

ZEITSCHRIFT
FÜR
EMPIRISCHE
HOCHSCHULFORSCHUNG



JG 7
HEFT 1/2023

THEMENHEFT: INNOVATIVE BILDUNGSTECHNOLOGIEN UND TECHNOLOGIEGESTÜTZTES LERNEN ALS GEGENSTAND MODERNER HOCHSCHULFORSCHUNG

MAXIMILIAN SAILER, JOHANNES ABEL, JUDITH SCHWEPPE *Bildungstechnologische Fragen in der Hochschullehre im Überblick*

FLORIAN TRAUTEN, CAROLIN EITEMÜLLER, MAIK WALPUSKI *Lernaufgaben mit automatisiertem Feedback in einem digitalen Hochschultutorium zur Vorlesung der Allgemeinen Chemie als Angebot zur Binnendifferenzierung*

CAROLINE GÖTZ, PAUL ERHARDT, SIMONE N. LÖFFLER
Selbstwirksamkeit in selbstregulierten Lernprozessen an Hochschulen mithilfe von Podcasts fördern – eine Interventionsstudie mit Interaktivem Ambulantes Assessment

LEA GROTEGUT, KATRIN B. KLINGSIECK
Kompetenzentwicklung in der Lehrkräftebildung: der Mehrwert einer digitalen Simulation

STEFAN BEHRENDT, JAN KÖLLNER, KRISTINA KÖGLER, CHRISTINE SÄLZER, ANDREAS JUST
Konstruktion und psychometrische Prüfung eines Tests zur Diagnostik mathematischer Studieneingangsleistungen

ELISABETH MAYWEG, NATALIE ENDERS, CARLA BOHNDICK, JANA RÜCKMANN *Online, blended oder Präsenz? Ein systematisches Literaturreview von Metaanalysen zur Effektivität hochschulischer Lehrformate*



Verlag Barbara Budrich

ISSN 2367-3044

IMPRESSUM

Zeitschrift für empirische Hochschulforschung (ZeHf)

HERAUSGEBERINNEN

Prof. Dr. Joachim Grabowski, Institut für Psychologie, Leibniz Universität Hannover

PD Dr. Otto Hüther, Institut für Soziologie, Universität Hamburg

Prof. Dr. Elke Wild, Leiterin der Arbeitseinheit Pädagogische Psychologie, Universität Bielefeld

WISSENSCHAFTLICHER BEIRAT

Prof. Dr. Stefan Fries, Fakultät für Psychologie und Sportwissenschaft, Abteilung Psychologie, Arbeitseinheit

Psychologie der Erziehung und Bildung, Universität Bielefeld; Prof. Dr. Anke Hanft, Institut für Pädagogik, Universität

Oldenburg; Prof. Dr. Michael Hölscher, Lehrstuhl für Hochschul- und Wissenschaftsmanagement, Deutsche Universität

für Verwaltungswissenschaften Speyer; Prof. Dr. Monika Jungbauer-Gans, Deutsches Zentrum für Hochschul- und

Wissenschaftsforschung, Leibniz Universität Hannover; Prof. Dr. Georg Krücken, Internationales Zentrum für Hoch-

schulforschung, Universität Kassel; Prof. Dr. Stephan Laske, Institut für Organisation und Lernen, Universität Innsbruck

ERSCHEINEN UND BEZUGSBEDINGUNGEN

Die Zeitschrift erscheint 2 x jährlich mit einem Jahresumfang von rd. 240 Seiten (Print und Online).

Einzelheft Print: 35,00 €

Jahresabonnement Print: Institutionen 96,00 €; Privat: 56,00 €

Jahresabonnement Online: Institutionen 145,00 €, Privat: 67,00 €

Jahresabonnement Print+Online: Institutionen 145,00 €; Privat 67,00 € Mit einem Online- oder einem Print+Online-

Abonnement haben Sie freien Zugriff auf das gesamte Volltext-Archiv der ZeHf, solange Ihr Abonnement besteht.

Print-Preise jeweils zzgl. Versandkosten.

Die Bezugspreise enthalten die gültige Mehrwertsteuer. Abonnements verlängern sich automatisch um ein Jahr.

Abonnement-Kündigungen bitte schriftlich an den Verlag. Kündigungsfrist bis drei Monate zum Jahresende.

Das digitale Angebot finden Sie auf: <https://zehf.budrich-journals.de>.

BESTELLUNGEN BITTE AN DEN BUCHHANDEL ODER AN

Verlag Barbara Budrich, Stauffenbergstr. 7, D-51379 Leverkusen-Opladen

Tel.: (+49) (0)2171 79491 50 Fax: (+49) (0)2171 79491 69, zeitschriften@budrich.de

www.budrich.de • www.budrich-journals.de • www.shop.budrich.de

Jg. 7, Heft 1 (2023)

© 2023 Verlag Barbara Budrich, Opladen • Berlin • Toronto

Die Zeitschrift sowie alle darin enthaltenen Beiträge sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen. Namentlich gekennzeichnete Beiträge geben nicht in jedem Fall die Meinung der Redaktion wieder.

Aktuelle Mediadaten/Anzeigenpreisliste: ZEHF23 vom 16.02.2023

Druck: paper & tinta, Warschau, printed in Europe. Gedruckt auf FSC®-zertifiziertem Papier.

Satz: Anja Borkam, Jena – kontakt@lektorat-borkam.de

Umschlaggestaltung: Bettina Lehfeldt, Kleinmachnow – www.lehfeldtgraphic.de

ISSN 2367-3044

ISSN Online 2367-3052

Inhalt

Themenheft: Innovative Bildungstechnologien und technologiegestütztes Lernen als Gegenstand moderner Hochschulforschung

Editorial	3
<i>Maximilian Sailer, Johannes Abel, Judith Schweppe</i> Bildungstechnologische Fragen in der Hochschullehre im Überblick	4
<i>Florian Trauten, Carolin Eitemüller, Maik Walpuski</i> Lernaufgaben mit automatisiertem Feedback in einem digitalen Hochschultutorium zur Vorlesung der Allgemeinen Chemie als Angebot zur Binnendifferenzierung	13
<i>Caroline Götz, Paul Erhardt, Simone N. Löffler</i> Selbstwirksamkeit in selbstregulierten Lernprozessen an Hochschulen mithilfe von Podcasts fördern – eine Interventionsstudie mit Interaktivem Ambulantem Assessment	34
<i>Lea Grotegut, Katrin B. Klingsieck</i> Kompetenzentwicklung in der Lehrkräftebildung: der Mehrwert einer digitalen Simulation	56
<i>Stefan Behrendt, Jan Köllner, Kristina Kögler, Christine Sälzer, Andreas Just</i> Konstruktion und psychometrische Prüfung eines Tests zur Diagnostik mathematischer Studieneingangsleistungen	74
<i>Elisabeth Mayweg, Natalie Enders, Carla Bohndick, Jana Rückmann</i> Online, blended oder Präsenz? Ein systematisches Literaturreview von Metaanalysen zur Effektivität hochschulischer Lehrformate	96

Editorial

Liebe Leserinnen und Leser,

der siebte Jahrgang der Zeitschrift für empirische Hochschulforschung (ZeHf) startet mit einem Themenheft zu einem spannenden und hochschuldidaktisch hoch relevanten Gegenstandsbereich.

Die Gastherausgeber*innen – Prof. Dr. Maximilian Sailer, Prof. Dr. Judith Schewpe und Dr. Johannes Abel von der Universität Passau – leiten in ihrem Eingangsbeitrag prägnant in die Bedeutung des Themas „Innovative Bildungstechnologien und technologiegestütztes Lernen als Gegenstand moderner Hochschulforschung“ ein und skizzieren die in den Einzelbeiträgen verfolgten Zugänge zur wissenschaftlichen Analyse der im Zuge der digitalen Transformation festzustellenden Veränderungen und Herausforderungen vor allem in der Lehre an Hochschulen. Aus diesem Grund beschränken wir uns an dieser Stelle auf eine kurze Schilderung der Entstehung dieses Themenhefts.

Das Gastherausgabeteam ist 2021 mit einem inhaltlich überzeugenden Konzept an die Herausgeber der ZeHf herangetreten, und so wurde zeitnah vereinbart, einen bundesweiten *Call for Papers* über viele einschlägige Verteiler zu schicken. Entsprechend gingen zahlreiche Skizzen ein, die sorgfältig von dem erweiterten Herausgabekreis im Hinblick auf ihre mutmaßliche Passung zum Thema sowie zu den Anforderungen der Zeitschrift eingeschätzt wurden. Einvernehmlich wurden letztlich fünf Autor*innengruppen zur Einreichung eines Manuskripts aufgefordert. Alle fünf Beiträge wurden nach Abgabe einem Peer-review-Verfahren unterzogen (wie immer in anonymisierter Form) und haben es nicht nur erfolgreich durchlaufen. Vielmehr haben Überarbeitungen aufgrund der gutachterlichen Rückmeldungen aus Sicht aller Beteiligten dazu beitragen, den jeweiligen Forschungsstand, die adressierten Fragestellungen, die methodischen Herangehensweisen und/oder die Darstellung und Einordnung der Ergebnisse noch klarer und auch für eine heterogene Leserschaft sehr gut nachvollziehbar darzulegen.

Das Themenheft ist insofern nicht nur für Personen aufschlussreich, die hochschuldidaktische beziehungsweise „bildungstechnologische“ Themen selbst beforschen, sondern auch für alle Hochschullehrenden, die in der Corona-Pandemie umfängliche Erfahrungen zum Einsatz technischer Medien in Bildungskontexten gewonnen haben und an wissenschaftlich fundierten Erkenntnissen zur Effektivität von technologisch-begründeten innovativen Lehransätzen an Hochschulen interessiert sind.

Unseren Leserinnen und Lesern dürfen wir also wieder eine spannende Lektüre wünschen mit hoffentlich interessanten Anregungen für die eigene Forschung und Lehre.

Ihr Herausgabeteam

Bildungstechnologische Fragen in der Hochschullehre im Überblick

Maximilian Sailer, Johannes Abel, Judith Schweppe

Zusammenfassung: In den vergangenen Jahrzehnten hat der technologische Fortschritt den Bereich der primären, sekundären und tertiären Bildung grundlegend verändert. Entsprechend etabliert sich auch in der Hochschulforschung der Begriff „Bildungstechnologie“. Der Beitrag skizziert den Forschungsstand zum technologiebasierten Lernen im Hochschulbereich und stellt ein deskriptives Modell vor, das vier verschiedene lernrelevante Kontexte und Komponenten einbezieht. Wissenschaftliche Arbeiten, die sich mit dem Einsatz von Bildungstechnologien in der Hochschullehre befassen, weisen eine erfreuliche Bandbreite auf und rücken eine Vielzahl von unterschiedlichen Fragestellungen und Forschungsinteressen in den Fokus. Dabei werden drei thematische Schwerpunkte besonders häufig fokussiert, die auch untereinander Bezüge aufweisen: (1) Instruktionsansätze und didaktische Vermittlungskonzepte, die umfassende Lernepisoden strukturieren; (2) Szenarien und Formate, die Lernprozesse unterstützen; (3) Systeme und Methoden, die Lernprozesse erfassen, formativ steuern und bewerten. Anschließend werden die einzelnen Beiträge des Themenhefts vorgestellt und entlang der skizzierten Schwerpunkte eingeordnet. Als Fazit kann festgehalten werden, dass die „neu entdeckte“ Aufmerksamkeit für Bildungstechnologien genutzt werden sollte, um dieses Forschungsfeld im deutschsprachigen Raum zu stärken.

Schlüsselwörter: Bildungstechnologien, Hochschulforschung, digitale Lehr/Lernszenarien

Educational technology related issues in higher education teaching at a glance

Summary: In recent decades, technological progress has fundamentally changed the field of primary, secondary and tertiary education. Accordingly, the term „educational technology“ is also becoming established in higher education research. The present paper outlines the state of research on technology-based learning in higher education and presents a descriptive model that incorporates four different contexts and components relevant to educational technologies. Scientific papers dealing with the use of technologies in higher education show a gratifying range and focus on a variety of different issues and research interests. In this context, three thematic focal points are emphasized with particular frequency: (1) instructional approaches and didactic mediation concepts that structure comprehensive learning episodes; (2) scenarios and formats that support learning processes; (3) systems and methods that record, formatively control and evaluate learning processes. Subsequently, the individual contributions of the special issue are presented and classified along the outlined focal points. As a conclusion it can be stated, that the „newly discovered“ attention for educational technologies should be used to strengthen this research field in German-speaking countries.

Keywords: instructional technology, educational technology, higher education, e-learning

1 Ausgangslage

Instructional Technology oder Educational Technology sind die englischen Bezeichnungen der Wissenschaftsdisziplin Bildungstechnologie, die sich derzeit im deutschsprachigen Raum zu etablieren versucht. Dabei haben Überlegungen zu Gestaltung und Verbesserung institutionalisierter Lern- und Bildungssettings unter Berücksichtigung bildungstechnologischer Aspekte eine lange Tradition und reichen bis in die 1960er-Jahre zurück (Tulodziecki, 2011). Im Gegensatz zum englischsprachigen Raum verläuft die Institutionalisierung der Wissenschaftsdisziplin Bildungstechnologie in Deutschland vergleichsweise schleppend (Kerres, 2018). Niegemann und Weinberger (2020) sehen einen Hauptgrund hierfür darin, dass sich im deutschsprachigen Bereich die geisteswissenschaftliche Pädagogik beziehungsweise Erziehungswissenschaft und die Pädagogische Psychologie viele Jahrzehnte nebeneinander entwickelten und eine genuin geisteswissenschaftliche Pädagogik eine ablehnende Haltung gegenüber Technik und technologischen Entwicklungen nachweislich hegte. Erst seit den späten 1960er-Jahren konnte sich neben der Pädagogischen Psychologie auch eine empirisch arbeitende pädagogische Forschung etablieren, in deren Rahmen auch bildungstechnologische Fragestellungen zunehmend aufgegriffen wurden, allerdings mit Schwerpunkt auf Schule und Unterricht.

In den vergangenen Jahrzehnten hat der technologische Fortschritt den Bereich der primären, sekundären und tertiären Bildung grundlegend verändert. Entsprechend etabliert sich auch in der Hochschulforschung der Begriff der „Bildungstechnologie“ (Opfermann et al., 2020) zur Kennzeichnung einer Forschungslinie, in der „unterschiedliche Arrangements von Lernbedingungen, die Unterstützung des Erwerbs von Wissen und Können, die Beeinflussung von Motiven und Emotionen sowie die Funktionalität [...] von Medien mit dem Ziel der Förderung von Lern- und Bildungsprozessen“ adressiert werden (Niegemann & Weinberger, 2020, S. 4).

Stark intensiviert hat sich das Interesse an Bildungstechnologien bekanntlich im Zuge der Corona-Pandemie: Anfang 2020 mussten schlagartig nahezu alle institutionalisierten Lehr-Lern-Angebote „über Nacht“ in den digitalen Raum verlagert werden. Im Hochschulkontext gelang dies vergleichsweise leichter als im Schulbereich und hat fraglos zu einem deutlichen Anstieg „digitaler Kompetenzen“ von Hochschullehrenden geführt. Allerdings mussten digitale bzw. hybride Lehr-Lern-Angebote binnen kürzester Zeit mithilfe der von Hochschulen lizenzierten Programme entwickelt werden. So dürfte der Fokus vielerorts ganz pragmatisch auf eine „Konvertierung“ bewährter Veranstaltungskonzepte gerichtet gewesen sein, die nur bedingt auf hochschuldidaktischen Erwägungen fußte. Mit anderen Worten ist hier ein doppelter wissenschaftlicher „Aufholbedarf“ zu konstatieren: zum einen hinsichtlich der Folgen dieser pandemiebedingten „Digitalisierung der Hochschullehre“ für die Kompetenzentwicklung von Studierenden sowie deren Studienerfolg und zum anderen bezüglich der Entwicklung von Lehrveranstaltungskonzepten, in denen Bildungstechnologien – vor, während und nach der Pandemie - hochschuldidaktisch reflektiert zum Einsatz kommen.

2 Stand der Forschung

Bereits in den 1970er-Jahren changierte der bildungspolitische, teilweise aber auch der wissenschaftliche Diskurs im Hinblick auf den Einsatz von Medien in Lehr-Lern-Szenarien zwischen euphorischer Überbetonung von Chancen und Nutzen und einer ernüchterten Wahrnehmung der Grenzen (Hüther & Podehl, 2005). Dies gilt im Wesentlichen auch für aktuell geführte Debatten, und so ist es angesichts hoher finanzieller Anschaffungskosten sowie kurzer Innovationszyklen immer wichtiger, die Vorteile und Begrenzungen eines Einbezugs technischer Neuerungen in der Hochschullehre auf den empirischen Prüfstand zu stellen.

In einer Metaanalyse untersuchten Yildiz et al. (2020) aktuelle Trends der „education technologies research worldwide“ zwischen 2015 und 2020. Trotz einer großen Varianz hinsichtlich des genauen Forschungsinteresses, der verwendeten Methoden sowie der beteiligten Fachrichtungen zeigten sich klare Muster. So war ein Großteil der Studien quantitativer Natur, wobei deskriptive Umfragen und Befragungen von Studierenden dominierten (Yildiz et al., 2020). Zur generellen Frage nach der Wirksamkeit von Bildungstechnologien haben Tamim et al. (2021) eine umfangreiche Metaanalyse zweiter Ordnung vorgelegt, in der sie 25 Meta-Analysen und 1.055 Primärstudien integriert haben. Hierbei konnte ein positiver, aber kleiner Effekt von digitalen Medien auf den kognitiven Lernerfolg nachgewiesen werden. In einer weiteren Studie untersuchten die Autor:innen die Qualität von 52 Metaanalysen, die zwischen 1988 und 2017 in dem Themenfeld „educational technology“ entstanden sind (Tamim et al., 2021). In dieser Studie fand der sogenannte „Meta-Analysis Methodological Reporting Quality Guide (MMRQG)“ Anwendung, ein Instrument, das dazu dient, die Qualität von Metaanalysen zu bestimmen. Tamim et al. (2021) stellten dabei einen inversen Zusammenhang zwischen dem MMRQG-Wert und der berichteten Effektstärke fest: Die Metaanalysen, deren Qualität als geringer eingeschätzt wurde, gingen also mit höheren Effektstärken einher (und umgekehrt). Daraus schließen sie, dass der Einfluss von Technologieeinsatz auf das Lernen eher überschätzt und sowohl Lehrenden als auch Politikerinnen und Politikern dadurch ein irreführender Eindruck vermittelt wurde. Dieses Ergebnis verdeutlicht die Schwierigkeit, generalisierende Aussagen über die Effektivität von digitalen Medien zu treffen. Somit kann es hilfreich sein, lediglich Einzelstudien, die den Einsatz von Bildungstechnologien thematisieren, in den wissenschaftlichen Fokus zu rücken – dies vermindert die Gefahr, dass Sachverhalte miteinander verglichen werden, die sich nur schwer miteinander vergleichen lassen, sofern keine theoretische Einordnung erfolgt. In einer weltweit angelegten Metaanalyse von Yu (2021) wurden 7 unterschiedliche Faktoren ermittelt, die sich auf den Erfolg von Online-Lehre auswirken können: (1) behavioral intention, (2) instruction, (3) engagement, (4) interaction, (5) motivation and self-efficacy, (6) performance, (7) satisfaction and self-regulation. Diese Auflistung zeigt, dass sich viele Faktoren sowie deren Zusammenspiel auf den Erfolg, aber auch auf den Misserfolg von Online-Lehre auswirken können, die zudem auf unterschiedlichen theoretischen Ebenen verankert sind. Sie verdeutlicht weiterhin eine Gefahr, die in Metaanalysen zweiter Ordnung liegt: Diese können – gerade durch zusammenfassende Auflistungen – einen rein induktiven Blick auf Faktoren suggerieren, die die Wirkung von Bildungstechnologien beeinflussen, der wenig theoriegetrieben beziehungsweise theoretisch verankert erscheint und eine von Rankings dominierte Rezeption solcher Studien nach sich ziehen kann. Um einen möglichst breiten, aber strukturierten Überblick über Bedingungsfaktoren des Einsatzes von Bildungstechnologien zu erhalten, bietet

es sich daher an, ein theoretisches Modell heranzuziehen, das lernrelevante Kontexte differenziert und miteinander in Verbindung setzt. Hierbei ist zu beachten, dass der Einsatz von Bildungstechnologien in Lehr-Lern-Arrangements generell stark von Einstellungen und Kompetenzen der Lehrenden abhängt (Gilch et al., 2019) und speziell im Hochschulkontext aufgrund des rechtsstaatlichen Prinzips der „Freiheit in Forschung und Lehre“ den Haltungen von Hochschullehrenden eine besondere Bedeutung für Art und Umfang des Einsatzes von Technik zu Lehrzwecken zukommt.

Ein deskriptives Modell, das vier verschiedene lernrelevante Kontexte und Komponenten einbezieht, die in technologiebasierten Lernszenarien zum Tragen kommen, bietet Tergan (2004). Er unterscheidet zwischen *individuellem Lernkontext* (Lernende), *Anwendungskontext* (Inhalt), *pädagogischem Kontext* (Konzepte und Methoden) und *Technologiekontext* (Technologie und Medien).

Der individuelle Lernkontext fokussiert die Seite der Rezipient:innen und umfasst die persönlichen Rahmenbedingungen wie Bildungsstand, soziale und finanzielle Ressourcen sowie den Zugang zu technischen Geräten. Zudem fallen auch weitere kontextuale Merkmale wie Motivation, Vorwissen und allgemeine kognitive Fähigkeiten in diesen Bereich. In Hinblick auf den individuellen Lernkontext würden wir noch zwischen *fakultativ* und *obligatorisch* unterscheiden wollen. Neben der Möglichkeit, sich selbstgesteuert und auf freiwilliger Basis Wissen anzueignen (fakultativ), existieren obligatorische Settings, also durch Schulen, Universitäten oder durch Arbeitgeber:innen organisierte digitale Lernkurse und Angebote (diese können sowohl formaler als auch nonformaler Natur sein).

Unter den Anwendungskontext fallen Eigenschaften des Lerngegenstandes wie beispielsweise das Thema, die sachliche Korrektheit, die Lernziele und die Authentizität. Der Anwendungskontext lässt sich nach unserer Sicht zusätzlich in die Unterkategorien *materiales* und *formales Wissen* in Anlehnung an Klafkis Bildungstheorie unterteilen (Klafki, 2007). Während materiales Wissen meist schnell per Internetsuche verfügbar ist, so erscheint die Verknüpfung und Anwendung von Wissen essenziell, um den Aufgaben und Herausforderungen der modernen (Arbeits-)Welt gerecht zu werden. Bildungstechnologien sehen sich somit immer mit der Frage konfrontiert, welche Art des Wissens sie zu fördern versuchen. Im Idealfall werden durch Bildungstechnologien beide Bereiche der Wissensaneignung bedient. Der pädagogische Kontext verweist auf die dem Lernszenario zugrunde liegende didaktische Methode, also bewusste instruktionale Maßnahmen zur Unterstützung des Lernens. Für diesen Kontext lässt sich ergänzend zu Tergan (2004) zwischen *Motivierung* und *kognitiver Aktivierung* unterscheiden, wenn Bildungstechnologien zum Einsatz kommen. Instruktionale Maßnahmen, die Bildungstechnologien integrieren, dienen zur Unterstützung des Lernens. Dabei wird entweder die motivationale oder kognitive Route bedient, im Idealfall auch beide. Aktives Lernen und kognitive Aktivierung können nach Chi und Wylie (2014) verschiedene Ausprägungen haben. Didaktische und instruktionale Maßnahmen sollten Bildungstechnologien dann einbinden, um eine aktive, konstruktive oder interaktive Auseinandersetzung mit Inhalten zu ermöglichen und die kognitive Verarbeitung zu unterstützen. Motivationale Routen sollten dann bedient werden, um andererseits das Interesse am Gegenstand (Inhalt) zu erhöhen und Lernenden ein Kompetenzerleben zu ermöglichen.

Der Technologiekontext umfasst schließlich alle technischen Aspekte des Lehr-Lernszenarios (Tergan, 2004). Wir würden diesen in die Unterkategorien *Bedienerfreundlichkeit* und *Nützlichkeit* (perceived usefulness) unterteilen. Gerade unter Einbezug des Technologiekontextes wird ersichtlich, dass Anwendungsfreundlichkeit und Nützlichkeit nicht zwingend auf-

einander bezogen sein müssen und es stets die Balance zwischen beiden Aspekten auszuloten gilt.

Die beschriebenen Lernkontexte sind nicht voneinander isoliert und austauschbar zu betrachten, vielmehr bedingen sie sich gegenseitig – denn nicht jede didaktische Methode ist für jede Zielgruppe geeignet, nicht jede Technik zur Aufbereitung eines jeden Inhalts sinnvoll, und nicht jede Zielgruppe kann mit jeder Technik umgehen. Somit beeinflussen neue technische Errungenschaften nicht nur den Technologiekontext, sondern wirken sich auf alle vier lernrelevanten Kontexte aus beziehungsweise die anderen Kontexte beeinflussen, welche Neuerung wie bei welchen Lernenden zu welchem Zweck eingesetzt wird.

Wissenschaftliche Arbeiten, die sich mit dem Einsatz von Bildungstechnologien in der Hochschullehre befassen, weisen eine erfreuliche Bandbreite auf und rücken eine Vielzahl von unterschiedlichen Fragestellungen und Forschungsinteressen in den Fokus. Ausgehend von der Funktion, die Bildungstechnologien in Lehr-Lernprozessen erfüllen, werden dabei drei thematische Schwerpunkte besonders häufig fokussiert, die auch untereinander Bezüge aufweisen:

Instruktionsansätze und didaktische Vermittlungskonzepte, die umfassende Lernepisoden (z.B. ein Seminar, ein Modul) strukturieren. Hierunter lassen sich Modifikationen traditioneller Hochschullehre wie im Flipped Classroom oder Inquiry Learning fassen, die Lernprozesse auf neue Art strukturieren und Studierende kognitiv aktivieren sollen. Der Erwerb von Wissen wird typischerweise stärker in Phasen des Eigenstudiums verlagert, während in Präsenzphasen das Wissen vertieft, geprüft und konsolidiert wird. Weiterhin werden beispielsweise interaktive Aufgaben und Feedbackstrukturen integriert, um diese Form des Blended Learning auch im Hinblick auf nicht-kognitive Faktoren, wie emotionale und motivationale Aspekte, auszulegen.

Szenarien und Formate, die Lernprozesse unterstützen. Vielfältige Szenarien und Formate werden erprobt, um herkömmliche Veranstaltungsformen (z.B. Vorlesungen) interaktiver sowie adaptiver und damit lernförderlicher zu gestalten. Dazu zählen etwa der Einsatz von Spielelementen in Lernkontexten (Gamification) und Lernspiele, digitale Simulationen, Virtual/Augmented Reality, Audience Response-Systeme, mobiles Lernen sowie weitere innovative, technologiegestützte Lernformate.

Systeme und Methoden, die Lernprozesse erfassen, formativ steuern und bewerten. Neben klassischen diagnostischen Verfahren werden zunehmend auch Methoden, die auf großen Datenmengen und auf künstlicher Intelligenz basieren, wie Learning Analytics, Educational Data Mining und Deep Learning, eingesetzt. Ein Beispiel hierfür sind Intelligente Tutorielle Systeme. Diese ermöglichen durch Auswertung von Lernaktivitäten personalisierte, wissensbasierte Empfehlungen an Lernende, die auch im Sinne von individuellem Feedback und adaptiver Unterstützung genutzt werden können. Aber auch KI-nahe Anwendungen wie Software zur Gesichtsausdruckserkennung kommen in diesem Themenschwerpunkt zum Tragen. Damit lassen sich Merkmale aufseiten der Dozierenden (wie deren Engagement) erfassen, die für die didaktische Qualität von Lehrveranstaltungen bedeutsam sind. Ferner können kognitive Group Awareness-Tools sowohl stabile (Vorwissen, Interessen) als auch situationale (Leistung, Engagement) Charakteristika von Lernenden erfassen und somit Lernprozesse auf individueller oder sozialer Ebene verbessern (Niegemann & Weinberger, 2020).

3 Einordnung der Beiträge

Das vorliegende Themenheft der Zeitschrift für empirische Hochschulforschung bündelt Erkenntnisse zu innovativen Lehr-Lernkonzepten, die auf bildungstechnologische Neuerungen zurückgreifen und sich entlang der skizzierten Schwerpunkte einordnen lassen. Es umfasst drei empirische Originalarbeiten, in denen die Wirksamkeit spezifischer Bildungstechnologien evidenzbasiert überprüft wurde. In einem weiteren Beitrag wird ein computerbasiertes Instrument zur Erfassung mathematischer Grundkenntnisse von Sekundarstufenschüler:innen vorgestellt, und Befunde zu dessen Güte werden erläutert. Der abschließende Beitrag adressiert mit Hilfe eines systematischen Literaturreviews detailliertere Fragen zur Effektivität von Online-Lehre.

Der erste Beitrag von Florian Trauten, Carolin Eitemüller und Maik Walpuski kombiniert ein didaktisches Vermittlungskonzept mit adaptivem Feedback zur Unterstützung und Steuerung von Lernprozessen und lässt sich somit allen genannten Schwerpunkten zuordnen. Naturwissenschaftliche Studiengänge im Allgemeinen und speziell in der Chemie weisen hohe Abbruchquoten auf. Die Autor:innen stellen daher ein neu entwickeltes Förderangebot für Chemiestudierende im ersten Semester zur Erhöhung des Studienerfolgs vor, mit dem insbesondere auch Defizite im Vorwissen von Studienanfänger:innen ausgeglichen werden sollen, die sonst die Wahrscheinlichkeit von Studienabbrüchen erhöhen. Hierzu entwickelten die Autor:innen Online-Lernaufgaben, die adaptives Feedback bereitstellen und basierend auf der individuellen Performance einen multiple-try Feedback Algorithmus nutzen. Insbesondere wurde der Einfluss des Vorwissens auf die Wirksamkeit zweier adaptiver Feedback-Algorithmen (elaboriert vs. korrektiv) untersucht. Bei Studierenden, die vorher keinen Chemie-Leistungskurs belegt haben, konnte ein signifikanter positiver Effekt für adaptives elaboriertes gegenüber korrektivem Feedback festgestellt werden.

Der Beitrag von Caroline Götz, Paul Erhardt und Simone Löffler nutzt im Sinne des zweiten genannten Schwerpunkts individualisiertes Feedback zur Unterstützung von Lernprozessen und fokussiert dabei die Rolle von Selbstwirksamkeit im selbstregulierten Lernprozess. Diese gilt als ein wichtiger Prädiktor für den Studienerfolg. In ihrer Studie wurden Interaktive Ambulante Assessments mit Podcasts kombiniert, um die Selbstwirksamkeit im selbstregulierten Lernprozess zu fördern. Hierfür haben Studierende (N = 106) drei Mal täglich Abfragen zu ihrem Lernprozess beantwortet, die ihnen über elektronische Tagebücher präsentiert wurden. Mittels Mehrebenenanalysen werden statistisch überzufällige Unterschiede zwischen der Interventionsgruppe mit täglichem individualisiertem Feedback und einer Kontrollgruppe ohne Feedback geprüft. Die Ergebnisse zeigen, dass die Selbstwirksamkeit der Interventionsgruppe innerhalb der Interventionsphase, im Vergleich zur Baselinephase und zur Kontrollgruppe, gefördert werden konnte.

Ebenfalls dem zweiten Schwerpunkt zuzuordnen ist der Beitrag von Lea Grotegut und Katrin B. Klingsieck, der digitale Simulationen als Format zur Unterstützung diagnostischer Kompetenzen von angehenden Lehrkräften nutzt. Die Autorinnen setzt an zwei zentralen Herausforderungen in der Lehrer:innenbildung an: Zum einen bleibt in großen Lehrveranstaltungen wenig Raum für die Anwendung theoretischen Wissens, zum anderen können Fehlentscheidungen in realen diagnostischen Situationen starke negative Folgen für Schüler:innen haben, weshalb die Sensibilisierung für die Bedeutung sowie die Vermittlung von Diagnosekompetenzen ein zentrales Ausbildungsanliegen sein muss. Um den diesbezüglichen

chen Nutzen digitaler Simulationen zu prüfen, wurden drei Lehr-Lern-Szenarien (digitaler Simulation vs. problemorientierte vs. vorlesungsbasierte Unterweisung) quasi-experimentell untersucht. Den Ergebnissen zufolge hat die digitale Simulation einen positiven Effekt auf die selbstbestimmte Motivation, weshalb in dieser Hinsicht von einem Mehrwert im Rahmen der universitären Ausbildung von Lehrer:innen auszugehen ist. Allerdings wurde auch ein negativer Effekt auf den Wissenszuwachs festgestellt, für den die Autorinnen weiterführende Erklärungen liefern.

Im vierten Beitrag wird von Stefan Behrendt, Jan Köllner, Kristina Kögler, Christine Sälzer und Andreas Just ein computerbasiertes Instrument zur Erfassung mathematischer Grundkenntnisse in der Sekundarstufe präsentiert. Am Ausgangspunkt des Beitrags steht die Beobachtung, dass mathematische Basisfähigkeiten auch in höheren Klassenstufen für den Lernerfolg bedeutsam sind, dies aber oft vernachlässigt wird, auch weil es an geeigneten diagnostischen Instrumenten mangelt. Umso verdienstvoller ist, dass das Autorenteam ein solches Diagnoseinstrument entwickelt hat und zeigen kann, dass dieses reliabel sowie inhaltlich und prognostisch valide ist.

Der abschließende Beitrag von Elisabeth Mayweg, Natalie Enders, Carla Bohndick und Jana Rückmann weitet die Perspektive und wirft auf Basis eines systematischen Literaturreviews von $N = 44$ Metaanalysen (Publikationszeitraum 2011–2022) einen übergeordneten Blick auf die vergleichende Effektivität der Lehrformate Onlinelehre, Blended Learning (einschließlich Flipped Classroom) und traditionelle Präsenzlehre. Ziel der Autorinnen ist es, eine evidenzbasierte Grundlage für Empfehlungen zum Einsatz von Lehrformaten in der Hochschullehre zu schaffen. Die in den recherchierten Metaanalysen betrachteten Indikatoren, welche zu kognitiven Lernergebnissen, praktischen Fertigkeiten und Studierendeneinschätzungen zusammengefasst werden, zeigen überwiegend kleine Effekte des Lehrformats. In einigen Fällen werden mittelgroße Effekte berichtet, wobei Blended-Learning-Formate geringfügig besser abschneiden. Die Höhe der Effektstärken hängt auch von den Einschlusskriterien der Metaanalysen ab. Moderatoranalysen zeigen, dass insbesondere durch Lehrkräfte veränderbare Gestaltungseigenschaften zusätzliche Variabilität aufklären können.

4 Ausblick

Dass in Folge der rasanten KI-Forschung bildungstechnologische Fragestellungen immer bedeutsamer werden, ist schon jetzt aufgrund der aktuellen Diskussion um Nutzen und Risiken von Programmen wie ChatGPT erkennbar (z.B. Ahlborn, 2023; Schelling et al., 2023; Stokel-Walker & Van Noorden, 2023; van Dis et al. 2023). Derzeit werfen sie gerade auch im Hochschulkontext unmittelbar relevante prüfungsrechtliche Fragen auf, verlangen aber mittel- und langfristig eine fortwährende Reflexion hochschulischer Lehr-Lernziele (wie der Befähigung zum wissenschaftlichen Schreiben) und deren Operationalisierung. Auf dieser Basis gilt es, in der zukünftigen Bildungsforschung grundlegend die kognitiven sowie motivationalen Auswirkungen eines veränderten Medienkonsums im Allgemeinen sowie die Konsequenzen des Rückgriffs auf intelligente Programme für das Lehr- und Lernverhalten im Besonderen zu untersuchen. Speziell der bildungstechnologischen Forschung kommt dabei die Aufgabe zu, die Übertragbarkeit bisheriger Erkenntnisse zu prüfen, das Lehr-Lern-Potential KI-basierter Tools zu analysieren und damit zur Weiterentwicklung der Lehr-Lern-Forschung durch neue Formen des Medieneinsatzes beizutragen.

Anhand dieser Diskussion lässt sich ein Punkt besonders gut verdeutlichen: Das Feld und die Rahmenbedingungen, in denen Bildungstechnologien zum Einsatz kommen, sind nicht statisch, sondern sie befinden sich in einem ständigen Wandel und Umbruch. Vor dem Hintergrund einer dynamischen technologischen (Bildungs-)Landschaft müssen wissenschaftliche Erkenntnisse in Hinblick auf ihre Gültigkeit ständig neu geprüft werden. So ist die „neu entdeckte“ Aufmerksamkeit, die Bildungstechnologien in den vergangenen drei Jahren erfahren haben, auch als Auftrag an die Wissenschaft zu begreifen, das Forschungsfeld Bildungstechnologie im deutschsprachigen Raum zu stärken – und dem Eindruck entgegenzuwirken, dass die Lehre in den „Corona-Semestern“ beziehungsweise „Corona-Schuljahren“ ein prototypisches digital unterstütztes Lehr-Lern-Szenario darstellt.

Literatur

- Ahlborn, J. (2023). KI – Kunst – Bildung. Wie komplexe algorithmische Systeme das Verhältnis von Kunst, Ästhetik und Bildung verschieben. In C., Leineweber, M. Waldmann & M. Wunder (Hrsg.), *Materialität – Digitalisierung – Bildung* (S. 192–209). Julius Klinkhardt.
- Chi, M.-T. H. & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243.
- Gilch, H., Beise, A. S., Krempkow, R., Müller, M., Stratmann, F. & Wannemacher, K. (2019). *Digitalisierung der Hochschulen. Ergebnisse einer Schwerpunktstudie für die Expertenkommission Forschung und Innovation. Studien zum deutschen Innovations-system*. https://www.e-fi.de/fileadmin/Assets/Studien/2019/StuDIS_14_2019.pdf
- Hüther, J., & Podehl, B. (2005). Geschichte der Medienpädagogik. In J. Hüther & B. Schorb (Hrsg.), *Grundbegriffe Medienpädagogik* (S. 116–127). Kopaed.
- Kerres, M. (2018). *Mediendidaktik. Konzeption und Entwicklung digitaler Lernangebote*. De Gruyter.
- Klafki, W. (2007). *Neue Studien zur Bildungstheorie und Didaktik. Zeitgemäße Allgemeinbildung und kritisch konstruktive Didaktik*. Beltz.
- Niegemann, H. & Weinberger, A. (2020). Was ist Bildungstechnologie? In H. Niegemann & A. Weinberger (Hrsg.), *Handbuch Bildungstechnologie. Konzeption und Einsatz digitaler Lernumgebungen* (S. 3–16). Springer VS.
- Opfermann, M., Höffler, T. N. & Schmeck, A. (2020). Lernen mit Medien: ein Überblick. In H. Niegemann & A. Weinberger (Hrsg.), *Handbuch Bildungstechnologie, Konzeption und Einsatz digitaler Lernumgebungen* (S. 17–30). Springer VS.
- Schelling, K., Schmohl, T., & Watanabe, A. (2023). *Künstliche Intelligenz in der Hochschulbildung Chancen und Grenzen des ki-gestützten Lernens und Lehrens*. transcript.
- Stokel-Walker, C. & Van Noorden, R. (2023). What ChatGPT and generative AI mean for science. *Nature*, 614(7947), 214–216. <https://doi.org/10.1038/d41586-023-00340-6>
- Tamim, R. M., Borokhovski, E., Bernard, R. M., Schmid, R. F., Abrami, P. C., & Pickup, D. I. (2021). A study of meta-analysis reporting quality in the large and expanding literature of educational technology. *Australasian Journal of Educational Technology*, 37(4), 100–115. <https://doi.org/10.14742/ajet.6322>
- Tergan, S.-O. (2004). Was macht Lernen erfolgreich? Die Sicht der Wissenschaft. In S.-O. Tergan und P. Schenkel (Hrsg.), *Was macht E-Learning erfolgreich? Grundlagen und Instrumente der Qualitätsbeurteilung* (S. 15–28). Springer.
- Tulodziecki, G. (2011). Zur Entstehung und Entwicklung zentraler Begriffe bei der pädagogischen Auseinandersetzung mit Medien. *MedienPädagogik*, 20, 11–39.
- van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R. & Bockting, C. L. (2023). ChatGPT: five priorities for Research. *Nature*, 614(7947), 224–226. <https://doi.org/10.1038/d41586-023-00288-7>

Yildiz, E., Alkan, A. & Cengel, M. (2020). Current trends in education technologies research worldwide: meta-analysis of studies between 2015 and 2020. *New Trends and Issues. Proceedings on Humanities and Social Sciences*, 7(1), 47–62. <https://doi.org/10.18844/prosoc.v7i1.4867>

Yu, Z. (2021). A meta-analysis and bibliographic review of the effect of nine factors on online learning outcomes across the world. *Education and Information Technologies*, 27(2), 2457–2482. <https://doi.org/10.1007/s10639-021-10720-y>

Kontakt

Prof. Dr. Maximilian Sailer · Dr. Johannes Abel

Lehrstuhl für Erziehungswissenschaft

Prof. Dr. Judith Schweppe

Professur für Psychologie mit Schwerpunkt Lehren und Lernen mit digitalen Medien

Universität Passau

Innstraße 41

94032 Passau

E-Mail: maximilian.sailer@uni-passau.de

E-Mail: johannes.abel@uni-passau.de

E-Mail: judith.schweppe@uni-passau.de

Lernaufgaben mit automatisiertem Feedback in einem digitalen Hochschultutorium zur Vorlesung der Allgemeinen Chemie als Angebot zur Binnendifferenzierung

Florian Trauten, Carolin Eitemüller, Maik Walpuski

Zusammenfassung: Die hohen Abbruchquoten in naturwissenschaftlichen Studiengängen im Allgemeinen und speziell in der Chemie wurden zum Anlass genommen, ein digitales, binnendifferenzierendes Förderangebot für Chemiestudierende im ersten Semester zur Erhöhung des Studienerfolgs zu entwickeln. Da die Überwindung von Vorwissensdefiziten entscheidend für den Verbleib im Studium ist (Heublein et al., 2017), wurden Online-Lernaufgaben entwickelt, die basierend auf der individuellen Performance in einem multiple-try Feedback Algorithmus automatisiert adaptives Feedback bereitstellen. In einem Prä-post-Vergleichsgruppendesign wurde zudem die Rolle des Vorwissens auf die Wirksamkeit der zwei adaptiven Feedback-Algorithmen (elaboriert vs. korrektiv) untersucht, da die aktuelle Befundlage diesbezüglich keine eindeutigen Schlüsse zulässt. Für Studierende, die zuvor keinen Chemie Leistungskurs belegt haben, konnte dabei ein signifikanter positiver Effekt für adaptives elaboriertes Feedback gefunden werden.

Schlüsselwörter: Adaptives Feedback, Digitalisierung, Online-Lehre, Chemie, Studieneingangsphase, Hochschuldidaktik

Learning tasks with automatic adaptive feedback in a digital university tutorial for a lecture in general chemistry as a tool for internal differentiation

Summary: High dropout rates in natural science and specifically in chemistry studies led us to provide online learning tasks for internal differentiation to foster freshmen's study success in chemistry study programs. Deficits in content knowledge have to be compensated during the study entry phase, otherwise the likelihood of student dropout is fuelled (Heublein et al., 2017). Thus, based on a multiple-try feedback algorithm various online learning tasks had been developed to provide adaptive feedback based on the learners' performance while processing the learning task. As preceding studies were not able to clarify the exact impact of prior content knowledge on the efficacy of guiding informative components on learning process, two types of feedback (elaborative vs. corrective) have been compared within a quantitative study with a pre-post comparison-group design. Here, the adaptive elaborated feedback was found to be significantly better for low-achieving students than corrective feedback alone.

Keywords: adaptiv feedback, digitalization, online teaching, chemistry, study entry phase, university didactics

1 Einleitung

Der Übergang von der Schule zur Hochschule stellt Studierende vor eine Vielzahl von Herausforderungen (Trautwein & Bosse, 2017). Die hohe Abbruchquote in Chemiestudiengängen deutscher Universitäten von aktuell 47% (Heublein et al., 2020) lässt darauf schließen, dass es nur einem Teil der Studierenden gelingt, diese Anforderungen zu meistern. Da Leistungsprobleme für 84% der Abbrechenden mindestens als wichtiger Grund aufgeführt werden (Heublein et al., 2017), kann man an den Universitäten grundsätzlich von einem *Performance-Problem* in Chemiestudiengängen sprechen.

Neben der Überwindung von Leistungsdefiziten müssen sich die Studierenden zu Beginn des Studiums in neue Sozialgefüge integrieren (Tinto, 1988), sodass das Verhältnis zu den Lehrpersonen und/oder Mitstudierenden gerade in der frühen Studieneingangsphase von eher distanzierterer Natur ist. Feedback muss zudem nun – anders als an der Schule – aktiv eingefordert werden (Heublein et al., 2017; Rost, 2018). Der Einfluss dieses *Feedback-Problems* auf das *Performance-Problem* erscheint fast als logische Folge, wenn man beachtet, dass Feedback an Hochschulen besonders selten von Studierenden aus der Gruppe der Studienabbrechenden in Anspruch genommen wird (Heublein et al., 2017), Feedback aber eine der wichtigsten Einflussgrößen auf Lernprozesse darstellt (Hattie & Timperley, 2007). Die Kontaktaufnahme zu anderen Studierenden und Dozierenden, um Feedback zu erhalten, kann so zu einem entscheidenden Faktor für den Verbleib an der Hochschule und für den Studienerfolg werden (Heublein et al., 2017; Sarcletti & Müller, 2011), es sei denn, es wird auf anderen Wegen Feedback bereitgestellt.

Für das Fach Chemie konnte gezeigt werden, dass Lücken im Vorwissen im Verlauf des ersten Semesters nicht aufgeholt werden (Averbeck, 2020). Über informatives tutorielles Feedback (ITF), das z.B. in einer Online-Lernumgebung Informationen zur Bewältigung einer Aufgabe zur Verfügung stellt, ohne unmittelbar die Lösung zu präsentieren, kann personenunabhängig effektiv Hilfestellung bereitgestellt werden (Narciss & Huth, 2004). Fraglich ist allerdings, ob elaboriertes Feedback das Vorwissen aller Studierenden konzeptionell berücksichtigen kann. Denn während Novizen tendenziell von erklärendem Feedback profitieren (Feedback Principle; Johnson & Priest, 2014; Moreno & Mayer, 2007), ist es naheliegend, dass dieser Vorteil bei zunehmender Expertise verschwindet (Expertise Reversal Effekt; Chen et al., 2017; Kalyuga, 2014). Ergebnisse aus Metastudien zur Wirksamkeit von Feedback im Allgemeinen (Bangert-Drowns et al., 1991; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Wisniewski et al., 2019) und Studien, die spezifisch den Vorwissenseinfluss untersuchen (Fyfe et al., 2012; Narciss, 2006; Smits et al., 2008), zeigen ein uneindeutiges und zum Teil kontroverses Bild von Effektstärken.

