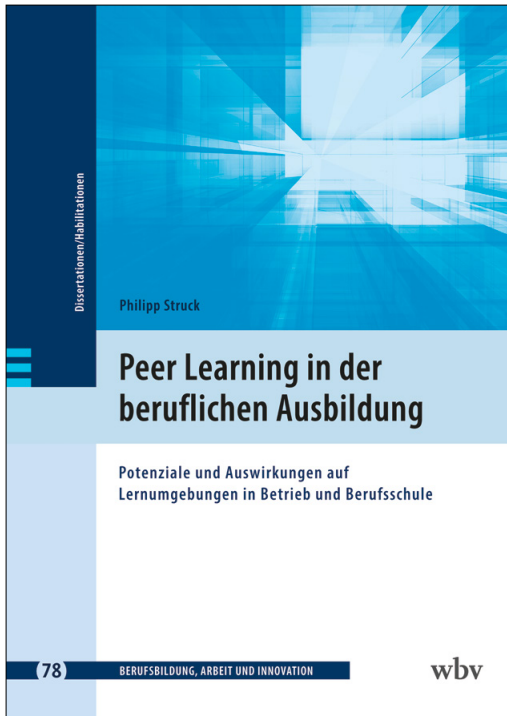
Claudia Müller, Jan Pranger, Jens Reißland (Hg.)

Nachhaltigkeitsorientierte Weiterbildungsdidaktik

Die doppelte Multiplikatorenqualifizierung

Der Sammelband basiert auf den Erkenntnissen und Ergebnissen des Projekts „TraNaxis - Transfer von Nachhaltigkeit in die berufliche Aus- und Weiterbildungspraxis“, welches im Rahmen des Förderschwerpunkts „BBNE Transfer 2020-2022“ aus Mitteln des BMBF durch das BIBB gefördert wurde. Ein zentrales Projektergebnis ist die doppelte Multiplikatorenqualifizierung als Ansatz einer nachhaltigkeitsorientierten Weiterbildungsdidaktik. In Teil I des Bandes erfolgt eine multiperspektivische theoretische Rahmung der Projektergebnisse im Kontext der Berufsbildungsforschung. Die Teile II, III und IV stellen die Entwicklung, Erprobung sowie die erfolgreiche Transferphase des Ansatzes dar. Teil II fokussiert das betriebliche Ausbildungs- und Weiterbildungspersonal. Eine theoretische Grundlegung für den Ansatz einer nachhaltigkeitsorientierten Weiterbildungsdidaktik erfolgt in Teil III. Der entwickelte Ansatz der doppelten Multiplikatorenqualifizierung wird in Teil IV dargestellt und reflektiert.

wbv.de/bai



Berufsbildung, Arbeit und Innovation –
Dissertationen und Habilitationen, 78
2023, 284 S., 49,90 € (D)
ISBN 978-3-7639-7617-1
E-Book im Open Access

Philipp Struck

Peer Learning in der beruflichen Ausbildung

Potenziale und Auswirkungen auf Lernumgebungen
in Betrieb und Berufsschule

Die Gestaltung von Lernsettings in der beruflichen
Ausbildung ist eine zentrale Aufgabe für schuli-
sches und betriebliches Bildungspersonal.

Die Arbeit möchte einen Forschungsbeitrag für
die berufliche Bildung leisten und den Nutzen des
Peer Learning in der beruflichen Ausbildung mit
vier forschungsmethodischen Zugängen ana-
lysieren. Des Weiteren sollen für die Gestaltung
pädagogischer Lernsituationen konkrete Hinweise
und Anregungen zur Implementierung des Peer
Learning in der beruflichen Ausbildung formuliert
werden.

Die Handlungsempfehlungen zur Gestaltung von
Lernsettings richten sich an betriebliches und
schulisches Bildungspersonal. Es wird gezeigt, dass
Peer Learning zu einer Förderung der Sozialkom-
petenz und der Fachkompetenz beitragen kann.
Zudem weisen die Ergebnisse auf eine steigende
Bereitschaft der Auszubildenden hin, mit anderen
Auszubildenden gemeinsam lernen zu wollen.

wbv.de/bai

In der Dissertation von Frau Dr.in Schmidt wird ausgehend von der zunehmenden Relevanz von Künstlicher Intelligenz (KI) im Rahmen digitaler Transformationsprozesse ein Strukturmodell für KI-bezogene Kompetenzfacetten (angehender) Lehrkräfte im berufsbildenden Bereich entwickelt. Das Wissen zu KI nimmt dabei in Anlehnung an die Professionalisierungsforschung eine zentrale Rolle ein. Im Rahmen der Arbeit wird der Frage nachgegangen, wie das Grundlagenwissen (angehender) Lehrkräfte theoretisch modelliert und empirisch erfasst werden kann. Das entwickelte Testinstrument wurde anhand eines quantitativen Studiendesigns umfassend validiert.

Die Reihe **Berufsbildung, Arbeit und Innovation** bietet ein Forum für die grundlagen- und anwendungsorientierte Berufsbildungsforschung. Sie leistet einen Beitrag für den wissenschaftlichen Diskurs über Innovationspotenziale der beruflichen Bildung.

Die Reihe wird herausgegeben von Prof.in Marianne Frieze (Justus-Liebig-Universität Gießen), Prof. Klaus Jenewein (Otto-von-Guericke-Universität Magdeburg), Prof.in Susan Seeber (Georg-August-Universität Göttingen) und Prof. Lars Windelband (Karlsruher Institut für Technologie).

Jacqueline Marie-Charlotte Schmidt war von 2018 bis 2023 wissenschaftliche Mitarbeiterin am Institut für Wirtschaftspädagogik der Universität Leipzig. Ihre Promotion schloss sie im September 2023 mit dem Gesamtpredikat summa cum laude an der Wirtschaftswissenschaftlichen Fakultät ab. Seit Januar 2024 ist sie Juniorprofessorin für Wirtschaftspädagogik, insbesondere: Digitalisierung in Bildungs- und Arbeitswelten an der Technischen Universität Dresden.