Bei den vorangestellten Studien variieren Stichprobe, Feedbackmaßnahme und/oder Aufgabenbereich, sodass aktuell keine gesicherten Erkenntnisse vorliegen. Ziel des hier vorgestellten Projektes war es daher zu ermitteln, ob mit einem binnendifferenzierenden Förderangebot, das die individuellen Voraussetzungen der Erstsemesterstudierenden in Bezug auf deren fachspezifische Vorkenntnisse durch adaptives Feedback stärker berücksichtigt, der Studienerfolg am Ende des ersten Semesters erhöht werden kann. Die Schlüsselrolle, die der Disziplin der Allgemeinen Chemie für das gesamte Chemiestudium zukommt, bietet daher berechtigte Gründe, genau dort eine Intervention anzusetzen.

2 Stand der Forschung

Eine Interventionsmaßnahme für eine vom Wissensstand her heterogene Lerngruppe, die das Vorwissen konzeptionell berücksichtigen soll, steht einer Reihe an Herausforderungen gegenüber, insbesondere, wenn Feedback automatisiert und fehlerspezifisch vom Computer ausgegeben werden soll. Für ein besseres Verständnis, wie Feedback wirkt, modelliert Narciss (2018) ausgewählte Lernprozesse in ihrem *Interactive two feedback-loops* (ITFL) Modell. Dieses weist deutliche Parallelen zu dem Modell des selbstregulierten Lernens von Butler und Winne (1995) auf. Es wird hier allerdings verstärkt Augenmerk auf die modellierten Subprozesse von internem und externem Feedback sowie die Einflussnahme von externem Feedback auf interne Informationsverarbeitungsprozesse gelegt. Das Modell in Abbildung 1 verdeutlicht auf anschauliche Weise die ablaufenden Prozessschritte der Informationsverarbeitung im Individuum bei der Bearbeitung einer Aufgabe und wie performancebasiertes Feedback entstehen sollte und wie Feedback auf diese wirken kann.

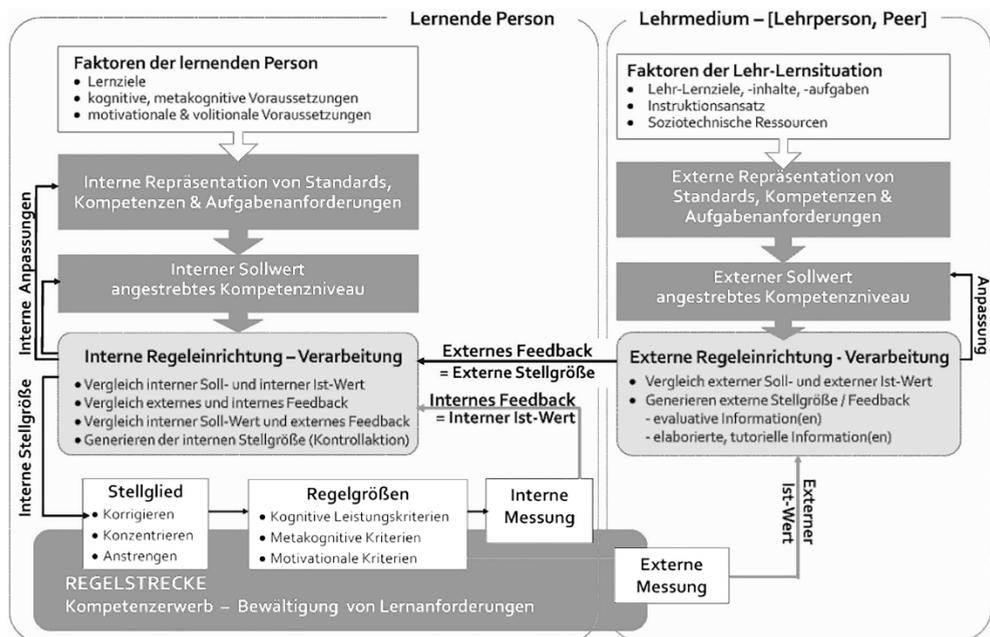


Abbildung 1: Interactive two feedback-loops model – ITFL-Modell (Narciss, 2018)

Daraus lassen sich weitere Anforderungen an die angestrebten digitalen Lernaufgaben ableiten. Wird mit der Bearbeitung einer Aufgabe begonnen, so bedingen die Faktoren der lernenden Person maßgeblich, wie die Aufgabeninformation verarbeitet werden. Das Vorwissen bedingt, ob z.B. Heuristiken für die spezifische Aufgabe herangezogen werden können oder nicht. Damit resultiert auf Personenebene – auf Basis einer individuellen Repräsentation der Aufgabe – eine von individuellen Faktoren bedingte Performance in einer Aufgabe. Diese Summe an individualitätsstiftenden Faktoren muss bei der Konzeption von Feedback bestmöglich berücksichtigt werden. Dieser systemtheoretische Ansatz geht davon aus, dass in dem jeweiligen Regelkreis (lernende Person oder Lehrmedium) bestimmte Regelgrößen fort-

laufend erfasst werden, sodass sich entsprechend des regelkreisspezifischen Blickwinkels ein *Ist-Wert* ergibt, der innerhalb des Regelkreises an die informationsverarbeitende Instanz rückgemeldet wird. Diese generiert dann entsprechend internes bzw. externes Feedback.

Der externe Regelkreis muss dabei nicht zwingend in einer Person stattfinden, sondern kann auch von einer automatisierten digitalen Lernaufgabe als Lernmedium vollzogen werden. Liegt eine Diskrepanz zwischen *Ist-* und *Soll-Wert* vor, werden dort, wie auch bei Butler und Winne (1995), Korrekturgrößen oder -maßnahmen spezifiziert (sogenannte *Stellgrößen*), um eine Veränderung in Richtung des Soll-Wertes herbeizuführen. Nimmt man also das beobachtbare Verhalten in Form der Performance in einer Aufgabe der Lernenden, so sind erfahrenere Lernende in der Lage, ihre Lösung eigenständig zu bewerten und gegebenenfalls auch zu korrigieren (internes Feedback). Wenn das interne Feedback kompetenzbedingt aber nicht ausreicht, um Selbstregulationsprozesse in Gang zu setzen, ist externe Unterstützung geboten. Auf Basis dieser externen Ist-Wert-Soll-Wert-Diskrepanz kann dann ein entsprechendes Feedback formuliert werden. Externes Feedback, das auf Basis von Fehlern Defizite aufzeigt, muss darauf aufbauend adaptive Hilfestellungen geben, die durch aktive Wissenskonstruktion Hilfe zur Selbsthilfe leisten. Denn das Ziel ist es, dass die Lernenden langfristig in die Lage versetzt werden, mittels internen Feedbacks ihre Performance selbst zu überwachen. Diese Korrekturen werden dann von einer korrigierenden Instanz (dem sogenannten *Stellglied*) in der lernenden Person bewertet und bestenfalls umgesetzt. Dabei wird ein neuer Ist-Zustand generiert, und der Loop beginnt erneut. Soll überprüft werden, ob das vorherige Feedback zur Verbesserung geführt hat, muss ein neuer Ist-Zustand erhoben werden. Folglich bedarf es eines multiple-try Feedbacks (Narciss, 2018). Für die Passung von Feedback zum Individuum spielt es daher eine große Rolle, wie detailliert und wie oft der Ist-Zustand erhoben wird. Feedback Preavailability darf nicht gegeben sein, denn die Performance eines Individuums zu adressieren, setzt mindestens eine vom Feedback unbeeinflusste Performance in einer Situation voraus, z.B. einen ersten Antwortversuch in einer Lernaufgabe. Soll das Feedback konkrete Hilfestellung geben, bieten ausschließlich ergebnisbezogene Feedbackarten (*knowledge of result, knowledge of performance*; vgl. Tabelle 1) nur wenig Unterstützung für die Wissenskonstruktion an, da sie relativ wenige Anhaltspunkte für Handlungs- bzw. Wissensoptimierung geben.

Feedback sollte folglich nicht nur evaluative Funktionen erfüllen, sondern vielmehr wie ein Tutor beziehungsweise eine Tutorin strategisch hilfreiche Informationen zur Überwindung von Hürden im Lernprozess liefern. Ein so beschaffenes Feedback wird als *informatives tutorielles Feedback* (ITF) bezeichnet. Dieses Feedback macht es sich zum Ziel, regulierend auf Lernprozesse zu wirken und Hilfestellung zu geben, sodass Lernende Wissens (strukturen) beziehungsweise Kompetenzen erwerben, um die Anforderungen der Lernsituation zu meistern, ohne dass ihnen dabei zuvor die Aufgabenlösung preisgegeben wurde (Narciss, 2006, 2014, 2018). Dazu bedient es sich je nach Zielsetzung verschiedener elaborierter Feedback-Typen (s. Tabelle 1).

Entlang verschiedener Metastudien (Bangert-Drowns et al., 1991; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Wisniewski et al., 2019) zeigen sich allerdings große Schwankungen in den Effektstärken hinsichtlich der Effektivität von Feedback-Typen im Allgemeinen. Aufgrund großer Varianz hinsichtlich der Effektstärken in Verbindung mit unterschiedlichen Forschungsdesigns kann die Feedbackforschung zurzeit auch keine gesicherten Erkenntnisse hinsichtlich des Vorwissenseinflusses auf die Wirksamkeit von Feedback unterschiedlichen Umfangs liefern (Fyfe et al., 2012; Narciss, 2006; Smits et al., 2008). Fyfe et al.

(2012) konnten dabei hinsichtlich des Erlernens von Äquivalenzrechnungen feststellen, dass bei steigendem Vorwissen der Vorteil elaborierter Hilfestellung zurückgeht und das eigenständige Finden von Fehlern zu höherem Lernerfolg führt. Bei Narciss (2006) wurde ebenfalls im mathematischen Kontext ein Vergleich zwischen rein korrekktivem und elaboriertem Feedback, hier explizit ITF, angestrebt. Hier konnten keine Treatment-Unterschiede unter Berücksichtigung des Vorwissens gefunden werden. Smits et al. (2008) konnten feststellen, dass rein evaluatives Feedback für Vorwissensstarke im Rahmen von Lerneinheiten zur Genetik zu besseren Leistungen führte. Bezüglich Studien, die den Zusammenhang von Vorwissen und Feedback explizit untersucht haben, zeigt sich also eine unklare Ergebnislage. Da für chemienahe Kontexte ebenfalls nur Indizien für einen solchen Zusammenhang gefunden werden konnten (Albacete & VanLehn, 2000; Moreno, 2004; Narciss & Huth, 2004), zeigt sich bereits für die domänenspezifische Feedbackforschung eine explizite Forschungslücke.

Tabelle 1: Klassifikation unterschiedlicher Feedback-Typen nach inhaltlichen Gesichtspunkten (Narciss, 2006, S. 23)

Bezeichnung ^a	Beispiele für Feedback-Inhalte	
Knowledge of Performance [KP]	<ul style="list-style-type: none"> • 15 von 20 Aufgaben richtig • 85% der Aufgaben gelöst 	
Knowledge of result/response [KR]	<ul style="list-style-type: none"> • Falsch/richtig • Nicht richtig/stimmt 	
Knowledge of correct response [KCR]	<ul style="list-style-type: none"> • Angabe der korrekten Antwort • Markierung der korrekten Antwort 	
Elaborierte Feedback-Typen	Knowledge on task constraints [KTC]	<ul style="list-style-type: none"> • Hinweise auf Art der Aufgabe • Hinweise auf Bearbeitungsregeln • Hinweise auf Teilaufgaben • Hinweise auf Aufgabenanforderungen
	Knowledge about mistakes [KM]	<ul style="list-style-type: none"> • Anzahl der Fehler • Ort der Fehler/ des Fehlers • Art der Fehler/ des Fehlers • Ursache/n des/r Fehler(s)
	Knowledge on how to proceed [KH] („know how“)	<ul style="list-style-type: none"> • Fehlerspezifische Korrekturhinweise • Aufgabenspezifische Lösungshinweise • Hinweise auf Lösungsstrategien • Leitfragen • Lösungsbeispiele
	Knowledge on meta-cognition [KMC]	<ul style="list-style-type: none"> • Hinweise auf meta-kognitive Strategien • Metakognitive Leitfragen

^a Bei der Wahl der Bezeichnungen wurde einerseits darauf geachtet, die in der Feedback-Forschung üblichen Bezeichnungen zu verwenden beziehungsweise weiter zu führen, andererseits sollte die inhaltliche Feedback-Facetten in der Bezeichnung deutlich werden.

Spezifisch für automatisiertes Feedback in Online-Lernumgebungen erstreckt sich hingegen über den gesamten Bereich der Naturwissenschaften eine eklatante Forschungslücke: Cavalcanti et al. (2021) betrachten beispielsweise in ihrem Literatur-Review 3,178 Online-Artikel, die über ein automatisiertes Suchverfahren anhand von Schlagwortgruppen erhalten wurden (u.a. von SpringerLink, Scopus und ScienceDirect). Durch ein anschließendes dreistufiges Review-Verfahren konnten daraus 63 Artikel klassifiziert werden, die sich hinreichend mit dem Thema automatisierten Feedbacks in Online-Lernumgebungen befassen. Mathematik war dabei mit lediglich drei Artikeln der einzige naturwissenschaftsnaher Vertreter. Über alle

Artikel konnte allerdings unabhängig von der Fachrichtung bei 32 Artikeln ein empirischer Nachweis für positive Effekte von automatisiertem Feedback auf die Performance von Nutzerinnen und Nutzern gefunden werden, wodurch sich tendenziell ein Trend ergibt, den es in chemischen Kontexten zu replizieren gilt. Auf den Einfluss von Vorwissen auf die Wirksamkeit von Feedback wurde in dieser Studie nicht eingegangen. Jenseits des Reviews liegen den technischen Stand betreffend auch nur wenige Anwendungsbeispiele für automatisiertes Feedback im Bereich chemiespezifischer Aufgaben vor. Diese beschränken sich auf kostenpflichtige Programme (z.B. Mastering Chemistry, ALEKS, Achieve), die überwiegend in den Vereinigten Staaten und Kanada genutzt werden. Weitere individuelle Online-Lösungen weisen hingegen funktionale Defizite auf, da sie z.B. Reaktionsgleichungen in der Chemie als offenen Aufgabentyp nicht berücksichtigen (für einen Überblick vgl. Eitemüller et al., 2023).

3 Fragestellung

Über einen systematischen Vergleich von zwei vom Informationsgehalt her unterschiedlichen Feedback-Typen, eingebettet in einen multiple-try Feedback Algorithmus, sollen weitere Erkenntnisse über den Einfluss des Vorwissens der Lernenden auf die lernförderliche Wirkung von Feedback-Typen gewonnen werden.

FF: Wie wirken sich zwei vom Feedbackumfang unterschiedliche Feedback Algorithmen (Error-Specific-Tutoring Feedback vs. korrekatives Feedback) in Abhängigkeit vom fachspezifischen Vorwissen auf den Studienerfolg von Erstsemesterstudierenden in der Allgemeinen Chemie aus?

Als Maß für den Studienerfolg wird in dieser Arbeit das Fachwissen (Post) – geprüft durch einen unabhängigen Fachwissenstest – sowie der Klausurerfolg am Ende des Semesters verstanden. Der Error-Specific-Tutoring Feedback Algorithmus (EST-Feedback) zeichnet sich durch verschiedene ITF-Typen aus. Diese scheinen besonders vielversprechend für Studierende mit Vorwissensdefiziten zu sein. Denn besonders diese Studierenden benötigen beim Auftreten von Fehlern elaboriertes Feedback (Krause et al., 2009; Mayer & Johnson, 2010; Moreno, 2004). So erfordert das selbstständige Finden von Rechenfehlern und/oder fehlerhaften Ansätzen im Lösungsprozess bereits ein gewisses Grundmaß an Vorwissen (Schuhmacher, 2008). Daraus ergibt sich folgende erste Hypothese:

H1: Studierende mit geringem fachspezifischem Vorwissen profitieren in ihrem Studienerfolg stärker durch ein Training mit einem EST-Feedback als Studierende mit hohem Vorwissen.

Bei zunehmender Expertise ist allerdings zu erwarten, dass der Nutzen von elaboriertem Feedback sinkt, sodass sich für Fortgeschrittene und Experten kein Vorteil mehr ergibt (Expertise Reversal Effect; Kalyuga, 2014) oder sogar Nachteile durch zu umfangreiches Feedback entstehen können (z.B. Smits et al., 2008). Ein korrekativer Feedback Algorithmus (KOR-Feedback), der nur auf eine fehlerhafte Lösung hinweist, scheint damit für Fortgeschrittene die bessere Wahl zu sein.

H2: Studierende mit hohem fachspezifischem Vorwissen profitieren in ihrem Studienerfolg stärker durch ein Training mit einem KOR-Feedback als Studierende mit geringem Vorwissen.

4 Material

4.1 Der Feedback Algorithmus

Wie bereits in Abschnitt 2 beschrieben, wird im Rahmen des ITFL-Modells eine vorherige Studierendenleistung benötigt, um fehlerspezifisches Feedback geben zu können. Das EST-Feedback soll aber zusätzlich zu einer fehlerspezifischen Rückmeldung ITF bereitstellen. Dieses soll passend zu bestimmten Fehlerkonstellationen automatisiert eine Hilfestellung enthalten, die den Lernenden helfen soll, die Aufgabe selbst zu vollenden, aber ohne die Lösung der Aufgabe zu präsentieren. Somit bedarf es einer Aufgabe, die einem multiple-try Schema mit mindestens zwei möglichen Antwortversuchen folgt. In Anlehnung an den Bug-Related-Tutoring Feedback Algorithmus (Narciss & Huth, 2004) wurde ein multiple-try Algorithmus mit drei möglichen Antwortversuchen entwickelt, der bei Lernaufgaben im ersten und zweiten Feedback neben korrektivem Feedback (KR) noch elaborierte Feedback-Typen in Form von ITF präsentiert (KM, KH; s. Tabelle 1). Das letzte Feedback enthielt neben dem Indizieren von spezifischen Fehlern (KM) zusätzlich die richtige Lösung (KCR). Die richtige Lösung wurde dabei feingliedrig mit Teilschritten (KH) präsentiert. Parallel zu diesem EST-Feedback Algorithmus wurde ein dreistufiger korrektiver Feedback Algorithmus (KOR-Feedback) entwickelt, der in der ersten und zweiten Rückmeldung nur KR und in der letzten Feedback Nachricht KCR präsentierte.

Durch zwei korrekte Antworten im ersten Lösungsversuch oder durch sechs Aufgaben mit beliebiger Performance gelangten die Studierenden zu einer Testaufgabe, mit der anteilig Bonuspunkte für die Klausur gesammelt werden konnten. Daraus resultierte nicht nur adaptives Feedback, sondern auch eine adaptive Aufgabenmenge. Zur Beantwortung der Forschungsfrage wurden zu zwanzig Lernzielen basierend auf diesem Schema insgesamt 120 Online-Aufgaben mit adaptivem fehlerspezifischem Feedback (EST-Feedback) beziehungsweise korrektivem Feedback (KOR-Feedback) in der Webapplikation JACK[®] (Pobel & Striewe, 2019; Striewe, 2016) realisiert (sechs je Lernziel). Die Testaufgabe am Ende diente zur Lernzielüberprüfung. Da sich diese interaktiven Aufgaben als externes Tool in die Lernplattform Moodle einbetten lassen, eigneten sie sich sehr gut für einen webbasierten Übungsbetrieb.

4.2 Lernzielbasierte Digitalisierung von Chemieaufgaben

Die Auswahl spezifischer Lernziele des gegebenen domänenspezifischen Wissens als anzustrebende Kompetenz sowie typischer Aufgaben, die diese kompetenzbildenden Lernziele bedienen, sind laut Narciss und Huth (2004) grundlegend für die Spezifizierung von Feedbackelementen. Chemie ist für viele ohne mathematische Formeln, Reaktionsgleichungen, Diagramme usw. überhaupt nicht denkbar.

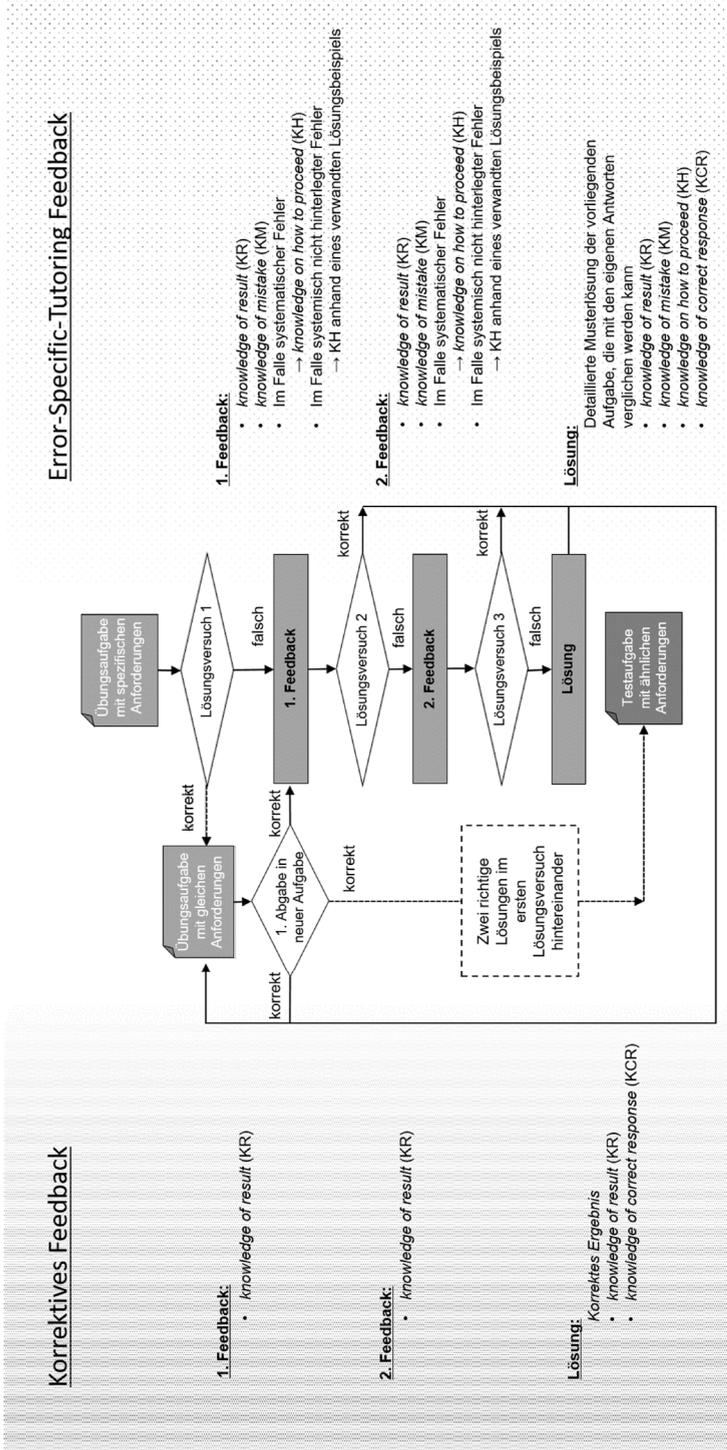


Abbildung 2: Gegenüberstellung des korrekativen Feedback Algorithmus (KOR-Feedback) mit dem Error-Specific-Tutoring-Feedback Algorithmus (EST-Feedback)

Eine Analyse von Lehrbüchern, die Studierende in der Eingangsphase des Chemiestudiums häufig zum Lernen nutzen, konnte zeigen, dass wenigstens 87% aller Lehrbuchseiten mindestens eine Visualisierung enthalten. Mit wenigstens 60% Verteilungsanteil stellen symbolische Repräsentationsformen dabei den Hauptanteil. Im Mittel sind 34% davon Reaktionsgleichungen (vgl. Dickmann, 2019). All diese Repräsentationsformen in domänenspezifischer Schreibweise müssen durch die Software adäquat abgebildet werden. Klassische Übungs- und Klausuraufgaben repräsentieren in diesem Zusammenhang die intendierten Lernziele. Eine Analyse von Übungsaufgaben, Klausuraufgaben und weiteren Modulinhalt nach Lernzielen lieferte daher im Rahmen dieser Arbeit entsprechende Operatoren und damit die nötigen Anforderungen an das Software-Tool. Parallel dazu wurden typische Fehlerquellen ermittelt und bei der Konzeption des Feedbacks berücksichtigt, wodurch sich hinsichtlich der automatisierten Fehlererkennung weitere Anforderungen an das Software-Tool ergaben. Dieser Prozess ist in Abbildung 3 verdeutlicht.

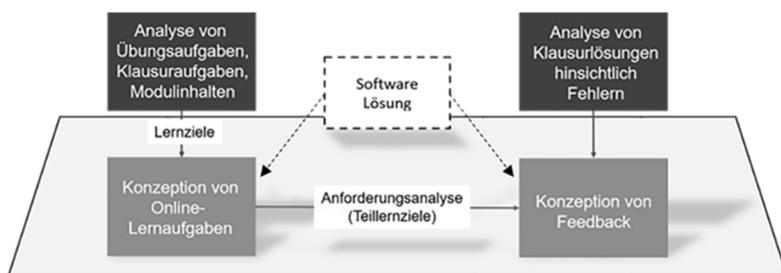


Abbildung 3: Überführung der Papier-Bleistift-Aufgaben in digitale Lernaufgaben unter Berücksichtigung der Einschränkungen durch die Software Lösung

Bei der Überführung von klassischen Paper-Pencil Aufgaben in ein Online-Format ist zudem zu vermuten, dass sich durch diesen Transfer die Anforderungen der Aufgaben verändern würden. Für chemische Kontexte konnte ein Effekt des Mediums allerdings widerlegt werden, solange die Lernziele zwischen analogen und digitalen Lernaufgaben konstant gehalten wurden. Einschränkungen konnten eindeutig nicht in der Aufgabenstellung, sondern bei der Eingabe der Lösung verortet werden (Ashton et al., 2005), z.B. wenn gewohnte Schreibweisen nicht möglich waren. Gesucht wurde daher eine Softwarelösung, die zugleich die Möglichkeit für eine chemiespezifische Schreibweise bot und zugleich eine automatisierte synchrone Evaluation von Antworten, unter anderem in dieser spezifischen Schreibweise, ermöglichte. In Zusammenarbeit mit dem Moodle-Kompetenzzentrum der Universität Duisburg-Essen sowie im interdisziplinären Austausch mit Doktorierenden und Professor/innen der Informatik, rückte das vom Ruhr Institute for Software Technology entwickelte Java-basierte E-Assessment System „JACK® – Lernen. Testen. Prüfen.“ in den Fokus.¹

Während Freitext und Zeichenaufgaben aufgrund von hochkomplexen Evaluationsstrukturen für fehlerspezifisches Feedback nicht erschlossen werden konnten, ist regelbasiertes Feedback zu verschiedensten mathematischen Aufgabentypen, die einen hohen Stellenwert in der Allgemeinen Chemie haben (Goldhausen, 2015), einfach zu realisieren. Zudem wurde

1 Sowohl formative als auch summative Assessments waren hiermit zu realisieren. Jede Einreichung einer Aufgabe wird dabei vollständig unabhängig von einer anderen getestet. Bis 2015 wurde JACK® im Informatik-kontext an 4 Universitäten in Deutschland und England eingesetzt und erhielt dabei sehr viel positiven Zuspruch seitens der Studierendenschaft (Striwe, 2016).

ein Editor für chemische Summenformeln und Reaktionsgleichungen (Eitemüller et al., 2023; Pobel & Striewe, 2019), sowie ein Eingabetool für Atomorbitaldiagramme (Striewe et al., 2020) fertiggestellt. So ergab sich für die domänenexklusiven Schreibweisen der Chemie mit ihrer spezifischen Semantik die nötige Fehlersensitivität für passgenaues Feedback.

5 Studiendesign, Stichprobe und Testinstrumente

5.1 Studiendesign und Stichprobe

Aufgrund der oben genannten Schlüsselrolle des Moduls der Allgemeinen Chemie wurden die Aufgaben in einem Online-Tutorium inhaltlich parallel zu den Vorlesungsinhalten arrangiert und hinsichtlich ihrer Lernwirksamkeit getestet. Als Teilstichprobe der Grundgesamtheit der Chemiestudierenden an deutschen Universitäten wurden die Erstsemesterstudierenden der Universität Duisburg-Essen mit der Fachwahl Bachelor Chemie gewählt. Aufgrund der inhaltlichen Parallelität und des gemeinsamen Besuchs der Vorlesungen in den ersten Semestern wurden Studierende des Studiengangs Bachelor Water Science mit hinzugezogen.

Die Einteilung der Studierenden in die Übungsgruppen hatte bei dieser Feldstudie nach der ersten Veranstaltung schnell innerhalb der ersten Semesterwoche zu erfolgen. In der Kürze der Zeit ist eine Auswertung von Fachwissensdaten nicht möglich gewesen. Daher wurden die Studierenden über die Kurswahl in der Oberstufe (Chemie Leistungskurs: ja/nein), die als guter Prädiktor für das Vorwissen gilt (Averbeck, 2020), gleichmäßig und randomisiert auf die Interventionsgruppen aufgeteilt (2×2 -Design, Kurswahl \times Feedback-Typ, vgl. Tabelle 2). Eine Poweranalyse mit G*Power (Faul et al., 2007) ermittelt zum Nachweis mittlerer Effektstärken ($f^2 = 0.15$, $\alpha = 0.05$, $Power = 0.80$) inklusive Korrektur für hierarchische Datenstrukturen eine notwendige Stichprobe von 42 Probanden pro Kategorie (LK/kein LK), also eine minimale Gesamtstichprobengröße von $N = 84$ Probanden. Da erfahrungsgemäß aufgrund des Dropouts über die Semester und des damit verbundenen Schwundes an Probanden damit zu rechnen war, dass die Gesamtstichprobengröße über ein Semester nicht ausreichend für die anschließende Auswertung sein könnte, wurden insgesamt drei Interventionsstudien über die Wintersemester (WiSe) 2018–2020 durchgeführt.

Tabelle 2: Interventionsgruppen im 2×2 -Design

	Feedback-Typ	
Kurswahl Chemie	EST	KOR
Kein Leistungskurs	EST _(kein LK)	KOR _(kein LK)
Leistungskurs	EST _(LK)	KOR _(LK)

Für die drei Interventionsjahrgänge konnten in der ersten Semesterwoche auf Basis der Kurswahl folgende Teilstichproben erhalten werden (s. Tabelle 3, linke Seite). Die Bearbeitung der Übungsaufgaben ist mit Bonuspunkten für die Klausur und des Weiteren mit einer Probandenvergütung von 50 € pro Person belohnt worden. Dazu mussten die Studierenden mindestens zwei Aufgaben zu einem Lernziel in Folge richtig lösen oder alle sechs Aufgaben eines Aufgabenpakets sowie eine abschließende Testaufgabe bearbeiten (s. Abbildung 2). Da zudem nur sehr wenige Studierende alle verfügbaren Aufgabensets im Sinne des Algorithmus

vollständig bearbeitet haben, wurde bereits im Rahmen der Pilotierungsstudie, in der sich dieses Verhalten bereits abzeichnete, die aktive Arbeit mit den Übungsaufgaben an die Regel geknüpft, dass die Hälfte der zehn in der Pilotierung verfügbaren Aufgabensets mit je zwei Lernaufgaben bearbeitet worden sein musste. Die verfügbaren Aufgabensets wurden dann sukzessiv auf 20 Lernziele erweitert. Um am Ende der Studie einen Effekt auf den Studien-erfolg der Studierenden untersuchen zu können, wurde diese Regel für die aktive Übungs-teilnahme auch für die Folgejahre beibehalten. So ergibt sich folgende Verteilung der „akti-ven“ Gesamtkohorte für die Evaluation des Fachwissenszuwachs (s. Tabelle 3, rechte Seite). Um tieferegehende Informationen über die Wirkweise der Lernaufgaben in Abhängigkeit von dem unterstützten Feedback-Typ gewinnen zu können, konnte zusätzlich auf eine Kontroll-ko-horte aus dem WiSe 2016 zurückgegriffen werden. Diese besuchte bei identischen Vorle-sungsinhalten *nur* eine reine Präsenzübung ohne interaktive Lernaufgaben.

Tabelle 3: Verteilung der „aktiv“ Studierenden pro Interventionsgruppe

DEMO	Interventionsgruppe		Σ	„aktiv“	Interventionsgruppe		Σ
	KOR	EST			KOR	EST	
Kein Leistungskurs	90	96	186	→ Kein Leistungskurs Leistungskurs Σ	28	33	61
Leistungskurs	50	55	105		30	31	61
Σ	140	151	291		58	64	122

Mithilfe eines Pre-Post-Vergleichsgruppendesigns sollten die Aufgaben hinsichtlich ihrer Lernwirksamkeit im Rahmen des folgenden Designs geprüft werden (s. Tabelle 4). Averbek (2020) konnte neben dem Chemie-Vorwissen noch weitere Einflussfaktoren (kognitive Fähigkeiten, mathematische Fähigkeiten, Studien- und Lern- und Leistungsmotivation) auf den Studienerfolg identifizieren.

Tabelle 4: Studiendesign für die Interventionsstudie

Semesterverlauf	EST _(kein LK) + EST _(LK)	KOR _(kein LK) + KOR _(LK)
Erste Semesterwoche		Demografie-Fragebogen Allgemeine Chemie (Pre) Kognitiver Fähigkeitentest Test für mathematische Fähigkeiten Tutorial für den Umgang mit JACK Studienmotivation (Erwartung, Wert, Kosten) Allgemeine Lern- und Leistungsmotivation im Studium
Semesterbegleitende Intervention		<i>Arbeiten mit den Lernaufgaben</i> empfundene Aufgabenschwierigkeit/ Denkanstrengung/Nützlichkeit des Feedbacks/ Nutzerfreundlichkeit der Aufgaben, time-on-task
Letzte Semesterwoche		Allgemeine Chemie (Post)
Vorlesungsfreie Zeit		Klausur/Nachschreibklausur

Diese wurden als Kontrollvariablen vor der Intervention miterhoben. Um den Arbeitsaufwand, der sich durch die Studie für die Testpersonen ergab, angemessen darstellen zu können, werden diese hier mit dargestellt, im Umfang dieses Papers aber nicht näher behandelt.

Die Pilotierung der interaktiven, durch Feedback gestützten Lernaufgaben wurde im selben Design wie die Hauptstudie durchgeführt und unterscheidet sich lediglich durch die Menge an verfügbaren JACK-Aufgaben von der Hauptstudie. Bedingt durch die COVID-19 Pandemie ergaben sich nur für die Kohorte des WiSe 2020/2021 (HS 2) Einschränkungen. Da dieses Semester an der Universität als reines Online-Semesters durchgeführt und zudem bedingt durch die Maßnahmen gegen das Virus verspätet in das Semester gestartet wurde, musste auf einige Testinstrumente verzichtet werden (mathematische Fähigkeiten, Studienmotivation, Lern- und Leistungsmotivation, Selbstkonzept und Lernstrategien).

Tabelle 5: Stichprobenbeschreibung

	Demografie			+	Allg. Chemie (Pre)			+	Allg. Chemie (Post)			+	Aktive Übungsteilnahme		
	N	%	%(\varnothing) ¹		N	%	%(\varnothing) ¹		N	%	%(\varnothing) ¹		N	%	%(\varnothing) ¹
Ges	291	100 ²	42.6		234	80.4 ²	42.7		149	51.2 ²	36.9		122	41.9 ²	40.2
Pil	84	100 ³	39.3		77	91.7 ³	39.0		54	64.3 ³	35.2		39	46.4 ³	43.6
HS1	92	100 ³	41.3		83	90.2 ³	42.2		58	63.0 ³	25.7		53	60.4 ³	24.5
HS2	115	100 ³	44.3		74	64.3 ³	47.3		37	32.2 ³	56.8		30	26.1 ³	63.3

Anmerkungen: ¹ Anteil weiblicher Studierender bezogen auf den jeweiligen MZP; ² bezogen auf die Gesamtstichprobe (Studierende, die den Demografie-Fragebogen bearbeitet haben); ³ bezogen auf die Teilstichprobe je WiSe (Studierende, die den Demografie-Fragebogen bearbeitet haben)

Über die drei Interventionsstudien wurde eine Gesamtkohorte von $N = 291$ Studierenden² ($M_{\text{Alter}} = 20.72$ Jahre, 57.4% männlich, 42.6 % weiblich) aus Erstsemesterstudierenden der Studiengänge Bachelor Chemie und Bachelor Water Science erhalten (s. Tabelle 5).

5.2 Testinstrumente

Damit ein Vergleich zwischen den Jahrgängen hinsichtlich der Leistungsdaten möglich wurde (allgemeine Chemie Fachwissen, mathematische Fähigkeiten, kognitive Fähigkeiten), wurden die Leistungsdaten mithilfe der Item Response Theorie (Rasch Modelle) ausgewertet. Durch Fixierung der Itemschwierigkeiten lassen sich auch sukzessiv entstandene Stichproben auf einer gemeinsamen Skala abbilden. Denn anders als die klassische Testtheorie, die von homogenen Itemschwierigkeiten ausgeht, ermittelt man in der IRT sowohl die Personenfähigkeit (θ) als auch die Itemschwierigkeit (σ) und setzt sie für jedes Item in Bezug zueinander, sodass sich letztlich beides auf einer eindimensionalen Logit-Skala einordnen lässt. Wenn der Personenparameter (θ) einem Itemparameter (σ) entspricht, beträgt dabei die Lösungswahrscheinlichkeit eines Items 50%. Da sich die Itemschwierigkeiten zwischen unterschiedlichen Stichproben fixieren (also gleichsetzen) lassen, liegen die in diesem Modell geschätzten Personenfähigkeiten (θ) auf derselben Skala und lassen sich so gleichwertig verwenden und vergleichen. Der WLE-Schätzer gilt dabei als geeignetster Schätzer für Personenfähigkeiten im Rahmen von IRT-Modellen (Walter, 2005). Im Folgenden wird daher nicht mit z-standardisierten Personenfähigkeiten, sondern mit den WLEs (Weighted Maxi-

2 Die Stichprobe hat einen Anteil von 17.6% Probanden, die nicht in Deutschland geboren wurden. Dabei haben 40.2% einen Migrationshintergrund, d.h., dass mindestens ein Elternteil nicht in Deutschland geboren wurde. Bei 29.1% der Probanden ist es bei keinem Elternteil der Fall. Insgesamt gaben 69.8% der Probanden die deutsche Sprache als Muttersprache an.

imum Likelihood Estimates; Warm, 1989), also Personenfähigkeitsparametern auf einer Logit-Skala gerechnet.

Die Itemschwierigkeiten wurden im Rahmen des interdisziplinären Projektes ALSTER (Akademisches Lernen und Studienerfolg in der Eingangsphase von naturwissenschaftlich-technischen Studiengängen) auf den Pre-Messzeitpunkt der Kontrollkohorte fixiert, da hier die größte Stichprobe für eine valide Schätzung gegeben war (Averbeck, 2020). Die Erhebung des Fachwissens in der Allgemeinen Chemie bildet im Rahmen dieses Prä-Post-Vergleichsgruppendesigns das zentrale Testinstrument zur Erfassung des Studienerfolgs und liegt somit im Fokus dieses Papers. Hierzu wurde auf den Fachwissenstest von Freyer (2013), modifiziert nach Averbeck (2020), zurückgegriffen. Er liefert eine akzeptable bis gute EAP/PV-Reliabilität von .765 zum ersten und von .802 zum zweiten Messzeitpunkt. Der Item-Fit aller 35 Items (Range: 0.76–1.275) liegt zudem innerhalb der von Bond et al. (2020) empfohlenen Spanne für gute Infit- und Outfit-Parameter für Item Mittelwertquadrate (Range: 0.7–1.3; S. 244). Für das allgemeine Chemie-Fachwissen liegt die WLE Reliabilität bei .752 für den ersten beziehungsweise .789 für den Post-Messzeitpunkt am Ende des ersten Semesters. Für die mathematischen Fähigkeiten zu Beginn des ersten Semesters zeigt sich mit .842 ebenfalls eine gute WLE Reliabilität über KG bis hin zur HS1.

Tabelle 6: Testinstrumente zur Erfassung von Kontrollvariablen

Skala	Quelle	N _{Items}	EAP/PV
Kognitive Fähigkeiten	Heller & Perleth (2000)	25	.77
Mathematische Fähigkeiten	Müller et al. (2018)	23	.86

6 Ergebnisse

6.1 Leistungsunterschiede im Fachwissen zwischen Interventionsgruppen und Kontrollgruppe

Über einen Vergleich mit der Kontrollkohorte wurde zunächst untersucht, ob ein Training mit EST- bzw. KOR-Feedback (in Abhängigkeit vom Vorwissen) zu signifikant besseren Leistungen im Fachwissen am Ende des ersten Semesters führt. Hierzu wurden die Daten des unabhängigen Fachwissenstests analysiert, die sowohl für alle drei Interventionsjahrgänge als auch für die Kontrollkohorte zur Verfügung standen. Der Einfluss der Interventionsbedingungen, bei sonst gleicher Lehre und gleichen Dozierenden sowie vergleichbarer Abiturnote ($F(1,1216) = 1,243, p = .265$), auf den Fachwissensstand am Ende des ersten Semesters wurde mithilfe einer multiplen linearen Regression untersucht, hierbei zeigen sich signifikante Effekte ($F(4, 224) = 106.8, p \leq .001$). Neben dem Fachwissen Chemie zu Beginn des Semesters und der Kurswahl in der Oberstufe wurden in einem ersten Modell zwei weitere Variablen hinzugezogen. Durch diese zwei Dummy-Variablen EST-Feedback (0 = nein, 1 = ja) und KOR-Feedback (0 = nein, 1 = ja) wird der Vergleich beider Feedback-Typen mit der Kontrollkohorte ermöglicht. Das Fachwissen zu Beginn des Semesters ($\beta = .740, p \leq .001$) und die Kurswahl ($\beta = .103, p \leq .05$) bieten erwartungskonform gute prädiktive Kraft. Zusätzlich trägt der Prädiktor, ob EST-Feedback erhalten wurde ($\beta = .107, p \leq .05$), signifikant zur Varianzaufklärung bei (vgl. Modell 1 in Tabelle 7). Der Erhalt von KOR-

Feedback ist hingegen unerheblich für den Fachwissensstand am Ende des Semesters (s. Modell 1, Tabelle 7). Ob die Studierenden unabhängig von ihrer Kurswahl vom EST-Feedback profitieren, wurde in einem weiteren Modell überprüft.

Tabelle 7: Hierarchische lineare Regressionen zur Vorhersage des Fachwissens (Post)

Modell	Modell 1			Modell 2			Modell 3		
	B	β	SE	B	β	SE	B	β	SE
Konstante	.453 ***		.076	.410 ***		.083	.408 ***		.078
Allg. Ch.	.930 ***	.734 ***	.055	.920 ***	.725 ***	.055	.920 ***	.730 ***	.055
Fachw. (pre)									
Kurswahl	.261 *	.107 *	.105	.408 *	.167 *	.158	.414 **	.169 **	.125
EST-Feedback	.285 *	.107 *	.116	.489 **	.184 **	.151	.491 **	.184 **	.148
KOR-Feedback	.165	.060	.120	.129	.047	.159	.135	.049	.120
Kurswahl × EST-Feedback				-.469 *	-.134 *	.235	-.474 *	-.136 *	.214
Kurswahl × KOR-Feedback				.015	.004	.242			
korr. R ²	.646			.650			.656		
ΔR^2				.004			.01 *		

Anmerkungen: *** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$

Ein zweites Modell (vgl. Tabelle 7), welches die Interaktion aus Feedback und Kurswahl in der Oberstufe berücksichtigt, zeigt allerdings einen signifikanten negativen Effekt für die Interaktion aus EST-Feedback und Kurswahl ($\beta = -.134$, $p \leq .05$). Da dieses Modell nicht signifikant besser misst als Modell 1 ($\Delta R^2 = .004$, $p = .089$) und der Interaktionseffekt von KOR-Feedback und Kurswahl in der Oberstufe mit $\beta = .004$ ($p = .170$) vernachlässigbar gering ausfällt, wurde ein weiteres Modell mit einem Interaktionsterm aus EST-Feedback und der Chemie-Kurswahl berechnet (vgl. Modell 3 in Tabelle 7). Es zeigt sich eine signifikante Verbesserung der Varianzaufklärung im Vergleich zu Modell 1 bei sonst gleichbleibenden Wirkzusammenhängen ($\Delta R^2 = .01$, $p \leq .001$). Der Vorteil von EST-Feedback ($\beta = .184$, $p \leq .01$) wird durch den Interaktionsterm nahezu aufgehoben ($\beta = -.136$, $p \leq .05$).

Dass sich nur ein Vorteil für diejenigen Studierenden durch EST-Feedback ergibt, die keinen Chemie-Leistungskurs in der Oberstufe besucht haben, zeigt sich auch rein deskriptiv in Abbildung 4. Studierende ohne Leistungskurs, die aber EST-Feedback bekommen haben (EST_(kein LK)), haben ein vergleichbares Fachwissen ($M = 1.18$, $SD = 1.16$) wie Studierende der Kontrollgruppe (KG_(LK)), die einen Chemie-Leistungskurs besucht haben ($M = 1.28$, $SD = 1.23$).

Zu Semesterbeginn lässt sich kein Unterschied zwischen den Subgruppen mit gleicher Kurswahl hinsichtlich des Vorwissens in der allgemeinen Chemie feststellen (ANOVA_(kein LK): $F(2, 112) = 1.152$, $p = .320$; ANOVA_(LK): $F(2, 73) = 0.127$, $p = .881$). Die entsprechende ANOVA, die den Fachwissensstand zum Ende des Semesters zwischen den Interventionsgruppen vergleicht, zeigt im post-hoc Test (Games-Howell) keinen signifikanten Unterschied zwischen der Interventionsgruppe ohne Leistungskurs mit EST-Feedback zu allen anderen Subgruppen, die einen Leistungskurs in der Oberstufe belegt hatten (EST_(LK), KOR_(LK) und KG_(LK)) ($F(5, 223) = 11.26$, $p \leq .001$). Die Gruppe EST_(kein LK) ist zudem signifikant besser ($p \leq .01$) als die Kontrollgruppe ohne Leistungskurs ($M = .30$, $SD = 1.23$) (vgl. Abbildung 4). Im Vergleich zur Gruppe KOR_(kein LK) ($M = 0.45$, $SD = 0.92$) ist zumindest ein möglicher Trend zugunsten der Gruppe EST_(kein LK) zu erkennen ($p = .082$).

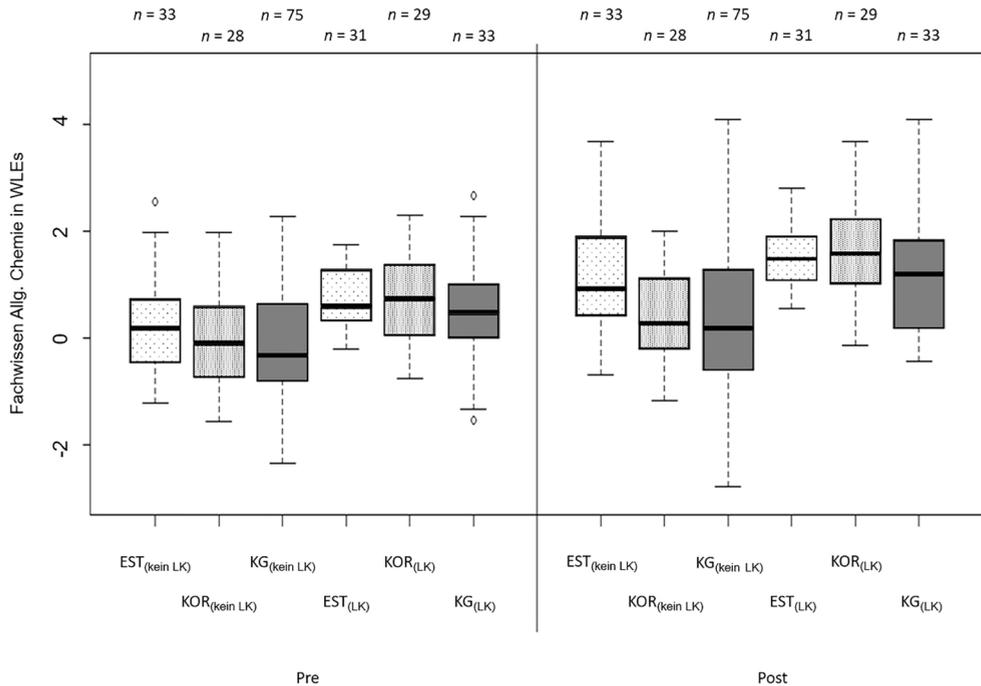


Abbildung 4: Entwicklung des Chemie Fachwissens nach Interventionsgruppe (Kohortenvergleich)

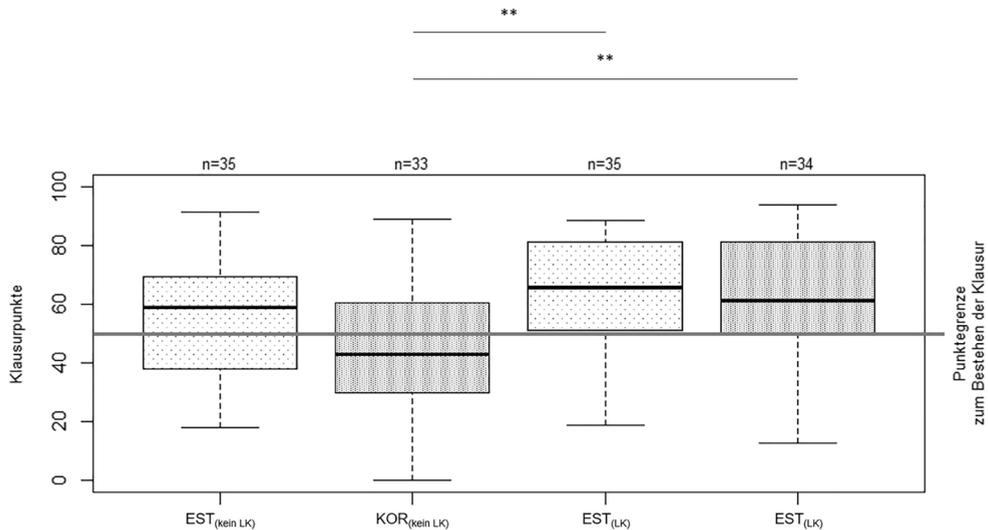
6.2 Unterschiede in den Klausurleistungen nach Interventionsgruppe

Während die Klausuren über die Pilotierung und Hauptstudie 1 in Absprache mit dem Dozenten am ersten und zweiten Versuchstermin identisch gehalten werden konnten, musste aufgrund der Pandemie im WiSe 2020 das Klausurformat kurzfristig geändert werden. Daher sind die Ergebnisse zum Klausurerfolg beschränkt auf die WiSe 2018 und 2019 ($N_{Post} = 93$). Aufgrund einer universitätsinternen Freiversuchsregelung wird der Klausurerfolg als die beste Punktzahl angesehen, die über die ersten beiden Klausur-Termine je Proband/in erreicht wurde. Durch mindestens 50 von 100 erzielten Punkten bei einer Teilnahme an mindestens einem der zwei Klausur-Termine, die in der vorlesungsfreien Zeit nach Semesterende innerhalb von vier Wochen zur Verfügung standen, konnte die Klausur bestanden werden. Über eine einfaktorische ANOVA ($F(3, 89) = 5.384, p \leq .01, \eta^2 = .154$) konnten für den Klausurerfolg im post-hoc Test (Tukey) signifikante Unterschiede zwischen den Interventionsgruppen ermittelt werden (s. Abbildung 5).

Dabei blieben Studierende der Gruppe mit KOR-Feedback ohne Leistungskurs ($M = 39.36, SD = 20.20$) signifikant hinter den Studierenden mit Leistungskurswahl zurück, unabhängig davon, welches Feedback sie bekommen haben. Zwischen EST- ($M = 60.52, SD = 17.44$) und KOR-Feedback ($M = 61.00, SD = 22.23$) findet sich kein Unterschied, wenn vormals ein Leistungskurs besucht wurde. Da sich zwischen der Gruppe EST_(kein LK) und KOR_(kein LK) keine Unterschiede hinsichtlich des Fachwissens (Pre) und mathematischen und kognitiven Fähigkeiten finden ließen, profitieren Studierende der Gruppe ohne Leistungskurs

am meisten vom EST-Feedback ($M = 55.61$, $SD = 21.98$), da sie hinsichtlich des Klausurerfolgs nicht von beiden Gruppen mit Leistungskurs hintergrund unterscheidbar sind und sich zudem ein Trend hinsichtlich eines Vorsprungs gegenüber der gleich beschulten Gruppe, die mit KOR-Feedback gelernt hat, abzeichnet.

Abbildung 5: Klausurpunkte in Abhängigkeit von der Interventionsgruppe WiSe 2018 und 2019 (** $p \leq .01$)



7 Diskussion und Fazit

Die Ergebnisse deuten grundsätzlich auf einen Vorteil des EST-Feedbacks hin, besonders scheinen aber Studierende mit geringerem Vorwissen vom EST-Feedback zu profitieren. Betrachtet man die Chemie-Kurswahl in der Oberstufe als zusätzliches Maß für das Vorwissen der Studierenden, so zeigt sich hinsichtlich des Klausurerfolgs für Studierende der Gruppe EST_(kein LK) eine Performance, die nicht mehr von der der Studierenden mit früherer Leistungskurswahl (EST_(LK) und KOR_(LK)) zu unterscheiden ist, während Studierende der Gruppe KOR_(kein LK) signifikant zurückbleiben und im Mittel mit knapp 39.5 Punkten nicht einmal die Klausur bestehen. Hier deutet sich also bereits konform zur Hypothese 1 ein Vorteil von EST-Feedback für Studierende mit geringem Vorwissen an.

Zwischen EST_(LK) und KOR_(LK) gibt es allerdings keinen Unterschied in den Klausurleistungen, weshalb es für Hypothese 2, also einen Nachteil für Vorwissenstarke resultierend aus EST-Feedback, zunächst einmal keine bestätigenden Indizien gibt. Betrachtet man mit Modell 2 nun die Regression zur Vorhersage des Fachwissens zum Post-Messzeitpunkt unter Kontrolle des Fachwissens (Pre), dann zeigt sich der Vorteil für EST-Feedback nur für Studierende ohne Leistungskurs (EST_(kein LK)). Wurde hingegen ein Leistungskurs Chemie in der Oberstufe belegt und mit EST-Feedback gelernt (EST_(LK)), so äußert sich in der Interaktion ein schwacher Expertise Reversal Effekt, der den Vorteil des EST-Feedbacks nahezu aufhebt. Auf Basis dessen war für Studierende der Gruppe KOR_(LK) ein Vorteil für das Lernen mit

KOR-Feedback erwartet worden. Der Interaktionsterm aus Kurswahl und KOR-Feedback weist zwar konform zu Hypothese 2 ein positives Vorzeichen auf, doch ist er weder signifikant noch von nennenswerter Effektstärke ($\beta = .004, p = .170$). Das Modell 3, das auf Basis dessen zur Überprüfung nur eine Interaktion für EST-Feedback annimmt, führt zudem zu einer signifikant besseren Varianzaufklärung.

Eine direkte Interaktion zwischen Feedback-Typ und den harten Vorwissensdaten aus dem Fachwissenstest ließ sich hingegen in keiner Analyse zeigen. Der Fachwissenstest von Averbeck (2020) besteht allerdings zu einem Drittel aus Items, die Faktenwissen, und zu zwei Dritteln aus Items, die Zusammenhänge abfragen, und von der inhaltlichen Komponente scheinen die 35 Items auch überwiegend Aspekte zu prüfen, die grundlegend für die Allgemeine Chemie sind und sich in den Lehrplänen von Grundkurs und Leistungskurs wiederfinden. Der Test weist gute Kennwerte hinsichtlich der Gütekriterien auf. Fraglich ist aber, ob er den gesamten Vorteil abbilden kann, der sich durch einen Leistungskurs ermöglicht.

Auf der Internetseite des Landesinstituts für Schule (QUA-LiS NRW) findet sich folgender Ausschnitt:

„[Als Bruttounterrichtszeit ergeben sich] ... für die Einführungsphase 90 Unterrichtsstunden, für den Grundkurs in der Q1 ebenfalls 90 und in der Q2 60 Stunden und für den Leistungskurs in der Q1 150 und für Q2 90 Unterrichtsstunden“ (QUA-LiS NRW, 2022).

Demnach ergibt sich für einen Chemie-Leistungskurs ein Zugewinn an Lernzeit, welcher der gesamten Lernzeit eines Grundkurses in der Q1 oder Einführungsphase entspricht. Im Rahmen eines Leistungskurses besteht folglich viel mehr Zeit, durch kumulatives Lernen ein Wissensnetz aufzubauen. Die geringere Lernzeit bietet im Grundkurs deutlich weniger Möglichkeiten, Zusammenhänge im gleichen Umfang wie in einem Leistungskurs zu vermitteln. Das EST-Feedback präsentiert aber gerade diese Zusammenhänge anwendungsbezogen u.a. in Form von Lösungsbeispielen. Da sich diesbezüglich aber nur für Studierende ein signifikanter Vorteil zeigt, die früher einen Grundkurs Chemie belegt hatten, ist es möglich, dass die Aufgaben mit EST-Feedback dabei helfen, bei gleicher Lernzeit fehlendes Zusammenhangswissen aufzuholen beziehungsweise vorhandenes (Insel-)Wissen miteinander zu verknüpfen, was sich in einer höheren Test- und Klausurperformance bei Studierenden der Gruppe $EST_{(\text{kein LK})}$ – verglichen mit anderen Studierenden mit einem Grundkurs in der Oberstufe – zeigt.

Einigt man sich darauf, das Konstrukt Vorwissen als Kombination aus dem Fachwissenstest (Pre) und der Chemie-Kurswahl in der Oberstufe zu beschreiben, dann lässt sich Hypothese 1 bestätigen, denn es scheinen laut Datenlage besonders Studierende ohne frühere Leistungskurswahl von den Aufgaben mit EST-Feedback zu profitieren. Wenngleich sich ein schwindender Vorteil der Unterstützung in der Interaktion aus EST-Feedback und Leistungskurswahl für $EST_{(\text{LK})}$ zeigt, lassen sich keine Indizien für Hypothese 2 finden, weshalb sie verworfen werden muss.

In einem klassischen Übungsbetrieb (vgl. $KG_{(\text{kein LK})}$ und $KG_{(\text{LK})}$) bietet es sich also an, diese Lernaufgaben mit EST-Feedback als zusätzliches Lernangebot und damit als binnendifferenzierendes Förderangebot einzusetzen, da aktives Lernen mit den Aufgaben zu einem signifikant besseren Fachwissensstand und Klausurerfolg am Ende des ersten Semesters führt. Damit liefert diese Studie hinsichtlich elaborierten Feedbacks vergleichbare Ergebnisse wie Fyfe et al. (2012). Die sehr komplexen Aufgaben, die die Allgemeine Chemie bereithält, können allerdings der Grund sein, warum Studierende der Gruppe $EST_{(\text{LK})}$ keine

Nachteile durch Lernen mit dem EST-Feedback erfahren. Daraus ergibt sich für die Studieneingangsphase eine spezifische Handlungsempfehlung für das Fach Chemie, etwaige digitale Lernmodule, die elaboriertes Feedback bereitstellen, einzurichten beziehungsweise (falls vorhanden) systematisch auszubauen. Ob sich durch eine systematische Intervention mit solchen digitalen Lernmodulen langfristig Abbruchquoten reduzieren lassen, gilt es allerdings noch zu prüfen.

Im Rahmen der Feedback-Forschung liefert diese Arbeit einen (vergleichsweise selten anzutreffenden) systematischen Vergleich zweier Feedback-Typen. Naturwissenschaftliche Fächer und insbesondere die Chemie als Forschungsbereich sind allerdings in der Feedback-Forschung unterrepräsentiert beziehungsweise hinsichtlich der Beforschung der Wirksamkeit von automatisiertem Feedback nach Cavalcanti et al. (2021) nicht vorhanden. Allerdings konnte auch diese Arbeit zeigen, dass automatisiertes (elaboriertes) Feedback grundsätzlich einen positiven Effekt auf die Performance hat. Angelehnt an den positiven Trend, der sich bei 32 der von Cavalcanti et al. (2021) ermittelten Artikel abzeichnet, liefert diese Arbeit damit nun auch einen empirischen Nachweis für positive Effekte automatisierten Feedbacks auf die Performance von Nutzerinnen und Nutzern für die Chemie und damit ebenfalls einen guten Ansatzpunkt für etwaige Replikationsstudien.

Anmerkung

Dieses Projekt wurde gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 397641476.

Literatur

- Albacete, P. L. & VanLehn, K. A. (2000). Evaluating the effectiveness of a cognitive tutor for fundamental physics concepts. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 22. Verfügbar unter <https://escholarship.org/uc/item/0166b7p0>
- Ashton, H. S., Beevers, C. E., Korabinski, A. A. & Youngson, M. A. (2005). Investigating the medium effect in computer-aided assessment of school chemistry and college computing national examinations. *British Journal of Educational Technology*, 36(5), 771–787. <https://doi.org/10.1111/j.1467-8535.2005.00501.x>
- Averbeck, D. (2020). *Zum Studienerfolg in der Studieneingangsphase des Chemiestudiums* (Studien zum Physik- und Chemielernen, Bd. 308). Logos.
- Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A. & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61(2), 213–238. <https://doi.org/10.3102/00346543061002213>
- Bond, T. G., Yan, Z. & Heene, M. (2020). *Applying the Rasch model. Fundamental measurement in the human sciences* (Fourth edition). Routledge. <https://doi.org/10.4324/9780429030499>
- Butler, D. L. & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245–281.
- Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y.-S., Gašević, D. et al. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, 2, 100027. <https://doi.org/10.1016/j.caeai.2021.100027>

- Chen, O., Kalyuga, S. & Sweller, J. (2017). The Expertise Reversal Effect is a variant of the more general Element Interactivity Effect. *Educational Psychology Review*, 29(2), 393–405. <https://doi.org/10.1007/s10648-016-9359-1>
- Dickmann, T. (2019). *Visuelles Modellverständnis und Studienerfolg in der Chemie. Zwei Seiten einer Medaille* (Studien zum Physik- und Chemielernen). Logos.
- Eitemüller, C., Trauten, F. Striewe, M. & Walpuski, M. (2023). Digitalization of multistep chemistry exercises with automated formative feedback. *Journal of Science Education and Technology*, 32, 453–467. <https://doi.org/10.1007/s10956-023-10043-2>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Freyer, K. (2013). *Zum Einfluss von Studieneingangsvoraussetzungen auf den Studienerfolg Erstsemestrierender im Fach Chemie* (Studien zum Physik- und Chemielernen, Bd. 156). Logos.
- Fyfe, E. R., Rittle-Johnson, B. & DeCaro, M. S. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology*, 104(4), 1094–1108. <https://doi.org/10.1037/a0028389>
- Goldhausen, I. (2015). *Mathematische Modelle im Chemieunterricht*. Dissertation, Universität Kassel. Uni-Edition GmbH.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (KFT 4-12+R)*. Göttingen: Beltz Test.
- Heublein, U., Ebert, J., Hutzsch, C., Isleib, S., König, R., Richter, J. et al. (2017). *Zwischen Studienerwartungen und Studienwirklichkeit. Ursachen des Studienabbruchs, beruflicher Verbleib der Studienabbrecherinnen und Studienabbrecher und Entwicklung der Studienabbruchquote an deutschen Hochschulen* (Forum Hochschule). Deutsches Zentrum für Hochschul- und Wissenschaftsforschung.
- Heublein, U., Richter, J. & Schmelzer, R. (2020). *Die Entwicklung der Studienabbruchquoten in Deutschland*. https://doi.org/10.34878/2020.03.DZHW_BRIEF
- Johnson, C. I. & Priest, H. A. (2014). The feedback principle in multimedia learning. In R. E. Mayer (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (S. 449–463). Cambridge University Press.
- Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (S. 576–597). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.028>
- Kluger, A. N. & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>
- Krause, U.-M., Stark, R. & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. *Learning and Instruction*, 19(2), 158–170. <https://doi.org/10.1016/j.learninstruc.2008.03.003>
- Mayer, R. E. & Johnson, C. I. (2010). Adding instructional features that promote learning in a game-like environment. *Journal of Educational Computing Research*, 42(3), 241–265. <https://doi.org/10.2190/EC.42.3.a>
- Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, 32(1/2), 99–113. <https://doi.org/10.1023/B:TRUC.0000021811.66966.1d>
- Moreno, R. & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19(3), 309–326. <https://doi.org/10.1007/s10648-007-9047-2>

- Müller, J., Stender, A., Fleischer, J., Borowski, A., Dammann, E., Lang, M. & Fischer, H. E. (2018). Mathematisches Wissen von Studienanfängern und Studienerfolg. *Zeitschrift für Didaktik der Naturwissenschaften*, 24, 183–199.
- Narciss, S. (2006). *Informatives tutorielles Feedback. Entwicklungs- und Evaluationsprinzipien auf der Basis instruktionspsychologischer Erkenntnisse* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 56). Waxmann.
- Narciss, S. (2014). Modelle zu den Bedingungen und Wirkungen von Feedback in Lehr-Lernsituationen. In H. Ditton (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 43–82). Waxmann.
- Narciss, S. (2018). Feedbackstrategien für interaktive Lernaufgaben. In H. M. Niegemann & A. Weinberger (Hrsg.), *Handbuch Bildungstechnologie. Konzeption und Einsatz digitaler Lernumgebungen* (Springer reference, S. 1–24). Springer. https://doi.org/10.1007/978-3-662-54373-3_35-1
- Narciss, S. & Huth, K. (2004). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction*, 16(4), 310–322. <https://doi.org/10.1016/j.learninstruc.2006.07.003>
- Pobel, S. & Striewe, M. (2019). Domain-specific extensions for an e-assessment system. In M. A. Herzog, Z. Kubincová, P. Han & M. Temperini (Hrsg.), *Advances in Web-Based Learning – ICWL 2019* (Bd. 11841, S. 327–331). Springer International Publishing.
- QUA-LiS NRW (Rüdiger Käuser, Hrsg.). (2022). *Schulentwicklung NRW – Lehrplannavigator S II – Gymnasiale Oberstufe – Chemie – Hinweise und Beispiele – Beispiel eines schulinternen Lehrplans für die gymnasiale Oberstufe im Fach Chemie*. Schulentwicklung NRW. <https://www.schulentwicklung.nrw.de/lehrplaene/lehrplannavigator-s-ii/gymnasiale-oberstufe/chemie/hinweise-und-beispiele/schulinterner-lehrplan/schulinterner-lehrplan.html>
- Rost, F. (2018). *Lern- und Arbeitstechniken für das Studium* (Lehrbuch, 8., vollständig überarbeitete und aktualisierte Auflage). Springer VS.
- Sarceletti, A. & Müller, S. (2011). Zum Stand der Studienabbruchforschung. Theoretische Perspektiven, zentrale Ergebnisse und methodische Anforderungen an künftige Studien. *Zeitschrift für Bildungsforschung*, 1(3), 235–248. <https://doi.org/10.1007/s35834-011-0020-2>
- Schuhmacher, R. (2008). Der produktive Umgang mit Fehlern. In R. Caspary (Hrsg.), *Nur wer Fehler macht, kommt weiter* (S. 49–72). Herder.
- Smits, M. H., Boon, J., Sluijsmans, D. M. & van Gog, T. (2008). Content and timing of feedback in a web-based learning environment: effects on learning as a function of prior knowledge. *Interactive Learning Environments*, 16(2), 183–193. <https://doi.org/10.1080/10494820701365952>
- Striewe, M. (2016). An architecture for modular grading and feedback generation for complex exercises. *Science of Computer Programming*, 129, 35–47. <https://doi.org/10.1016/j.scico.2016.02.009>
- Striewe, M., Trauten, F. & Eitemüller, C. (2020). Aufgaben mit automatischem Feedback zu chemischen Atom-Orbitalmodellen. In R. Zender, D. Ifenthaler, T. Leonhardt & C. Schumacher (Hrsg.), *DELFI 2020. Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V., Fachtagung vom 14.-18. September 2020 Online* (GI-Edition. Proceedings, Bd. 308, S. 109–119). Gesellschaft für Informatik.
- Tinto, V. (1988). Stages of student departure: Reflections on the longitudinal character of student leaving. *Journal of Higher Education*, 59(4), 438. <https://doi.org/10.2307/1981920>
- Trautwein, C. & Bosse, E. (2017). The first year in higher education – critical requirements from the student perspective. *Higher Education*, 73(3), 371–387. <https://doi.org/10.1007/s10734-016-0098-5>
- Walter, O. (2005). *Kompetenzmessung in den PISA-Studien. Simulationen zur Schätzung von Verteilungsparametern und Reliabilitäten*. Pabst.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427–450. <https://doi.org/10.1007/BF02294627>

Wisniewski, B., Zierer, K. & Hattie, J. (2019). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*, 1–14. <https://doi.org/10.3389/fpsyg.2019.03087>

Kontakt

Florian Trauten · Dr. Carolin Eitemüller · Prof. Dr. Maik Walpuski
Universität Duisburg-Essen
Didaktik der Chemie
Schützenbahn 70
45127 Essen
E-Mail: florian.trauten@uni-due.de
E-Mail: carolin.eitemueller@uni-due.de
E-Mail: maik.walpuski@uni-due.de

Selbstwirksamkeit in selbstregulierten Lernprozessen an Hochschulen mithilfe von Podcasts fördern – eine Interventionsstudie mit Interaktivem Ambulantes Assessment

Caroline Götz, Paul Erhardt, Simone N. Löffler

Zusammenfassung: Selbstwirksamkeit (SW) im Selbstregulierten Lernprozess (SRLP) gilt als wichtiger Faktor, um Studienerfolg an Hochschulen vorauszusagen. Wir verwendeten Interaktives Ambulantes Assessment in Kombination mit Podcasts, um die SW im SRLP zu fördern. Dafür bearbeiteten Studierende ($N = 106$) drei Mal täglich Abfragen zu ihrem Lernprozess, die ihnen über elektronische Tagebücher präsentiert wurden. Unterschiede zwischen der Kontrollgruppe (KG) ohne Feedback und der Interventionsgruppe (IG) mit täglichem individualisiertem Feedback, basierend auf den Tagebucheinträgen, in Kombination mit Podcasts, wurden mittels Multilevelanalysen geprüft. Die Ergebnisse zeigen, dass die SW im SRLP der IG innerhalb der Interventionsphase, in Relation zu einer Baselinephase und im Vergleich zur KG, gefördert werden konnte. Entsprechend konnte im Prä-Post-Vergleich eine größere Steigerung der SW-Erwartung (vgl. Schwarzer & Jerusalem, 1999) bei der IG (im Vergleich zur KG) nachgewiesen werden.

Schlüsselwörter: Selbstwirksamkeit, Selbstreguliertes Lernen, Interaktives Ambulantes Assessment, Echtzeit-Intervention, Podcasts, Multilevel Analysen

Promoting self-efficacy in self-regulated learning processes in higher education with the help of podcasts – an intervention study with interactive ambulatory assessment

Abstract: Self-efficacy (SE) in self-regulated learning processes (SRLP) is considered an important factor to predict student success in higher education. We used interactive ambulatory assessment in combination with podcasts to promote SE in SRLP. For this purpose, students ($N = 106$) completed queries regarding their learning process three times a day, which were presented to them via electronic diaries. Differences between the control group (CG) without feedback and the intervention group (IG) with daily individualized feedback, based on the entries in the electronic diaries, in combination with podcasts, were tested using multilevel analyses. The results show that SE in SRLP was promoted in the IG within the intervention phase compared to the CG and the baseline phase. In the pre/post comparison, a stronger increase in SE-expectancy (cf., Schwarzer & Jerusalem, 1999) was revealed for the IG, in comparison to the CG.

Keywords: Self-efficacy, self-regulated learning, interactive ambulatory assessment, real-time intervention, podcasts, multilevel analyses

1 Einleitung

Der technologische Wandel an Hochschulen macht es möglich, akademisches Lernen und Lehren stärker zu individualisieren. Durch Educational Data Mining und den Einsatz von Learning Analytics können Lernprozesse abgebildet, analysiert und Förderbedarfe aufgedeckt werden, auf deren Basis individuelle Empfehlungen zur Gestaltung und Modifizierung von Lernprozessen ableitbar sind. Zur Anpassung der eigenen Lernprozesse an äußere und persönliche Ansprüche benötigt es Soft Skills, wie die Fähigkeit zur Selbstregulation (SR). Gerade an Hochschulen sind SR-Fähigkeiten für ein erfolgreiches Studium unabdingbar (Krapp, 1993; Schmidt et al., 2011; Zimmerman, 2002). Die Umstellung der Hochschulen auf die digitale Lehre durch die COVID-19-Pandemie hat den Einsatz und die Entwicklung von SR-Fähigkeiten zusätzlich erschwert (Breitenbach, 2021; Marczuk et al., 2021). Jedoch auch ohne pandemisches Geschehen findet eine zunehmende Individualisierung und dadurch eine Verlagerung der Verantwortung für Lernprozesse statt (Pflegig & Gerhardt, 2013). Dies stellt eine besondere Herausforderung für Studierende dar. In diesem Kontext könnte sich studentische Selbstwirksamkeit (SW) als besonders relevant erweisen.

SW beschreibt das Erleben einer Person, dass sie Herausforderungen durch eigene Fähigkeiten bewältigen kann. In Abgrenzung dazu wird die Selbstwirksamkeitserwartung (SWE) als die Erwartung und Zuversicht einer Person angesehen, schwierige und herausfordernde Situationen und Aufgaben mithilfe der eigenen Fähigkeiten und Erfahrungen bewältigen zu können (Bandura, 1994). Eine hohe SWE kann den Studienverlauf positiv beeinflussen (Gaylon et al., 2012; Robbins et al., 2004; Roick & Ringeisen, 2017; Schnell et al., 2015; Schwarzer & Jerusalem, 2002; Sitzmann et al., 2012). Der vorliegenden Studie liegt die Annahme zugrunde, dass das Konstrukt eine zeitlich (relativ) stabile Komponente aufweist (die SWE), die sich durch Unterschiede zwischen Personen abbilden lässt, die erlebte SW im Selbstregulierten Lernprozess (SRLP) jedoch zusätzlich einen dynamischen Verlauf hat, der sich in intraindividuellen Fluktuationen über die Zeit hinweg zeigt (vgl. Bandura, 1994; Kitsantas et al., 2008).

2 Selbstreguliertes Lernen

Der Begriff SRLP beschreibt die Aktivität einer lernenden Person, die für ihren Lernprozess Ziele (Soll-Wert) definiert, den aktuellen Stand ihres Lernprozesses beziehungsweise ihrer Kenntnisse erfasst (Ist-Wert) und Maßnahmen ergreift, um die Lernziele zu erreichen (Pintrich, 2000; Schmitz & Schmidt, 2007). Dabei werden kognitive, motivationale und behaviorale Vorgänge selbst beobachtet und aktiv, entsprechend der Herausforderungen der aktuellen Aufgabe, reguliert (Zimmerman et al., 1992). Die Lernmotivation und der damit verbundene engagierte Einsatz von SR-Strategien entspringt dem Bedürfnis nach Kompetenzerleben, welches auf die Selbstbestimmungstheorie nach Ryan und Deci (2000) zurückgeht. Das Modell von Zimmerman (2002), auf das die vorliegende Studie konzeptionell basiert, unterteilt den Lernprozess in drei zeitlich aufeinander folgende Phasen.

In der präaktionalen Phase erfolgt die Aufgabenanalyse und strategische Planung; hierbei sind Selbstmotivationsprozesse von Bedeutung. Dabei sollte die lernende Person ihre Ziele so formulieren, dass sie den eigenen Fähigkeiten entsprechen und durch Anstrengung

und Erfahrung bewältigt werden können (Schwarzer & Jerusalem, 2002; Zimmerman, 2000). In der aktionalen Phase kommen kognitive und metakognitive Lernstrategien zum Einsatz. Lernstrategien beschreiben die Aktivitäten, die eingesetzt werden, um Lernprozesse zu planen, zu beobachten, zu gestalten und zu regulieren. Die postaktionale Phase enthält Selbstreflexionsprozesse sowie die daraus resultierenden affektiven Reaktionen, die wiederum die präaktionale Phase der nächsten Lerneinheit beeinflussen. So nehmen Emotionen Einfluss auf den Lernstrategieinsatz, die Informationsspeicherung und die lern- und leistungsrelevante Motivation (Hasselhorn & Gold, 2017). Starke Emotionen, wie sie z.B. im Zusammenhang mit Prüfungsangst beobachtet werden, können die kognitiven Kapazitäten der Person und die Anpassungen des eigenen Lernverhaltens, die aus der Reflexion innerhalb der postaktionalen Phase entstehen, negativ beeinflussen (Lowe, 2021). Die Phasen des Lernprozessmodells laufen idealtypisch zeitlich nacheinander ab. Je nach Anforderung kann es jedoch nützlich sein, die Phasen parallel zu durchlaufen oder in eine vorherige Phase zurückzukehren (Zimmermann, 2002).

Die Lernzufriedenheit beschreibt die subjektive Einschätzung der lernenden Person hinsichtlich des Lernergebnisses beziehungsweise des Erreichens des Lernziels (Gabriel et al., 2007). Vor allem der bedarfsgerechte und kombinierte Einsatz von kognitiven, metakognitiven und ressourcenbezogenen Lernstrategien sowie eine Mischung aus Oberflächen- und Tiefenverarbeitung scheint zu positiven Lernergebnissen und Zufriedenheit mit dem Lernprozess zu führen (Kállay, 2012; Loeffler, Bohner et al., 2019; Ludwig et al., 2013; Martin & Nicolaisen, 2015). Neben dem effektiven Einsatz von Lernstrategien gibt es weitere Faktoren (habituelle Eigenschaften der Person und situative Merkmale), die den Erfolg eines SRLP beeinflussen. So konnte in verschiedenen Studien ein positiver Zusammenhang von Interesse und (Lern-)Leistung nachgewiesen werden (Blüthmann, 2012; Krapp, 1993; Müller, 2006).

Die Lernzeit ist fest in die SRL-Modelle verankert. Liborius et al. (2019) postulieren einen linearen positiven Zusammenhang zwischen der Investition von Lernzeit und der Lernzufriedenheit. In der Prokrastinationsforschung wird das Aufschieben von akademischen Lernaktivitäten und der dadurch aufkommende Zeitdruck mit negativen Emotionen, einem erhöhten Stresslevel und geringeren akademischen Leistungen in Verbindung gebracht (Flett et al., 1995; Kim & Seo, 2015; Pychyl et al., 2000; Steel, 2007).

3 Selbstwirksamkeit und Selbstwirksamkeitserwartung

Das Konzept der SWE geht auf die sozial-kognitive Theorie nach Bandura (1994) zurück. Selbstwirksame Personen sind zuversichtlich, dass sie über die notwendigen Fähigkeiten zur Aufgabenbewältigung verfügen oder diese erlernen können. Durch diese Form des Optimismus sind sie in der Lage, auf sich verändernde Umstände adäquat zu reagieren und sich nicht durch ängstliche Gedanken oder Stress zu blockieren. Die SWE wirkt sich darauf aus, wie Personen denken, sich selbst motivieren, fühlen und Herausforderungen auswählen. Personen, die über eine hohe SWE verfügen, zeigen sich auch in krisenhaft erlebten Situationen ausdauernd und gehen aus diesen gestärkt hervor, sie orientieren sich am eigenen Prozess und an dessen Fortschritten und weniger daran, wie erfolgreich sie im Gegensatz zu anderen Personen sind (Bandura, 1982; Schwarzer & Jerusalem, 2002). Die SWE kann als zeitlich stabil beschrieben werden, Kontext und Inhalt der (Leistungs-) Situation haben jedoch einen

erheblichen Einfluss auf die Einschätzung und das Erleben der eigenen Fähigkeiten und somit auf die erlebte SW (Bandura, 1994; Kitsantas et al., 2008).

Im SRLP nimmt die SW eine selektierende Funktion ein: Je nach Ausprägung der SW werden komplexe oder einfache Lernziele formuliert, unterschiedliche Lernstrategien eingesetzt und die Ausprägung der Lernmotivation beeinflusst (Bandura & Locke, 2003; Borgognoni & Pokropek, 2019; Roick & Ringeisen, 2018; Schwarzer & Jerusalem, 2002; Sitzman et al., 2010; Usher & Pajares, 2008a). Alghambi et al. (2020) definieren die SW im SRLP als die Zuversicht Lernender, unterschiedliche SRL-Strategien sinnvoll einzusetzen, um akademische Erfolge zu erzielen. Schwarzer und Jerusalem (2002) beschreiben die SW als Schlüssel zur kompetenten SR. Dabei werden die Einflüsse der SW auf die SR weitgehend unabhängig von den tatsächlichen Fähigkeiten der Person angesehen (Schwarzer & Jerusalem, 2002). Dass es einen positiven Zusammenhang zwischen SW und Fähigkeiten der SR sowie der Lernleistung gibt, ist jedoch hinreichend belegt (Galyon et al., 2012; Robbins et al., 2004; Roick & Ringeisen, 2017; Schnell et al., 2015; Schwarzer & Jerusalem, 2002; Sitzmann et al., 2010). Demnach lohnt es sich auch aus pädagogischer Perspektive, SW als übergeordnete SR-Fähigkeit in Bildungseinrichtungen zu festigen und zu fördern (Roick & Ringeisen, 2018; Schnell et al., 2015; Schwarzer & Jerusalem, 2002).

4 Förderung von Selbstwirksamkeit im Selbstregulierten Lernprozess

Bandura (1994) beschreibt vier Quellen, durch die SW gefördert werden kann. So wird SW vor allem dann gefördert, wenn eine Person sich als wirksam erlebt, d.h. wenn sie Erfolgserlebnisse hat, die sie auf eigenes Können und Handeln zurückführen kann (Bandura, 1994; Usher & Pajares, 2008b). Eine weitere Quelle ist das Lernen an (positiven) Modellen durch Beobachtung (Bandura, 1994; Schwarzer & Jerusalem, 2002; Usher & Pajares, 2008b). Hierbei findet nicht nur ein Abgleich der Fähigkeiten und des Durchhaltevermögens statt. Es ist auch möglich, dass die Personen voneinander lernen, wie eine konkrete Situation bewältigt werden kann (Bandura, 1986). SW kann des Weiteren durch verbale Ermutigung sowie durch Reduktion negativer Gefühle gefördert werden (Bandura, 1994; Usher & Pajares, 2008b). Die SW speist sich demnach aus eigenen Erfahrungen, dem Lernen mit und durch andere Personen, gutem Zureden und der Steuerung von Emotionen.

Bei den bisherigen Interventionsansätzen zur Förderung des SRL handelt es sich um Face-to-Face Trainings, Lerntagebücher, Feedback-Interventionen oder um Veränderungen des Lehr-Lernarrangements. Interventionen, die dazu dienen, gezielt die SW zu fördern, existieren bisher nur wenige. Vielversprechende Ergebnisse zur Förderung der SW im SRLP an Hochschulen liefern die Studien von Hawe et al. (2019), Bellhäuser et al. (2016) und Morris (2010). Hawe et al. (2019) intervenierten mithilfe des Einsatzes von Vorbildern beziehungsweise exemplarischen Arbeiten innerhalb eines Hochschulkurses für Wirtschaftsstudierende ($N = 7$) an einer neuseeländischen Universität. Dazu wurden Ausarbeitungen vorheriger Semester genutzt, um den Studierenden zu verdeutlichen, welche Anforderungen an sie gestellt werden und wie sie diesen Erwartungen gerecht werden können. Die Studierenden berichten in den qualitativen Interviews nach der Intervention, dass sie sich motivierter und wirksamer fühlten, die Aufgabenstellung besser verstanden hatten und ihr Wissen erweitern konnten, die Erwartungen an sie klarer und sie selbst reflektierter wurden. Bellhäuser et al. (2016)

testeten die Wirkung von Lerntagebüchern und eines webbasierten Trainings zur Förderung von SRL-Fähigkeiten an einer relativ großen Studierendenstichprobe ($N = 211$). Durch das webbasierte Training konnten SRL-Fähigkeiten und auch die SW gesteigert werden. Für den Einsatz der Lerntagebücher ohne Training konnten keine Interventionseffekte nachgewiesen werden. Da die Studienteilnahme sich jedoch als sehr zeitaufwändig gestaltete, ist durch die hohe Arbeitsbelastung mit Einschränkungen der Compliance zu rechnen, was auch die Abbruchrate von fast 25% nahelegt. In einer Studie von Morris (2010) schnitt die Interventionsgruppe (IG) nach einer Intervention, bestehend aus elf aufgezeichneten Vorlesungen in Form von Audio-Podcastepisoden sowie einer Lernerfolgskontrolle mittels Multiple-Choice Fragen (MC-Fragen) und Feedback, in einer Probeklausur rund 6% besser ab als die Kontrollgruppe (KG).

5 Zielsetzungen

In bisherigen Interventionsstudien zum SRL wurden meist einzelne Phasen des SRLP oder ausgewählte Lernstrategien gefördert und die Effekte auf objektive Kriterien wie z.B. Klausurnoten untersucht (Alghambi et al., 2020; Bergey et al., 2019; Morris, 2010; Roick & Ringeisen, 2017; Salanova et al., 2012; Schnell et al., 2015). Die notwendige SW/SWE, um neue Herausforderungen im SRLP zu bewältigen, wurde in gewisser Weise implizit vorausgesetzt, jedoch in der Regel nicht explizit erfasst. Die Studien, die SW/SWE untersuchen, greifen meist auf Laborsettings zurück oder setzen voraus, dass SW/SWE zeitlich relativ stabil ist (vgl. z.B. Roick & Ringeisen, 2018; Salanova et al., 2012). Somit werden Schwankungen während eines Lernprozesses nicht gemessen und Wechselwirkungen mit dem Lernstrategieinsatz finden keine Berücksichtigung.

Die vorliegende Untersuchung nimmt nun die Dynamik der erlebten SW innerhalb des SRLP (orientiert an den drei Phasen des Prozessmodells von Zimmerman, 2002) explizit in den Fokus, indem die Teilnehmenden innerhalb ihrer natürlichen Lernsituationen Daten aus ihrem aktuellen SRLP in Echtzeit durch Eintragungen in das im Alltag mitgeführte Studienschmartphone übermitteln und die Datenauswertung mittels linearer Mehrebenenanalysen erfolgt. Der Einsatz von Mehrebenenanalysen ermöglicht die Differenzierung zwischen Innersubjekt- und Zwischensubjekteffekten (vgl. Nezlek, 2012) und somit die Abbildung von Fluktuationen im Verlauf des individuellen Lernprozesses. Durch die Verwendung der elektronischen Tagebücher können außerdem Verzerrungen der Daten, z.B. durch Erinnerungseffekte oder fehlende Eintragungen, reduziert werden (vgl. Reuschenbach & Funke, 2011). Zur Erfassung der erlebten SW wurde ein Index (SW-Index) bestehend aus verschiedenen Komponenten studentischer SW im SRLP speziell für die in situ Erfassung mit Hilfe elektronischer Tagebücher entwickelt. Zusätzlich wurde die SWE als Trait-Komponente mittels Fragebogen erhoben.

Die übergeordnete Zielsetzung der Intervention bestand darin, die erlebte SW der Studierenden, in Anlehnung an die vier Quellen der SW nach Bandura (1997), in den drei Phasen des SRLP explizit zu unterstützen und darüber hinaus Wissen um SRL und adäquate SRL-Strategien zu vermitteln. Hierfür verwendeten wir Interaktives Ambulantes Assessment (IAA) mit Feedback, als eine Form von Learning Analytics. Die Studierenden erhielten in der IAA-Interventionsphase automatisiertes und individualisiertes Feedback in Form von Po-

dcastepisoden basierend auf ihren individuellen Einträgen in die elektronischen Lerntagebücher auf den Studiensmartphones. Wenn ein Proband beim abendlichen Ausfüllen des Lerntagebuchs angab, dass er/sie bei der Arbeit am Lernziel einer Herausforderung begegnet war, wurde über das Studiensmartphone eine zu der entsprechenden Herausforderung passende Podcastepisode bereitgestellt.

Bisher wurden Podcasts im Lernkontext meist eingesetzt, um Interesse an einem Thema oder einer Veranstaltung zu wecken (Fernandez et al., 2009; Lee & Chan, 2007), um wichtige Inhalte noch einmal zusammenzufassen (Fernandez et al., 2009; Lee & Chan, 2007; Morris, 2010), Tipps zu bestimmten Inhalten oder Arbeitsaufträgen zu vermitteln (Lee & Chan, 2007; Powell & Mason, 2013) oder einen Überblick über eine bevorstehende Veranstaltung zu geben (Powell & Mason, 2013). Die vorliegende Untersuchung ist unseres Wissens die erste, die Audiopodcasts im Rahmen der Schlüsselqualifikationslehre einsetzt, um Studierenden Fähigkeiten in ihrem Lernalltag zu vermitteln und ihre SW im SRLP zu fördern. Da der Podcast als Medium akademischer Lehre zu den studierendenzentrierten Lehransätzen gezählt wird, eignete er sich besonders für den Einsatz im Rahmen eines rein technikgestützten Interventionsansatzes während der digitalen Semester (vgl. Gnaur & Huttel, 2016). Da das IAA so konzipiert war, dass eine vorläufige Datenanalyse und daraus abgeleitetes Feedback in Echtzeit direkt innerhalb der zu untersuchenden Situation möglich war, konnten die Teilnehmenden, trotz der Kontaktbeschränkungen im Kontext der COVID-19-Pandemie, unmittelbar in ihrem Lernalltag unterstützt werden.

Konkret sollte in der vorliegenden Interventionsstudie geprüft werden, ob sich die IG in ihrer erlebten SW im SRLP durch die IAA-Interventionsphase steigern konnte. Zusätzlich sollte untersucht werden, ob sich die SWE der IG nach der Interventionsphase verbessert hatte. Aufgrund der Vorteile, die das IAA für Interventionen im SRLP bietet, wie z.B. konkrete Unterstützung unmittelbar während des Lernens im Alltag, Förderung des Strategietransfers in die Lernroutine und individualisiertes Feedback (vgl. Loeffler, Bohner, et al., 2019; Loeffler, Stupp et al., 2019), wurde angenommen, dass die zeitlich begrenzte, aber kompakte Intervention zu einer positiveren SW-Entwicklung führt; auch die Studie von Morris (2010) legt nahe, dass sich Feedback in Verbindung mit Podcasts zur Förderung des SRLP Studierender eignen könnte. Hieraus ergaben sich folgende gerichtete Hypothesen:

- H1: Es wird angenommen, dass sich die erlebte SW (SW-Index, *state*) der IG in der Interventionsphase, im Vergleich zur KG und in Relation zur Baseline, steigert.
- H2: Wir erwarten, dass die SWE (*trait*) der IG im Prä/Post Vergleich und im Vergleich zur KG gefördert wird.

Explorativ sollte untersucht werden, auf welche einzelnen *Komponenten* der erlebten SW im SRLP die Intervention einen förderlichen Einfluss zeigte beziehungsweise auf welche sich möglicherweise kein positiver Effekt der Maßnahmen nachweisen ließ. Zusätzlich sollte geprüft werden, ob sich die Intervention förderlich auf die Lernzufriedenheit und den subjektiven Lernerfolg der IG im Vergleich zur KG und zur Baseline auswirkte. Zur Differenzierung dynamischer und statischer Anteile der SW wurden Intraklassenkorrelationskoeffizienten (ICCs) im Rahmen von Mehrebenenanalysen berechnet.

6 Methodik

6.1 Prozedur

Die Datenerhebung fand innerhalb eines Online-Seminars, das asynchron im Bereich der Schlüsselqualifikationslehre an einer deutschen Technischen Universität angeboten wurde, statt. Der Onlinekurs „Selbststudium erfolgreich gestalten“ diente als Akquisemöglichkeit für die Studienteilnehmenden sowie als Plattform der Wissensvermittlung, um die Studierenden angemessen im Seminarverlauf zu begleiten. Dazu wurde auf der Lernplattform ILIAS der Onlinekurs erstellt und modular aufgebaut. Modul 1 lieferte Informationen über die Kursorganisation sowie die Lernziele der Veranstaltung. Modul 2 vermittelte Grundlagenwissen zur Formulierung von Lernzielen. Im Anschluss wurde die SMART Methode (Doran, 1981) vorgestellt, die dabei unterstützt, Lernziele möglichst konkret und realistisch zu formulieren. Solche Nahziele sorgen dafür, dass die Studierenden ihre Lernziele als erfolgsversprechend und bewältigbar einschätzen (Schwarzer & Jerusalem, 2002). Das Modul wurde mit einer Lernzielkontrolle in Form eines Tests abgeschlossen. Modul 3 bereitete die Studierenden auf den Ablauf der Datenerhebung vor. Modul 4 diente der Organisation der obligatorischen Abschlussgespräche für die KG. Die IG fand im Modul 4 weiterführende Informationen zu den Podcastepisoden. Diese waren mit einem Passwort geschützt und nur zugänglich, wenn die Teilnehmenden die entsprechende Podcastepisode im Rahmen des individualisierten Feedbacks der IAA-Interventionsphase empfohlen bekommen hatten. Modul 5 bereitete die Studierenden auf das Anfertigen des Abschlussberichts (als Seminarleistung für die Vergabe von Leistungspunkten) vor.

Der Kurs startete erstmals im Wintersemester 2019/2020 und wurde bis zum Ende des Wintersemesters 2020/2021 im April 2021 für Studierende aller Studiengänge und angestrebten Abschlüsse angeboten. Die Studienteilnahme erfolgte freiwillig, nach informierter schriftlicher Zustimmung, und war keine zwingende Voraussetzung zur Erlangung von Leistungspunkten. Die Datenerhebung fand aufgrund der pandemischen Lage kontaktlos statt. Alle Daten wurden pseudonymisiert erhoben und verarbeitet.

Abbildung 1 zeigt den Studienablauf für beide Gruppen. In der Baselinephase (Phase 0) war die Datenerhebung mittels AA für beide Gruppen identisch. In der Kontroll- bzw. Interventionsphase (Phase 1) unterschieden sich die täglichen Abfragen der Gruppen. Die KG beantwortete, wie in Phase 0, an jedem Lerntag die Items des elektronischen Lerntagebuchs (AA). Für die IG wurden die täglichen Abfragen um das Feedback mit Podcastepisoden und die Kontrollfragen zur empfohlenen Podcastepisode ergänzt (IAA). Mit dem Ziel, die Compliance der Probanden möglichst auch in der Kontrollphase aufrecht erhalten zu können, wurde Phase 1 für die KG auf sechs Lerntage verkürzt, während für die IG eine Interventionsphase (Phase 1) von elf Lerntagen vorgesehen war.

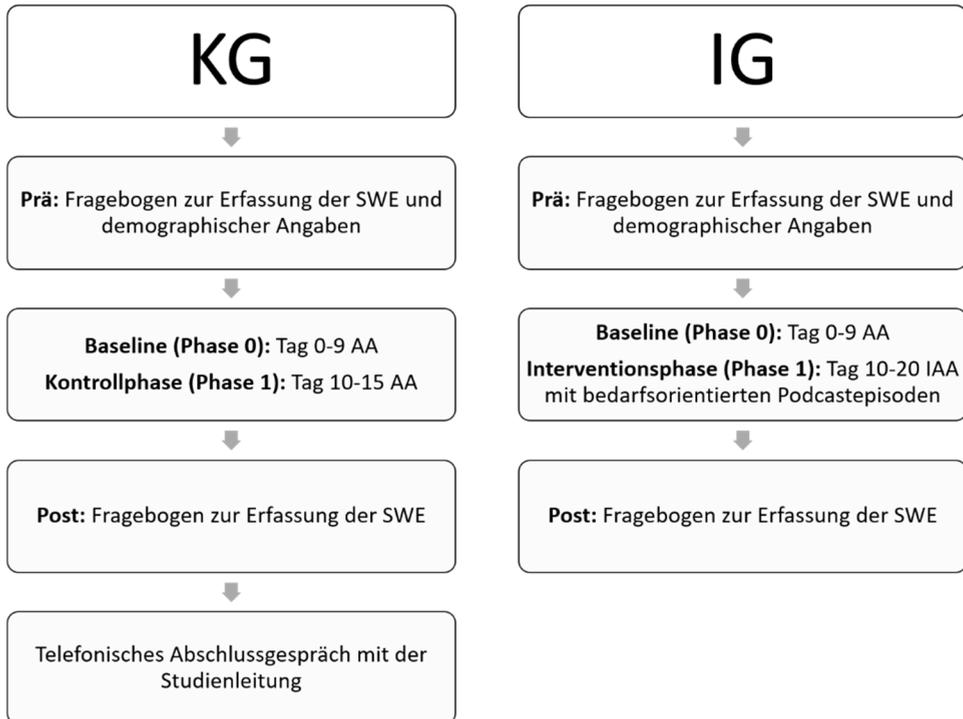


Abbildung 1: Prozedur. Auf der linken Seite ist der Ablauf für die Kontrollgruppe, auf der rechten Seite für die Interventionsgruppe abgebildet.

6.1.1 Ambulantes Assessment

Die Studierenden konnten selbst bestimmen, in welchem Zeitraum des Semesters sie das elektronische Lerntagebuch führen wollten, und dementsprechend einen beliebigen Tag zur Ausgabe der Studienunterlagen und des Studiensmartphones buchen. Am Tag nach der Geräteausgabe konnten sie mit dem AA (d.h. den täglichen Eintragungen in das Lerntagebuch) im Alltag beginnen. Die Datenerhebung erfolgte mit der Applikation movisensXS (movisens GmbH, Karlsruhe), die auf den Smartphones installiert war. Da die Studie etwas breiter angelegt war, finden sich in der folgenden Darstellung des AA auch Konstrukte des SRL, die im vorliegenden Artikel keine Berücksichtigung finden.

Aufgrund der technischen Gegebenheiten der Applikation konnte die Erinnerungsfunktion zur Eingabe der Daten nur bei der Verwendung fester Zeitfenster für die Dateneingabe aktiviert werden. Somit war es notwendig, feste Eingabezeiten für die elektronischen Tagebücher zu definieren. Die Erfassung der präaktionalen Phase war täglich von 8:00 bis 11:30 Uhr vorgesehen. Zunächst wurden die Studierenden aufgefordert, für den bevorstehenden Lernprozess (in freier Texteingabe) ein Lernziel zu formulieren. Des Weiteren wurden die emotionale Befindlichkeit, die Erfolgserwartung, die erwartete SW im SRLP (bezogen auf die Arbeit am Lernziel) und die geplante Lernzeit erfasst. Die Eingaben zur aktionalen Phase sollten die Studierenden zu Beginn ihrer Mittagspause (zwischen 11:31 und 16:00 Uhr) tätigen. Hier wurden emotionale Befindlichkeit, Ausdauer, Volition, die Lernzeit sowie die

physische und soziale Lernumgebung erhoben. Angaben zur postaktionalen Phase waren zwischen 16:09 und 22:59 Uhr möglich. Hier wurde die aktuelle emotionale Befindlichkeit, die erlebte SW im SRLP, Lernzufriedenheit und Lernerfolg erfragt. Wurden die Eingaben nicht zum Startzeitpunkt der Abfragen getätigt, folgte alle 30 Minuten eine Erinnerung, bis die Abfrage bearbeitet oder der Zeitraum zur Bearbeitung abgelaufen war. Die Teilnehmenden konnten einstellen, ob diese Erinnerungen ein akustisches, ein sensorisches oder kein Signal beinhalten sollte. Wurde die jeweilige Abfrage gestartet, musste sie innerhalb von 30 Minuten ausgefüllt werden.

Für die KG folgte nach Beendigung der beiden Registrierphasen (Phase 0 und Phase 1) ein telefonisches Abschlussgespräch (60 Minuten) mit der Studienleitung, um über individuelle Herausforderung während der Selbstlernphase zu sprechen und neue Lernstrategien kennenzulernen. Die individuelle Lernberatung fungierte als Ausgleich für das automatisierte und individualisierte Feedback, das die IG über das IAA in Kombination mit den Podcastepisoden erhielt, fand jedoch in der Datenauswertung der Studie keine Berücksichtigung.

6.1.2 Interaktives Ambulantes Assessment mit Podcastepisoden

In Phase 1 wurden der IG weiterhin die AA-Items (siehe Abschnitt 6.1.1) zur täglichen Einschätzung vorgegeben. Zusätzlich wurden die jeweiligen Podcastepisoden über das Smartphone bereitgestellt, sofern die Studierenden in der Abfrage am Abend berichtet hatten, dass sie bei der Arbeit an ihrem Lernziel einer Herausforderung begegnet waren. Dazu sollte die IG in ihren täglichen Tagebucheinträgen angeben, ob eine oder mehrere der in Tabelle 1 aufgelisteten Aussagen eher zutreffend oder nicht zutreffend war. Sofern die Person einer oder mehreren dieser Aussagen zustimmte, konnte sie über das Smartphone die entsprechende Podcastepisode aufrufen (s. Tabelle 1), um so Unterstützung bei der persönlichen Herausforderung zu erhalten.

Die Konzeption der Podcasts basierte auf der Förderung der vier Quellen der SW nach Bandura (1997). Es wurden insgesamt elf Episoden entwickelt, deren Länge zwischen 4:02 und 18:29 Minuten variierte. Die Inhalte der Episoden eins bis zehn orientierten sich des Weiteren an den Lernstrategien, die Martin und Nicolaisen (2015) definieren; dabei enthielt jede Episode eine Kombination aus kognitiven und metakognitiven beziehungsweise ressourcenorientierten Lernstrategien, die entlang der Phasen eines SRLP nach Zimmerman (2002) benötigt werden. Episode 0 wurde allen Studierenden der IG am ersten Tag des IAA präsentiert und bestand aus einer kurzen Einführung zur Interventionsphase.

Die passenden Podcastepisoden wurden mit einer entsprechenden Instruktion unmittelbar nach dem Ausfüllen der Abendabfrage für den gesamten Zeitraum der Interventionsphase auf dem Studiensmartphone bereitgestellt. Im Anschluss an die Wiedergabe der Podcastepisode sollten die Studierenden drei bis fünf MC-Fragen zur Lernkontrolle der Inhalte sowie zwei bis drei Transferfragen zur Anwendung des neu Gelernten für den nächsten SRLP beantworten. Basierend auf den Ergebnissen vorheriger Arbeiten (Bellhäuser et al., 2016; Chan & Lam, 2010; Gessnitzer et al., 2016; Loeffler, Stumpp et al., 2019; Morris, 2010; Roelle et al., 2011) zu positiven Einflüssen direkten Feedbacks auf den Lernprozess erschien unmittelbar nach der Bearbeitung der MC-Fragen eine Rückmeldung auf dem Smartphone und machte den Studierenden transparent, ob sie die Fragen richtig oder falsch beantwortet hatten. Wurde die Frage falsch beantwortet, erfolgte im Anschluss an das Feedback die Präsentation der korrekten Antwort. Wenn die Fragen nicht zeitnah bearbeitet wurden, wurde

die Abfrage automatisch geschlossen. In der morgendlichen Abfrage wurden die Studierenden gefragt, ob ihnen am Vorabend eine oder mehrere Podcastepisode(n) empfohlen worden war(en), und was die Inhalte dieser Podcastepisode(n) waren; gegebenenfalls sollten ein oder zwei Stichwörter pro Episode genannt werden. Des Weiteren wurde erfragt, wie die Inhalte aus der/den Podcastepisode(n) für den aktuellen Lerntag nutzbar sei(en); hier sollten die Teilnehmenden konkrete Maßnahmen nennen oder angeben, welches Ziel sie sich dafür setzten.

Tabelle 1: Auflistung der Aussagen zu aktuellen Herausforderungen während des SRLP sowie Titel der zugeordneten Podcastepisode und der darin enthaltenen Lern- und SR-Strategien

Item	Podcastepisode	Lern- und SR-Strategien
Ich hatte noch keine Erfahrung mit dem Bearbeiten solcher Lernziele.	Vorerfahrung	Mindmapping, Ressourcenmindmap
Ich hatte noch kein Vorwissen zu dem Thema.	Elaboration	Mindmapping, Strukturlegetechnik, Aufklärung zum Thema
Ich hatte meine Zeit nicht gut geplant.	Zeitplanung	ALPEN Methode, Pausenplanung, 60:40-Regel
Ich wusste nicht, wen ich um Hilfe fragen kann, als ich Schwierigkeiten hatte, mein Lernziel erfolgreich umzusetzen.	Hilfe suchen	Vorstellung der KIT-Anlaufstellen (im Interview)
Ich war nicht am Thema meines Lernziels interessiert.	Interesse	Visualisierung, Zielsetzung, Einsatz von Belohnung
Ich war nicht motiviert, mein Lernziel erfolgreich umzusetzen.	Motivation	Zielsetzung, Belohnung, Ablenkung reduzieren
Der Zeitdruck war zu groß, um mein Lernziel erfolgreich umzusetzen.	Metakognitive Organisation	Planungsstrategien, Eisenhower- Methode, Paretoprinzip
Die Angst war zu groß, um mein Lernziel erfolgreich umzusetzen.	Emotionsregulation	Aufklärung zu Stressmechanismen, mentale Entspannungstechniken
Ich habe mich nicht genug angestrengt, um mein Lernziel erfolgreich umzusetzen.	Aufwandssteuerung	Powerpose, Meilensteinplanung, Erfolge dokumentieren
Ich hatte noch nicht alle Fähigkeiten und Fertigkeiten, um mein Lernziel erfolgreich umzusetzen.	Kompetenzerleben	Mindmapping mit Erfolgen, Lebenslauf erstellen

6.2 Operationalisierung

6.2.1 SW (state)

Die SW wurde entlang des SRLP (an jedem Lerntag in Phase 0 und Phase1) während der Bearbeitung der von den Probanden selbstformulierten Lernziele erfasst. Die Studierenden schätzten an jedem Lerntag in der Abendabfrage (postaktionale Phase) ihre erlebte SW im SRLP unter Nutzung der elektronischen Tagebücher ein (s. Abschnitt 6.1.1). Hierzu wurde, angelehnt an die Empfehlungen von Bandura (2003) zur Erfassung der SW, zum Einstieg die übergeordnete Frage gestellt „Woran machen Sie Ihre Einschätzung, wie erfolgreich Sie Ihr Lernziel umsetzen konnten, fest?“. Die Beantwortung der Frage erfolgte über die Bewertungen von zehn Aussagen zur erlebten SW beim Einsatz von SRL-Strategien (s. Tabelle 2) auf siebenstufigen Likertskalen (Ankerpunkte „1 = gar nicht“ und „7 = vollkommen“). Als Komponenten der SW im SRLP wurden metakognitive, ressourcenbezogene, motivationale

und volitionale Strategien berücksichtigt. Hierfür wurden Items aus standardisierten Messinstrumenten, basierend auf den Überlegungen von Rost et al. (2008), für die Feld-Forschung ökonomisch sinnvolle und präzise Ein-Item Skalen zu entwickeln, übernommen und für die verhaltensnahe Erfassung mittels AA modifiziert. Der SW-Index wurde auf der Basis des Motivated Strategies for Learning Questionnaire (MSLQ) und des daraus resultierenden Fragebogens zu Lernstrategien im Studium (LIST) konzipiert (Pintrich et al., 1991; Wild & Schiefele, 1994). Die Skalen wurden auf je ein Item reduziert, und die Formulierung wurde für die Bewertung im Rahmen eines aktuellen Lernprozesses mit selbst formuliertem Lernziel angepasst. Zur Indexbildung wurden die Skalenwerte aller zehn Angaben aufsummiert, wobei die Items zu metakognitiver Organisation und Emotionsregulation negativ formuliert waren und vorher umgepolt werden mussten. Der SW-Index kann Ausprägungen zwischen 10 und 70 annehmen. Ein hoher Indexwert steht für eine hohe Ausprägung der aktuell erlebten SW im SRLP. Die interne Konsistenz des Index ist zufriedenstellend ($\alpha = .73$).

Tabelle 2: Kategorien der SW im SRLP und das jeweils zugehörige Item

Kategorie	Item
Vorerfahrung	Ich hatte schon Erfahrung mit dem Bearbeiten solcher Lernziele.
Elaboration	Ich hatte schon viel Wissen zu dem Thema.
Zeitplanung	Ich habe meine Zeit gut geplant.
Hilfe suchen	Ich habe jemanden um Hilfe gefragt, als ich Schwierigkeiten hatte, mein Lernziel erfolgreich umzusetzen.
Interesse	Ich war sehr am Thema meines Lernziels interessiert.
Motivation	Ich war motiviert, mein Lernziel erfolgreich umzusetzen.
Metakognitive Organisation	Ich musste mein Lernziel erfolgreich umsetzen, da ich unter Zeitdruck stehe.
Emotionsregulation	Ich hatte Angst, mein Lernziel nicht erfolgreich umzusetzen.
Aufwandssteuerung	Ich konnte mein Lernziel erfolgreich umsetzen, weil ich mich genug angestrengt habe.
Kompetenzerleben	Ich konnte mein Lernziel erfolgreich umsetzen, weil ich alle dazu nötigen Fähigkeiten und Fertigkeiten habe.

6.2.2 SWE (trait)

Die SWE wurde vor Beginn von Phase 0 und nach Ende von Phase 1 bei KG und IG über die Skala zur Allgemeinen SWE von Schwarzer und Jerusalem (1999) erhoben. Die Allgemeine SWE Skala erfasst eine allgemeine optimistische Selbstüberzeugung, besteht aus zehn Items und geht auf die Theorie von Bandura zurück (Bandura, 1982; Schwarzer & Jerusalem, 1999). Die Validität der Skala wurde in mehreren deutschen Stichproben überprüft (Hinz et al., 2006). Die Cronbachs alpha Werte liegen zwischen .78 und .79. Zur Selbsteinschätzung wird ein vierstufiges Antwortformat mit den Antwortmöglichkeiten „stimmt nicht“, „,,stimmt kaum“, „stimmt eher“, „stimmt genau“ vorgegeben. Daraus ergibt sich ein Gesamtwert zwischen zehn und 40 Punkten. Eine hohe Punktzahl steht für eine optimistische Selbstüberzeugung.

6.2.3 Lernerfolg und Lernzufriedenheit im SRLP

Zum erlebten Lernerfolg beantworteten die Probanden an jedem Lerntag unter Nutzung der elektronischen Tagebücher die folgende Frage über eine siebenstufige Ratingskala: „Wie erfolgreich haben Sie heute an Ihrem Lernziel gearbeitet?“ Die Ankerpunkte waren beschriftet mit „1 = gar nicht erfolgreich“ und „7 = sehr erfolgreich“. Ihre subjektive Zufriedenheit mit der Arbeit an ihrem Lernziel bewerteten sie (jeweils für den aktuellen Lerntag) über eine siebenstufigen Likertskala mit den Ankerpunkten „1 = gar nicht zufrieden“ und „7 = vollkommen zufrieden“.

6.3 Stichprobe

Von den ursprünglichen 114 Studienteilnehmenden konnten drei Personen der IG nicht berücksichtigt werden, da aus technischen Gründen keine Interventionen durch das IAA ausgelöst wurden. Fünf weitere Datensätze konnten nicht in die Auswertung einbezogen werden, da drei Studierende ihr Einverständnis zur Datennutzung nicht gegeben hatten und zwei weitere Personen die Studiensmartphones nicht im vorgesehenen Zeitfenster zurückgebracht hatten. Von den 106 in der Stichprobe verbleibenden Studierenden waren die Daten vollständig. 51 Personen (25 Frauen, 26 Männer) wurden der IG und 55 Personen (26 Frauen, 29 Männer) der KG randomisiert zugeordnet. Die Studierenden waren zwischen 18 und über 30 Jahre alt. Am häufigsten war die Altersgruppe zwischen 22 und 24 Jahren vertreten. Die meisten Studierenden kamen aus dem Bereich Lehramt (s. Tabelle 3).

Tabelle 3: Verteilung der Teilnehmenden nach Studienabschluss

Abschluss	Kontrollgruppe	Interventionsgruppe
Bachelor of Arts/Science	25%	33%
Master of Arts/Science	22%	23%
Bachelor of Education	15%	14%
Master of Education	25%	18%
Staatsexamen	13%	12%

6.4 Statistische Analysen

Die Datenauswertung erfolgte mit der Statistiksoftware IBM SPSS Statistik 27 für Windows. Die AA Daten wurden (zur Überprüfung von Hypothese 1 und für die explorativen Analysen) mittels linearer Mehrebenenanalysen über zwei Ebenen modelliert. Die Anwendung von Mehrebenenanalysen liefert adäquate Modelle zum Umgang mit fehlenden Daten und ermöglicht den Einbezug mehrerer Erfassungszeitpunkte pro Person, wobei sich die Anzahl der Messungen zwischen den Personen unterscheiden darf und keine Äquidistanz der Messzeitpunkte vorausgesetzt wird (vgl. Hoffman & Rovine, 2007; Nezlek, 2012; Singer & Willett, 2003). Die Lerntage bildeten in der vorliegenden Studie die erste Ebene (Level 1), die einzelnen Personen die zweite Ebene (Level 2). Die Gruppenzugehörigkeit (KG/IG) und die Erhebungsphase (Phase 0/Phase 1) wurden als Prädiktoren eingesetzt. Zur Analyse der Interventionseffekte auf die SWE (trait) wurde (zur Überprüfung von Hypothese 2) eine zweifak-

torielle ANOVA mit Messwiederholung mit den Faktoren Gruppe (IG/KG) und Messzeitpunkt (prä/post) durchgeführt.

7 Ergebnisse

7.1 Compliance, deskriptive Angaben zur Intervention, ICCs

Alle Probanden erfüllten die vorgegebene Registrierdauer. Diese betrug mindestens 16 Lern- tage AA (Phase 0 und Phase 1) für die KG und zehn Lern- tage für die IG (Phase 0); alle Probanden der IG absolvierten zusätzlich die elftägige Interventionsphase mittels IAA (Phase 1). Die Podcastepisode zum Thema *Ressourcen aktivieren* wurde keiner Person vorgeschlagen und somit nicht angehört. Die anderen Episoden wurden von der IG zwischen elf Mal (Episode zum Kompetenzerleben) und 60 Mal (Episode zur Elaboration) angehört; prinzipiell war es möglich, eine Episode mehrmals anzuhören. Circa 20% der MC-Fragen wurden nicht beantwortet. Die Quote korrekter Antworten bei den MC-Fragen lag im Schnitt bei 73%.

Die ICCs des un konditionierten Modells wurden für die Baselinephase (Phase 0) über alle Probanden berechnet, um für die dynamisch erfassten Konstrukte die intraindividuelle Varianz von den Varianzanteilen, die aufgrund zeitlich relativ stabiler Personenmerkmale zustande kamen, zu differenzieren (s. Tabelle 4).

Tabelle 4: Inter- und intraindividuelle Varianz für die SW in SRLP

Konstrukt	Interindividuelle Varianz	Intraindividuelle Varianz
Index SW im SRLP	42%	58%
Vorerfahrung	36%	64%
Elaboration	24%	76%
Zeitplanung	28%	72%
Hilfe suchen	42%	58%
Interesse	39%	61%
Motivation	30%	70%
Metakognitive Organisation	53%	47%
Emotionsregulation	54%	46%
Aufwandssteuerung	28%	72%
Kompetenzerleben	34%	66%
Lernzufriedenheit	18%	82%
Lernerfolg	17%	83%

7.2 Interventionseffekte

Die Ergebnisse der Mehrebenenanalyse bestätigen eine positive Wirkung der Intervention auf die SW im SRLP. In den statistischen Kennwerten (s. Tabelle 5) zeigt sich der Interventionseffekt durch den signifikanten Interaktionseffekt der Prädiktoren Gruppe \times Phase und den Anstieg der SW bei der IG in Phase 1, in Relation zur KG und zu Phase 0, was sich am positiven Vorzeichen des Koeffizienten des SW-Index ablesen lässt. Hypothese 1 konnte somit angenommen werden.

Die Varianzanalyse mit Messwiederholung zur Prüfung von Hypothese 2 zeigt für die SWE (trait) einen signifikanten Anstieg bei der gesamten Stichprobe (Haupteffekt Messzeitpunkt) bei der Erfassung nach der Intervention ($MW = 29.17$; $SD = 4.11$) im Vergleich zur Erfassung vor der Intervention ($MW = 28.29$; $SD = 5.22$), $F(1,6936) = 147.95$; $p < .0001$; $\eta^2 = .02$. Zusätzlich ergab sich ein signifikanter Interaktionseffekt Gruppe \times Messzeitpunkt, $F(1,6936) = 111.79$; $p < .0001$; $\eta^2 = .02$. Die deskriptiven Daten bilden eine stärkere Verbesserung der SWE bei der Prä-Post-Messung der IG (Prä: $MW = 27.79$; $SD = 4.19$; Post: $MW = 28.91$; $SD = 3.72$) im Vergleich zu den Messungen der KG (Prä: $MW = 29.29$; $SD = 4.70$; Post: $MW = 29.37$; $SD = 4.71$) ab und lassen somit (in Kombination mit dem signifikanten Interaktionseffekt) auf eine positive Wirkung der Interventionen schließen. Hypothese 2 gilt folglich als belegt.

7.3 Explorative Analysen

Zur Auswertung der Effekte der Intervention auf die einzelnen Komponenten der SW wurden Mehrebenenanalysen durchgeführt. Die Ergebnisse zeigen, dass die SW bezüglich der Anwendung der folgenden SRL-Strategien bei der IG innerhalb der Interventionsphase (in Relation zur KG und Baselinephase) gefördert werden konnte: Elaboration, Zeitplanung, Nutzen von Hilfsangeboten, Wecken von Interesse, Motivation, Aufwandssteuerung und Kompetenzerleben (vgl. signifikante Interaktionseffekte Gruppe \times Phase und positives Vorzeichen der Koeffizienten in Tabelle 5). Für metakognitive Organisation, Emotionsregulation und Vorerfahrung wurden keine bedeutsamen Veränderungen festgestellt (nicht signifikante Interaktionseffekte in Tabelle 5). Des Weiteren konnten die Studierenden der IG im Vergleich zur KG ihre Lernzufriedenheit und ihren Lernerfolg in der Interventionsphase in Relation zur Baselinephase bedeutsam steigern (s. signifikante Interaktionseffekte Gruppe \times Phase und positives Vorzeichen der Koeffizienten in Tabelle 5).

Tabelle 5: Ergebnisse der Mehrebenenanalysen

AV	Prädiktor	Koeffizient	SE	t	p
Index SW im SRLP	Gruppe	-2.102	1.319	-1.594	.133
	Phase	0.136	0.508	0.268	.789
	Gruppe \times Phase	3.610	0.725	4.957	<.0001
Vorerfahrung	Gruppe	0.112	0.190	0.588	.557
	Phase	-0.067	0.081	-0.827	.409
	Gruppe \times Phase	0.018	0.116	0.157	.875
Elaboration	Gruppe	-0.260	0.165	-1.575	.117
	Phase	0.077	0.087	0.881	.379
	Gruppe \times Phase	0.379	0.125	3.029	.002
Zeitplanung	Gruppe	-0.100	0.184	-0.543	.588
	Phase	-0.024	0.090	-0.264	.791
	Gruppe \times Phase	0.500	0.129	3.873	<.0001
Hilfe suchen	Gruppe	-0.271	0.288	-0.941	.349
	Phase	0.079	0.112	0.707	.480
	Gruppe \times Phase	0.769	0.161	4.769	<.0001
Interesse	Gruppe	-0.436	0.201	-2.165	.032
	Phase	-0.079	0.082	-0.971	.332
	Gruppe \times Phase	0.780	0.118	6.627	<.0001

AV	Prädiktor	Koeffizient	SE	t	p
Motivation	Gruppe	-0.116	0.191	-0.606	.545
	Phase	-0.781	0.090	-0.896	.385
	Gruppe × Phase	0.450	0.129	3.483	.001
Metakognitive Organisation	Gruppe	-0.314	0.303	-1.037	.302
	Phase	0.088	0.097	0.906	.365
	Gruppe × Phase	-0.168	0.139	-1.211	.226
Emotionsregulation	Gruppe	-0.297	0.249	-1.190	.237
	Phase	0.133	0.079	1.694	.091
	Gruppe × Phase	-0.052	0.113	-0.456	.649
Aufwandssteuerung	Gruppe	0.301	0.192	0.159	.874
	Phase	0.080	0.094	0.847	.397
	Gruppe × Phase	0.427	0.135	3.157	.002
Kompetenzerleben	Gruppe	-0.016	0.178	-0.088	0.930
	Phase	-0.073	0.079	-0.922	0.357
	Gruppe × Phase	0.454	0.113	4.009	<.0001
Lernzufriedenheit	Gruppe	-0.579	0.173	-3.345	.0001
	Phase	0.079	0.105	0.749	0.454
	Gruppe × Phase	0.482	0.141	3.410	0.001
Lernerfolg	Gruppe	-0.614	0.179	-3.433	0.001
	Phase	0.175	0.110	0.159	0.874
	Gruppe × Phase	0.509	0.147	3.455	0.001

8 Diskussion

8.1 Interventionseffekte

Die vorliegende Studie verfolgte das Ziel, die erlebte SW Studierender während des SRL unmittelbar im Lernalltag zu untersuchen und durch Interventionen mit Podcastepisoden zu fördern. Die Mehrebenenanalyse belegt einen Anstieg der aktuell erlebten SW im Lernprozess (SW-Index) für die IG in der Interventionsphase im Vergleich zur Baselinephase und in Relation zur KG (vgl. Hypothese 1). Im Einklang mit den Ergebnissen zur State-Komponente zeigten die Interventionen eine positive Wirkung auf die studentische SWE (trait) im Prä-Post-Vergleich der beiden Untersuchungsgruppen (vgl. Hypothese 2). Die IG wies zu Studienbeginn eine geringere SWE als die KG auf. Dieser Gruppenunterschied konnte durch die Intervention signifikant verringert werden, indem die IG ihre SWE stärker verbessern konnte als die KG. Die bei der KG beobachtete leichte Verbesserung der SWE im Prä-Post-Vergleich ist möglicherweise mit einer Intensivierung von Selbstreflexionsmechanismen aufgrund der wiederholten Tagebucheingaben, die kognitive und behaviorale Veränderungen nach sich ziehen können, erklärbar. Dieser Effekt wurde bereits in früheren Studien mit elektronischen Tagebüchern beobachtet (Dignath-van Ewijk et al., 2015; Loeffler, Bohner et al., 2019). Führt man beide Ergebnisse zusammen, zeichnen sich die positiven Effekte der Intervention sowohl für das stabile Persönlichkeitsmerkmal als auch für die kontextspezifische SW ab.

In der vorliegenden Arbeit wurden in Anlehnung an das Konzept der SW im SRLP, das die Zuversicht zum sinnvollen Einsatz von SRL-Strategien als Voraussetzung erachtet, um akademische Erfolge zu erzielen, zusätzlich die Auswirkungen der Interventionen auf die

erlebte SW beim Einsatz verschiedener SRL-Strategien geprüft. Positive Effekte konnten für Elaboration, Zeitplanung, Hilfe suchen, Interesse wecken, Motivation steigern, Aufwandssteuerung und Kompetenzerleben nachgewiesen werden.

Die positiven Interventionseffekte auf die SW/SWE stehen im Einklang mit den Ergebnissen vorheriger Arbeiten (Bandura, 1994; Bergey et al., 2019; Chan & Lam, 2010; Gessnitzer et al., 2016; Hawe et al., 2019; Jerusalem & Hopf, 2002; Schnell et al., 2015; Usher & Pajares, 2008b). Nach unserem Wissen ist die vorliegende Studie jedoch die erste, die Podcasts als Intervention zur expliziten Förderung von SW im SRLP mithilfe von IAA einsetzt, was eine technikgestützte, automatisierte und dennoch individualisierte Unterstützung unmittelbar im Lernalltag der Studierenden ermöglicht.

Die beiden SW-Komponenten *metakognitive Organisation* und *Emotionsregulation* konnten nicht durch die Intervention gefördert werden. Dies kann daran liegen, dass die Bewältigung von Stress und negativen Emotionen durch komplexe Zusammenhänge der eigenen Biographie entstanden ist und nicht hinreichend durch den Einsatz kurzer technikgestützter Interventionen korrigiert werden kann. Gerade bei Prüfungsangst kann von einer vielschichtigen Herausforderung gesprochen werden, bei der die kognitiven Prozesse nur eine Ebene des Problems darstellen (Lowe, 2021). Bei einer tiefgreifenden Prüfungsangst reicht eine einmalige Intervention in Form einer Podcastepisode ohnehin nicht aus. Der niederschwellige technikgestützte Ansatz kann in diesem Fall eher als Anhaltspunkt für die betreffende Person dienen, dass möglicherweise ein individuelles Problem vorliegt, an dem zukünftig (gegebenenfalls mit professioneller Unterstützung) gearbeitet werden sollte. Das Fehlen adäquater Strategien beim Umgang mit Zeitdruck (*metakognitive Organisation*) kann mit akademischer Prokrastination einhergehen und benötigt möglicherweise spezielle, stärker ausdifferenzierte Trainingsangebote (vgl. Loeffler, Stumpp et al., 2019). Darüber hinaus ist es auch möglich, dass die Erfassung von Interventionseffekten auf diese beiden Konstrukte durch jeweils ein Item nicht ausreichend war.

Als Bandura das Konzept der SW im Jahr 1994 vorstellt, weist er bereits auf die Dynamik und die besondere kontextübergreifende Relevanz des Konstruktes hin (Bandura, 1994). In den 90er- und 2000er-Jahren wurde SW häufig im Zusammenhang mit Lernerfolg und Lernzufriedenheit untersucht. Die Ergebnisse zeigen, dass eine hohe SW positive Effekte auf Lernerfolg und Lernzufriedenheit hat, wobei gleichzeitig die Orientierung an guten Lernleistungen an Relevanz verliert (Jerusalem & Hopf, 2002; Kitsantas et al., 2008; Usher & Pajares, 2008b). Negative Einflüsse auf die Lernleistung, wie sie bei Vancouver und Kendall (2006) vorgestellt wurden, konnten jedoch nicht repliziert werden. In den bisher durchgeführten Untersuchungen wurde die dynamische Komponente der SW meist nicht beachtet. In der vorliegenden Studie wurde nun explizit die Dynamik des SRLP berücksichtigt; hierbei ergaben sich ebenfalls positive Effekte der Interventionen zur Förderung der SW auf die Lernzufriedenheit und den subjektiven Lernerfolg der Studierenden. Darüber hinaus zeigen die ICCs, dass es sich bei 46% bis 83% der Varianz in den untersuchten Konstrukten um intraindividuelle Varianz handelt, was die Relevanz situativer Einflüsse belegt und Forschungs- und Interventionsansätze in Echtzeit im Alltag der Probanden nahelegt.

8.2 Limitationen

Aufgrund der signifikanten Interventionseffekte sowohl hinsichtlich der State- als auch der Trait-Komponente der SW/SWE im Kontrollgruppendesign mit Prä-Post- beziehungsweise Verlaufsmessung und randomisierter Gruppenzuordnung sowie aufgrund der sehr zufriedenstellenden Compliance-Werte der über 100 teilnehmenden Studierenden gehen wir davon aus, dass die zeitlich begrenzte kompakte Intervention tatsächlich zu einer positiveren SW/SWE-Entwicklung (entsprechend Hypothese 1 und 2) der IG geführt hat.

Nichtsdestotrotz sind folgende Limitationen der vorliegenden Studie zu nennen. Mit dem Ziel, die Compliance der KG (trotz fehlender Interventionen) in der zweiten Registrierphase aufrechterhalten zu können, wurde die Registrierdauer der Kontrollphase für die KG verkürzt. Aufgrund der Auswertung mit Mehrebenenanalysen ist dies aus methodischer Sicht zwar unproblematisch, trotzdem ist nicht auszuschließen, dass die unterschiedliche Phasenlänge sich auf die Eintragungen in die elektronischen Tagebücher in irgendeiner Form ausgewirkt haben könnte.

Durch den verwendeten Interventionsansatz kann nicht eindeutig geklärt werden, ob die SWE/SW durch das Ansprechen der vier Quellen zur Förderung der SW (Bandura, 1994) gesteigert wurde. Es ist auch möglich, dass das Lernstrategietraining und/oder das Feedback zu den MC-Fragen zur Intensivierung der SWE/SW geführt hat. Die Vermutung liegt nahe, dass eine Kombination der drei Komponenten, vor allem auch aufgrund der angenommenen Wechselwirkungen, am wirkungsvollsten sein könnte: Die Studierenden erleben sich durch die Förderung der SW, das Strategietraining und die Feedbackmaßnahmen als selbstwirksam, was sich positiv auf den Strategieeinsatz und den Lernerfolg auswirken kann, was wiederum die SW und SWE fördert. Um die Effekte experimentell differenzieren zu können, wäre ein Ansatz mit mehreren Interventionsgruppen notwendig gewesen, was aufgrund der Probandenanzahl in der vorliegenden Arbeit nicht realisiert werden konnte.

Ob die Inhalte der Podcastepisoden den Probanden tatsächlich präsent waren, wurde mithilfe von MC-Fragen und Transferfragen überprüft. Die MC-Fragen wurden vorab nicht pilotiert, sodass hohe Fehlerquoten auch ein Resultat eines zu hohen Schwierigkeitsgrads sein können. Circa 20% der MC-Fragen wurden nicht beantwortet. Dies kann ein Hinweis darauf sein, dass die Podcasts nicht aufmerksam gehört, sondern lediglich während einer anderen Aktivität abgespielt wurden. Ob die Anzahl der gehörten Podcastepisoden oder die Anzahl korrekt beantworteter Fragen pro Episode einen Einfluss auf die Interventionseffekte hatte, konnte aufgrund der zu geringen Gruppengrößen der resultierenden Subgruppen in der vorliegenden Studie nicht geprüft werden.

In der ersten Podcastepisode wurde das Thema, bisherige Erfahrungen mit ähnlichen Lernzielen/Aufgaben in den aktuellen Lernprozess einfließen zu lassen, behandelt. Keine der befragten Personen gab dies als Herausforderung für den aktuellen Lernprozess an, sodass davon ausgegangen werden kann, dass einerseits alle Studierenden bereits Erfahrungen mit ihren selbst definierten Lernzielen gemacht hatten, andererseits jedoch, aufgrund der Konzeptionierung des Algorithmus, auch keine Anregungen für neue Lernerfahrungen in diesem Bereich erhielten.

9 Fazit

Die Podcastepisoden sind als niederschwelliger technikgestützter Ansatz anzusehen, der mittels eines vertrauten und alltagsnahen Mediums Studierenden Unterstützung bei der Gestaltung des eigenen Lernprozesses bieten kann. Der Einsatz von IAA als eine Form von Learning Analytics ermöglicht darüber hinaus automatisierte, dynamische und personalisierte Interventionen in Echtzeit. Damit geht dieser Ansatz über die bereits bestehenden technikgestützten Programme hinaus, die entweder in künstlichen Settings eingesetzt wurden oder den Transfer von SR-Strategien, welche zuvor in Einzel- oder Gruppentrainings erlernt wurden, im Alltag der Studierenden (z.B. durch Prompting-Maßnahmen) unterstützten. Insgesamt kann durch die vielfältigen positiven Effekte, die bei der IG beobachtet wurden, davon ausgegangen werden, dass es sich hier um einen innovativen und vielversprechenden Ansatz zur Förderung der SW in studentischen Lernprozessen handelt.

Danksagung

Die Studien wurden durch das *Bundesministerium für Bildung und Forschung* (BMBF) gefördert (01PL12004).

Literatur

- Alghamdi, A., Karpinski, A. C., Lepp, A. & Barkley, J. (2020). Online and face-to-face classroom multitasking and academic performance: Moderated mediation with self-efficacy for self-regulated learning and gender. *Computers in Human Behavior*, 102, 214–222. <https://doi.org/10.1016/j.chb.2019.08.018>
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122–147.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive view*. Prentice-Hall. <https://doi.org/10.5465/amr.1987.4306538>
- Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Hrsg.), *Encyclopedia of human behaviour* (Vol. 4) (S. 71–81). Academic Press.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
- Bandura, A. & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology*, 88(1), 87–99. <https://doi.org/10.1037/0021-9010.88.1.87>
- Bellhäuser, H., Lösch, T., Winter, C. & Schmitz, B. (2016). Applying a web-based training to foster self-regulated learning – Effects of an intervention for large numbers of participants. *The Internet and Higher Education*, 31(4), 87–100. <https://doi.org/10.1016/j.iheduc.2016.07.002>
- Bergey, B. W., Parrila, R. K., Laroche, A. & Deacon, S. H. (2019). Effects of peer-led training on academic self-efficacy, study strategies, and academic performance for first-year university students with and without reading difficulties. *Contemporary Educational Psychology*, 56, 25–39. <https://doi.org/10.1016/j.cedpsych.2018.11.001>
- Blüthmann, I. (2012). Individuelle und studienbezogene Einflussfaktoren auf die Zufriedenheit von Bachelorstudierenden. *Zeitschrift Für Erziehungswissenschaft*, 15(2), 273–303. <https://doi.org/10.1007/s11618-012-0270-3>

- Borgonovi, F. & Pokropek, A. (2019). Seeing is believing: Task-exposure specificity and the development of mathematics self-efficacy evaluations. *Journal of Educational Psychology*, 111(2), 268–283.
- Breitenbach, A. (2021). *Digitale Lehre in Zeiten von Covid-19: Risiken und Chancen*. <https://doi.org/10.25656/01:21274>
- Chan, J. C. Y. & Lam, S. (2010). Effects of different evaluative feedback on students' self-efficacy in learning. *Instructional Science*, 38(1), 37–58. <https://doi.org/10.1007/s11251-008-9077-2>
- Dignath-van Ewijk, C., Fabriz, S. & Buettner, G. (2015). Fostering self-regulated learning among students by means of an electronic learning diary: A training experiment. *Journal of Cognitive Education and Psychology*, 14, 77–97.
- Doran, G. T. (1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review*, 70, 35–36.
- Fernandez, V., Simo, P. & Sallan, J. M. (2009). Podcasting: A new technological tool to facilitate good practice in higher education. *Computers & Education*, 53(2), 385–392. <https://doi.org/10.1016/j.compedu.2009.02.014>
- Flett, G. L., Blankstein, K. R. & Martin, T. R. (1995). Procrastination, negative self-evaluation, and stress in depression and anxiety: A review and preliminary model. In J. R. Ferrari, J. L. Johnson & W. G. McCown (Hrsg.), *Procrastination and task avoidance* (S. 137–167). Springer US.
- Gabriel, R., Gersch, M., Weber, P. & Venghaus, C. (2007). Blended Learning Engineering: Der Einfluss von Lernort und Lernmedium auf Lernerfolg und Lernzufriedenheit – Eine evaluationsgestützte Untersuchung. In M. H. Breitner, B. Bruns & F. Lehner (Hrsg.), *Neue Trends im E-Learning* (S. 75–92). Physica-Verlag HD. https://doi.org/10.1007/978-3-7908-1922-9_6
- Galyon, C. E., Blondin, C. A., Yaw, J. S., Nalls, M. L. & Williams, R. L. (2012). The relationship of academic self-efficacy to class participation and exam performance. *Social Psychology of Education*, 15(2), 233–249. <https://doi.org/10.1007/s11218-011-9175-x>
- Gessnitzer, S., Schulte, E.-M. & Kauffeld, S. (2016). „I am going to succeed“: The power of self-efficient language in coaching and how coaches can use it. *Consulting Psychology Journal: Practice and Research*, 68(4), 294–312. <https://doi.org/10.1037/cpb0000064>
- Gnauer, D. & Huttel, H. (2016). *Podcasting for teaching and learning in higher education* (Higher Education Practices Series No. 2). Aalborg Universitetsforlag.
- Hasselhorn, M. & Gold, A. (2017). *Pädagogische Psychologie: Erfolgreiches Lernen und Lehren* (4. Aufl.). W. Kohlhammer.
- Hawe, E., Lightfoot, U. & Dixon, H. (2017). First-year students working with exemplars: promoting self-efficacy, self-monitoring and self-regulation. *Journal of Further and Higher Education*, 11(2), 1–15. <https://doi.org/10.1080/0309877X.2017.1349894>
- Hinz, A., Schumacher, J., Albani, C., Schmid, G. & Brähler, E. (2006). Bevölkerungsrepräsentative Normierung der Skala zur Allgemeinen Selbstwirksamkeitserwartung. *Diagnostica*, 52(1), 26–32. <https://doi.org/10.1026/0012-1924.52.1.26>
- Hoffman, L. & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, 39, 101–117. <https://doi.org/10.3758/BF03192848>
- Jerusalem, M. & Hopf, D. (2002). Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen. *Zeitschrift für Pädagogik* (44. Beiheft).
- Kállay, E. (2012). Learning strategies and metacognition awareness as predictors of academic achievement in a sample of Romanian second-year students. *Cognition, Brain, Behavior. An Interdisciplinary Journal*, 16, 369–385.
- Kim, K. R. & Seo, E. H. (2015). The relationship between procrastination and academic performance: A meta-analysis. *Personality and Individual Differences*, 82, 26–33. <https://doi.org/10.1016/j.paid.2015.02.038>
- Kitsantas, A., Winsler, A. & Huie, F. (2008). Self-regulation and ability predictors of academic success during college: A predictive validity study. *Journal of Advanced Academics*, 20(1), 42–68.

- Krapp, A. (1993). Die Psychologie der Lernmotivation: Perspektiven der Forschung und Probleme ihrer pädagogischen Rezeption. *Zeitschrift für Pädagogik*, 39(2), 187–206.
- Lee, M. J. & Chan, A. (2007). Reducing the effect of isolation and promoting inclusivity for distance learners through podcasting. *Turkish Online Journal of Distance Education*, 8(1), 85–105.
- Liborius, P., Bellhäuser, H. & Schmitz, B. (2019). What makes a good study day? An intraindividual study on university students' time investment by means of time-series analyses. *Learning & Instruction*, 60, 310–321. <https://doi.org/10.1016/j.learninstruc.2017.10.006>
- Loeffler, S. N., Bohner, A., Stumpp, J., Limberger, M. F. & Gidion, G. (2019). Investigating and fostering self-regulated learning in higher education using interactive ambulatory assessment. *Learning and Individual Differences*, 71(4), 43–57. <https://doi.org/10.1016/j.lindif.2019.03.006>
- Loeffler, S. N., Stumpp, J., Grund, S., Limberger, M. F. & Ebner-Priemer, U. W. (2019). Fostering self-regulation to overcome academic procrastination using interactive ambulatory assessment. *Learning and Individual Differences*, 75(4), 101760. <https://doi.org/10.1016/j.lindif.2019.101760>
- Lowe, P. A. (2021). The test anxiety measure for college students-short form: Development and examination of its psychometric properties. *Journal of Psychoeducational Assessment*, 39(2), 139–152. <https://doi.org/10.1177/0734282920962947>
- Ludwig, P. H., Finkbeiner, C. & Knierim, M. (2013). Effects of the adequacy of learning strategies in self-regulated learning settings: A video-based micro-analytical lab study. *Journal of Cognitive Education and Psychology*, 12(3), 374–390.
- Marczuk, A., Multrus, F. & Lörz, M. (2021). *Die Studiensituation in der Corona-Pandemie. Auswirkungen der Digitalisierung auf die Lern- und Kontaktsituation von Studierenden*. Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). https://doi.org/10.34878/2021.01.DZHW_BRIEF
- Martin, P.-Y. & Nicolaisen, T. (2015). Einführung und Grundlagen. In P.-Y. Martin & T. Nicolaisen (Hrsg.), *Lernstrategien fördern* (S. 9–70). Beltz Juventa.
- Morris, N. P. (2010). Podcasts and mobile assessment enhance student learning experience and academic performance. *Bioscience Education*, 16(1), 1–7. <https://doi.org/10.3108/beej.16.1>
- Müller, Florian, H. (2006). Interesse und Lernen. *Report – Zeitschrift für Weiterbildungsforschung*, 29(1), 48–62.
- Nezlek, J.B. (2012). Multilevel modeling analyses of diary-style data. In M.R. Mehl & T.S. Conner (Hrsg.), *Handbook of research methods for studying daily life* (S. 357–383). Guilford.
- Pfleging, S. & Gerhardt, C. (2013). Ausgebrannte Studierende: Burnout-Gefährdung nach dem Bologna-Prozess. *Journal of Business and Media Psychology*, 4(1), 1–12.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Hrsg.), *Handbook of self-regulation* (S. 451–502). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50043-3>
- Pintrich, P. R., Smith, D. A. F., Garcia, T. & Mckeachie, W. J. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. University of Michigan, Ann Arbor. Verfügbar unter: <http://files.eric.ed.gov/fulltext/ED338122.pdf>
- Powell, C. B. & Mason, D. S. (2013). Effectiveness of podcasts delivered on mobile devices as a support for student learning during general chemistry laboratories. *Journal of Science Education and Technology*, 22(2), 148–170. <https://doi.org/10.1007/s10956-012-9383-y>
- Pychyl, T. A., Lee, J. M., Thibodeau, R. & Blunt, A. (2000). Five days of emotion: An experience sampling study of undergraduate student procrastination. *Journal of Social Behavior and Personality*, 15, 239–254.
- Reuschenbach, B. & Funke, J. (2011). Ambulantes Assessment. In N. Birbaumer, D. Frey, J. Kuhl, W. Schneider & R. Schwarzer (Hrsg.), *Enzyklopädie der Psychologie* (S. 529–594). Hogrefe.
- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R. & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130(2), 261–288. <https://doi.org/10.1037/0033-2909.130.2.261>

- Roelle, J., Berthold, K. & Fries, S. (2011). Effects of feedback on learning strategies in learning journals. *International Journal of Cyber Behavior, Psychology and Learning*, 1(2), 16–30. <https://doi.org/10.4018/ijcbpl.2011040102>
- Roick, J. & Ringeisen, T. (2017). Self-efficacy, test anxiety, and academic success: A longitudinal validation. *International Journal of Educational Research*, 83, 84–93. <https://doi.org/10.1016/j.ijer.2016.12.006>
- Roick, J. & Ringeisen, T. (2018). Students' math performance in higher education: Examining the role of self-regulated learning and self-efficacy. *Learning and Individual Differences*, 65, 148–158. <https://doi.org/10.1016/j.lindif.2018.05.018>
- Rost, D. H., Sparfeldt, J. R. & Buch, S. R. (2008). Kann den Kürze Sünde sein? Erfassung schulfachspezifischer Interessen mit nur einem Item. In F. Hofmann, C. Schreiner & J. Thonhauser (Hrsg.), *Qualitative und quantitative Aspekte. Zu ihrer Komplementarität in der erziehungswissenschaftlichen Forschung* (S. 225–237). Waxmann.
- Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- Salanova, M., Lorente, L. & Martínez, I. M. (2012). The dark and bright sides of self-efficacy in predicting learning, innovative and risky performances. *The Spanish Journal of Psychology*, 15(3), 1123–1132. https://doi.org/10.5209/rev_sjop.2012.v15.n3.39402
- Schmidt, K., Allgaier, A., Lachner, A., Stucke, B., Rey, S., Frömmel, C. & Nückles, M. (2011). Diagnostik und Förderung selbstregulierten Lernens durch Self-Monitoring-Tagebücher. *Zeitschrift für Hochschulentwicklung*, 6(3), 246–269.
- Schmitz, B. & Schmitz, M. (2007). Einführung in die Selbstregulation. In M. Landmann & B. Schmitz (Hrsg.), *Selbstregulation erfolgreich fördern* (S. 9–18). W. Kohlhammer.
- Schnell, K., Ringeisen, T., Raufelder, D. & Rohrmann, S. (2015). The impact of adolescents' self-efficacy and self-regulated goal attainment processes on school performance – Do gender and test anxiety matter? *Learning and Individual Differences*, 38, 90–98. <https://doi.org/10.1016/j.lindif.2014.12.008>
- Schwarzer, R. & Jerusalem, M. (Hrsg.) (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen: Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Freie Universität Berlin.
- Schwarzer, R. & Jerusalem, M. (2002). Das Konzept der Selbstwirksamkeit. *Zeitschrift für Pädagogik*, 44, 28–53.
- Singer, J.D. & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.
- Sitzmann, T., Ely, K., Brown, K. G. & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education*, 9(2), 169–191.
- Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, 133(1), 65–94. <https://doi.org/10.1037/0033-2909.133.1.65>
- Usher, E. L. & Pajares, F. (2008a). Self-efficacy for self-regulated learning. *Educational and Psychological Measurement*, 68(3), 443–463. <https://doi.org/10.1177/0013164407308475>
- Usher, E. L. & Pajares, F. (2008b). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research*, 78(4), 751–796. <https://doi.org/10.3102/0034654308321456>
- Vancouver, J. B. & Kendall, L. N. (2006). When self-efficacy negatively relates to motivation and performance in a learning context. *Journal of Applied Psychology*, 91(5), 1146–1153. <https://doi.org/10.1037/0021-9010.91.5.1146>
- Wild, K. P. & Schiefele, U. (1994). Lernstrategien im Studium. Ergebnisse zur Faktorstruktur und Reliabilität eines neuen Fragebogens. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 15, 185–200.

- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Hrsg.), *Handbook of self-regulation* (S. 13–39). Academic Press.
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–70.
- Zimmerman, B. J., Bandura, A. & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, 29(3), 663–676.

Kontakt

Dr. Caroline Götz · Paul Erhardt
Lernlabor, House of Competence
PD Dr. Simone N. Löffler
Lernlabor, House of Competence und Institut für Schulpädagogik und Didaktik
Karlsruher Institut für Technologie
Fritz-Erler-Straße 23
76133 Karlsruhe
E-Mail: kontakt@caroline-goetz.de
E-Mail: paul.erhardt9@kit.edu
E-Mail: simone-nadine.loeffler@kit.edu

Kompetenzentwicklung in der Lehrkräftebildung: der Mehrwert einer digitalen Simulation

Lea Grotegut, Katrin B. Klingsieck

Zusammenfassung: Die Förderung diagnostischer Kompetenz in der Lehrer*innenbildung birgt zwei Herausforderungen: Erstens bleibt in großen Lehrveranstaltungen wenig Raum für die notwendige Anwendung theoretischen Wissens, zweitens können Fehlentscheidungen in realen diagnostischen Situationen gravierende negative Folgen für Schüler*innen haben. Eine digitale Simulation kann beiden Herausforderungen begegnen. In diesem Beitrag wurde eine digitale, problemorientierte Simulation zur Förderung diagnostischer Kompetenz von Lehramtsstudierenden evaluiert und mit einer textbasierten, problemorientierten und einer lehrendenzentrierten Kontrollgruppe verglichen. Die Ergebnisse zeigen positive Effekte für die digitale Simulation hinsichtlich der selbstbestimmten Motivation, aber einen geringeren Wissenszuwachs im Vergleich zum lehrendenzentrierten Setting. Die Selbstwirksamkeit nahm in allen drei Gruppen zu. Eine digitale Simulation bietet in der untersuchten Form somit einen, wenn auch geringen, Mehrwert für die Lehrkräftebildung im Bereich der Kompetenzentwicklung.

Schlüsselwörter: Lehrkräftebildung, diagnostische Kompetenz, digitale Simulation, problemorientiertes Lernen

How can a digital simulation enrich educational processes in teacher education? An empirical study

Summary: Developing diagnostic competence is a central task in teacher education but faces two major challenges: First, there is little room in large university courses for the application of theoretical knowledge necessary for competence acquisition. Second, practicing diagnostic actions in real situations can have serious negative consequences in case of diagnostic errors. A digital simulation can meet both challenges and thereby enhance educational processes. This study evaluated a digital simulation to promote diagnostic competence in pre-service teachers, comparing it to two control groups: a problem-based setting and a lecture-based setting. Results show positive effects for the digital simulation regarding self-determined motivation, but a lower knowledge gain compared to the lecture. Self-efficacy increased equally in all three learning environments. Thus, in the form investigated in this study, a digital simulation offers at least a small amount of added value to teacher education in the area of competency development.

Keywords: teacher education, diagnostic competence, digital simulation, problem-based learning

1 Einleitung

In der universitären Lehrer*innenbildung ist der Erwerb professioneller Handlungskompetenz als Voraussetzung für professionelles Handeln im Unterricht (Baumert et al., 2011; Blömeke et al., 2022) eine zentrale Aufgabe. Als wesentlicher Aspekt professioneller Handlungskompetenz von Lehrkräften gilt die diagnostische Kompetenz (Südkamp & Praetorius, 2017), also die Fähigkeit, lern- und leistungsbezogene Merkmale von Schüler*innen korrekt einzuschätzen (F.-W. Schrader, 1989; Spinath, 2005) und relevante diagnostische Informationen zu ermitteln, um zu einem fundierten Urteil über verschiedene kognitive, affektive und verhaltensbezogene Aspekte zu gelangen (Praetorius & Südkamp, 2017). Diagnostische Urteile von Lehrkräften beeinflussen neben der schulischen auch die persönliche Entwicklung sowie die akademische und berufliche Laufbahn der Lernenden (Südkamp et al., 2012). Die Tragweite diagnostischer Entscheidung verdeutlicht die hohe Bedeutung, die der Ausbildung diagnostischer Kompetenz angehender Lehrkräfte zukommt. Diese gestaltet sich jedoch schwierig: Zum einen bleibt in Lehrveranstaltungen mit häufig großer Studierendenzahl wenig Raum für die für den Kompetenzerwerb notwendige Anwendung des theoretischen Wissens (Römer et al., 2018; Seifert et al., 2018). Zum anderen kann die Erprobung diagnostischer Handlungen in Praktika im Falle diagnostischer Fehlentscheidungen gravierende negative Konsequenzen für Schüler*innen nach sich ziehen.

Dieser Beitrag stellt eine digitale, problemorientierte Simulation zur Förderung diagnostischer Kompetenz von Lehramtsstudierenden vor. Der Einsatz einer solchen Simulation kann beiden Herausforderungen begegnen, indem er das Erproben diagnostischer Handlungen in möglichst realitätsnahen Situationen, jedoch ohne reale pädagogische Konsequenzen ermöglicht. Ziel des Beitrags ist es, die Effektivität der Simulation mit der eines textbasierten, problemorientierten Settings sowie eines lehrendenzentrierten Settings anhand der Ergebnisse eines Quasi-Experiments zu vergleichen. So soll die Frage beantwortet werden, inwiefern eine digitale Simulation einen Mehrwert für die Lehrkräftebildung im Bereich der Kompetenzentwicklung bieten kann.

2 Konzeptioneller Rahmen

2.1 Die Notwendigkeit der Förderung diagnostischer Kompetenz

Diagnostische Entscheidungen können weitreichende Konsequenzen unter anderem für die Zuweisung von Qualifikationen, die Bewilligung von Fördermaßnahmen sowie das Selbstkonzept und die akademische und persönliche Entwicklung von Schüler*innen haben (Südkamp et al., 2012). Umso erschreckender ist, dass diagnostische Urteile von Lehrkräften oft fehleranfällig sind (Karing & Artelt, 2013; Urhahne & Wijnia, 2021). Um Lehrkräfte zur Bildung möglichst akkurater Urteile über unterschiedliche Schüler*innenmerkmale zu befähigen, ist die Förderung von Dispositionen diagnostischer Kompetenz wie dem professionellen Wissen (Bartel & Roth, 2020) ein vielversprechender Ansatz. Es steht in engem Zusammenhang mit der Urteilsakkuratheit (Kramer et al., 2021) und bildet die Grundlage für die spätere Anwendung von Wissen (Heitzmann et al., 2018). Genau dort setzt der hier vorgestellte Ansatz an.

2.2 Didaktische Ansätze zur Kompetenzförderung

Der Erwerb deklarativen Wissens als Grundlage für den Aufbau diagnostischer Kompetenz kann, insbesondere wenn Lernende nur über ein geringes Vorwissen verfügen, besonders gut durch Methoden der direkten Instruktion erfolgen (Atkinson et al., 2000). Kennzeichnend für die direkte Instruktion ist, dass Lernende Problemlösungen nicht selbst generieren, sondern anhand von Beispielen oder Musterlösungen lernen, wie es in Vorlesungen der Fall ist. Vorlesungen und andere Methoden der direkten Instruktion scheinen indirekter Instruktion, bei der Lernende selbstständig Problemlösungen generieren, hinsichtlich des Wissenserwerbs überlegen zu sein (Pourghaznein et al., 2015). Ein weiterer Vorteil von Methoden direkter Instruktion ist, dass ihre Umsetzung im Vergleich zu Methoden indirekter Instruktion effizienter ist und weniger Ressourcen erfordert (McLaren et al., 2008). Sie eignen sich allerdings weniger gut als Methoden indirekter Instruktion dazu, das Behalten von Wissen (Klegeris & Hurren, 2011; Pourghaznein et al., 2015), Problemlösefähigkeiten (Hmelo, 1998) und kreatives Denken zu fördern (Martin et al., 2007). Außerdem scheinen Lernende, die bereits über Vorwissen verfügen, eher von Methoden indirekter Instruktion zu profitieren (Atkinson et al., 2000). Eine verbreitete Form indirekter Instruktion ist das problemorientierte Lernen (POL), bei dem Lernende selbstreguliert in Kleingruppen realitätsnahe Problemstellungen bearbeiten (Reusser, 2005). Neben den bereits genannten Vorteilen indirekter Instruktion wirkt sich POL im Vergleich zu direkter Instruktion positiv auf Einstellungen und Motivation der Lernenden aus (Demirel & Dağyar, 2016; Klegeris & Hurren, 2011). Motivation gilt als Schlüsselfaktor für erfolgreiches Lernen und hängt unmittelbar mit Lern- und Testergebnissen zusammen (Hsieh, 2014). Lernumgebungen, die autonomes Lernen unterstützen, fördern außerdem selbstbestimmte Formen von Motivation und werden mit besseren Lernergebnissen in Verbindung gebracht (Ryan & Deci, 2000; Vansteenkiste et al., 2004). Eine Herausforderung von POL ist die Generierung authentischer Problemstellungen, die für den erfolgreichen Einsatz von POL notwendig sind (van Merriënboer, 2013). Diese Schwierigkeit trifft vor allem auf POL mit textbasierten Problemdarstellungen zu, die im Vergleich zu echten Situationen oft zu strukturiert und einfach sind (Dolmans & Schmidt, 1996). Eine Möglichkeit, dieser Herausforderung zu begegnen, bietet die Umsetzung von POL in einer digitalen Simulation. Eine digitale Simulation ist die vereinfachte Abbildung eines realen Systems, beispielsweise einer Schule, die dessen Komponenten auf physischer (z.B. die visuelle Darstellung eines Klassenzimmers) und informationsbezogener Ebene (z.B. Informationen über Schüler*innenleistungen) abbildet und die durch Handlungen der Lernenden beeinflusst werden kann (Sauvé et al., 2007). Digitale Simulationen bieten somit die Möglichkeit, Probleme und Fallbeispiele komplexer und authentischer darzustellen als textbasierte Umsetzungen von POL. Im Gegensatz zu echten Praxissituationen sind sie ökonomischer einsetzbar und ermöglichen es Lernenden, dieselbe Handlung unter denselben Rahmenbedingungen mehrfach zu durchlaufen (Kaufman & Ireland, 2016).

Digitale Simulationen kommen bereits in verschiedenen Kontexten in der Lehrkräftebildung (Christensen et al., 2011; Sullivan et al., 2020) zum Einsatz. Sie wirken positiv auf die Selbstwirksamkeit sowie internale Kontrollüberzeugungen angehender Lehrkräfte (Christensen et al., 2011). Insgesamt besteht jedoch noch ein großer Bedarf an Kontrollgruppenstudien. Dieser Beitrag stellt eine digitale, problemorientierte Simulation zur Förderung diagnostischer Kompetenz von Lehramtsstudierenden vor und vergleicht sie mit einem textbasierten, problemorientierten sowie einem lehrendenzentrierten Setting.

2.3 Die digitale Simulation zur Förderung der diagnostischen Kompetenz

Die für diese Studie entwickelte digitale Simulation soll das deklarative Wissen als zentrale Komponente diagnostischer Kompetenz im Sinne der professionellen Handlungskompetenz (Baumert et al., 2011) fördern. Sie stellt eine dreidimensionale Schulumgebung dar, in der die Studierenden die Rolle der Lehrkraft einer sechsten Klasse einer Gesamtschule übernehmen. Der Schwerpunkt der digitalen Simulation liegt auf dem Sammeln und Systematisieren von Informationen über die Schüler*innen der virtuellen Klasse. Basierend auf dem Prinzip der Siebensprungmethode des problemorientierten Lernens (Reusser, 2005) werden die Studierenden anhand von Arbeitsaufträgen in der digitalen Simulation durch den diagnostischen Prozess geleitet (Hesse & Latzko, 2011) und nutzen sowohl Informationen über Schüler*innen in der Simulation als auch bereitgestellte Fachliteratur, um ein begründetes diagnostisches Urteils über die Lernvoraussetzungen einzelner Schüler*innen zu entwickeln. Als Informationsquellen in der Simulation dienen ihnen dabei unter anderem Zeugnisse, Arbeitshefte, Ergebnisse psychometrischer Tests, Lehrer*innennotizen, Elternsprechtagsdokumente und Gespräche mit Schüler*innen, Kolleg*innen und Eltern (s. Abbildung 1). Zudem können die Studierenden die Klasse in einer virtuellen Schulstunde beobachten (s. Abbildung 2). Einige Schüler*innen, die als virtuelle Fallbeispiele und Referenzpunkt der Arbeitsaufträge fungieren, zeigen unterschiedliche Auffälligkeiten im Bereich des Lernens (z.B. Probleme beim Lesen, Rechtschreiben, Rechnen) oder Verhaltens (z.B. Aufmerksamkeit, Affekt).

Abbildung 1: Elterngespräch in der digitalen Simulation



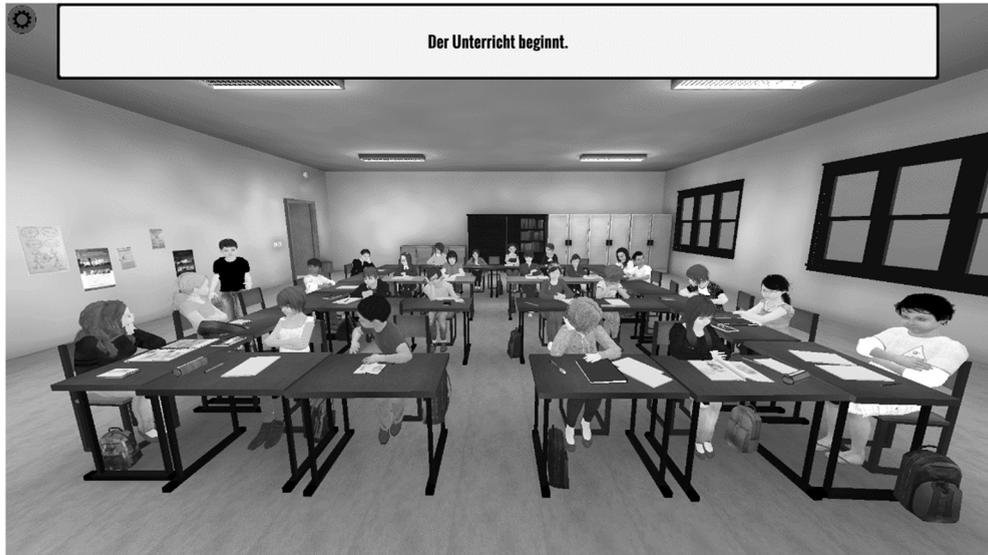


Abbildung 2: Beobachtung einer Schulstunde in der digitalen Simulation

2.4 Die vorliegende Studie

Um die Effektivität der digitalen Simulation und somit den potenziellen Mehrwert dieser Form technologiegestützten Lernens für den Kompetenzerwerb in der Lehrer*innenbildung zu ermitteln, wird in dieser Studie die digitale, problemorientierte Simulation (DS) mit einem textbasierten, problemorientierten (POL) und einem traditionellen, lehrendenzentrierten Setting (LZ) verglichen. Dazu wurde eine aktive Kontrollgruppenstudie mit drei Messzeitpunkten in einem quasi-experimentellen Design durchgeführt.

Bei dem Vergleich stehen die Effekte der drei Settings auf den Wissenserwerb, die Motivation und die Selbstwirksamkeit zum Diagnostizieren von Lernvoraussetzungen am Beispiel von Symptomen der Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung (ADHS) im Fokus. Die Selbstwirksamkeit kann dabei auch als Annäherung an die Erfassung der tatsächlichen diagnostischen Kompetenz verstanden werden. So soll die Frage beantwortet werden, inwiefern eine digitale Simulation einen Mehrwert für die Förderung diagnostischer Kompetenz von Lehramtsstudierenden bieten kann.

3 Methode

3.1 Stichprobe

Es wurden Daten von $N = 155$ Lehramtsstudierenden zu drei Messzeitpunkten analysiert. Alle Studierenden waren in einem Lehramtsstudiengang an einer mittelgroßen deutschen Universität eingeschrieben und nahmen an einem obligatorischen Seminar zu pädagogisch-psychologischer Diagnostik und Förderung teil. Jedes Seminar wurde entweder der Experi-

mentalgruppe oder einer von zwei Kontrollgruppen zugeordnet. So waren 55 Studierende der Experimentalgruppe (DS), 49 der Kontrollgruppe 1 (POL) und 51 der Kontrollgruppe 2 (LZ) zugewiesen. Insgesamt waren 81.3% der Teilnehmenden weiblich. Das Durchschnittsalter betrug 21.6 Jahre ($SD = 2.91$, range = 19–39) und 78.7% gaben Deutsch als ihre Erstsprache an. Im Durchschnitt waren die Studierenden seit 4.1 Semestern eingeschrieben ($SD = 1.3$, range = 1–7) und hatten eine Abiturnote von 2.6 ($SD = .56$, range = 1.0–3.7).

3.2 Studiendesign

In der achten und neunten Semesterwoche nahmen alle Studierenden an zwei aufeinanderfolgenden, 90-minütigen Seminarsitzungen (= Intervention) teil, in der die Lernvoraussetzung *Aufmerksamkeit* thematisiert wurde. T1 fand vor der Intervention, T2 unmittelbar nach der Intervention und T3 etwa sieben Wochen nach der Intervention statt. Die Online-Umfragen zu T1 und T2 wurden während der Seminarsitzungen bearbeitet, während die Online-Umfrage zu T3 von den Studierenden eigenverantwortlich zu Hause ausgefüllt wurde.

Tabelle 1: Vergleich der drei Fördermaßnahmen

	Lehrendenzentrierte Maßnahme	Problemorientierte Maßnahme	Digitale Simulation
Inhalt	Grundlagen zur Theorie, Diagnostik und Förderung der ADHS	Grundlagen zur Theorie, Diagnostik und Förderung der ADHS	Grundlagen zur Theorie, Diagnostik und Förderung der ADHS
Methodische Konzepte	<ul style="list-style-type: none"> • Diagnostischer Prozess (Hesse und Latzko, 2011) 	<ul style="list-style-type: none"> • Diagnostischer Prozess (Hesse und Latzko, 2011) • Siebensprungmethode (Reusser, 2005) 	<ul style="list-style-type: none"> • Diagnostischer Prozess (Hesse und Latzko, 2011) • Siebensprungmethode (Reusser, 2005) (adaptiert an virtuelles Fallbeispiel)
Taxonomiestufen der Lernziele (Anderson & Krathwohl, 2001)	<ul style="list-style-type: none"> • Erinnern • Verstehen 	<ul style="list-style-type: none"> • Erinnern • Verstehen • Anwenden • Analysieren • Evaluieren 	<ul style="list-style-type: none"> • Erinnern • Verstehen • Anwenden • Analysieren • Evaluieren
Arbeitsformen	<ul style="list-style-type: none"> • Lehrendenvortrag • Kleingruppenarbeit 	<ul style="list-style-type: none"> • Arbeit am textbasierten Fall • Kleingruppenarbeit 	<ul style="list-style-type: none"> • Arbeit an Fall in Simulation • Kleingruppenarbeit
Materialien	<ul style="list-style-type: none"> • 90-minütiger Vortrag durch Dozent*in • Online-Materialsammlung 	<ul style="list-style-type: none"> • Fallbeispiel in Textform • Arbeitsaufträge zur Fallarbeit • Online-Materialsammlung 	<ul style="list-style-type: none"> • Fallbeispiel umgesetzt in der digitalen Simulation • Arbeitsaufträge zur Fallarbeit • Online-Materialsammlung
Dauer insgesamt	<ul style="list-style-type: none"> • 180 Minuten 	<ul style="list-style-type: none"> • 180 Minuten + Selbststudium 	<ul style="list-style-type: none"> • 180 Minuten + Selbststudium
Ergebnissicherung in der Gruppenarbeitsphase	<ul style="list-style-type: none"> • Erstellung eines Posters zu Hintergrund, Diagnostik und Förderung der ADHS 	<ul style="list-style-type: none"> • Erstellung eines Posters zu Hintergrund, Diagnostik und Förderung der ADHS 	<ul style="list-style-type: none"> • Erstellung eines Posters zu Hintergrund, Diagnostik und Förderung der ADHS
Annäherung an reale diagnostische Situationen	•○○	•●○	•••

Vor der Intervention konzentrierten sich die Seminarinhalte in allen Seminaren auf Grundkenntnisse der pädagogisch-psychologischen Diagnostik und Förderung und waren für alle drei Gruppen basierend auf einem seit mehreren Jahren etablierten Curriculum einheitlich gestaltet. Je nach Bedingung waren die Interventions Sitzungen methodisch unterschiedlich gestaltet. Während der Intervention arbeiteten die Studierenden entweder mit einem virtuellen Schüler, der ADHS-Symptome zeigt, in der Simulation (DS, Experimentalgruppe), an einem textbasierten, problemorientierten Fallbeispiel eines Schülers mit ADHS-Symptomen (POL, Kontrollgruppe 1) oder aber sie erhielten einen Kurzvortrag zum Thema ADHS und vertieften ihr Wissen anschließend mithilfe von Literatur (LZ, Kontrollgruppe 2). In allen Interventionen arbeiteten die Studierenden in Kleingruppen und nutzten dieselbe bereitgestellte Literaturbasis. Alle Gruppen verfolgten das Lernziel, die Hintergründe der Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung (ADHS) einschließlich Symptomen, Klassifikation und Ätiologie sowie der Diagnostik und Interventions- und Fördermaßnahmen zu verstehen. Zur Ergebnissicherung wurde in jeder Kleingruppe ein Poster erstellt. Für die DS- und die POL-Gruppe, nicht aber für die LZ-Gruppe, war eine Selbstlernphase von ca. 90 Minuten zwischen den beiden Seminarsitzungen obligatorisch. Die Gruppen wurden von drei verschiedenen Dozent*innen betreut, die jeweils für eine Untersuchungsbedingung zuständig waren. Alle Dozent*innen wurden intensiv in die jeweilige Instruktionmethode eingewiesen und waren mit dem den Seminaren zugrundeliegenden Curriculum sehr vertraut. Die Gemeinsamkeiten und Unterschiede zwischen den drei Interventionen werden im Detail in Tabelle 1 dargestellt (vgl. Grotegut & Klingsieck, 2022).

3.2.1 Die drei Versuchsbedingungen

In der DS-Gruppe arbeiteten die Studierenden nach einer kurzen Einführung in die digitale Simulation selbstständig in Kleingruppen mit der Simulation. Pro Gruppe wurden in der Regel ein bis zwei Laptops genutzt, an denen die Studierenden gemeinsam arbeiteten.

In der POL-Gruppe bekamen die Studierenden zu Beginn der ersten Sitzung eine Anleitung zum problemorientierten Lernen inklusive der zugehörigen Arbeitsschritte sowie ein textbasiertes Fallbeispiel ausgehändigt, das inhaltlich mit dem virtuellen Fallbeispiel in der digitalen Simulation vergleichbar war. Die Materialien wurden bereits in früheren Seminaren erfolgreich erprobt. Die Anleitung enthielt acht an den diagnostischen Prozess angelehnte Aufgaben, die mit den Arbeitsaufträgen in der digitalen Simulation vergleichbar waren.

In der LZ-Gruppe erhielten die Studierenden in der ersten Seminarsitzung einen 90-minütigen Vortrag, in dem Grundlagen zu den Themen Aufmerksamkeit und ADHS sowie der Diagnostik und Förderung von ADHS vermittelt wurden. In der zweiten Seminarsitzung arbeiteten die Studierenden in Kleingruppen und vertieften die Inhalte durch die Erarbeitung des Posters.

3.3 Messinstrumente für die abhängigen Variablen

3.3.1 Wissen

Um das deklarative diagnostische Wissen zu erfassen, wurden 15 Items aus einem in der Arbeitsgruppe bewährten Multiple-Choice-Fragebogen verwendet. Es sind Fragen sowohl zu

allgemeinen diagnostischen Inhalten („Was ist der Unterschied zwischen einer systematischen Diagnose und einer Alltagsdiagnostik?“) als auch zu ADHS („Mit welchen Maßnahmen kann eine Lehrkraft ein Kind mit ADHS im Unterricht unterstützen?“) enthalten. Die Itemschwierigkeiten waren für alle drei Messzeitpunkte zufriedenstellend (.39 – .85).

3.3.2 Motivation

Es wurden zwei Skalen zur Messung der situativen Motivation der Studierenden zu T2 verwendet. Die situative Motivation wurde als ein von der allgemeinen Motivation getrenntes Konstrukt betrachtet und steht somit in direktem Zusammenhang mit den gerade zu lösenden Aufgaben. Zum Einsatz kam die *Situational Motivation Scale* (SIMS, Guay et al., 2000), die auf der Selbstbestimmungstheorie (Deci & Ryan, 2014) basierend 16 Items umfasst, die die vier Subskalen intrinsische Motivation, identifizierte Regulation, externale Regulation und Amotivation bilden. Für die vorliegende Studie wurden nur die Subskalen intrinsische Motivation („Weil diese Tätigkeit Spaß macht“, $\alpha_{T2} = .86$), identifizierte Regulation („Weil ich es zu meinem eigenen Wohl tue“, $\alpha_{T2} = .78$) und externale Regulation („Weil ich es tun soll“, $\alpha_{T2} = .83$) verwendet. Die Items werden auf einer 7-stufigen Likert-Skala beantwortet, wobei die Optionen von *trifft gar nicht zu* bis *trifft genau zu* reichen. Ursprünglich war jedem Item die Frage „Warum üben Sie diese Tätigkeit gerade aus?“ vorangestellt. Da die Studierenden die Items jedoch erst unmittelbar nach Beendigung der Intervention beantworteten, wurde diese Frage geändert zu „Bei der Bearbeitung der Seminareinheit hat es mich angetrieben, dass ...“, und die Items wurden syntaktisch angepasst.

Zusätzlich zur SIMS wurde die Subskala Interesse aus dem *Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen* (FAM, Rheinberg et al., 2001) genutzt. Das Interesse wurde als Operationalisierung der intrinsischen Motivation angesehen und daher zur Absicherung der Ergebnisse der SIMS einbezogen. Die Subskala Interesse umfasst vier Items, die auf einer 7-stufigen Likert-Skala von *trifft nicht zu* bis *trifft zu* bewertet wurden ($\alpha_{T2} = .82$; „Nach dem Lesen der Instruktion erschien mir die Seminareinheit sehr interessant.“).

3.3.3 Selbstwirksamkeit zum Diagnostizieren von Lernvoraussetzungen

Zur Erfassung der Selbstwirksamkeit zum Diagnostizieren von Lernvoraussetzungen wurde die Skala *Selbstwirksamkeit zum Diagnostizieren lern- und leistungsrelevanter Merkmale in der Schule* (SWK-DS, Klingsieck, 2017) eingesetzt ($\alpha_{T1} = .93$, $\alpha_{T2} = .95$ und $\alpha_{T3} = .95$). Die SWK-DS umfasst 21 Items, die auf einer 11-Punkte-Skala in Schritten von jeweils 10 Prozent bewertet werden (0%, 10%, 20% usw.). Alle Items bestehen aus einer Aussage in der Form „Ich traue mir zu, den Ausprägungsgrad ... richtig einzuschätzen“ (z.B. „Ich traue mir zu, den Ausprägungsgrad der Motivation eines Schülers richtig einzuschätzen“).

3.3.4 Arbeitsintensität während der Selbststudienphase

Um zu beurteilen, wie intensiv die Studierenden in den DS- und POL-Gruppen während der obligatorischen Selbststudienphase gearbeitet haben, wurde ein Item in den Fragebogen zu T2 aufgenommen, mithilfe dessen die Studierenden die Intensität ihrer Arbeit zwischen den

beiden Seminarsitzungen auf einer 5-Punkte-Skala von *überhaupt nicht* bis *sehr intensiv* bewerten sollten.

3.4 Statistische Analysen

Es wurden gemischte Kovarianzanalysen (ANCOVAs) mit Messwiederholung durchgeführt, um die Entwicklung des deklarativen Wissens und der Selbstwirksamkeit in Abhängigkeit von den drei Versuchsbedingungen zu bestimmen. Eine multivariate Kovarianzanalyse (MANCOVA) wurde durchgeführt, um mögliche Gruppenunterschiede in Bezug auf die motivationalen Variablen aufzuzeigen. Da die Stichproben in allen Gruppen größer als 40 waren, wurde von einer Normalverteilung ausgegangen (Field, 2011). Bei ANCOVAs wurde die Huynh-Feldt-Korrektur verwendet. Die Homogenität der Fehlervarianzen wurde mit dem Levene-Test ($p > .05$) für alle ANCOVAs und MANOVAs geprüft. Alle statistischen Analysen wurden mit IBM SPSS Statistics Version 25.0 durchgeführt.

4 Ergebnisse

4.1 Deskriptive Analysen

Hinsichtlich der demografischen Variablen wurden keine signifikanten Unterschiede in Bezug auf das Alter ($F(2, 152) = .18, p = .84$), das Geschlecht ($F(2, 152) = 1.1, p = .34$) und das Fachsemester ($F(2, 152) = .36, p = .70$) festgestellt. Außerdem unterschieden sich die Gruppen nicht hinsichtlich ihres Vorwissens ($F(2, 151) = .24, p = .79$) und der Selbstwirksamkeit ($F(2, 152) = 1.79, p = .17$) zu T1. Die Gruppen unterschieden sich jedoch geringfügig hinsichtlich der Abiturnote ($F(2, 152) = 3,19, p = .04$): Die POL-Gruppe wies signifikant bessere Abiturnoten auf als die DS-Gruppe, weshalb die Abiturnote in alle Analysen als Kovariate aufgenommen wurde.

Die deskriptiven Analysen zeigen eine Zunahme des deklarativen Wissens für alle Gruppen von T1 bis T3 (Tabelle 2). Auch die Selbstwirksamkeit ist in allen Gruppen gestiegen. Das Interesse war zu T2 eher gering ausgeprägt. Die Studierenden berichteten zu T2 jeweils ähnlich hohe Werte für intrinsische Motivation, identifizierte Regulation und externale Regulation.

Von den Studierenden der DS- und POL-Gruppen gaben insgesamt 8.5% an, während der Selbstlernphase überhaupt nicht an den Aufgaben gearbeitet zu haben, 24%, ein wenig gearbeitet zu haben, 48.8%, mäßig gearbeitet zu haben, 15.5%, intensiv gearbeitet zu haben und 3.1%, während der Selbstlernphase sehr intensiv gearbeitet zu haben. Insgesamt scheint die DS-Gruppe während der Selbstlernphase etwas intensiver gearbeitet zu haben als die POL-Gruppe: Hier gaben 81.2% an, zumindest mäßig gearbeitet zu haben, verglichen mit 67.7% in der POL-Gruppe.

Tabelle 2: Deskriptive Kennwerte

	T1		T2		T3	
	M	SD	M	SD	M	SD
Deklaratives Wissen						
DS	2.04	.43	2.16	.46	2.26	.44
POL	2.08	.41	2.25	.36	2.41	.40
LZ	2.08	.43	2.35	.31	2.48	.32
Selbstwirksamkeit						
DS	64.06	11.86	66.63	11.72	68.48	12.88
POL	61.32	10.56	66.26	10.48	70.01	11.32
LZ	62.67	13.09	64.71	14.58	69.52	12.82
Interesse						
DS			2.65	1.11		
POL			2.10	1.06		
LZ			2.26	1.26		
Intrinsische Motivation						
DS			4.97	1.02		
POL			4.59	1.05		
LZ			4.70	1.24		
Identifizierte Regulation						
DS			4.76	1.03		
POL			4.47	1.16		
LZ			4.80	1.14		
Externale Regulation						
DS			4.20	1.36		
POL			5.15	1.24		
LZ			4.92	1.32		
Arbeitsintensität während der Selbstlernphase						
DS			2.90	.91		
POL			2.75	1.19		
LZ			.a	.a		

^a Da eine Selbstlernphase für die LZ-Gruppe nicht verpflichtend war, sind hier keine Werte angegeben.

4.2 Kovarianzanalysen

4.2.1 Wissen

Eine ANCOVA mit Messwiederholung wurde berechnet, um die Auswirkungen von Gruppe und Messzeitpunkt auf das deklarative Wissen der Studierenden zu bestimmen. Als Kontrollvariable wurde die Abiturnote einbezogen.

Die Ergebnisse zeigen eine signifikante Interaktion zwischen Zeit und Gruppe (Huynh-Feldt $F(3.92, 294.19) = 3.13, p = .02, \eta_p^2 = .04$). Die LZ-Gruppe erzielte zu T2 signifikant höhere Werte als die DS-Gruppe ($p < .01$) und POL-Gruppe ($p = .02$), zu T3 erzielte die LZ-Gruppe höhere Werte als die DS-Gruppe ($p < .01$). Die Haupteffekte für Zeit (Huynh-Feldt $F(1.96, 294.19) = 2.87, p = .06, \eta_p^2 = .02$) und Gruppe ($F(2, 150) = 1.93, p = .15, \eta_p^2 = .03$) waren nicht signifikant. Die Hinzunahme der Abiturnote veränderte die Ergebnisse nicht.

4.2.2 Motivation

Für das Interesse und die extrinsische und intrinsische Motivation wurde eine MANCOVA durchgeführt, in der die Abiturnote als Kontrollvariable einbezogen wurde. Es gab einen signifikanten Gruppen-Haupteffekt (Wilk's Lambda = .88, $F(8, 296) = 2.37$, $p = .02$, $\eta_p^2 = .06$) auf univariater Ebene für Interesse ($F(2, 151) = 3.69$, $p = .03$, $\eta_p^2 = .05$) und externe Regulation $F(2, 151) = 5.19$, $p < .01$, $\eta_p^2 = .06$). Konkret zeigte die DS-Gruppe ein größeres Interesse ($p = .01$) und eine geringere externe Regulation als die POL-Gruppe ($p < .01$). Die Hinzunahme der Abiturnote veränderte die Ergebnisse nicht.

4.2.3 Selbstwirksamkeit zum Diagnostizieren von Lernvoraussetzungen

Eine ANCOVA mit der Abiturnote als Kontrollvariable zeigte für die Selbstwirksamkeit keine signifikante Interaktion zwischen Zeit und Gruppe (Huynh-Feldt $F(3.75, 282.88) = 1.89$, $p = .12$, $\eta_p^2 = .02$). Ebenfalls fand sich kein signifikanter Haupteffekt für die Gruppe ($F(2, 151) = .10$, $p = .90$, $\eta_p^2 = .001$), was darauf hindeutet, dass die Zuordnung zu einer der drei Interventionen keinen Einfluss auf die Entwicklung der Selbstwirksamkeit hatte. Es gab jedoch einen signifikanten Effekt für Zeit (Huynh-Feldt $F(1.87, 282.88) = 4.09$, $p = .02$, $\eta_p^2 = .03$): Die Studierenden haben über alle drei Bedingungen hinweg im Laufe des Semesters an Selbstwirksamkeit in Bezug auf diagnostische Kompetenz gewonnen. Die Abiturnote beeinflusste die Ergebnisse nicht.

5 Diskussion

Ziel dieser Studie war es, den potenziellen Mehrwert einer digitalen Simulation für die Lehrkräftebildung im Bereich der Kompetenzentwicklung zu ermitteln. Dazu wurde eine digitale, problemorientierte Simulation zur Förderung diagnostischer Kompetenz mit einem textbasierten, problemorientierten und einem lehrendenzentrierten Setting verglichen. In einem Quasi-Experiment mit drei Messzeitpunkten wurden die Effekte der drei Lernsettings in Bezug auf deklaratives Wissen, Motivation und Selbstwirksamkeit zum Diagnostizieren von Lernvoraussetzungen am Beispiel von ADHS untersucht. Insgesamt zeigt die vorliegende Studie, dass eine digitale Simulation einen Mehrwert im Vergleich zu den anderen Lernsettings insbesondere im Bereich der selbstbestimmten Motivation bieten kann. Deklaratives Wissen wurde jedoch durch das lehrendenzentrierte Setting besser erworben als durch die digitale Simulation. Die Selbstwirksamkeit konnte durch alle drei Lernsettings gleichermaßen gefördert werden.

In Bezug auf die Motivation zeigen die deskriptiven Ergebnisse eine höhere intrinsische Motivation beim Lernen in der DS-Gruppe als in den beiden anderen Gruppen, wobei der größte Unterschied zwischen der DS- und der POL-Gruppe besteht. Dies ist insbesondere bemerkenswert, da beiden Settings derselbe indirekte Instruktionsansatz – das problemorientierte Lernen – zugrunde liegt. Darüber hinaus berichteten die Studierenden in der DS-Gruppe ein signifikant größeres Interesse und eine geringere externe Regulation als die Studierenden in der POL-Gruppe. Die deskriptiven Ergebnisse deuten zudem auch auf ein größeres Interesse und eine geringere externe Regulation in der DS-Gruppe im Vergleich

zur LZ-Gruppe hin. Zusammengefasst legen diese Befunde nahe, dass in erster Linie die digitale Umsetzung der Simulation einen Effekt auf motivationale Variablen hat. In Anbetracht der kurzen Dauer der Interventionen ist dieser Befund besonders hervorzuheben. Ein größeres Interesse im Rahmen eines Lernsettings ist nicht nur per se erstrebenswert, sondern kann auch mit besseren Lernergebnissen verbunden sein (Schiefele et al., 1992). Die digitale Simulation scheint also vor allem mit selbstbestimmten Formen von Motivation assoziiert zu sein. Diese Ergebnisse stimmen mit früheren Studien überein, die Vorteile digitaler Simulationen und gamebasierten Lernens in Bezug auf die intrinsische Motivation zeigen (Papastergiou, 2009; Tüzün et al., 2009).

Die Selbstwirksamkeit zum Diagnostizieren von Lernvoraussetzungen kann durch eine digitale Simulation ebenso gut gefördert werden wie in einem textbasierten, problemorientierten Setting und einem lehrendenzentrierten Setting. In allen drei Gruppen fühlten sich die Studierenden nach Abschluss der Interventionen und bis zum Semesterende selbstwirksamer in Bezug auf das Diagnostizieren von Lernvoraussetzungen, wobei sich keine Gruppenunterschiede zeigten. Ein Grund dafür könnte sein, dass die Selbstwirksamkeit zum Diagnostizieren gegebenenfalls weniger mit dem Lernmedium als mit dem eigentlichen Lerninhalt zusammenhängt, der in allen drei Gruppen gleich war.

In Bezug auf den Wissenserwerb profitieren Studierende vom Einsatz einer digitalen Simulation weniger als von einer lehrendenzentrierten Intervention. So zeigten die Studierenden der LZ-Gruppe im Verlauf des Semesters einen größeren Wissenserwerb als die Studierenden in der DS-Gruppe. Dieser Befund deckt sich mit den Ergebnissen früherer Studien zu digitalen Simulationen und gamebasiertem Lernen, die einen geringeren Wissenserwerb in gamebasierten Settings im Vergleich zu Kontrollgruppen zeigen (Girard et al., 2013; Vlachopoulos & Makri, 2017). Eine mögliche Erklärung findet sich in der Theorie der kognitiven Belastung: Digitale Simulationen und gamebasierte Lernumgebungen könnten die kognitive Belastung der Lernenden erhöhen, wodurch weniger Ressourcen für die Verarbeitung neuer Informationen zur Verfügung stehen (C. Schrader & Bastiaens, 2012). Außerdem können von Lernansätzen mit eher indirekter Instruktion (Koedinger & Aleven, 2007), wie dem problemorientierten Lernen, das dem POL- und DS-Setting zugrunde lag, insbesondere Lernende mit größerem Vorwissen profitieren, während Lernende mit geringerem Vorwissen eher von direkter Instruktion wie im LZ-Setting profitieren (Renkl & Atkinson, 2003). Da die Teilnehmenden in dieser Studie über ein eher geringes Vorwissen verfügten, könnte auch darin ein möglicher Grund für die Befunde hinsichtlich des Wissenserwerbs liegen. Befunde, die einen größeren Erwerb von deklarativem diagnostischem Wissen für Lehramtsstudierende in einem problemorientierten Setting zeigen (Wedel et al., 2020), konnten in der vorliegenden Studie nicht repliziert werden.

Zusammenfassend kann eine digitale Simulation einen Mehrwert für die Lehrkräftebildung im Bereich der Kompetenzentwicklung insbesondere auf Ebene der selbstbestimmten Motivation während des Lernens bieten. Sie trägt außerdem genauso gut wie ein textbasiertes, problemorientiertes und ein lehrendenzentriertes Setting zur Förderung der Selbstwirksamkeit bei und unterstützt den Erwerb deklarativen Wissens. In diesem Punkt profitieren Studierende jedoch noch mehr von einem lehrendenzentrierten Setting. Da deklaratives Wissen ein zentraler Bestandteil professioneller Kompetenz ist, sind die motivationalen Vorteile der digitalen Simulation nur von eingeschränkter Bedeutung.

5.1 Limitationen

Das zugrundeliegende Studiendesign bringt einige Einschränkungen mit sich, auf die an dieser Stelle eingegangen werden soll. Erstens: Die drei Lernsettings unterscheiden sich hinsichtlich ihres instruktionalen Ansatzes und sind daher nur schwer vergleichbar. So nutzt das LZ-Setting die direkte Instruktion, während das POL- und DS-Setting indirekte Instruktion nutzen. Beide Instruktionsformen können abhängig von verschiedenen Einflussfaktoren, wie zum Beispiel dem Vorwissen der Lernenden, grundsätzlich effektiv sein, weshalb statt der Form der Instruktion vor allem die Implementierungsqualität der Maßnahmen als ausschlaggebend für deren Erfolg gilt (Renkl, 2015) und der „bessere“ instruktionale Ansatz kaum ermittelt werden kann. In zukünftigen Studien sollten daher auch Komponenten innerhalb der digitalen Simulation variiert und hinsichtlich ihrer Wirksamkeit untersucht werden. Im Fokus dieser Studie stand jedoch der Vergleich der digitalen Simulation mit in der Praxis etablierten Lernsettings mit dem Ziel, mögliche Mehrwerte dieser neu entwickelten Maßnahme gegenüber bereits genutzten Lernsettings zu ermitteln. Der Anwendungsbezug dieses Studiendesigns ist somit hoch und rechtfertigt zumindest teilweise den Vergleich unterschiedlicher instruktionaler Ansätze (vgl. Salomon, 2002). Dennoch sollten die Ergebnisse dieser Studie vor diesem Hintergrund primär als erster Schritt für zukünftige Forschung betrachtet werden. Zweitens: Da die Teilnehmenden aus Seminaren rekrutiert wurden, waren die drei Gruppen nicht randomisiert. In den Kovarianzanalysen zeigt sich jedoch, dass es keine signifikanten Unterschiede zwischen den Gruppen in Bezug auf soziodemografische Merkmale oder Vorwissen gab. Die Abiturnote, die sich zwischen den Gruppen signifikant unterschied, wurde als Kontrollvariable mit einbezogen und beeinflusste die Ergebnisse nicht. Damit zusammenhängend muss die Möglichkeit eines Versuchsleitereffekts in Betracht gezogen werden, da jede Versuchsbedingung von jeweils einer/einem Dozent*in betreut wurde. Drittens wurde das prozedurale Wissen als Aspekt professioneller Handlungskompetenz (Baumert et al., 2011) nicht als abhängige Variable herangezogen. Der Erwerb prozeduralen Wissens wird als wesentlicher Vorteil von POL gegenüber anderen Lehr-Lernformaten angesehen, so dass die Erfassung prozeduralen Wissens ein Aspekt ist, der in zukünftigen Studien berücksichtigt werden sollte. Erfasst wurde jedoch die Selbstwirksamkeit, die weithin als Prädiktor für das tatsächliche Verhalten und Kompetenz angesehen wird (Chesnut & Burley, 2015; MacNab & Worthley, 2008). Sie kann daher als Annäherung für die Erfassung des prozeduralen Wissens verstanden werden. Hinsichtlich der Motivation stand in dieser Studie die Erfassung der Motivation während des Lernens in der jeweiligen Fördermaßnahme im Vordergrund. Aus diesem Grund wurde lediglich die situative Motivation unmittelbar nach Durchführung der Fördermaßnahme erfasst. Um mögliche Vorabunterschiede zwischen den Gruppen auszuschließen und eine eventuelle zeitliche Veränderung der Motivation feststellen zu können, sollte Motivation in zukünftigen Studien jedoch auch zum Prätest und Follow-Up erfasst werden. Eine weitere Einschränkung entstand durch die Selbstlernphase zwischen den beiden Seminareinheiten, die nur für die DS- und die POL-Gruppe obligatorisch war, nicht aber für die LZ-Gruppe. Da die LZ-Gruppe die Selbstlernmaterialien während der Seminarsitzungen und nicht zu Hause erarbeitete, war die Wahrscheinlichkeit einer intensiven Auseinandersetzung mit den entsprechenden Inhalten höher als in den anderen beiden Gruppen. Dies könnte den geringeren Erwerb deklarativen Wissens in der DS-Gruppe im Vergleich zur LZ-Gruppe erklären. In zukünftigen Studien sollte eine Selbstlernphase für alle drei Gruppen obligatorisch sein, um Vergleichbarkeit hinsichtlich

des Wissenserwerbs zu gewährleisten. Problematisch ist zudem, dass alle drei Interventionen von sehr kurzer Dauer (weniger als vier Stunden) waren, was die Erzeugung möglicher signifikanter Gruppenunterschiede erschwert haben könnte. Viertens: Die Ursachen für die Unterschiede im Erwerb deklarativen Wissens wurden in dieser Studie nicht weiter untersucht. Sie könnten jedoch dazu beitragen, die Wirksamkeit von Interventionen zur Förderung diagnostischer Kompetenz zu erklären. In künftigen Studien sollte daher zum Beispiel das Ausmaß verschiedener Arten des Cognitive Load erfasst werden, da ein höherer Extrinsic Cognitive Load zu einem geringeren Wissenserwerb beitragen könnte.

5.2 Ausblick

Insgesamt scheint eine digitale, problemorientierte Simulation eine vielversprechende Instruktionsmethode darzustellen, die insbesondere in Kombination mit anderen Methoden das Potenzial hat, einen Mehrwert für die Förderung diagnostischer Kompetenz von Lehramtsstudierenden zu bieten. Es bedarf jedoch noch weiterer Forschung, um Erkenntnisse hinsichtlich unterschiedlicher Einflussfaktoren auf die Wirksamkeit digitaler Simulationen und die Effektivität einzelner Komponenten digitaler Simulationen zu erlangen. Eine digitale Simulation könnte sich in der hier untersuchten Form zum Beispiel als Ergänzung zu traditionellen Vorlesungsformaten in Form eines Flipped Classroom-Konzepts (Strelan et al., 2020) eignen: So kann das selbstregulierte Lernen durch eine digitale Simulation unterstützt und gleichzeitig ein möglichst umfangreiches Fachwissen aufgebaut werden. Im Rahmen dieser Studie wurde ein Fallbeispiel zum Thema ADHS in der problemorientierten, digitalen Simulation erprobt. Die digitale Simulation beinhaltet darüber hinaus weitere Fallbeispiele, sodass ihre potenzielle Nützlichkeit auf die Diagnostik weiterer Lernvoraussetzungen ausgeweitet werden kann. Neben der Kombination mit anderen Instruktionsansätzen sollte zukünftig auch die (Weiter-)Entwicklung digitaler Simulationen im Hinblick auf den Wissenserwerb angestrebt werden: Potenzial bietet hier zum Beispiel, basierend auf den Prinzipien des testbasierten Lernens, die Implementierung eines Quiz in eine digitale Simulation, welches durch den gezielten Abruf von Fachwissen (Adesope et al., 2017) das Lernen unterstützt. Schon jetzt stellen digitale Simulationen neuartige und innovative Lehr-Lernformate im universitären Kontext dar und sorgen für Abwechslung in einem Bereich, der weitgehend von traditionellen Vorlesungen und Seminaren geprägt ist. Ihr Potential, einen Mehrwert für die Lehrkräftebildung im Bereich der Kompetenzentwicklung zu bieten, ist derzeit jedoch bei Weitem noch nicht in vollem Umfang erforscht und ausgeschöpft.

Hinweis

Dieses Manuskript ist im Rahmen des kumulativen Dissertationsprojektes der Erstautorin zum Thema „GEProS – Entwicklung und Evaluation einer digitalen Simulation zur Förderung diagnostischer Kompetenz von Lehramtsstudierenden“ entstanden.

Literatur

- Adesope, O. O., Trevisan, D. A. & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659–701. <https://doi.org/10.3102/0034654316689306>
- Anderson, L. W. & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Complete ed.). Longman. <https://eduq.info/xmlui/handle/11515/18345>
- Atkinson, R. K., Derry, S. J., Renkl, A. & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, 70(2), 181–214. <https://doi.org/10.3102/00346543070002181>
- Bartel, M.-E. & Roth, J. (2020). Video- und Transkriptvignetten aus dem Lehr-Lern-Labor – die Wahrnehmung von Studierenden. In B. Priemer & J. Roth (Hrsg.), *Lehr-Lern-Labore: Konzepte und deren Wirksamkeit in der MINT-Lehrpersonenbildung* (S. 299–315). Springer Spektrum. https://doi.org/10.1007/978-3-662-58913-7_19
- Baumert, J., Kunter, M., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (2011). Professionelle Kompetenz von Lehrkräften, kognitiv aktivierender Unterricht und die mathematische Kompetenz von Schülerinnen und Schülern (COACTIV) – ein Forschungsprogramm. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 7–25). Waxmann.
- Blömeke, S., Jentsch, A., Ross, N., Kaiser, G. & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and Instruction*, 79, 101600. <https://doi.org/10.1016/j.learninstruc.2022.101600>
- Chesnut, S. R. & Burley, H. (2015). Self-efficacy as a predictor of commitment to the teaching profession: A meta-analysis. *Educational Research Review*, 15, 1–16. <https://doi.org/10.1016/j.edurev.2015.02.001>
- Christensen, R., Knezek, G., Tyler-Wood, T. & Gibson, D. (2011). SimSchool: An online dynamic simulator for enhancing teacher preparation. *International Journal of Learning Technology*, 6(2), 201–220.
- Deci, E. L. & Ryan, R. M. (2014). *Intrinsic motivation and self-determination in human behavior* (1. Aufl.). *Perspectives in social psychology*. Springer Science+Business Media.
- Demirel, M. & Dağyar, M. (2016). Effects of problem-based learning on attitude: A meta-analysis study. *EURASIA Journal of Mathematics, Science & Technology Education*, 12(8), 2115–2137. <https://doi.org/10.12973/eurasia.2016.1293a>
- Dolmans, D. H. J. M. & Schmidt, H. T. (1996). The advantages of problem-based curricula. *Postgraduate Medical Journal*, 72(851), 535–538. <https://doi.org/10.1136/pgmj.72.851.535>
- Field, A. (2011). *Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll)* (3. Aufl.). Sage.
- Girard, C., Ecalle, J. & Magnan, A. (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning*, 29(3), 207–219. <https://doi.org/10.1111/j.1365-2729.2012.00489.x>
- Grotegut, L. & Klingsieck, K. B. (2022). Wie können unterschiedliche Aspekte diagnostischer Kompetenz gefördert werden? Drei Maßnahmen im Vergleich. *Zeitschrift für Pädagogische Psychologie*. <https://doi.org/10.1024/1010-0652/a000352>
- Guay, F., Vallerand, R. J. & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion*, 24(3), 175–213.
- Heitzmann, N., Fischer, F. & Fischer, M. R. (2018). Worked examples with errors: when self-explanation prompts hinder learning of teachers diagnostic competences on problem-based learning. *Instructional Science*, 46(2), 245–271. <https://doi.org/10.1007/s11251-017-9432-2>

- Hesse, I. & Latzko, B. (2011). *Diagnostik für Lehrkräfte*. Barbara Budrich.
- Hmelo, C. E. (1998). Problem-based learning: Effects on the early acquisition of cognitive skill in medicine. *The Journal of the Learning Sciences*, 7(2), 173–208.
- Hsieh, T.-L. (2014). Motivation matters? The relationship among different types of learning motivation, engagement behaviors and learning outcomes of undergraduate students in Taiwan. *Higher Education*, 68(3), 417–433. <https://doi.org/10.1007/s10734-014-9720-6>
- Karing, C. & Artelt, C. (2013). Genauigkeit von Lehrpersonenurteilen und Ansatzpunkte ihrer Förderung in der Aus- und Weiterbildung von Lehrkräften. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31(2), 166–173.
- Kaufman, D. & Ireland, A. (2016). Enhancing teacher education with simulations. *TechTrends*, 60(3), 260–267. <https://doi.org/10.1007/s11528-016-0049-0>
- Klegeris, A. & Hurren, H. (2011). Impact of problem-based learning in a large classroom setting: student perception and problem-solving skills. *Advances in Physiology Education*, 35(4), 408–415. <https://doi.org/10.1152/advan.00046.2011>
- Klingsieck, K. B. (2017). Selbstwirksamkeit zum Diagnostizieren lern- und leistungsrelevanter Merkmale in der Schule - Entwicklung und Erprobung einer Skala. *Lehrerbildung auf dem Prüfstand*, 10(1), 55–66.
- Koedinger, K. R. & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), 239–264. <https://doi.org/10.1007/s10648-007-9049-0>
- Kramer, M., Förtsch, C., Boone, W. J., Seidel, T. & Neuhaus, B. J. (2021). Investigating pre-service biology teachers' diagnostic competences: Relationships between professional knowledge, diagnostic activities, and diagnostic accuracy. *Education Sciences*, 11(3), 1–23. <https://doi.org/10.3390/educsci11030089>
- MacNab, B. R. & Worthley, R. (2008). Self-Efficacy as an intrapersonal predictor for internal whistleblowing: A US and Canada examination. *Journal of Business Ethics*, 79(4), 407–421. <https://doi.org/10.1007/s10551-007-9407-3>
- Martin, T., Rivale, S. D. & Diller, K. R. (2007). Comparison of student learning in challenge-based and traditional instruction in biomedical engineering. *Annals of Biomedical Engineering*, 35(8), 1312–1323. <https://doi.org/10.1007/s10439-007-9297-7>
- McLaren, B. M., Lim, S.-J. & Koedinger, K. R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In B. C. Love, K. McRae & V. M. Sloutsky (Hrsg.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (S. 2176–2181). Cognitive Science Society.
- Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education*, 52(1), 1–12. <https://doi.org/10.1016/j.compedu.2008.06.004>
- Pourghaznein, T., Sabeghi, H. & Shariatinejad, K. (2015). Effects of e-learning, lectures, and role playing on nursing students' knowledge acquisition, retention and satisfaction. *Medical Journal of the Islamic Republic of Iran*, 29, 162.
- Praetorius, A.-K. & Südkamp, A. (2017). Eine Einführung in das Thema der diagnostischen Kompetenz von Lehrkräften. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Pädagogische Psychologie und Entwicklungspsychologie: Bd. 94. Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (S. 13–18). Waxmann.
- Renkl, A. (2015). Different roads lead to Rome: the case of principle-based cognitive skills. *Learning: Research and Practice*, 1(1), 79–90. <https://doi.org/10.1080/23735082.2015.994255>
- Renkl, A. & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist*, 38(1), 15–22. https://doi.org/10.1207/S15326985EP3801_3
- Reusser, K. (2005). Problemorientiertes Lernen. – Tiefenstruktur, Gestaltungsformen, Wirkung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 23(2), 159–182.

- Rheinberg, F., Vollmeyer, R. & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen (Langversion, 2001). *Diagnostica*, 2, 57–66.
- Römer, J., Rothland, M. & Straub, S. (2018). Pädagogische Vorerfahrungen und ihre Bedeutung für die Kompetenzeinschätzung und das Flow-Erleben beim Unterrichten im Praxissemester. In J. König, M. Rothland & N. Schaper (Hrsg.), *Learning to practice, learning to reflect? Ergebnisse aus der Längsschnittstudie LtP zur Nutzung und Wirkung des Praxissemesters in der Lehrerbildung* (S. 223–240). Springer VS. https://doi.org/10.1007/978-3-658-19536-6_8
- Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Salomon, G. (2002). Technology and pedagogy: Why don't we see the promised revolution? *Educational Technology*, 42(2), 71–75. <https://www.jstor.org/stable/44428740>
- Sauvé, L., Renaud, L., Kaufman, D. & Marquis, J.-S. (2007). Distinguishing between games and simulations: A systematic review. *Educational Technology & Society*, 10(3), 247–256.
- Schiefele, U., Krapp, A. & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K. A. Renninger, S. Hidi & A. Krapp (Hrsg.), *The role of interest in learning and development* (S. 197–226). Psychology Press. <https://doi.org/10.4324/9781315807430-18>
- Schrader, C. & Bastiaens, T. J. (2012). The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior*, 28(2), 648–658. <https://doi.org/10.1016/j.chb.2011.11.011>
- Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Lang.
- Seifert, A., Schaper, N. & König, J. (2018). Bildungswissenschaftliches Wissen und Kompetenzeinschätzungen von Studierenden im Praxissemester: Veränderungen und Zusammenhänge. In J. König, M. Rothland & N. Schaper (Hrsg.), *Learning to practice, learning to reflect? Ergebnisse aus der Längsschnittstudie LtP zur Nutzung und Wirkung des Praxissemesters in der Lehrerbildung* (S. 325–347). Springer VS. https://doi.org/10.1007/978-3-658-19536-6_12
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19(1/2), 85–95. <https://doi.org/10.1024/1010-0652.19.12.85>
- Strelan, P., Osborn, A. & Palmer, E. (2020). The flipped classroom: A meta-analysis of effects on student performance across disciplines and education levels. *Educational Research Review*, 30, 100314. <https://doi.org/10.1016/j.edurev.2020.100314>
- Südkamp, A., Kaiser, J. & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762.
- Südkamp, A. & Praetorius, A.-K. (Hrsg.). (2017). *Pädagogische Psychologie und Entwicklungspsychologie: Bd. 94. Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen*. Waxmann.
- Sullivan, F., Hillaire, G., Larke, L. & Reich, J. (2020). Using teacher moments during the COVID-19 pivot. *Journal of Technology and Teacher Education*, 28(2), 303–313.
- Tüzün, H., Yılmaz-Soylu, M., Karakuş, T., İnal, Y. & Kızılkaya, G. (2009). The effects of computer games on primary school students' achievement and motivation in geography learning. *Computers & Education*, 52(1), 68–77.
- Urhahne, D. & Wijnia, L. (2021). A review on the accuracy of teacher judgments. *Educational Research Review*, 32, 100374. <https://doi.org/10.1016/j.edurev.2020.100374>
- van Merriënboer, J. J. (2013). Perspectives on problem solving and instruction. *Computers & Education*, 64, 153–160. <https://doi.org/10.1016/j.compedu.2012.11.025>
- Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M. & Deci, E. L. (2004). Motivating learning, performance, and persistence: the synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology*, 87(2), 246–260.

- Vlachopoulos, D. & Makri, A. (2017). The effect of games and simulations on higher education: a systematic literature review. *International Journal of Educational Technology in Higher Education*, 14(1), 14–22. <https://doi.org/10.1186/s41239-017-0062-1>
- Wedel, A., Müller, C. R., Pfetsch, J. & Ittel, A. (2020). Entwicklung diagnostischer Kompetenz in der Lehramtsausbildung – Effekte problemorientierten Lernens mit Textfällen. In I. Gogolin, B. Hannover & A. Scheunpflug (Hrsg.), *Edition ZfE: Bd. 4. Evidenzbasierung in der Lehrkräftebildung* (S. 95–121). Springer VS. https://doi.org/10.1007/978-3-658-22460-8_5

Kontakt

Dr. Lea Grotegut
Zentrum für Bildungsforschung und Lehrerbildung – PLAZ-Professional School
Prof. Dr. Katrin B. Klingsieck
Fakultät für Kulturwissenschaften, Fach Psychologie
Universität Paderborn
Warburger Str. 100
33098 Paderborn
E-Mail: grotegut@plaz.upb.de
E-Mail: katrin.klingsieck@upb.de

Konstruktion und psychometrische Prüfung eines Tests zur Diagnostik mathematischer Studieneingangsleistungen

Stefan Behrendt, Jan Köllner, Kristina Kögler, Christine Sälzer, Andreas Just

Zusammenfassung: Eingangsvoraussetzungen im Bereich Mathematik sind für den Erfolg in der Studieneingangsphase von MINT-Studiengängen von übergeordneter Bedeutung. Dennoch werden mathematische Basisfähigkeiten aus der Sekundarstufe I in Wiederholungs- und Unterstützungsmaßnahmen häufig vernachlässigt. Gleichmaßen fehlen geeignete diagnostische, qualitätsgeprüfte Instrumente für diesen Zweck. Der Beitrag stellt ein reliables und sowohl inhaltlich als auch differenziell sowie prognostisch valides computerbasiertes Instrument zur Diagnose dieser Fähigkeiten vor. Eine Papier- und eine Online-Version messen dasselbe Konstrukt. Die Online-Version enthält ein Instant-Feedback, welches sowohl Leistungs- als auch Verbesserungsrückmeldungen integriert. Im Ausblick werden noch ausstehende, zentrale Entwicklungs- und Prüfungsschritte thematisiert.

Schlüsselwörter: Mathematik der Sekundarstufe I, Online-Self-Assessment, Instant-Feedback, Studieneingangstest, MINT-Studiengänge, Studieneingangsphase

Development and psychometric verification of a diagnostic test for mathematical study entry performance

Summary: Entry-level prerequisites in mathematics are of paramount importance for success in STEM undergraduate programs. Nevertheless, basic mathematical skills from lower secondary school are often neglected in repetition and support measures. Similarly, there is a lack of appropriate diagnostic, quality-controlled instruments for this purpose. This paper presents a reliable and both content as well as differentially and prognostically valid computer-based instrument for diagnosing these skills. A paper and an online version measure the same construct. The online version includes instant feedback that integrates both performance and improvement feedback. In the outlook, outstanding key development and testing steps are addressed.

Keywords: lower secondary school mathematics, online self-assessment, instant feedback, university entrance test, STEM courses, study entry phase

1 Ausgangslage und Zielsetzung

1.1 Relevanz mathematischer Basisfähigkeiten in der Studieneingangsphase

Ein vor dem Abschluss abgebrochenes Studium ist sowohl auf individueller Ebene als auch aus volkswirtschaftlicher Perspektive nicht wünschenswert: Es ist kostenintensiv und transportiert das Etikett des Scheiterns (in einer Übersicht von Neugebauer et al., 2019). Hochschulen investieren Ressourcen in Form von Lehrveranstaltungen und Infrastruktur in abge-

brochene Studien, die sich letztlich nicht in einer entsprechenden Anzahl von Absolvierenden auszahlen. Tatsächlich steigen die Studienabbruchquoten jedoch in den letzten Jahren in den meisten Studiengängen kontinuierlich bis auf jüngst 43% in Mathematik und Naturwissenschaften – was der höchsten Quote aller Fächer entspricht – und 35% in den Ingenieurwissenschaften (Heublein et al., 2020) – eine pädagogische wie hochschuldidaktische Herausforderung, der mit geeigneten Maßnahmen zu begegnen ist.

Die Ursachen für einen Studienabbruch können dabei vielfältig sein und variieren teilweise je nach Studiengang. Dabei ist nach Heublein et al. (2017) mangelndes fachliches Vorwissen für drei Viertel aller Studienabbrüche mitverantwortlich. Auch Petri (2020) sieht die Abiturnote – und damit mitunter das mangelnde Vorwissen – neben motivationalen und emotionalen Merkmalen als hauptsächlichen. Insbesondere die Mathematikleistung stellt sich als direkt abhängig vom Vorwissen dar (Krawitz, 2020). Dabei sind mathematische Basisfähigkeiten gerade in den MINT-Studiengängen (Fächer der Gebiete Mathematik, Informatik, Naturwissenschaften und Technik) unerlässliches Grundhandwerk (Neumann et al., 2017; CoSH-Gruppe, 2021), welches entgegen dem im Abiturzeugnis enthaltenen Kompetenzversprechen jedoch von den Studienanfängenden nicht immer ausreichend sicher beherrscht wird. Die Hochschulen reagieren auf diese Passungsprobleme mit deutlichen curricularen Anpassungen (Bausch et al., 2014a). Dabei werden die Inhalte und Fähigkeiten der Sekundarstufe II mehr oder weniger intensiv in den entsprechenden Lehrveranstaltungen wiederholt. Die curricular vorgesehenen Inhalte in den Lehrveranstaltungen der ersten Semester der MINT-Studiengänge begrenzen jedoch die Möglichkeiten zum Wieder- und Nachholen, sodass eigentlich als bekannt vorauszusetzende Inhalte der Sekundarstufe I meist nur implizit integriert werden. Gleichzeitig werden Orientierungs- und Lernangebote auch schon in die Zeit vor Studienbeginn verlegt (Bausch et al., 2014a), welche sich aber ebenfalls auf die Sekundarstufe II fokussieren und die Sekundarstufe I nur implizieren. Dies wiederum erfordert eine zuverlässige individuelle Diagnostik, welche unmittelbar von den Studierenden selbst interpretiert und genutzt werden kann, damit individuell festgestellte Lücken auch ohne Anleitung durch Dozierende gezielt geschlossen werden können (Karst et al., 2017).

1.2 Potenziale technologiebasierter IRT-skaliertes Diagnoseinstrumente

Im Kontext der Diagnostik studienfeldspezifischer Eingangsfähigkeiten spielen Online-Self-Assessments als Instrumente der Studienorientierung eine bedeutende Rolle, die von vielen Hochschulen auch im Bereich der Mathematik in der Phase der Studienwahl angeboten werden, aber in der Regel keinerlei Verbindlichkeit mit sich bringen. Sie dienen vielmehr lediglich dem Abgleich individueller Neigung und Eignung mit dem jeweiligen Studiengangsprofil. Damit sollen sie „Auskunft über die Passung der Erwartungen zum angestrebten Studienfach geben und Selbstselektionsprozesse anstoßen, indem die studienfachbezogenen Einstellungen verändert werden“ (Karst et al., 2017, S. 205). Die Ergebnismeldungen sind häufig sehr einfach gehalten, etwa in Gestalt numerischer Summenscores und kurzer Erläuterungen, die zumeist nicht auf komplexen Skalierungsverfahren beruhen, sondern ähnlich wie bei schulischen Klausuren den Anteil korrekter Antworten rückmelden (Brunner, 2017). Individuelle Defizite, spezifischer Nachholbedarf oder festzustellende Kompetenzprofile werden damit selten erkannt. Entsprechend werden konkrete Maßnahmen mit Relevanz für die Studieneingangsphase im Sinne einer frühzeitigen Diagnostik individueller Defizite und Ablei-

tung geeigneter Fördermaßnahmen damit in aller Regel nicht verknüpft – es zeichnet sich somit mit Blick auf die sich verschärfende Problematik der frühen Studienabbrüche Handlungsbedarf ab.

Einen oftmals relevanten Baustein zum erfolgreichen Übergang in die Hochschule bilden Onlinebrückenkurse, die häufig von entsprechenden Einstufungstests flankiert werden, um die Studierenden entsprechend ihren Fähigkeiten bestmöglich zu fördern (Hanft et al., 2015). Die Qualität der Instrumente und des Feedbacks an Testteilnehmende variiert dabei deutlich. Zu kritisieren ist dabei insbesondere auch die unzureichend verfügbare Dokumentation der psychometrischen Qualität der Instrumente (Brunner, 2017), was darauf hinweisen kann, dass diese auch nur bedingt geprüft wurde. Skalierungsverfahren auf Basis der Item-Response-Theorie (IRT) bieten nicht nur die Möglichkeit, Personenfähigkeit und Itemschwierigkeit simultan auf einer Skala zu schätzen (Wilson, 2005), sondern eignen sich insbesondere auch für technologiebasierte Diagnoseinstrumente mit umfangreichen Feedbackformaten, da sie hohe Automatisierungspotenziale mit sich bringen und auch im Large-Scale-Bereich bei großen Stichproben effizient einsetzbar sind. Die technologiebasierte Diagnostik von Studieneingangsleistungen auf Basis komplexer Skalierungsverfahren ist ungeachtet dessen im Bereich der Mathematik noch nicht etabliert – ein Umstand, der mit dem hier vorgestellten Instrument adressiert werden sollte.

Hier setzt dieser Beitrag an und stellt ein selbst entwickeltes Instrument zur Diagnostik mathematischer Eingangsvoraussetzungen zu Studienbeginn vor, das auch Feedbackelemente enthält. Gerade die Feedbackkonzeption ist dabei bei Online-Self-Assessments besonders herausfordernd, da die Interpretation nicht von psychometrischen Fachleuten begleitet, sondern lediglich von den Testteilnehmenden gelesen und interpretiert wird. Der Gesamterfolg eines Assessments darf somit nicht lediglich auf die Instrumentenqualität zurückgeführt werden, sondern muss auch die Qualität des Feedbacks berücksichtigen (Kubinger et al., 2012). Dabei identifizieren Brunner et al. (2015) in einem Vergleich von über 40 verschiedenen Mathematik-Self-Assessment-Tools aus dem deutschsprachigen Raum „noch viel Potential für die Weiterentwicklung des Feedbacks“ (Brunner et al., 2015, S. 159). Wie genau sie dies quantifizieren und wo genau sie dies sehen, bleibt allerdings offen.

Aus Umfangsgründen sind die Ausgestaltung und erste Erkenntnisse zur Rezeption des Feedbacks nicht Teil dieses Beitrags. Ein zentrales Element des Feedbacks, welches hier aufgegriffen wird, ist eine einfach zu interpretierende Leistungsskala auf Basis eines Niveaumodells (Beaton & Allen, 1992). Dieses bietet die Möglichkeit, einerseits den dritten Feedbackschritt nach Hattie und Timperley (2007), das Feed Forward, zu integrieren und andererseits auch neben der Aufgabenebene die Konstruktebene (Belcadhi, 2016) einzubeziehen.

1.3 Entwicklungsbedarf psychometrisch fundierter Testinstrumente

Zur Testentwicklung wurden im Vorfeld relevante Rahmenbedingungen näher in den Blick genommen. Dazu wurden bereits bestehende Testinstrumente und Lernmaterialien, insbesondere für den Übergang zwischen Schule und Studium, begutachtet, beispielsweise Schäfer et al. (1997), Knorrenschild (2004) und Dürrschnabel et al. (2019). Die daraus resultierenden Inhalte wurden unter Berücksichtigung des Mindestanforderungskatalogs an ein WiMINT-Studium (MINT-Fächer und Fächer der Wirtschaftswissenschaften) der CoSH-Gruppe kritisch gesichtet, und entsprechend wurde eine Auswahl der Themen vorgenommen (CoSH-

Gruppe, 2021). Inhaltlich fokussiert der Test Themen der Sekundarstufe I, welche explizit nicht Gegenstand universitärer Lehre sind und daher bereits vor Studienbeginn von Anfangenden beherrscht werden müssen.

Die Berücksichtigung bestehender Testinstrumente wie die Vergleichsstudien TIMSS (Schwippert et al., 2020) oder PISA (Reiss et al., 2019) erwies sich aufgrund der Konzeption dieser Instrumente als nur eingeschränkt sinnvoll. Die Recherche bestehender Tests zeigte den Bedarf zur Erstellung eigener Tests aus zwei zentralen Gründen auf: (1) Gängige Aufgabensätze fokussieren hierbei eher offene Antwortformate mit Rechenwegen (beispielsweise Neubrand et al., 2004). Dies macht allerdings eine automatische Auswertung zur Generierung eines Instant-Feedbacks schwierig und fehleranfällig. (2) In kontextbehafteten Aufgaben findet schnell eine Verschiebung des Fokus weg von mathematischen Schwierigkeiten hin zu einer nicht immer eindeutigen Umsetzung von Texten in mathematische Fragestellungen statt (Jahnke et al., 2014). Um diese Form mathematischer Modellierung zu umgehen, liegt der Fokus vorwiegend auf handwerklichen Fähigkeiten und innermathematischen Problemen. Die Aufgaben sollten kleinschrittige Probleme widerspiegeln, die in kurzer Zeit ohne Hilfsmittel bearbeitbar sind. Dies folgt auch dem Ansatz etablierter Brückenkursprojekte wie Ve&MINT (Bausch et al., 2014b) bzw. OMB+ (Krunke et al., 2012), welche von der Allianz führender Technischer Universitäten in Deutschland (TU9) auch ausdrücklich zur Studienvorbereitung empfohlen werden (TU9, 2020).

Aus diesen Gründen entwickeln viele Hochschulen entsprechende eigene Eingangs- und Eignungstests, welche aber häufig nicht umfangreich auf ihre psychometrische Güte geprüft sind (Brunner, 2017). Das hier vorgestellte Instrument bezieht bestehende Aufgabenstellungen aus den Eingangstests der Mathematik-Vorkurse und der Prüfungsaufgaben an der Universität Stuttgart ein. Letztere dienen dazu, Besonderheiten des späteren Prüfungsformats als erste Lernhilfe aufzuzeigen. Die Aufgaben legen hierbei besonderen Wert auf eine in den Grundlagenthorlesungen üblichen Fachsprache und Notation.

1.4 Zielsetzung und Fragestellung des Beitrags

Vor dem Hintergrund der gewürdigten Argumente und Desiderate wird deutlich, dass ein feedbackgenerierender computerbasierter Eingangstest bezüglich der mathematischen Anforderungen der MINT-Studiengänge in einer expliziten Ausrichtung auf die Sekundarstufe I dazu beitragen könnte, mittelfristig Studienabbrüche effektiv zu reduzieren. Ungeachtet einer breiten Verfügbarkeit mathematischer Online-Self-Assessments konnte für den betreffenden Inhaltsbereich bislang kein Instrument identifiziert werden, das die skizzierten Anforderungen an individualisierte und psychometrisch basierte Ergebnis- und Förderrückmeldungen einzulösen vermochte, wodurch eine Eigenentwicklung angestrebt wurde. Zur Sicherstellung der diagnostischen Eignung wurden im Zuge dessen die folgenden drei Ziele definiert, die mit dem Instrument erreicht werden sollten:

- (1) Selbstreflektionsanlass für Studierende bezüglich ihrer individuellen Eignung,
- (2) Identifikation und adressatengerechte Rückmeldung individueller Nachholbedarfe,
- (3) Monitoring der Zielgruppe und von Kohortenveränderungen im Zeitverlauf.

Dabei wird bewusst ein inhaltlicher Schwerpunkt auf die relevanten Inhalte und Fähigkeiten der Mathematik der Sekundarstufe I gelegt, da diese häufig in der Studieneingangsphase

nicht in der unter den oben thematisierten Umständen gebotenen Ausführlichkeit wiederholt werden. Demgegenüber werden Inhalte und Fähigkeiten aus der Sekundarstufe II (zum Beispiel die Differential- und die Integralrechnung sowie die Vektor- und Matrizenalgebra) erneut grundlegend aufgegriffen und teilweise auch axiomatisch neu eingeführt (Meyberg & Vachenaer, 2001). Der Beitrag fokussiert dabei nachstehende Fragestellungen:

- (1) Welche Struktur und welche Parametrisierung des Item-Response-Modells ist für eine reliable und valide Interpretation geeignet?
- (2) Ist die Skala fair in Bezug auf das Geschlecht sowie invariant bezüglich des Darreichungsmodus und der Version?
- (3) Unterscheiden sich die Testleistungen der Teilnehmenden in Abhängigkeit ihres gewählten Studiengangs?
- (4) Kann aus der modellierten Skala ein inhaltlich valide interpretierbares Niveaumodell generiert werden?
- (5) Haben die gemessenen Leistungen – über die Hochschulzugangsberechtigungsnote hinaus – eine Vorhersagekraft für den Erfolg im ersten Studiensemester?

2 Methode

2.1 Aufbau des Testinstruments

Das Testinstrument gliedert sich in vier Teile mit insgesamt 17 Aufgaben zu folgenden Teilgebieten. Beispielaufgaben sind in den Abbildungen 2, 3 und 4 (s. unten Abschnitt 3.2) dargestellt.

- (1) Bruch- und Potenzrechnung (3 Aufgaben; 6 Minuten): Einfache Terme mit Zahlen; Binomische Formeln; Mustererkennung in Folgen.
- (2) Termumformung (5 Aufgaben; 10 Minuten): Brüche mit Variablen; Binomische Formeln; Techniken zur Lösung gemischt quadratischer Gleichungen; quadratisches Ergänzen.
- (3) Aspekte der Geometrie (4 Aufgaben; 10 Minuten): Räumliche und ebene Abstandprobleme; räumliche Vorstellung; Lineare Gleichungssysteme.
- (4) Elementare Funktionen (5 Aufgaben; 10 Minuten): Definitionen und Rechenregeln der Exponential-, Logarithmus- und trigonometrischen Funktionen.

Dabei wird für jeden Testteil in Abhängigkeit der angenommenen Itemschwierigkeit und der studienfeldspezifischen Anforderungen eine Höchstbearbeitungszeit festgelegt, da eine ausreichend zügige Bearbeitung dieser Basisanforderungen relevant ist, um entsprechende Routinen dieser Basisfähigkeiten annehmen zu können.

Es wird die in den (mathematischen) Eingangsveranstaltungen verwendete Fachsprache, auch mit nichtverbalisierten Formelzeichen, verwendet (zum Beispiel „Für $t \in \mathbb{R}$ besitzt das Lineare Gleichungssystem [...] eine eindeutige Lösung (x_1, x_2) . Wie lautet diese?“), anstatt die teilweise vereinfachenden Ausdrücke der Schule (zum Beispiel „Bestimmen Sie die Lösung des Linearen Gleichungssystems [...] in Abhängigkeit des Parameters t .“), um die implizite Orientierungsleistung des Tests zu erhöhen (Karst et al., 2017). Gleichzeitig wird im Sinne der *Opportunity to Learn* (AERA et al., 2014, 56f.) darauf geachtet, dass die Fachspra-

che auch mit dem curricular abgebildeten Vorwissen verstanden werden kann. Auf eine Kontextualisierung der Items wird auf Basis der Argumentation von Bach (2016) verzichtet, der prozessbezogene Aufgabenstellungen lediglich als eine Meta-Ebene der inhaltsbezogenen Aufgabenstellungen herausstellt.

Der Test existiert in drei parallelisierten Formen, um primär ein Abschreiben in Präsenzadministration zu vermeiden und sekundär auch Versionen für wiederholende Testungen verfügbar zu machen. Diese Formen unterscheiden sich meist lediglich in den verwendeten Zahlenwerten (die Beispielaufgabe in Abbildung 4 wird mit den Kombinationen $x = \log 2$ und $\log 8$; $x = \log 3$ und $\log 9$; $x = \log 2$ und $\log 4$ gestellt) oder in der Verwendung verwandter Definitionen (z.B. ist zu entscheiden, ob die Definition einer trigonometrischen Funktion korrekt ist, dabei werden $c \cdot \sin \alpha = b$; $c \cdot \cos \alpha = a$; $a \cdot \cos \alpha = c$ verwendet). Die stärksten Unterschiede zwischen den drei Versionen sind in den Abbildungen 2 und 3 (s. unten) dargestellt. Weiterhin existieren sowohl ein Darreichungsmodus als automatisch einlesbarer Papier-Test als auch als Online-Test in einer eigenentwickelten Plattform. Aufgabenstellungen und -zusammensetzungen sind für die Modi identisch.

Die Aufgabenformate umfassen Single-Choice-Aufgaben mit im Allgemeinen vier Antwortalternativen, komplexe Multiple-Choice-Aufgaben mit vier bis fünf zu entscheidenden Aussagen sowie ganzzahlig numerische Antwortformate.

2.2 Testentwicklung und analytisches Vorgehen

2.2.1 Testentwicklung

Der Test wurde von Hochschullehrenden der Mathematik unter Rückgriff auf bestehende Aufgaben und Instrumente erfahrungsgelernt entwickelt. Psychometrische und testtheoretische Expertise wurde erst nach der ersten Datenerhebung einbezogen. Dies führt zu der im Folgenden sehr exploratorisch ausgelegten Methodenauswahl.

Die fachwissenschaftliche Strukturierung des Tests erfolgte vor demselben erfahrungsgelernten Hintergrund. Die Auswahl der Themen und Aufgaben adressiert bekannte Probleme in der Studieneingangsphase und ist durch die Aufteilung in kurze, zeitbegrenzte Blöcke so ausgelegt, dass diese perspektivisch auch erweitert werden kann. Die aufgegriffenen Themen decken sich mit wichtigen Themenfeldern, die in unterschiedlichen Betrachtungen der Studieneingangsphase identifiziert werden (vgl. die MaLeMINT-Studie; Neumann et al., 2017; und den Mindestanforderungskatalog; CoSH-Gruppe, 2021). Dabei wird, wie in Abschnitt 1.4 begründet, der Schwerpunkt auf die Mathematik der Sekundarstufe I gelegt und hierin aus Testumfangsgründen weiter selektiert.

Die Pilotierung mit $n = 1004$ Studienanfängenden wurde mit vergleichbaren Methoden wie im Folgenden skizziert analysiert. Es wurden insgesamt 6 von 16 Aufgaben identifiziert, die die psychometrischen Anforderungen nicht erfüllten – meist aufgrund einer deutlich zu hohen Schwierigkeit für die Zielgruppe. Diese wurden im folgenden Durchgang durch geeignetere Aufgaben ersetzt. Der grundlegende Aufbau und die Struktur des Tests konnten bestätigt werden.

2.2.2 Itemkodierung

Die Basiselemente aller Aufgaben (die Auswahl bei Single-Choice, der Eintrag in einem numerischen Freitextfeld sowie die Bewertung einer einzelnen Aussage bei komplexen Multiple-Choice) werden dichotom kodiert: 1 für die richtige Antwort und 0 für falsche Antworten. Fehlende Antworten werden dabei durchgängig als falsche Antworten gewertet, da diese im Sinne der Testkonstruktion als fähigkeitsmindernd zu interpretieren sind (Lüdtke et al., 2007). Diese Einzelbetrachtung wird gewählt, um zu vermeiden, dass unterscheidbare Fähigkeiten in einer Aufgabe zusammengefasst wurden und damit nicht sauber differenziert werden können (Tripp & Tollefson, 1985).

Zur Ermittlung sinnvoll skalierbarer Items werden diese Einzelkodierungen schrittweise zusammengefasst. Dabei werden unter Verwendung des 1PL-IRT-Modells sowohl die Itemfitstatistiken (Infit, Outfit sowie Trennschärfe; Wilson, 2005) als auch die paarweisen $Q_{3,*}$ -Statistiken (Chen & Thissen, 1997) herangezogen und diejenigen Items zusammengefasst, welche die höchsten $Q_{3,*}$ -Werte aufweisen, um die Grenzwertproblematik dieser Statistik (Chen & Thissen, 1997) zu umgehen. Dabei wird eine dichotome Kodierung beibehalten, wobei 1 für vollständig richtige Antworten steht.

Die mittlere Bearbeitungsdauer liegt bei 69% der verfügbaren Bearbeitungszeit (Rechnen 66%, Terme 71%, Geometrie 74% und Funktionen 63%). Weniger als 5% der Teilnehmenden hat die jeweils verfügbare Bearbeitungszeit ausgenutzt. Da außerdem nicht davon ausgegangen werden kann, dass die Teilnehmenden die Aufgaben in der gezeigten Reihenfolge bearbeiten, wurde auf eine spezielle Kodierung für fehlende Antworten aus Zeitgründen verzichtet.

2.2.3 Skalierung

Zur Skalierung werden Methoden der Item-Response-Theorie zu Grunde gelegt (Wilson, 2005). Der resultierende Itempool wird mittels Likelihood-Ratio-Tests bezüglich der notwendigen Parametrisierung (1PL, 2PL oder 3PL) sowie Between-Item-Multidimensionalität bezüglich der vier Testlets, also die abgegrenzt administrierten theoretisch konstruierten Themenbereiche (Rechnen, Terme, Geometrie und Funktionen) geprüft (Reise, 2015). Weiterhin wird das Informationskriterium nach Akaike unter Anpassung auf kleine Stichprobengrößen (da $N/p = 33.5 < 60$; Burnham & Anderson, 2002) verwendet. Alle hier angegebenen Parameter beziehen sich auf die *Difficulty*-Parametrisierung, also $\alpha_i \cdot (\theta - \beta_i)$. Ergänzend wird noch auf Itemebene die Notwendigkeit des Rateparameters der 3PL-Parametrisierung geprüft. Diese Prüfung scheint angebracht, da einzelne Items rein theoretisch eine hohe Ratewahrscheinlichkeit besitzen können. Die Passung des Modells auf die Daten wird mittels gängiger Kenn- und Grenzwerte überprüft: Itemfit-Statistiken ($\in [0.7, 1.3]$; Wilson, 2005), EAP/PV-Reliabilität ($> .7$; Bond & Fox, 2015) und SRMSR ($< .08$; Maydeu-Olivares, 2013).

2.2.4 Prüfung der Parallelität und der Fairness

Die Fairness des Tests, d.h. das Fehlen einer systematischen Benachteiligung einzelner Gruppen, sowie die Parallelität bezüglich der drei Versionen und der beiden Modi wird auf Basis von Methoden des *Differential Item Functioning* (DIF) geprüft. Da bezüglich der Fairness der drei Testversionen mehr als zwei Gruppen miteinander verglichen werden, werden statt

Hypothesentests zwei Kennwerte je Gruppe betrachtet: (1) Die Kommunalität h^2 einer einfaktoriellem Hauptkomponentenanalyse (Jolliffe, 2002) der Itemschwierigkeitsparameter mit Grenzwert $> .90$ sowie (2) die *Mean Absolute Deviation* (MAD) gegenüber der Itemschwierigkeitsparameter des Gesamtmodells mit Grenzwert < 0.25 (Grisay & Monseur, 2007). Die Itemdiskriminisationsparameter werden dabei gruppenunabhängig geschätzt und somit nicht berücksichtigt. Zusätzlich wird das zweite Kriterium erweitert: Da ein extremer Ausreißer durch viele gute Items zu akzeptablen Kennwerten führen kann, wird die absolute itembezogene Abweichung einbezogen und lediglich bis 0.4 akzeptiert.

Da im Fall der Versionen und Modi DIF identifiziert wird, werden die betroffenen Items unter Verwendung unterschiedlicher Itemschwierigkeitsparameter im Modell integriert. Auf dieser Basis erfolgt ein anschließender Vergleich der Schwierigkeiten der einzelnen Testzusammensetzungen mittels zweifaktorieller Varianzanalyse unter Verwendung des Zusammenhangsmaßes partielles η^2 .

2.2.5 Niveaumodellierung

Das absolute Leistungsfeedback erfordert eine individuell interpretierbare Skala. Aus diesem Grund wird die Personenfähigkeitsskala mittels Methoden von Beaton und Allen (1992) in ein Niveaumodell überführt. Die Niveaugrenzen werden dabei auf Häufungspunkte der Itemschwierigkeiten gesetzt. Die Beschreibung der Fähigkeiten erfolgte mittels qualitativer Anforderungsanalysen durch fünf Fachleute aus den Bereichen Mathematik, Mathematiklehramt sowie Erziehungswissenschaft. Dabei werden jeweils die Items herangezogen, welche von allen Studierenden des Niveaus n ausreichend sicher korrekt bearbeitet werden können ($P(X_i = 1 | \theta_{n,\min}) > .50$), aber nicht von den Studierenden des Niveaus unterhalb ($P(X_i = 1 | \theta_{n-1,\max}) < .65$). Die Verständlichkeit und Nutzbarkeit der Beschreibungen wurden durch Vorstellungen gegenüber unterschiedlichen Zielgruppen aus der Schulbildung, der universitären Bildung sowie der Fachwissenschaften überprüft und anhand der Rückmeldungen verbessert.

2.2.6 Gruppierung und Vergleich der Studiengänge

Das Gruppieren vergleichbarer Studiengänge erfolgt manuell, indem auf Basis der mittleren Fähigkeiten der einzelnen Studiengänge inhaltlich zusammengehörige Gruppen (z.B. Maschinenbau sowie Fahrzeug- und Motorentchnik) gebildet werden, innerhalb derer mittels einfaktoriellem Varianzanalyse die Zusammengehörigkeit identifiziert und unter Verwendung des Zusammenhangsmaßes η^2 quantifiziert wird.

2.2.7 Prognostische Validität

Zur Prüfung der Vorhersageeffekte werden Modelle der logistischen linearen Regression bestimmt. Dabei wird als Kriterium der Erwerb des Scheins der Lehrveranstaltung Höhere Mathematik 1 herangezogen, welcher angesichts der üblichen Bestehensquoten von lediglich um die 70% eine kritische Hürde im ersten Studiensemester darstellt. Als standardisierte Effektgröße dient zur unmittelbaren Vergleichbarkeit der Varianzaufklärungskoeffizient nach Nagelkerke (1991) beziehungsweise dessen Differenz bei Modellvergleichen. Das maximale Erklärungsmodell wird über schrittweises Hinzufügen nach größter Erhöhung des Vari-

anzaufklärungskoeffizienten nach Nagelkerke ermittelt, bis keine statistisch signifikante Modellverbesserung im Likelihood-Ratio-Test mehr stattfindet (Madsen & Thyregod, 2010).

2.3 Stichprobe

2.3.1 Gesamtstichprobe

Der Test adressiert Studienanfänger in MINT-Studiengängen dreier baden-württembergischer Universitäten aus dem zugrundeliegenden Projektkontext. Er soll perspektivisch aber auch Studienanfänger aller WiMINT-Studiengänge in Deutschland zwischen Erwerb der Hochschulzugangsberechtigung und Beginn des Studiums adressieren können. Durch das Vorhandensein spezifischer Förderangebote schon vor Studienbeginn sowie die bedingte Erreichbarkeit bereits vor Hochschuleintritt wird eine Gütebestimmung am ersten Vorkurstag durchgeführt. Dabei werden drei baden-württembergische Universitäten sowie in der zweiten Kohorte die Duale Hochschule Baden-Württemberg einbezogen. Dabei wird auf die Nennung und Darstellung hochschulspezifischer Ergebnisse verzichtet, um nicht den Eindruck eines (nicht belastbaren) Rankings entstehen zu lassen.

Die Erhebungen starteten zum Wintersemester 2018/19 mit einem Pilotinstrument. Seit dem Wintersemester 2019/20 ist dieses einheitlich gestaltet. Zum Wintersemester 2019/20 wurde rein in Papier-Form gemessen. Zum Wintersemester 2020/21 aufgrund der coronabedingten Umstellungen erfolgte die Erhebung ausschließlich online. Da zu diesem Zeitpunkt aber noch von relativ geringen Auswirkungen auf die Fähigkeiten bezüglich der Mathematik der Sekundarstufe I auszugehen ist, verwenden wir diese Besonderheit zur Prüfung der Invarianz der Modi. Die drei Versionen wurden jeweils gleichverteilt zufällig zugewiesen.

Das Verhältnis der Anzahl der Vorkursteilnehmenden zu Studienanfängernden unterliegt einer Unschärfe, da die Zahl der Angemeldeten und die der tatsächlich Teilnehmenden in den Vorkursen teils deutliche Diskrepanzen aufweist. Im Jahr 2019 nahmen alle Anwesenden des ersten Vorkurstags an der Erhebung teil. Im Jahr 2020 zeigt sich hier ein deutlicher Einbruch an einer Hochschule, wohingegen die beiden anderen Hochschulen hinzugewannen. Inwieweit die Stichproben repräsentativ für die Vorkursteilnehmenden stehen, kann mangels Vergleichsdaten nicht bestimmt werden. Es sind insgesamt 84 verschiedene Bachelor-Studiengänge berücksichtigt.

In Summe können für die Skalierung sowie die Modell- und Güteprüfungen $N = 3,819$ Studienanfängernde berücksichtigt werden. Dabei sind $n = 1,730$ Studienanfängernde (45%) im Jahr 2019 in Papier-Form und $n = 2,089$ Studienanfängernde im Jahr 2020 in Online-Form integriert. $n = 2,929$ Teilnehmende gaben ihr Geschlecht an, wovon $n = 976$ weiblich sind (33%), was einen durchaus realistischen Anteil in MINT-Studiengängen darstellt. $n = 2,885$ Teilnehmende gaben ihr Geburtsjahr an, das mittlere Alter zum Testzeitpunkt beträgt 19.5 Jahre ($SD = 1.7$ Jahre). Lediglich 55% der Teilnehmenden begannen das betroffene Studium direkt nach dem Abitur, was das vergleichsweise hohe mittlere Alter erklärt. Ob bei den anderen Teilnehmenden zwischen Abitur und Studienbeginn ein abgebrochenes Studium, eine Berufsausbildung oder andere Tätigkeiten lagen, wurde nicht abgefragt.

2.3.2 Stichprobe zur Bestimmung der prognostischen Validität

Zur Bestimmung der prognostischen Validität werden zwei Stichproben zusammengezogen. Mittels ausreichender itembasierter Verankerung des Pilotinstruments mit dem finalen Instrument ($\approx 65\%$) können die Pilotdaten auf dieselbe Skala gebracht (Reise, 2015) und somit die Leistungsdaten zweier Gruppen vereinigt werden: Für eine der Hochschulen konnten – auf freiwilliger Basis – Leistungsdaten für $n = 363$ Studierende bezüglich des ersten Studiensemesters in den Kohorten 2018 und 2019 einbezogen werden. Danach war aufgrund der fehlenden Passung des Datenschutzkonzepts auf die Corona-Situation keine Erhebung von Leistungsdaten mehr möglich.

Tabelle 1 nennt die Elemente der Hochschulzugangsberechtigung (HZB), die neben der mittels EAP-Schätzern (*Expected A-Posteriori*) modellierten Fähigkeit einbezogen wurden, und greift grundlegende deskriptive Maße auf.

Tabelle 1: Elemente der Hochschulzugangsberechtigung (HZB)

Merkmal	Codes	<i>N</i>	<i>M</i> (<i>SD</i>)
Gesamtnote	1.0 ... 3.9	353	2.17 (0.66)
Gesamtpunktzahl in Mathematik	≤ 4 (kodiert als 4) 5 ... 15	158	10.5 (3.2)
Schulart, an der die HZB erworben wurde	1: Allgemeinbildendes Gymnasium (AG) 0: Berufliche Gymnasien	250 100	.71
Bundesland, in dem die HZB erworben wurde	1: Baden-Württemberg (BW) 0: Andere Bundesländer	291 56	.84
Zeitlicher Abstand zwischen Erwerb der HZB und Beginn des Studiums	1: direkter Einstieg 0: mindestens ein Jahr Abstand	171 187	.48
Vertiefungskurs, wenn die HZB am AG in BW erworben wurde	1: teilgenommen 0: nicht teilgenommen	82 190	.30

3 Ergebnisse und Interpretation

3.1 Struktur- und Parameterprüfung

Zur Struktur- und Parameterprüfung wird das Informationskriterium AIC_c herangezogen. Dabei ergeben sich für die vier relevanten Struktur-Parameter-Kombinationen die folgenden Werte. 1-faktoriell 1PL: $AIC_c = 99013$; 1-faktoriell 2PL: $AIC_c = 98416$; 4-faktoriell 1PL: $AIC_c = 98719$; 4-faktoriell 2PL: $AIC_c = 98188$. Somit weist das 4-faktorielle 2PL-Modell den niedrigsten Wert auf, wobei die Differenz durchgängig bedeutsam ist (> 10 ; Burnham & Anderson, 2002). Die paarweisen Likelihood-Ratio-Tests dieses Modells mit dem 4-faktoriellen 1PL-Modell ($\chi^2(30) = 593.2$; $p < .001$) sowie mit dem 1-faktoriellen 2PL-Modell ($\chi^2(6) = 239.9$; $p < .001$) bestätigen diese Präferenz. Ein Vergleich mit der 3PL-Parameterschätzung erfolgt lediglich für die 1-faktorielle Modellierung, da ansonsten zu ungenauen Parameterschätzungen zu erwarten sind: $\chi^2(34) = 8.9$; $p = 1.000$. Wird jeweils ein Ratepara-

meter frei geschätzt und die anderen auf 0 fixiert, ergibt sich bei einem Item ein Parameter $\gamma_i = .031$, für alle anderen Items gilt $\gamma_i \leq .007$.

Zur einfacheren Kommunizierbarkeit der Ergebnisse sowie zur Reduktion der Ungenauigkeit aufgrund sehr geringer Testlängen soll zusätzlich eine Gesamtfähigkeit modelliert werden. Da deren Existenz nicht theoretisch begründet werden kann, wird ein hierarchisches Konstrukt abgelehnt, sodass ein *Bifactor*-Modell (Reise, 2012) in der 2PL-Parametrisierung gewählt wird. Hierfür werden Personenfähigkeiten mittels EAP-Schätzern gebildet, da WLE-Schätzer (*Weighted Likelihood Estimates*) keine Within-Item-Multidimensionalität unterstützen (Robitzsch & Steinfeld, 2018). Die Gütekennwerte dieses finalen Modells liegen innerhalb der oben genannten Grenzen: $Q_{3,*} \in [-0.37, 0.18]$; Item-Outfit $\in [0.80, 1.14]$; Item-Infit $\in [0.97, 1.01]$; SRMSR = .036; EAP/PV-Reliabilität des gemeinsamen Faktors = .77. Die Itemparameter der Difficulty-Parametrisierung liegen in den Bereichen $\alpha_{g,i} \in [0.31, 3.01]$ sowie $\beta_i \in [-4.6, 2.8]$. Abbildung 1 zeigt beispielhaft die zu Version A in der Papierform gehörige WrightMap unter Kennzeichnung der Zugehörigkeit zu den vier Residualfaktoren bezüglich der vier Inhaltsbereiche.

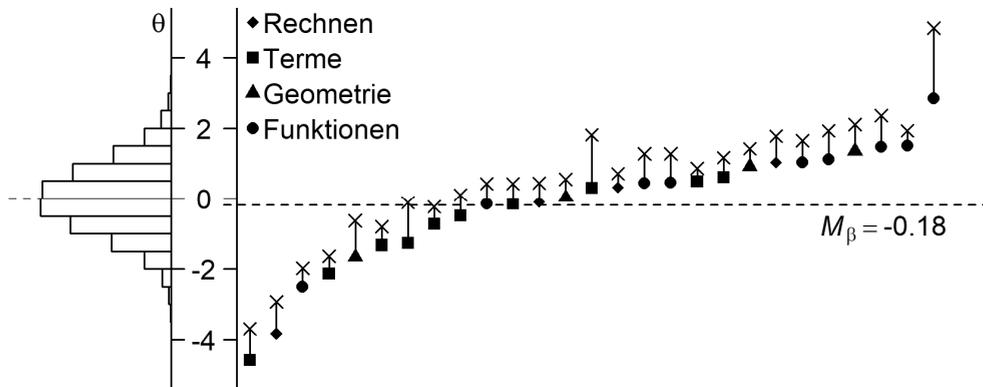


Abbildung 1: WrightMap des finalen Bifactor-Modells für Version A *Papier* unter Kennzeichnung der Zugehörigkeit zu den vier Residualfaktoren.

Anmerkung: Die beiden Itemparameter werden mit Hilfe des Bereichs der Lösungswahrscheinlichkeit von 50% bis 65% visualisiert.

Betrachtet man die vierfaktorielle Modellierung ohne den gemeinsamen Faktor, ergeben sich die in Tabelle 2 dargestellten Reliabilitäten sowie latenten Zusammenhänge. Gerade für individualdiagnostische Zwecke scheint diese Genauigkeit – insbesondere auf Basis der vier identifizierten Faktoren – noch deutlich optimierbar. Da die verfügbare Testzeit aus hochschulpolitischen Gründen kaum ausbaufähig erscheint, verbleiben drei Möglichkeiten: (1) Verbesserung der Genauigkeit des Itempools – unter Gefahr der Reduktion inhaltlicher Validität. (2) Entwicklung eines Feedbacksystems, welches die Messgenauigkeit kompensieren kann. Diese Optimierung wurde für das Wintersemester 2021/22 umgesetzt, ist aber nicht Teil dieses Beitrags. (3) Einsatz von Methoden des *Computerized Adaptive Testing* (van der Linden & Ren, 2019). Diese Optimierung wurde ebenfalls für das Wintersemester 2021/22 für den Teil zu Bruch- und Potenzrechnung umgesetzt.

Tabelle 2: EAP/PV-Reliabilität und latente Korrelationen im vierfaktoriellen Modell

Faktor	EAP/PV-Reliabilität	Latente Korrelation		
		Rechnen	Terme	Geometrie
Rechnen	.66			
Terme	.75	.76		
Geometrie	.66	.64	.81	
Funktionen	.71	.74	.86	.76

3.2 Invarianz bezüglich Modi und Versionen sowie Fairness bezüglich Geschlecht

Mangels inhaltlicher Vergleichbarkeit wurden zwei Items in den verschiedenen Versionen als unterschiedlich angesehen und aus der DIF-Prüfung ausgeschlossen. Hierbei handelt es sich um Aufgaben im komplexen Multiple-Choice-Format, wobei die einzelnen Entscheidungen nicht auf dieselben mathematischen Theoreme zurückführbar sind, obwohl der Inhalt der Aufgaben jeweils ähnlich ist. Ein Zusammenfassen allein auf Basis vergleichbarer Itemparameter erscheint nicht gerechtfertigt. Abbildung 2 zeigt ein Beispiel hierfür. Zwei weitere Items zeigen in jeweils einer Version eine deutliche Abweichung ihres Schwierigkeitsparameters vom mittleren Parameter und werden somit als DIF-auffällig klassifiziert. Diese Abweichung tritt in Geometrie-Aufgaben auf und kann wahrscheinlich auf unterschiedliche räumliche Komplexität oder unterschiedliche Komplexität des Lösungsweges oder des Lösungswertes begründet werden. Abbildung 3 zeigt ein Beispiel hierfür. Unter Berücksichtigung dieser Auffälligkeiten in der Modellierung resultieren die Kennwerte $h^2 = .99$ für alle drei Gruppen und $MAD \in [0.01, 0.12]$. Bezüglich der Modi sind ebenfalls zwei Items auffällig, was bei Berücksichtigung zu den Kennwerten $h^2 = .98$ und $MAD \in [0.13, 0.19]$ führt. Hier stellte sich bei Sichtung der aktuellen Bildungspläne in Baden-Württemberg (KM BW & ZSL, 2016) heraus, dass der Logarithmus zwar noch Bestandteil schulischer Lehre ist, aber nicht mehr zwingend alle Rechenregeln hierzu. Somit scheint hier die unbeaufsichtigte Administration der Online-Form zu Vorteilen und somit einer Reduzierung des Schwierigkeitsparameters zu führen. Abbildung 4 zeigt ein Beispiel hierfür.

Trotz der hieraus resultierenden leicht unterschiedlichen itembezogenen Zusammensetzungen zwischen den Versionen sowie den Modi zeigt Tabelle 3, dass diese verschiedenen Zusammensetzungen keinen nennenswerten Einfluss auf die jeweilige mittlere Testschwierigkeit haben. Somit müssen diese Effekte im Folgenden nicht berücksichtigt werden, und es kann von einer gemeinsamen, interpretierbaren Skalierung ausgegangen werden.

Sei $x \neq -1$ eine reelle Zahl.
 Welches ist eine richtige Umformung von $\frac{x^2 - 1}{x + 1}$?

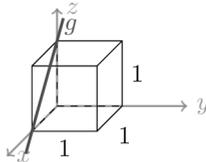
	richtig	falsch
$x - 1$	<input checked="" type="checkbox"/>	<input type="checkbox"/>
$x + 1$	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Seien x, y reelle Zahlen, $x \neq 0$ und $y \neq 0$.
 Welches ist eine richtige Umformung von $\frac{1}{\frac{1}{x} + \frac{1}{y}}$?

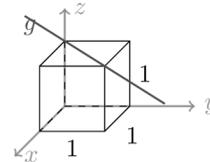
	richtig	falsch
$x + y$	<input type="checkbox"/>	<input checked="" type="checkbox"/>
$\frac{1}{x+y}$	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Abbildung 2: Teil einer Beispielaufgabe aus dem Bereich Terme mit unterschiedlichen zugrundeliegenden Theoremen

Die Gerade g verläuft durch die Punkte $(1, 0, 0)$ und $(0, 0, 1)$ (vgl. Skizze), die Gerade h verläuft durch die Punkte $(1, 0, 1)$ und $(1, 1, 1)$. Wie groß ist der Abstand der beiden Geraden g und h ?



Die Gerade g verläuft durch die Punkte $(0, 0, 1)$ und $(1, 1, 1)$ (vgl. Skizze), die Gerade h verläuft durch die Punkte $(1, 0, 0)$ und $(0, 1, 0)$. Wie groß ist der Abstand der beiden Geraden g und h ?



- 1
 1/2
 $\sqrt{2}$
 $\sqrt{2}/2$
 1
 1/2
 $\sqrt{2}$
 $\sqrt{2}/2$

Abbildung 3: Beispielaufgabe aus dem Bereich Geometrie mit versionsabhängiger Auffälligkeit im Differential Item Functioning

Sei $x = \log 3$. Drücken Sie $\log 9$ mit Hilfe von x aus.

- x
 $2x$
 $3x$
 $4x$

Abbildung 4: Beispielaufgabe aus dem Bereich Funktionen mit modusabhängiger Auffälligkeit im Differential Item Functioning

Tabelle 3: Varianzanalyse der Itemschwierigkeit bezogen auf Version und Modus

Merkmal	Varianzanalyse	Effektgröße
Version	$F(2) = 0.19; p = .829$	$\eta_{\text{part}}^2 = .002$
Modus	$F(1) = 0.10; p = .753$	$\eta_{\text{part}}^2 = .001$
Version * Modus	$F(2) = 0.00; p = 1.000$	$\eta_{\text{part}}^2 = .000$

Anmerkung: Die Anzahl der Items beträgt jeweils 28, außer für Version A lediglich 27.

Bezüglich des Geschlechts findet sich in den Kennwerten $h^2 = .98$ und $MAD \in [0.18, 0.25]$ kein Hinweis auf DIF. Somit können der Test und die Skala als fair bezüglich des Geschlechts angenommen werden.

3.3 Studiengangsunterschiede

Das Gruppieren der Studiengänge ergibt insgesamt 17 inhaltlich untereinander abgrenzbare Studiengangsgruppen. Dabei resultieren Gruppenmittelwerte im Intervall $\bar{\theta} \in [-0.65, 0.70]$ sowie ein Zwischengruppenvarianzanteil bezüglich der verschiedenen Gruppen an der Personenfähigkeit von $\eta^2 = .195$ ($F(16, 3440) = 52.1$; $p < .001$). Innerhalb der Studiengangsgruppen liegen die Zwischengruppenvarianzanteile bezüglich der einzelnen Studiengänge im Bereich $\eta^2 \in [.000, .050]$. Die Signifikanz dieser Varianzanalysen liegt im Bereich $p \in [.087, .969]$.

Die Zwischengruppenvarianzanteile sprechen dafür, dass die Zusammenfassung in die Gruppen zu interpretierbaren Vergleichen führt. Die Zwischengruppenvarianzanteile in den einzelnen Studiengangsgruppen zeigen dabei keine messbaren Unterschiede. Die Skala kann also gut zwischen verschiedenen Leistungen differenzieren, wobei ein nicht unerheblicher Varianzanteil auf individueller Ebene verbleibt. Eine Analyse, inwieweit dies den Anforderungen der verschiedenen Studiengänge entspricht, ist noch ausstehend und wird im Ausblick aufgegriffen. Da gerade die Grundlagenveranstaltungen in Höherer Mathematik gerne über Studiengänge hinweg zusammengefasst werden, ergeben sich hieraus zu berücksichtigende Aspekte bezüglich der Leistungsheterogenität der Studierenden.

3.4 Niveaumodellierung

Es resultieren vier auf der Logit-Skala äquidistante Schwellen für das Niveaumodell bei $\theta_1 = -2.1$, $\theta_2 = -0.9$, $\theta_3 = 0.3$ sowie $\theta_4 = 1.5$. Einschließlich des nicht beschreibbaren Niveaus können somit fünf Niveaustufen generiert werden. Aus den Verteilungsparametern der Personenfähigkeit ($\theta \sim \mathcal{N}(0,1)$) folgen somit lediglich 1% der Personen auf dem nicht beschreibbaren Niveau X, 17% der Personen auf Niveau I, 43% der Personen auf Niveau II, 32% der Personen auf Niveau III sowie 7% der Personen auf Niveau IV. Die Beschreibung der Niveaus gestaltet sich wie folgt:

- Studierende auf Niveau I können ausreichend sicher ...
 - Grundwissen der Sekundarstufe I wiedererkennen.
 - Grundtechniken der Sekundarstufe I anwenden.
- Studierende auf Niveau II können ausreichend sicher ...
 - einschrittige Probleme lösen.
 - Algorithmen der Sekundarstufe I anwenden, wenn die Lösungsidee offensichtlich ist.
- Studierende auf Niveau III können ausreichend sicher ...
 - mehrschrittige Probleme vollständig lösen.
 - Algorithmen der Sekundarstufe II anwenden, wenn die Lösungsidee trainiert wurde.
- Studierende auf Niveau IV können ausreichend sicher ...
 - mehrschrittige Probleme vollständig lösen.
 - Algorithmen mehrerer Wissensbereiche verknüpfen, auch wenn die Lösungsidee unbekannt ist.

Diese Abstufungen zeigen Parallelen zum in den PISA-Studien identifizierten Niveaumodell (OECD, 2019, S. 115). Dort findet sich ab Stufe 3 die Fähigkeit, „sequentielle Entscheidungen“ zu berücksichtigen, was hier mit der Mehrschrittigkeit der Problemlösung aufgenommen wird. Allerdings erreichen im Jahr 2018 54% der PISA-Teilnehmenden diese Stufe, in dieser Modellierung lediglich 39% der Teilnehmenden. Die Verknüpfung mehrerer Wissens-

bereiche findet sich auf Stufe 6 („Sie können verschiedene Informationsquellen und Darstellungen miteinander verknüpfen“) wieder. Diese wird von lediglich 2% der PISA-Teilnehmenden, aber von 7% der Teilnehmenden in dieser Modellierung erreicht.

Insbesondere die unteren Niveaustufen scheinen problematisch. Es fehlt im Allgemeinen weniger das Wissen über die Mathematik der Sekundarstufe I, sondern das Anwenden und Verknüpfen dieses Wissens, um problemhaltige Situationen lösen zu können. Gerade dies ist aber grundlegender Bestandteil von WiMINT-Studiengängen.

3.5 Prognostische Validität

Das finale Modell zur Erklärung des Scheinerwerbs unter Einbezug aller relevanter Merkmale wird in Tabelle 4 dargestellt. Die einbezogene Stichprobengröße beträgt $n = 335$. Der Anteil erklärter Varianz liegt bei 51.3% ($\chi^2(4) = 155.3; p < .001$).

Tabelle 4: Modell der logistischen linearen Regression zur Erklärung des Scheinerwerbs

exogene Variable	Odds-Ratio	$\Delta R^2_{\text{Nagelkerke}}$	Modellvergleichstest
Gesamtnote	0.1	19.3%	$\chi^2(1) = 67.0; p < .001$
Testskala	2.5	4.5%	$\chi^2(1) = 17.0; p < .001$
Schulart	3.5	3.9%	$\chi^2(1) = 14.7; p < .001$
Zeitlicher Abstand	0.4	3.1%	$\chi^2(1) = 11.6; p < .001$

Anmerkung: Die beiden letzten Spalten beziehen sich auf einen Modellvergleich zwischen einem Modell mit allen vier exogenen Variablen sowie einem Modell mit den drei anderen verbleibenden exogenen Variablen. Somit ist hiermit die alleinige Erklärungskraft unter Berücksichtigung des Restmodells dargestellt.

Obwohl die Gesamtpunktzahl in Mathematik der HZB ebenfalls einen erheblichen univariaten Erklärungsanteil zum Erfolg aufweist (36%), bleibt sie im finalen Modell unberücksichtigt. Dies liegt darin begründet, dass die Gesamtnote der HZB allein eine Varianzaufklärung an der Punktzahl Mathematik in Höhe von $R^2 = .62$ sowie gemeinsam mit der Testskala in Höhe von $R^2 = .65$ liefert. Es liegt nahe, dass genau diese gemeinsame Varianz in die Aufklärung des Erfolgs eingeht und die alleinige Mathematikleistung der HZB ihre Vorhersagekraft damit verliert. Dass dabei aber die Testskala im Modell erhalten bleibt, verstärkt die Interpretation, dass hier Leistungen gemessen werden, die nur bedingt in die einbezogenen HZB-Maße Eingang finden.

In Abbildung 5 wird der Effekt der Testskala visualisiert. Hierbei wird statt dem intervallskalierten EAP-Schätzer ($R^2_{\text{Nagelkerke}} = 23.5\%; \chi^2(1) = 67.4; p < .001$) das ordinalskalierte erreichte Niveau verwendet, um die visuelle Aussagekraft, auch aus Sicht der betroffenen Interpretierenden, zu erhöhen. Es zeigen sich disjunkte Konfidenzintervalle auf dem 95%-Niveau.

Es zeigt sich somit, dass die Maße der Hochschulzugangsberechtigung ein bedeutsamer Prädiktor für den Studienerfolg sind, insbesondere da diese Leistungsmerkmale über einen längeren Zeitraum hinweg einbeziehen und weitere Merkmale, wie zum Beispiel Anstrengungsbereitschaft, integrieren. Das hier vorgestellte Instrument kann trotzdem mit geringem Aufwand einen ergänzenden Beitrag leisten und insbesondere im Hinblick auf die einfach zu interpretierende Skala Möglichkeiten zu individuellen Interventionen aufzeigen.

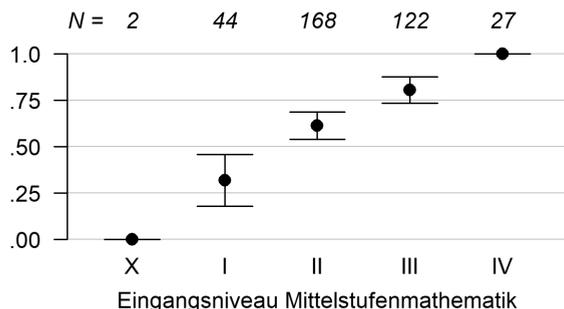


Abbildung 5: Wahrscheinlichkeit des Scheinerwerbs in Höhere Mathematik 1 nach erreichtem Eingangsniveau

Anmerkung: Die Fehlerbalken visualisieren das 95%-Konfidenzintervall der Mittelwertsschätzung.

4 Diskussion und Ausblick

4.1 Zusammenfassung der Ergebnisse

Mangels geeigneter Instrumente zur Überprüfung der mathematischen Basisfähigkeiten von Studierenden in der Studieneingangsphase wurde ein neues Instrument entwickelt, das sich insgesamt als reliabel und valide interpretierbar darstellt. Unter Verwendung eines 2PL-Bifactor-Modells mit vier Residualfaktoren (differenziert nach den inhaltlichen Testblöcken) kann die Gesamtfähigkeit der Studierenden modelliert werden, ein Rateparameter ist nicht notwendig (Fragestellung 1). Dieses kann mittels eines Niveaumodells auch für statistisch und psychometrisch Ungeübte einfach und verständlich interpretiert werden (Fragestellung 4).

Hinsichtlich der Darreichungsmodi sowie der Testvarianten ergeben sich nur marginale Unterschiede, welche in der Modellierung berücksichtigt werden. Im Vergleich der beiden Modi stellte sich ein Konstruktionsfehler heraus: Die Kennwerte des DIF identifizierten die beiden Logarithmus-Aufgaben als auffällig, was nach Betrachtung der Bildungspläne auf deren mangelnde curriculare Verankerung zurückzuführen ist. Darüber hinaus konnten aber keine Auffälligkeiten identifiziert werden, obwohl die Online-Bearbeitung ohne Aufsicht erfolgte. Die Zeitbegrenzung sowie die Freiwilligkeit der Teilnahme führen somit zu keinen Verzerrungen des gemessenen Konstrukts. Insgesamt sind alle Testformen auch in ihrer Schwierigkeit vergleichbar. Testfairness bezüglich des Geschlechts ist gegeben (Fragestellung 2).

Die Leistungsdifferenzierung nach Studiengängen sowie das Gruppieren in aussagekräftige Studienganggruppen zeigt erwartungskonforme und interpretierbare Ergebnisse. Eine differentielle Validität des Instruments kann somit impliziert werden (Fragestellung 3). Auch bezüglich der Vorhersagekraft zeigen sich eindeutige Effekte. Die Abiturnote hat zwar nach wie vor – als einziges der einbezogenen Merkmale – noch eine deutlich höhere Erklärungskraft an der erfolgreichen Teilnahme der Höheren Mathematik im ersten Studiensemester, trotzdem ergänzt der Test einen substanziellen zusätzlichen Anteil (Fragestellung 5). Dieser zusätzliche Anteil kann entscheidend sein für die Abwägung eines Studienabbruchs aufgrund fehlender Basisfähigkeiten, da er im Gegensatz zur allgemeineren Hochschulzugangsberech-

tigung ein gezielt zugeschnittenes Kompetenzspektrum erfasst und darüber hinaus mit ziel führendem Feedback ergänzt.

Übergreifend zeigt sich, dass – obwohl das Instrument aus einer rein fachwissenschaftlichen Perspektive entwickelt wurde – eine hohe psychometrische Güte erreicht werden konnte. In einem nächsten Schritt sind nun, neben der Klärung der im Ausblick beschriebenen offenen Fragestellungen, konkrete praktische Implikationen abzuleiten. Dies betrifft insbesondere die Gestaltung der Hochschullehre in den Grundlagenveranstaltungen. Die identifizierten, individuellen Förderbedarfe könnten durch ergänzende, niedrighschwellige Zusatzangebote adressiert werden. Gleichzeitig könnte eine Diskussion über die gängige Praxis, viele Studiengänge in gemeinsamen Lehrveranstaltungen zur Höheren Mathematik zu bündeln, angestoßen werden.

4.2 Limitationen und Ausblick

Die Güte der Modellierung des Tests spricht für ein großes Potenzial dieses Instruments, wirksam gegen spätere Studienabbrüche zu sein. Dennoch sind einige Limitationen festzuhalten.

Erstens ist der Themenumfang aufgrund der begrenzten Testzeit klar eingeschränkt. Zwar scheinen die Themen adäquat ausgewählt, wofür die hohe prädiktive Kraft spricht. Eine Überinterpretation im Sinne einer allgemeinen mathematischen Fähigkeit ist jedoch keinesfalls zulässig. Rückmeldungen durch betroffene Dozierende zeigen zugleich, dass diese Interpretationseinschränkung der praktischen Bedeutung der Resultate für die Gestaltung der Lehrveranstaltungen nur wenig entgegenwirkt. Die durch die Technologiebasierung fehlende direkte Interaktion mit den Testteilnehmenden macht eine finale Bewertung dieser Facette insbesondere auf Ebene der Testteilnehmenden jedoch schwierig. Inwieweit die Testteilnehmenden also das Feedback auch im Sinne einer konkreten Nachholempfehlung annehmen und umsetzen, bleibt derzeit offen.

Zweitens sind in den Stichproben spezifische Selektionseffekte zu vermuten: Die Mathematik-Vorkurse sind an allen Standorten ein freiwilliges Zusatzangebot vor Studienbeginn. Zur Frage, wie sich die hier stattfindenden Selbstselektionen auf die Zusammensetzung dieser Gruppe bezüglich aller betroffenen Studienanfängenden auswirkt, gibt es nur wenig Evidenz. Karaponos und Pelz (2021) sehen hier lediglich geringe Effekte, eine Übertragbarkeit der Erkenntnisse auf die vorliegende Stichprobe ist aber zweifelhaft. Dies macht sich insbesondere auch bezüglich der im Folgenden aufgegriffenen ausstehenden Fragestellungen bemerkbar.

Nichtsdestotrotz können mit dem verwendeten technologie- und IRT-basierten Ansatz einige Vorteile gegenüber klassischem Testen erreicht werden: Es steht direkt ein Feedback für die Testteilnehmenden sowie für weitere betroffene Personengruppen (zum Beispiel Dozierende) zur Verfügung, welches nicht nur auf der Aufgabenebene oder einer summarischen Ebene, sondern direkt auf der Konstruktebene interpretiert werden kann. Dieser Ansatz befreit aber nicht von der Verpflichtung, ein Feedback zu entwickeln, welches qualitativ hochwertig ist. Im Gegenteil werden durch die fehlende Möglichkeit der Interaktion mit den Testteilnehmenden besonders hohe Anforderungen an die Test- und die Feedbackqualität notwendig. Die hier genannten Limitationen führen zu vier zentralen Implikationen für die Weiterentwicklung des Testinstruments und der analytischen Vorgehensweise:

(1) Die in diesem Beitrag berichteten Reliabilitätsmaße zeigen noch ein deutliches Verbesserungspotential. Insbesondere die Verwendung als individualdiagnostisches Instrument verlangt, dass hierauf weiterhin ein besonderes Augenmerk gelegt wird. Es fanden bereits erfolgversprechende Ansätze Anwendung, die einerseits durch *Computerized Adaptive Testing* eine passgenauere Itemauswahl ermöglichen und andererseits durch die verwendeten Feedbackskalen die Bedeutsamkeit des Messfehlers für die Interpretation reduzieren. Die Erkenntnisse hierzu werden sich aus Umfangsgründen in einem weiteren Beitrag finden.

(2) In diesem Beitrag wurden die Ergebnisse erster Validitätsüberprüfungen berichtet. Allerdings sind weitere Prüfungen unerlässlich, um sicherstellen zu können, dass das entwickelte Leistungsmaß zuverlässig und zielgerichtet interpretiert werden kann. Einerseits sind Betrachtungen der inkrementellen Validität gegenüber alternativen Instrumenten und Skalen für eine bessere Einordnung notwendig. Andererseits muss eine repräsentative Stichprobe zur Prüfung der prognostischen Validität gefunden werden. Darüber hinaus muss der Test zu verschiedenen anderen Zeitpunkten, idealerweise längsschnittlich, eingesetzt werden, um einerseits eine Unabhängigkeit von den Selektionseffekten der Vorkurszusammensetzung und andererseits eine Eignung für die Studienwahlphase analysieren zu können.

(3) Inwieweit die differentiellen Unterschiede in den verschiedenen Studienganggruppen auch die tatsächlichen Anforderungen dieser Studiengänge widerspiegeln, ist noch nicht geklärt. Gleichzeitig scheint überaus interessant, welche Mindestniveaus für verschiedene Studiengänge oder Lehrveranstaltungen erreicht werden müssen. Hierzu ist eine Befragung der betroffenen Dozierenden auf Itemebene geplant, welche diese Informationen zuverlässig herausarbeiten kann. Gleichzeitig können hieraus noch Potentiale identifiziert werden, in welchen Skalenbereiche auf welche Art weitere Differenzierungen beziehungsweise Aufgaben integriert werden müssen, um die Aussagekraft der Interpretation weiter erhöhen zu können.

(4) Belastbare Aussagen zur Qualität und zu Optimierungsbedarfen des entwickelten Feedbacks sowie zu dessen praktischen Implikationen stehen ebenfalls noch aus. So sind jeweils Interviews geplant, welche sowohl die konkreten Fehlerursachen bei der Testbearbeitung als auch die konkreten Interpretationen der Studierenden bezüglich ihres Feedbackverständnisses erfassen sollen, um Informationen zu erhalten, welche Feedbackelemente zu Fehlinterpretationen führen. Dass die Ergebnisse bei den Dozierenden Reflektionsprozesse anstoßen, wurde von dieser Zielgruppe vielfach rückgemeldet. Inwieweit das Instrument auch Verhaltensänderungen bei den Testteilnehmenden und den Dozierenden auslöst sowie Studienabbrüchen entgegenwirkt, ist eine bislang offene empirische Frage.

Hinweis

Die hier vorgestellten Entwicklungen und Ergebnisse entstanden in einem Forschungsprojekt, welches vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg in Förderlinie 4 des Fonds *Erfolgreich Studieren* gefördert wurde.

Literatur

- AERA, APA & NCME (2014). *Standards for educational and psychological testing*. AERA.
- Bach, V. (2016). Kompetenzorientierung und Mindestanforderungen. *Mitteilungen der Deutschen Mathematiker-Vereinigung*, 24(1), 30–32. <https://doi.org/10.1515/dmvm-2016-0015>.
- Bausch, I., Biehler, R., Bruder, R., Fischer, P. R., Hochmuth, R., Koepf, W., Schreiber, S. & Wassong, T. (Hrsg.). (2014a). *Mathematische Vor- und Brückenkurse. Konzepte, Probleme und Perspektiven*. Springer Spektrum. <https://doi.org/10.1007/978-3-658-03065-0>.
- Bausch, I., Biehler, R., Bruder, R., Fischer, P. R., Hochmuth, R., Koepf, W. & Wassong, T. (2014b). VEMINT – Interaktives Lernmaterial für mathematische Vor- und Brückenkurse. In I. Bausch, R. Biehler, R. Bruder, P. R. Fischer, R. Hochmuth, W. Koepf, S. Schreiber & T. Wassong (Hrsg.), *Mathematische Vor- und Brückenkurse (Konzepte und Studien zur Hochschuldidaktik und Lehrerbildung Mathematik)*. Springer Spektrum. https://doi.org/10.1007/978-3-658-03065-0_18.
- Beaton, A. E. & Allen, N. L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics*, 17(2), 191–204. <https://doi.org/10.2307/1165169>.
- Belcadhi, L. C. (2016). Personalized feedback for self assessment in lifelong learning environments based on semantic web. *Computers in Human Behavior*, 55(A), 562–570. <https://doi.org/10.1016/j.chb.2015.07.042>.
- Bond, T. & Fox, C.M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge. <https://doi.org/10.4324/9781315814698>.
- Brunner, S. (2017). *Online-Self-Assessments*. Koordinierungsstelle der Begleitforschung des Qualitatspaktes Lehre (KoBF).
- Brunner, S., Ranft, A. & Wittig, W. (2015). Online-Self-Assessments: die Bedeutung von Feedback und Implikationen fur die (Weiter-)Entwicklung von Verfahren fur beruflich qualifizierte Studieninteressierte. In A. Hanft, O. Zawacki-Richter & W. B. Gierke (Hrsg.), *Herausforderung Heterogenitat beim Ubergang in die Hochschule* (S. 145–162). Waxmann.
- Burnham, K. P. & Anderson, D. R. (2002). *Model selection and multimodel inference. A practical information-theoretic approach*. Springer. <https://doi.org/10.1007/b97636>.
- Chen, W. & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22, 265–289. <https://doi.org/10.2307/1165285>.
- CoSH-Gruppe (Hrsg.). (2021). *Mindestanforderungskatalog Mathematik*. Version 3.0. https://lehrerfortbildung-bw.de/u_matnatech/mathematik/bs/bk/cosh/katalog/makv3.0.pdf
- Durrschnabel, K., Durr, R., Erben, W., Gercken, M., Lunde, K., Wurth, R. & Zimmermann, M. (2019). *So viel Mathe muss sein! Gut vorbereitet in ein WiMINT-Studium*. Springer Spektrum. <https://doi.org/10.1007/978-3-662-57951-0>.
- Grisay, A. & Monseur, C. (2007). Measuring the equivalence of item difficulty in the various versions of an international test. *Studies in Educational Evaluation*, 33(1), 69–86. <https://doi.org/10.1016/j.stueduc.2007.01.006>.
- Hanft, A., Zawacki-Richter, O. & Gierke, W. B. (Hrsg.). (2015). *Herausforderung Heterogenitat beim Ubergang in die Hochschule*. Waxmann.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>.
- Heublein, U., Ebert, J., Hutzsch, C., Isleib, S., Konig, R., Richter, J. & Woisch, A. (2017). *Motive und Ursachen des Studienabbruchs an baden-wurttembergischen Hochschulen und beruflicher Verbleib der Studienabbrecherinnen und Studienabbrecher*. DZHW Projektbericht.
- Heublein, U., Richter, J. & Schmelzer, R. (2020). Die Entwicklung der Studienabbruchquoten in Deutschland. *DZHW BRIEF*, 3, 1–12.
- Jahnke, T., Klein, H. P., Kuhnel, W., Sonar, T. & Spindler, M. (2014). Die Hamburger Abituraufgaben im Fach Mathematik. Entwicklung von 2005 bis 2013. *Mitteilungen der Deutschen Mathematiker-Vereinigung*, 22(2), 115–122. <https://doi.org/10.1515/dmvm-2014-0046>.

- Jolliffe, I. T. (2002). *Principal component analysis*. Springer. <https://doi.org/10.1007/b98835>.
- Karapanos, M. & Pelz, R. (2021). Wer besucht Mathematikvorkurse? *Zeitschrift für Erziehungswissenschaft*, 24(5), 1231–1252. <https://doi.org/10.1007/s11618-021-01035-2>.
- Karst, K., Ertelt, B.-J., Frey, A. & Dickhäuser, O. (2017). Studienorientierung durch Self-Assessments: Veränderung von Einstellungen zum Studienfach während der Bearbeitung eines Selbsttests. *Journal für Bildungsforschung Online*, 9(2), 205–227. <https://doi.org/10.25656/01:14935>.
- Knorrenschild, M. (2004). *Vorkurs Mathematik. Ein Übungsbuch für Fachhochschulen*. Carl-Hanser.
- Kubinger, K. D., Frebort, M. & Müller, C. (2012). Self-Assessment im Rahmen der Studienberatung: Möglichkeiten und Grenzen. In K. D. Kubinger, M. Frebort, L. Khorramdel & L. Weitensfelder (Hrsg.), *Self-Assessment: Theorie und Konzepte* (S. 9–24). Pabst Science Publishers.
- Krawitz, J. (2020). *Vorwissen als nötige Voraussetzung und potentieller Störfaktor beim mathematischen Modellieren*. Springer Spektrum. <https://doi.org/10.1007/978-3-658-29715-2>.
- Krunke, S. O., Roegner, K., Schüler, L., Seiler, R. & Stens, R. L. (2012). Der Online-Mathematik-Brückenkurs OMB. Eine Chance zur Lösung der Probleme an der Schnittstelle Schule/Hochschule. *Mitteilungen der Deutschen Mathematiker-Vereinigung*, 20(2), 115–120. <https://doi.org/10.1515/dmvm-2012-0048>.
- Lütke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Praxis. *Psychologische Rundschau*, 58(2), 103–117. <https://doi.org/10.1026/0033-3042.58.2.103>.
- Madsen, H. & Thyregod, P. (2010). *Introduction to General and Generalized Linear Models*. CRC Press.
- Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models. *Measurement: Interdisciplinary Research and Perspectives*, 11, 71–137. <https://doi.org/10.1080/15366367.2013.831680>.
- Meyberg, K. & Vachenaer, P. (2001). *Höhere Mathematik 1*. Springer. <https://doi.org/10.1007/978-3-642-56654-7>.
- Ministerium für Kultus, Jugend und Sport Baden-Württemberg (KM BW) & Zentrum für Schulqualität und Lehrerbildung (ZSL) (Hrsg.). (2016). *Bildungspläne Baden-Württemberg. Gymnasium – Mathematik*. https://www.bildungsplaene-bw.de/site/bildungsplan/get/documents/lsw/export-pdf/depot-pdf/ALLG/BP2016BW_ALLG_GYM_M.pdf
- Nagelkerke, N. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78, 691–692. <https://doi.org/10.1093/biomet/78.3.691>.
- Neubrand, M., Biehler, R., Blum, W., Cohors-Fresenborg, E., Flade, L., Knoche, N., Lind, D., Löding, W., Möller, G. & Wynands, A. (2004). Eine systematische und kommentierte Auswahl von Beispielaufgaben des Mathematiktests in PISA 2000. In M. Neubrand (Hrsg.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland* (Bd. 23) (S. 259–270). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-80661-1_13.
- Neugebauer, M., Heublein, U. & Daniel, A. (2019). Studienabbruch in Deutschland: Ausmaß, Ursachen, Folgen, Präventionsmöglichkeiten. *Zeitschrift für Erziehungswissenschaft*, 22, 1025–1046. <https://doi.org/10.1007/s11618-019-00904-1>.
- Neumann, I., Pigge, C. & Heinze, A. (2017). *Welche mathematischen Lernvoraussetzungen erwarten Hochschullehrende für ein MINT-Studium?* IPN.
- Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) (2019). *PISA 2018 Ergebnisse*. Band I. Was Schülerinnen und Schüler wissen und können. wbv Media. <https://doi.org/10.3278/6004763w>.
- Petri, P. S. (2020). *Ein Prozessmodell des Studieneinstiegs. Differentielle Aspekte studiumsbezogener Kognitionen und deren Effekte auf Studienerfolg und Studienabbruch*. Dissertation, Justus-Liebig-Universität Gießen.
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667–696. <https://doi.org/10.1080/00273171.2012.715555>.

- Reise, S. P. (Hrsg.). (2015). *Handbook of item response theory modeling. Applications to typical performance assessment*. Routledge.
- Reiss, K., Weis, M., Klieme, E. & Köller, O. (2019). *PISA 2018. Grundbildung im internationalen Vergleich*. Waxmann. <https://doi.org/10.31244/9783830991007>.
- Robitzsch, A. & Steinfeld, J. (2018). Item response models for human ratings: Overview, estimation methods, and implementation in R. *Psychological Test and Assessment Modeling*, 60(1), 101–139.
- Schäfer, W., Georgi, K. & Trippler, G. (1997). *Mathematik-Vorkurs. Übungs- und Arbeitsbuch für Studienanfänger*. Teubner. <https://doi.org/10.1007/978-3-322-97616-1>.
- Schwippert, K., Kasper, D., Köller, O., McElvany, N., Selter, C., Steffensky, M. & Wendt, H. (2020). *TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Waxmann.
- Tripp, A. & Tollefson, N. (1985). Are complex multiple-choice options more difficult and discriminating than conventional multiple-choice options? *Journal of Nursing Education*, 24(3), 92–98. <https://doi.org/10.3928/0148-4834-19850301-04>.
- TU9 (Allianz führender Technischer Universitäten in Deutschland) (2020). *Starthilfe fürs Studium: Online-Brückenkurs Physik. TU9-zertifiziertes kostenfreies digitales Lernangebot*.
- van der Linden, W. J. & Ren, H. (2019). A fast and simple algorithm for Bayesian adaptive testing. *Journal of Educational and Behavioral Statistics*, 45(1), 58–85. <https://doi.org/10.3102/1076998619858970>.
- Wilson, M. (2005). *Constructing measures. An item response modeling approach*. Routledge.

Kontakt

Stefan Behrendt
Schwenninger Str. 21
78083 Dauchingen
E-Mail: behrendt@bestetistics.de

Jan Köllner
Universität Stuttgart
Institut für Analysis, Dynamik und Modellierung
Pfaffenwaldring 57
70569 Stuttgart
E-Mail: jan.koellner@mathematik.uni-stuttgart.de

Prof. Dr. Kristina Kögler
Universität Stuttgart
Institut für Erziehungswissenschaft
Geschwister-Scholl-Str. 24D
70174 Stuttgart
E-Mail: koegler@bwt.uni-stuttgart.de

Prof. Dr. Christine Sälzer · Andreas Just
Universität Stuttgart
Institut für Erziehungswissenschaft
Azenbergstr. 16
70174 Stuttgart
E-Mail: christine.saelzer@ife.uni-stuttgart.de
E-Mail: just@bwt.uni-stuttgart.de

Online, blended oder Präsenz? Ein systematisches Literaturreview von Metaanalysen zur Effektivität hochschulischer Lehrformate

Elisabeth Mayweg, Natalie Enders, Carla Bohndick, Jana Rückmann

Zusammenfassung: Aktuelle Überlegungen zur Gestaltung effektiver Hochschullehre drehen sich vermehrt auch um die Wahl eines geeigneten Lehrformats, d.h. wie und in welchem Ausmaß soll Lehre (weiterhin auch) im Digitalen stattfinden. Der vorliegende Beitrag basiert auf einem systematischen Literaturreview von N = 44 Metaanalysen (Publikationszeitraum 2011–2022) zur lernbezogenen Effektivität der Lehrformate *Onlinelehre* und *Blended Learning* (inklusive *Flipped Classroom*) im Vergleich mit traditioneller *Präsenzlehre*. Die Ergebnisse liefern eine evidenzbasierte Grundlage für Empfehlungen zum Einsatz der Formate in der Hochschullehre. Es zeigen sich überwiegend kleine positive Effekte der Lehrformate auf kognitive Lernergebnisse, praktische Fertigkeiten und Studierendeneinschätzungen. Seltener werden mittlere Effekte berichtet, wobei sich Blended-Learning-Formate gegenüber reinen Online- und Präsenzformaten leicht im Vorteil zeigen. Die Höhe der Effektstärken ist zudem von den Einschlusskriterien der Metaanalysen abhängig. Moderatorenanalysen zeigen an, dass insbesondere über durch Lehrende veränderbare Gestaltungsmerkmale zusätzliche Varianz aufgeklärt wird.

Schlüsselwörter: Lehrformate, Effektivität, Onlinelehre, Blended-Learning, Flipped Classroom, Review

Online, blended or in presence: A systematic review of meta-analyses on the effectivity of university teaching formats

Summary: Current considerations on the design of effective teaching in higher education frequently revolve around the choice of a suitable format, i.e., how and to what extent teaching should (also) continue to take place in the digital realm. This paper is based on a systematic literature review of N = 44 meta-analyses (publication period 2011–2022) on the learning-related effectiveness of the teaching formats online teaching, blended learning (including flipped classroom) in comparison with traditional face-to-face teaching. The results provide an evidence-based foundation for recommendations on the use of the formats in higher education teaching. Predominantly small positive effects of the teaching formats on cognitive learning outcomes, practical skills, and student assessments are found. Rarely, medium effects are reported, with blended learning formats showing a slight advantage over online and traditional face-to-face teaching. The magnitude of effect sizes also depends on the inclusion criteria of the meta-analyses. Furthermore, the moderator analyses indicate that additional variance is explained in particular by design features that can be modified by teachers.

Keywords: teaching formats, effectiveness, online teaching, blended learning, flipped classroom, review

1 Einleitung

Mehrere Semester pandemiebedingter *Onlinelehre* an deutschen Hochschulen haben eine rege öffentliche Debatte über die zukünftige Gestaltung von Vorlesungen und Seminaren entfacht (Heinemann & Funke, 2020) und auf politischer Ebene einen Digitalisierungsschub angestoßen (HRK, 2021). So diskutieren sowohl Lehrende als auch Instituts- und Hochschulleitungen und Politiker*innen die Frage, wie und in welchem Ausmaß Studierende zukünftig weiterhin digital (unterstützt) lernen sollen. Welche Impulse aus der pandemischen Zeit haben eine positive Wirkkraft auf die Weiterentwicklung von Hochschullehre und sollen entsprechend nachhaltig weiter verankert und ausgebaut werden? Zum Umgang mit dieser Frage werden häufig stark durch subjektive Eindrücke beteiligter Akteur*innen geprägte Vergleiche der Onlinelehre der vergangenen Semester mit der Präsenzlehre vor der Corona-Pandemie angestellt. Allerdings liegen zu dieser Fragestellung bereits Ergebnisse aus Befragungen der beteiligten Akteur*innen vor, die die subjektiven Eindrücke objektivieren können. Deutlich wird dabei, dass vor allem die räumliche (und für asynchrone Formate auch zeitliche) Flexibilität der Onlinelehre allgemein als Vorteil gesehen wird, wobei die Studierenden jedoch gleichzeitig den mangelnden Austausch mit anderen Studierenden und Lehrenden sowie ein zu geringes Vorhandensein aktiver Lernmöglichkeiten betonten (Schmidberger et al., 2022). Geringe Erfahrung der Lehrenden und die durch den abrupten Wechsel bedingte unzureichende Zeit für die Vorbereitung digitaler Angebote führten dazu, dass insbesondere zu Beginn der Pandemie vor allem rein frontale und unidirektionale Formate umgesetzt wurden (Horstmann et al., 2021; Lörz et al., 2020;). In der Konsequenz ist es kaum verwunderlich, dass sich die Zufriedenheit der Studierenden mit der digitalen Lehre während der Pandemie eher im mittleren Bereich bewegt (Lörz et al., 2020) und stark von personenbezogenen Variablen wie z.B. Motivation beeinflusst ist (Schley & Buchwald, 2022). Vor dem Hintergrund dieser Ergebnisse ist eine Gleichsetzung des pandemiebedingten *Emergency Remote Teaching* (Hodges et al., 2020) mit didaktisch gut durchdachter und wissenschaftlich-fundierter Onlinelehre („high-quality online instruction“, Adedoyin & Soykan, 2020) aus wissenschaftlicher Perspektive hoch problematisch. Um die Erfahrungen aus der pandemiebedingten Hochschullehre fruchtbar für die zukünftige, postpandemische Hochschullehre zu nutzen, muss auch das evidenzbasierte Wissen aus der Zeit vor der Pandemie berücksichtigt werden, d.h. Wissen darüber, wie Lehre unter Einbezug des digitalen Raums als Lernort gestaltet werden kann und sollte (Knaus et al., 2022). Aus dem Kontext der (psychologischen) Lehr-Lern-Forschung liegen zahlreiche Metaanalysen vor (vgl. Abschnitt 4 in diesem Beitrag), die Hinweise zur (vergleichenden) Effektivität digital-gestützter Lehrformate, d.h. Onlinelehre und/oder Mischformen wie *Blended-Learning* (BL), mit der traditionellen *Präsenzlehre* bereitstellen. Leider werden diese wertvollen Erkenntnisse bisher zu selten in den laufenden Diskurs um die zukünftige Gestaltung der Hochschullehre eingebunden. In diesem Beitrag wird der Versuch unternommen, anhand eines systematischen Literaturreviews bereits existierender Metaanalysen zur Effektivität dreier verschiedener Lehrformate (reine Onlinelehre, BL und traditionelle Präsenzlehre) hierfür eine evidenzbasierte Grundlage zu schaffen. Dabei wird zum einen die Frage adressiert, wie sich die Formate auf unterschiedliche Determinanten des Lernerfolgs von Studierenden auswirken und welches der Formate dabei am förderlichsten ist. Zudem wird herausgestellt, welche weiteren Faktoren die Effekte des Formats auf den Lernerfolg beeinflussen.

2 Theoretischer Hintergrund und Forschungsfragen

Vergleiche zwischen digitalen und analogen Lehr-Lern-Formen in Bezug auf deren Auswirkungen auf das Lernen und die Leistung von Studierenden wurden in den letzten zwei Dekaden bereits vielfach vorgenommen und metaanalytisch ausgewertet (Bernard et al., 2014; Means et al., 2013). So ist eine umfangreiche empirische Grundlage vorhanden, die es erlaubt, systematisch und evidenzbasiert Aussagen über die Effektivität unterschiedlicher Formate abzuleiten. Als ein zentrales Kriterium zur Erfassung effektiver Lehre wird in vielen Studien der Lernerfolg der Studierenden herangezogen (Schneider & Preckel, 2017; Chaudhary & Singh, 2022), der über das erlangte deklarative und prozedurale Wissen in Form von Leistungsbeurteilungen abgebildet werden kann (Steinmayr et al., 2014). Digitale Lernumgebungen zeigen hier ein hohes Potential zur Förderung des Verständnisses fachspezifischer Konzepte (Merchant et al., 2014) und für die Weiterentwicklung fachspezifischer sowie fachübergreifender Fachkompetenzen (Chernikova et al., 2020; Cook et al., 2011). Darüber hinaus eignen sich auch Maße zur Bewertung von Lerngelegenheiten, wie die Zufriedenheit von Studierenden in Evaluationen (Reichert, 2019).

Es muss berücksichtigt werden, dass die in der einschlägigen Literaturlandschaft vorrangig genutzten Definitionen zu den Formaten (Präsenz, Online, BL/Flipped Classroom) oft nicht trennscharf sind und Begrifflichkeiten unterschiedlich verwendet werden. Bei der Gegenüberstellung von Online- und Präsenzunterricht wird vor allem räumliche und physische Anwesenheit als Zuordnungskriterium benannt: So betont Galvis (2018), dass Präsenzunterricht durch eine Face-to-face-Situation gekennzeichnet ist, in welcher Lernende und Lehrende in persona an einem physischen Lernort (z.B. im Seminarraum) interagieren. Lehrende und Lernende stehen dabei in einer direkten beziehungsweise räumlichen Beziehung zueinander, und es müssen – der jeweiligen Lehrmethode entsprechend – die notwendigen Ressourcen (z.B. Materialien, Technik) verfügbar sein. Demgegenüber zeichnet sich Online-Unterricht der Definition von Amaka und Goeman (2017) folgend dadurch aus, dass Lernprozesse ausschließlich über das Internet, beziehungsweise über digitale Systeme vermittelte Kommunikations- und Interaktionsprozesse, angeregt werden. Des Weiteren wird von mehreren Autor*innen betont, dass nicht-webbasierte, rein computergestützte Anwendungen (z.B. Lernprogramme) nicht der Onlinelehre zugeordnet werden sollten, diese aber häufig flankieren und eine strikte Trennung kaum noch möglich ist (Allen et al., 2004; Means et al., 2009).

Präsenzlernen und Online-Unterricht werden im BL-Format zusammengeführt (Amaka & Goeman, 2017; Anthony et al., 2022; Bernard et al., 2014; Vo et al., 2017). Geblendete Ansätze sind sicherlich das heutzutage gängigste Lehrformat, da so gut wie alle Universitäten ein Learning Management System (LMS) nutzen, über das zumindest Lernmaterialien zur Verfügung gestellt und organisatorische Anliegen geklärt werden können. BL wird deshalb oft als das *new normal* oder das *new traditional model* bezeichnet (Dziuban et al., 2018). So beziehen sich die meisten Konzeptionalisierungsversuche zunächst auf die Zusammensetzung von Online und Präsenz, wie beispielsweise in der vielfach zitierten Definition von Graham (2005), der BL als „combination of face-to-face instruction with computer-mediated instruction“ beschreibt (S. 5). Dieser Ansatz wird von einigen Autor*innen weitergeführt, indem die genauen Anteile von Präsenz- und Onlinelehre als Kriterium für die Zuordnung zum Format Blended Learning hervorgehoben werden (z.B. bei Bernard et al., 2014; Mül-

ler & Mildenerger, 2021). Diese Rahmung ist jedoch mitunter kritisch zu betrachten, da weniger der zeitliche Aspekt als vielmehr die sich in dem jeweiligen Lernraum verbundenen (didaktischen) Gestaltungsmöglichkeiten entscheidend für die Beschreibung des Formats sein sollten. So ist sicherlich fraglich, inwieweit die reine Bereitstellung von Lernmaterialien über ein LMS ohne didaktisches Begleitkonzept bereits als BL bezeichnet werden kann. Einer didaktischen Begründung zufolge kann BL als „a combination of a number of pedagogical approaches, which is not necessarily dependent on the use of learning technologies“ (Amaka & Goeman, 2017, S. 99) verstanden werden und erfordert „a thoughtful integration of classroom face-to-face learning experiences with online learning experiences“ (Garrison & Kanuka, 2004, S. 5). Somit können BL-Formate eine Kombination aus unterschiedlichen Lehrmethoden und Vorgehensweisen adressieren, die nicht notwendigerweise an die jeweilige Lernumgebung gebunden sein müssen (Kaur, 2013).

Es stellt sich die Frage, wie sich eine „thoughtful integration“ (Garrison & Kanuka, 2004, S. 5) gestalten kann beziehungsweise wie Online- und Präsenzlehre didaktisch begründet zusammengeführt werden und somit die Potentiale geblendeter Formate entfaltet werden können. Diese Forderung adressiert unter anderem das Format des *Flipped-* oder *Inverted Classrooms* (FC), bei dem es sich um eine spezielle Umsetzungsvariante des BL handelt. Die Besonderheit des FC gegenüber anderen Umsetzungsvarianten des BL liegt dabei in der Umkehrung der klassischen Funktionen der asynchronen Selbstlern- und synchronen Präsenzphase. So definieren Chen et al. (2018) den Ansatz durch „reversing lecture and homework elements of a course“ (S. 910). Die Umkehrung erfolgt dadurch, dass in der Selbstlernphase neue Inhalte (mit dem Ziel des Wissenserwerbs) von den Studierenden weitestgehend eigenständig erarbeitet werden, sodass in der gemeinsamen Präsenzphase die durch den Wegfall der Informationsvermittlung gewonnene gemeinsame Lernzeit für die Bearbeitung anspruchsvollerer Lernziele (insbesondere die Anwendung und die kritische Reflexion des Lernstoffs) aufgewendet werden kann. In der Konsequenz wird der FC als sehr studierendenzentriertes Lehrformat beschrieben, das insbesondere aktives Lernen befördern soll (z.B. in Aydin et al., 2021; Chen et al., 2018; Farmus et al., 2020; Hu et al., 2018; Talan & Batdi, 2020). Zumeist wird dann die Selbstlernphase durch technologiegestütztes Online-Lernen mit neuen Medien implementiert (u.a. Chen et al., 2018; Hew et al., 2021; Karagöl & Esen, 2018). Als Lernziele dieses Formats hervorgehoben werden ein tieferes Verständnis des Lernstoffs sowie die Förderung kritischen Denkens und interpersoneller Kompetenzen (Bredow et al., 2021; Gillette et al., 2018; Jang & Kim, 2020).

Die Unterteilung der Lehrformate „Präsenz“, „Online“ und „Blended“ verdeutlicht, dass auf dieser Ebene zunächst die Darbietungsform der Lehre im Vordergrund steht und so empirisch zunächst die Wirkung des jeweiligen „Lernraums“ in Bezug auf den Wissens- und Kompetenzerwerb von Studierenden herausgestellt werden kann. Der didaktisch orientierte Fokus bei der Beschreibung des FC wirft nun die Frage nach dem Einfluss weiterer Faktoren auf, welche die komplexen Dynamiken digitaler Lernumgebungen einfangen und erfolgreiches Lernen beeinflussen (Martin et al., 2020). Neben Merkmalen der Lernenden und der Spezifika der jeweiligen Fachdisziplin müssen insbesondere die Lernaktivitäten von Studierenden berücksichtigt werden (Knezek & Christensen, 2016; Stegmann et al., 2016). Einen umfassenderen Überblick über mögliche einflussreiche Faktoren geben Sailer et al. (2021) in ihrem Cb-Modell. Sie betonen die Rolle der Lehrenden sowie konzeptueller Faktoren, beispielsweise wie Lehrende digitale Technologien nutzen oder welche Lernaktivitäten für Lernende bereitgestellt werden. So sind insbesondere Aspekte der didaktischen Gestaltung für

die Effektivität hochschulischer Lehre relevant. Diese Gestaltungsmöglichkeiten lassen sich überwiegend zwei Kategorien des Cb-Modells zuordnen: (1) Nutzung digitaler Technologien durch Lehrende und (2) Lernaktivitäten mit digitalen Technologien, wobei diese in direktem Zusammenhang stehen (Sailer et al., 2021). In der ersten Kategorie geht es primär um das „Wie oft?“ der Nutzung digitaler Technologien, wobei das „Wie?“ stets mitgedacht wird. In dieser Kategorie steht im Fokus, die Lernenden zu stimulieren und ihnen Lerngelegenheiten zu ermöglichen. In der zweiten Kategorie steht die Ermöglichung von Lernen im Vordergrund (ebd.). Entlang des ICAP-Modells von Chi und Wylie (2014) können digitale Technologien dabei entweder interaktive, konstruktive, aktive oder passive Lernaktivitäten anstoßen, welche mit einem unterschiedlich stark ausgeprägten kognitiven Engagement der Studierenden zusammenhängen. Abgeleitet aus dem dargestellten Forschungsstand werden folgende Fragestellungen adressiert:

- (1) Inwieweit werden als Indikatoren für effektive Lehre kognitive Lernergebnisse, Fähigkeiten und Zufriedenheitsmaße berücksichtigt?
- (2) Inwieweit haben die verschiedenen digitalen Lehrformate einen Effekt im Vergleich zu Präsenzlehre?
- (3) Ist ein Lehrformat den anderen hinsichtlich der untersuchten Effektivitätskriterien statistisch signifikant überlegen?
- (4) Welche Moderatorvariablen erhöhen die Effektivität der Lehre?

3 Methode

Die Darstellung der Methode der systematischen Literaturreview erfolgt gemäß den Anforderungen des Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statements.

3.1 Auswahlkriterien

Eingeschlossen wurden publizierte und begutachtete Metaanalysen (keine systematischen Reviews), die Präsenzlehre mit Onlinelehre, Blended Learning oder Flipped Classroom vergleichen und dabei subjektiven oder objektiven Lernerfolg als abhängige Variable inkludieren. Hierbei sollten mindestens ein Teil der zusammengefassten Studien aus dem Hochschulkontext stammen. Hochschulkontexte sind international weiter gefasst als im deutschsprachigen Raum, vor allem weil die Ausbildung von Gesundheitsberufen international meist vollakademisiert ist.

Ausgeschlossen wurden Arbeiten, die nicht in englischer Sprache veröffentlicht wurden. Auch Arbeiten, die vor dem Jahr 2011 publiziert wurden, wurden ausgeschlossen, da diese mindestens in der einbezogenen umfassenden Metaanalyse von Means et al. (2013) enthalten sein müssten und so Dopplungen vermieden werden.

3.2 Informationsquellen und Suche

Im Februar 2022 wurden systematische Recherchen in den Datenbanken FIS Bildung, ERIC, APA PsycInfo und Education Source durchgeführt. Hier wurden die folgenden Begriffe mit dem Operator „OR“ verknüpft gesucht: digital learning; digital teaching; online learning; online teaching; hybrid learning; hyflex; inverted classroom; flipped classroom; blended learning; e-learning; flipped classrooms; flipped learning; digital classroom; online classroom; hybrid class; blended teaching; digital education; distance education; intelligent tutoring; technology-enhanced AND education.

Die Auswahlkriterien wurden erst in einem weiteren Schritt angewendet, da weitere Einschränkungen in der Suche (wie meta*) zu viele relevante Titel ausgeschlossen hätten. Außerdem wurden Übersichtsarbeiten genutzt, um weitere relevante Referenzen zu extrahieren. Im Anschluss an solche neu gefundenen Titel wurden die Stichwörter so angepasst, dass diese Referenzen mit der Suchstrategie auch auffindbar waren. So kommt es zu den spezifischen Begriffen in der Suche, und es wurde die Wahrscheinlichkeit minimiert, relevante Metaanalysen zu übersehen. Es konnten alle Studien beschafft werden.

3.3 Prozess der Datengewinnung

Die in die Analyse inkludierten Artikel wurden von den Autorinnen kodiert. Hierbei wurden insbesondere die abhängigen Variablen der Studien zu den Kategorien „kognitive Lernergebnisse“, „Skills“ und „Studierendeneinschätzung“ zugeordnet und die Effektstärken der Unterschiede zur Präsenzlehre nach Cohen (1988) in kein Effekt, kleiner Effekt ($d > .2$), mittlerer Effekt ($d > .5$) und großer Effekt ($d > .8$) eingeteilt. Außerdem wurden Besonderheiten und Einschränkungen der Interpretierbarkeit festgehalten.

Die in den Metaanalysen untersuchten Moderatoren wurden kodiert und in einem ersten Schritt in die vier Kategorien Studierendendemographie, Disziplin, Instruktion/Kontext/Praktik und Methodik eingeteilt. Für alle vier Kategorien gab es Beispiele, auf deren Grundlage kodiert wurde. Die für unsere Fragestellungen relevanten Moderatoren der dritten Kategorie wurden in einem nächsten Schritt inhaltlich nach dem Cb-Model (Sailer et al., 2021) gruppiert. Hier wurden die Effektstärken aller Kontraste extrahiert und signifikante Unterschiede zwischen den Kontrasten markiert. Bei zwei Arbeiten lagen die notwendigen Informationen nicht auf der benötigten Aggregationsebene vor (Özdemir & Senturk, 2021; van Alten et al., 2019), weshalb sie im Weiteren nicht berücksichtigt wurden. Die Entwicklung der Kodierschemata sowie die Zuordnung zu den Kategorien wurde innerhalb der Autorinnengruppe abgestimmt und vorgenommen. Uneindeutigkeiten wurden in der Autorinnengruppe diskutiert und gemeinsam aufgelöst.

4 Ergebnisse

4.1 Ergebnisse der Recherche

Die Datenbankrecherche führte zu 1,173 Treffern. Die Ergebnisse wurden in einem ersten Schritt von zwei studentischen Hilfskräften danach durchgesehen, ob es sich um Metaanaly-

sen handelte. Ferner wurden die verbleibenden Treffer dahingehend sortiert, ob die Lehr-Lern-Formate Onlinelehre, BL oder Flipped Classroom im Zentrum standen. Die Sortierung wurde in der weiteren Auswertung jeweils von einer zweiten Person bestätigt. Insgesamt wurden 1,108 Titel ausgeschlossen. Im zweiten Schritt wurden Artikel ausgeschlossen, bei denen kein Vergleich der Formate mit Präsenzlehre stattfand oder die nicht im Hochschulkontext angesiedelt waren. Weiterhin wurden nur die Titel eingeschlossen, bei denen als abhängige Variable der Lernerfolg einbezogen wurde. Insgesamt wurden so 44 Artikel in die Analyse inkludiert (s. Abbildung 1).

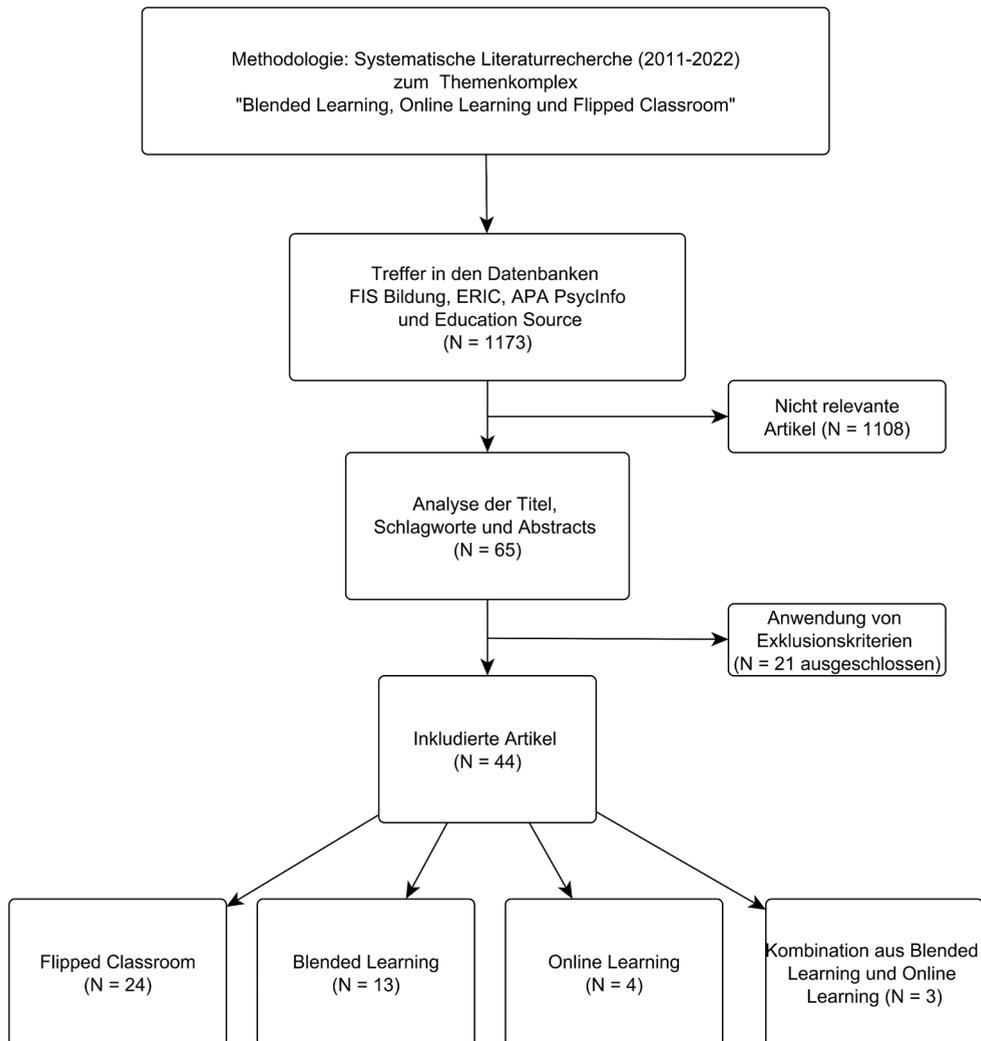


Abbildung 1: Flussdiagramm zur Darstellung der Ergebnisse der Recherche

4.2 Inwieweit werden als Indikatoren für effektive Lehre kognitive Lernergebnisse, Fähigkeiten und Zufriedenheitsmaße berücksichtigt?

In der Gesamtschau der $n = 44$ Metaanalysen konzentrierte sich die Analyse der Auswirkungen von Onlinelehre (OL), Blended-Learning (BL) beziehungsweise der Kombination von Onlinelehre und Blended Learning im Vergleich zum Präsenzunterricht auf den Lernerfolg der Studierenden: (1) Objektive Leistungsmaße kognitiver Lernergebnisse (z.B. Klausurergebnisse, Kursnoten, Punkte in Leistungstests, Bestehen/Durchfallen), (2) Praktische Tests/Skills (z.B. OSCE) und (3) Studierendeneinschätzung (z.B. Lernerfolg, Zufriedenheit, Evaluationsergebnisse). Wie die folgende Abbildung 2 verdeutlicht, zeigen sich in den Studienergebnissen Unterschiede: Während in nahezu allen betrachteten Studien Ergebnisse zu kognitiven Lernergebnissen berichtet wurden, waren Ergebnisse zu Skills in deutlich weniger Studien enthalten, je nach Bereich in 26% bis 43% der Studien. Ergebnisse zu Studierendeneinschätzungen liegen eher selten vor, je nach Bereich in 13% bis 26% der Studien.

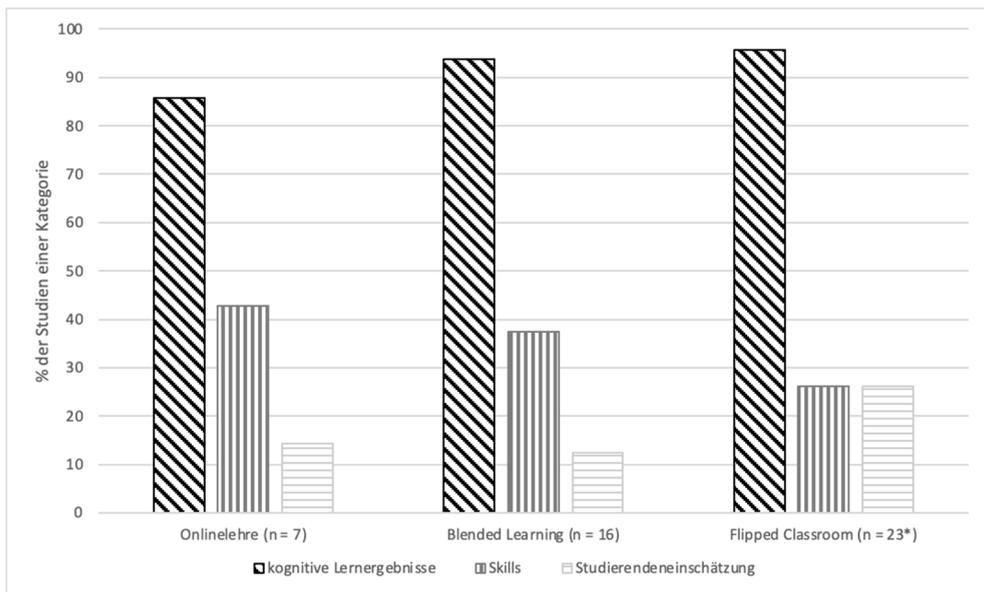


Abbildung 2: Abhängige Variablen in den Metaanalysen (* Metaanalyse zweiter Ordnung von Hew et al. (2021) hier nicht eingeschlossen, vgl. dazu Abschnitt 4.3)

Die Werte spiegeln wider, wie häufig die abhängigen Variablen in den jeweiligen Metaanalysen untersucht wurden. Auffällig ist, dass in einigen Metaanalysen die abhängigen Variablen (AVs) aus mehreren Kategorien untersucht wurden. Sobald aus einer Kategorie mehr als ein Effekt berichtet wurde, wurde die Kategorie einmal pro Studie gezählt; in diesen Fällen wurden die Studien damit mehrfach zugeordnet. In den folgenden Abschnitten werden die Ergebnisse für die jeweiligen Lehrformate genauer betrachtet.

4.2.1 Onlinelehre

Die Verteilung der in den $n = 7$ recherchierten Metaanalysen erfassten abhängigen Variablen kann Tabelle 1 entnommen werden. Kognitive Lernergebnisse wurden in den meisten ($n = 6$) Studien erhoben, gefolgt von Skills ($n = 2$) und Studierendeneinschätzungen ($n = 1$).

Tabelle 1: Abhängige Variablen und Effektstärken in den Metaanalysen ($n = 7$) zur Onlinelehre

Effektstärke*	Anzahl der Effektstärken			Besonderheiten
	kognitive Lernergebnisse	Skills	Studierendeneinschätzung	
kein Effekt	4	1	–	–
klein ($d > .2$)	2	1	–	Bei einer Studie OL = OL + BL
mittel ($d > .5$)	–	–	1	–
groß ($d > .8$)	–	1	–	Health Science, OL = OL + BL, keine Trennung von <i>knowledge</i> und <i>skills</i>

Anmerkung: * nach Cohen (1988). In einer Studie wurden z.T. mehrere abhängige Variablen gleichzeitig betrachtet; OL = Onlinelehre, BL = Blended Learning (BL).

4.2.2 Blended Learning

Die Ergebnisse der für BL recherchierten $n = 16$ Metaanalysen können Tabelle 2 entnommen werden. Die abhängige Variable Kognitive Lernergebnisse wurde in allen Studien erhoben, gefolgt von den Variablen Skills ($n = 6$) und Studierendeneinschätzungen ($n = 2$).

Tabelle 2: Abhängige Variablen und Effektstärken in den Metaanalysen ($n = 16$) zum Blended Learning

Effektstärke*	Anzahl der Effektstärken			Besonderheiten
	kognitive Lernergebnisse	Skills	Studierendeneinschätzung	
kein Effekt	2	–	–	–
klein ($d > .2$)	8	3	1	–
mittel ($d > .5$)	2	2	1	Vier Effekte aus dem Bereich Health Science
groß ($d > .8$)	4	1	–	–

Anmerkung: * nach Cohen (1988). In einer Studie wurden z.T. mehrere abhängige Variablen gleichzeitig betrachtet.

4.2.3 Flipped Classroom

Die Verteilung der in den $n = 23$ recherchierten Metaanalysen erfassten abhängigen Variablen kann Tabelle 3 entnommen werden. Kognitive Lernergebnisse wurden in den meisten ($n = 23$) Studien erhoben, gefolgt von Skills (z.B. OSCE, $n = 6$) und Studierendeneinschätzungen ($n = 6$).

Tabelle 3: Abhängige Variablen und Effektstärken in den Metaanalysen ($n = 23$) zum Flipped Classroom

Effektstärke*	Anzahl der Effektstärken			Besonderheiten
	kognitive Lernergebnisse	Skills	Studierenden- einschätzung	
kein Effekt	2	–	1	–
klein ($d > .2$)	10	1	2	–
mittel ($d > .5$)	6	–	2	–
groß ($d > .8$)	4	5	1	Schwerpunkt nursing education in China

Anmerkung: * nach Cohen (1988). In einer Studie wurden z.T. mehrere abhängige Variablen gleichzeitig betrachtet.

4.3 Inwieweit haben die verschiedenen digitalen Lehrformate einen Effekt im Vergleich zu Präsenzlehre?

4.3.1 Onlinelehre

Wie Tabelle 1 entnommen werden kann, zeigen die Metaanalysen für Onlinelehre für die Hälfte der Studien keinen Effekt im Vergleich zur Präsenzlehre. Für die andere Hälfte der Studien liegen die Effektstärken überwiegend im kleinen Bereich ($n = 3$), wobei sich der Effekt häufiger im Bereich der kognitiven Lernergebnisse zeigt. Lediglich eine Studie berichtet eine große Effektstärke; hierbei handelt es sich allerdings um die „pooled mean difference“, die auf einer Skala von 0–100 (Voutilainen et al., 2017) abgebildet wird. Problematisch hinzu kommt bei dieser Studie, dass sie weder zwischen Onlinelehre und BL noch zwischen kognitiven Lernergebnissen und Skills unterscheidet. Aufgrund der Definition der Autorinnen wurde diese ausschließlich für die Onlinelehre aufgenommen.

4.3.2 Blended Learning

Tabelle 2 zeigt die Befunde der Analysen zu BL. Die Auswertung der Metaanalysen verdeutlicht die sehr heterogenen Definitionen und Ansätze dieses Lehr-Lern-Formats. Insgesamt zeigt sich über die betrachteten Studien in den Effektstärken etwas Varianz, überwiegend liegen die Werte im kleinen Bereich. Vier Studien berichten hohe Effektstärken im Bereich der kognitiven Lernergebnisse und Skills (z.B. Liu et al. 2016 für Lernende in den Gesundheitsberufen). Hingegen zeigen zwei Studien keinen Effekt für das Blended-Learning-Format (z.B. Müller & Mildnerberger, 2021).

4.3.3 Flipped Classroom

Wie Tabelle 3 entnommen werden kann, liegen die Effektstärken für die Metaanalysen zum FC überwiegend im kleinen Bereich (Cohen, 1988). Allerdings zeigt sich in den Ergebnissen auch etwas Varianz zwischen den Studien. Insbesondere stechen $n = 6$ Studien mit hohen Effektstärken heraus. Bei genauerer Betrachtung lässt sich feststellen, dass diese Metaanalysen überwiegend aus dem chinesischen Sprachraum stammen: Vier dieser sechs Metaanaly-

sen (Hu et al., 2018; Li et al., 2020; Tan et al., 2017; Xu et al., 2019) schließen ausschließlich chinesische Originalstudien ein, in denen zudem das untersuchte Studienfach auf „nursing education“ eingeschränkt ist. Die zwei weiteren Studien mit auffallend hohen Effektstärken stammen von Özdemir und Senturk (2021) mit dem Untersuchungsfeld Naturwissenschaft und Mathematik und Chen et al. (2018) mit Studien aus dem Gesundheitsbereich.

4.4 Weitere Ergebnisse zum FC: Metaanalyse höherer Ordnung

In einer kürzlich erschienenen Metaanalyse zweiter Ordnung von Hew et al. (2021) wurden $n = 15$ Metaanalysen mit between-group research design eingeschlossen, die im Publikationszeitraum von 2017 bis 2020 erschienen sind. Da in dieser Metaanalyse zweiter Ordnung die Ergebnisse erneut aggregiert und gewichtet wurden, werden ihre Resultate separat von den anderen Treffern unserer Recherche berichtet. Die in diese Metaanalyse inkludierten Studien umfassen alle Fächer, verschiedene Ausbildungsstufen (inklusive Hochschulbildung) und wurden in drei verschiedenen Sprachen (englisch, chinesisch und türkisch) publiziert. Das Analysekorpus deckt sich weitestgehend mit den in diese Veröffentlichung eingeschlossenen Studien zum FC, mit wenigen Einschränkungen: Zwei der von Hew et al. (2021) eingeschlossenen Metaanalysen wurden mit der in diesem Artikel beschriebenen Suchstrategie nicht gefunden, wobei es sich bei einem Treffer um eine (vermutlich unveröffentlichte) Abschlussarbeit handelt. Andersherum umfasst die Meta-Metaanalyse von Hew et al. (2021) aufgrund anderer Einschlusskriterien (s.u.) nicht alle $n = 23$ Treffer der Recherche zum vorliegenden Beitrag.

Die abhängige Variable in der Metaanalyse von Hew et al. (2021) waren studentische Lernergebnisse, die über Indikatoren für den Wissenserwerb der Studierenden („teacher-developed or standardised tests and exams“, S. 134) erfasst wurden. Die Effektstärke liegt mit .45 im kleinen Bereich und bleibt auch in diesem, wenn die Ergebnisse um den Publication Bias bereinigt werden ($ES = .37$).

4.5 Welche Moderatorvariablen erhöhen die Effektivität der Lehre?

Eine weitere Fragestellung für die Analysen bezog sich auf die in den Metaanalysen betrachteten Moderatoren. Insgesamt haben $n = 27$ der $n = 44$ Studien Moderatoren einbezogen. Die folgende Tabelle 4 zeigt die Kategorien für die Moderatoren und Beispiele.

Insgesamt 158 Moderatoren wurden in den Studien untersucht. Die meisten untersuchten Moderatoren bezogen sich auf methodische Aspekte der Studien wie das Publikationsjahr oder auch Aspekte des Forschungsdesigns. Beeinflussbar waren insgesamt 52 Moderatoren, die weiter analysiert wurden. Tabelle A.1 im Anhang zeigt die Ergebnisse dieser Analysen, nach denen die Moderatoren entsprechend den beiden Kategorien des Cb-Modells (Sailer et al., 2021) „Higher Education Teachers‘ Digital Technology Use“ und „Students‘ Learning Activities Involving Digital Technology“ zugeordnet wurden. Weitere gefundene Unterkategorien waren „Didaktische Elemente mithilfe digitaler Technologie“, „Zeitliche Aufteilung der digitalen Lerngelegenheiten“, „Gesamtzeitraum der Intervention“, „Funktionen der eingesetzten Technologien“, „Interaktionen und Kommunikation“, „Lernangebote nach dem Unterricht“, „Instruktionsansatz“, „Feedback und Assessment“, „Zugänglichkeit“.

Tabelle 4: Kategorien für analysierte Moderatorvariablen und Beispiele

Kategorie (Beispiele)	Häufigkeit Moderatoren	Metaanalysen (Anzahl der Moderatoren in der Kategorie)
<i>Demografische Angaben zu den Studierenden</i> (z.B. Bildungsabschluss, Kurslevel)	17	Aydin et al. (2021) (1). Bernhard et al. (2014) (1). Borokhovski et al. (2016) (1). Bredow et al. (2021) (1). Cheng et al. (2019) (1). He et al. (2020) (1). Karagöl & Esen (2019) (1). Kazu & Yalcin (2022) (1). Lo & Hew (2019) (1). Låg & Sæle (2019) (1). Means et al. (2013) (1). Strelan et al. (2020) (1). Talan & Batdi (2020) (1). Tamim et al. (2011) (1). Van Alten et al. (2019) (1). Zheng et al. (2020) (1). Özdemir & Senturk (2021) (1).
<i>Inhaltsbereich/Disziplin</i> (z.B. Fachgebiet, MINT)	23	Bernhard et al. (2014) (1). Borokhovski et al. (2012) (1). Borokhovski et al. (2016) (1). Bredow et al. (2021) (1). Cheng et al. (2019) (1). Farmus et al. (2020) (1). He et al. (2020) (1). Hew et al. (2021) (1). Hew & Lo (2018) (1). Jang & Kim (2020) (2). Kazu & Yalcin (2022) (1). Låg & Van Alten et al. (2019) (1). Means et al. (2013) (1). Müller & Mildnerberger (2021) (1). Shi et al. (2020) (1). Schmid et al. (2014) (1). Strelan et al. (2020) (1). Låg & Sæle (2019) (1). Talan & Batdi (2020) (1). Tamim et al. (2011) (1). Vo et al. (2017) (1). Zheng et al. (2020) (1). Özdemir & Senturk (2021) (1).
<i>Unterrichtssituation/ Kontext/Praktiken</i> (z.B. Bedingungen im Kursraum, Größe der Klasse, Dauer der Intervention, Hilfsmittel, Kommunikation, Art der Technologie, Einsatz der Technologie, Zweck der Technologie)	52	Bernhard et al. (2014) (3). Borokhovski et al. (2012) (1). Borokhovski et al. (2016) (2). Bredow et al. (2021) (3). Farmus et al. (2020) (2). He et al. (2020) (3). Hew & Lo (2018) (2). Karagöl & Esen (2019) (1). Låg & Sæle (2019) (2). Means et al. (2013) (12). Müller & Mildnerberger (2021) (3). Shi et al. (2020) (2). Schmid et al. (2014) (2). Talan & Batdi (2020) (1). Tamim et al. (2011) (2). Van Alten et al. (2019) (1). Vo et al. (2017) (1). Zheng et al. (2020) (7). Özdemir & Senturk (2021) (1).
<i>Methodik</i> (z.B. Publikationsjahr, Forschungsdesign, Art der Veröffentlichung)	56	Aydin et al. (2021) (2). Bernhard et al. (2014) (1). Borokhovski et al. (2012) (1). Bredow et al. (2021) (4). Cheng et al. (2019) (3). He et al. (2020) (2). Hew et al. (2021) (8). Hew & Lo (2018) (3). Jang & Kim (2020) (3). Karagöl & Esen (2019) (3). Kazu & Yalcin (2022) (2). Låg & Sæle (2019) (4). Means et al. (2013) (7). Müller & Mildnerberger (2021) (2). Shi et al. (2020) (4). Strelan et al. (2020) (2). Tamim et al. (2011) (2). Van Alten et al. (2019) (1). Özdemir & Senturk (2021) (2).

Außerdem ergab sich eine kleine Sonstiges-Kategorie. Dargestellt sind jeweils die kleinste und die größte Effektstärke der verschiedenen Kontraste, außerdem ist angegeben, ob die Unterschiede zwischen den Kontrasten signifikant waren oder nicht. Es zeigt sich, dass der Großteil der untersuchten Variablen keine Moderatoren darstellt, da die Unterschiede zwischen den Kontrasten nicht signifikant sind. Drei Bereiche stechen hierbei heraus: Kursaktivitäten, Funktionen der eingesetzten Technologien und Instruktionsansätze. Alle drei Bereiche zeigen signifikante Unterschiede zwischen den Kontrasten über mehrere Studien hinweg und sind daher ganz besonders zu beachten. Im Bereich Kursaktivitäten zeigen sich beispielsweise Wiederholungen zu Beginn der Stunden in Präsenz beim FC besonders vorteilhaft für die Lernenden (Hedge's $g = 0.409$ im Vergleich zu 0.261 bei Quizzes; Lo & Hew, 2019). Werden Technologien beim Online-Lernen eingesetzt, um die Lernenden auch kognitiv zu unterstützen (0.63), zeigen sich deutlich größere Effekte, als wenn lediglich die Kommunikation (0.32) oder kognitive Aktivitäten (0.14) unterstützt werden (Borokhovski et al., 2016). Ein Beispiel für einen Instruktionsansatz, der die Effektivität von Online-Lernen moderiert, ist der pädagogisch-didaktische; dozierenden-geleitete Ansätze zeigen hier die stärksten Effekte (0.386) im Vergleich zu kooperativen (0.249) und selbstständigen (0.050) Ansätzen (Means et al., 2013).

5 Diskussion und Ausblick

Auf der Grundlage von $N = 44$ Metaanalysen wird in diesem Beitrag die Wirksamkeit unterschiedlicher Lehr-Lern-Formate in Bezug auf den Lernerfolg von Studierenden herausgestellt. Der direkte Vergleich von Präsenz- mit reiner Onlinelehre zeigt keine Überlegenheit von einem Format gegenüber dem anderen. Hingegen sind die gemischten Formate (BL, FC) gegenüber der Präsenzlehre effektiver, obgleich die in den hier einbezogenen Metaanalysen berichteten Effektstärken – bis auf wenige Studien – als klein einzuordnen sind. Dies deutet darauf hin, dass weniger das Lehrformat an sich – und somit die Entscheidung für Präsenz-, Online- oder Blended Learning – ausschlaggebend für die Qualität von Hochschullehre ist. Vielmehr müssen andere Faktoren berücksichtigt werden, die sich vor allem auf die Gestaltung der Formate beziehen. Da insgesamt alle Effektstärken über die unterschiedlich definierten Vergleichsgruppen in den Studien hinweg im kleinen Bereich liegen, sollte ein zukünftiger Analysefokus vor allem auf die Gestaltungsmerkmale von Hochschullehre gelegt werden, d.h. inwieweit die spezifischen Voraussetzungen und Möglichkeiten der Formate auch lernwirksam genutzt werden (z.B. in Bezug auf die (A-)synchronizität der Lernsettings). Hinweise darauf liefern die in den Artikeln identifizierten Moderatorvariablen, wobei die drei Bereiche Kursaktivitäten, Funktionen der eingesetzten Technologien und Instruktionsansätze besonders beachtenswert sind. Insgesamt verdeutlichen die Ergebnisse noch einmal die hohe Relevanz des wohlgedacht didaktisch-begründeten Einsatzes von Technologien mit dem Ziel, das aktive Lernen bei den Studierenden bestmöglich auszulösen. Entsprechend bringt die Auswertung der Moderatoren hervor, dass vor allem die durch Lehrende „veränderbaren“ Variablen (*Teaching Factors*, s. auch Chaudhary & Singh, 2022) einen Einfluss haben. So ist die Art des Lernangebots entscheidend für die angestoßene Lernaktivität, die wiederum auf den Lernerfolg wirkt.

Bei Betrachtung der einbezogenen Studien stellt sich die Frage nach der Vergleichbarkeit der Gegenüberstellungen der Formate in den einzelnen Metaanalysen, d.h. wie mit den unterschiedlichen Definitionen von Online-Learning, BL und Flipped Classroom und den unterschiedlich konzipierten Vergleichsgruppen umgegangen werden kann. So wird in einigen Metaanalysen bereits ab einem Online-Anteil von über 79% von Onlinelehre gesprochen (z.B. Müller & Mildenerger, 2021) oder BL wird als eine Form von Onlinelehre verstanden (z.B. Means et al., 2013). Dabei ist nicht immer klar beschrieben, welche Anteile bei dieser Zuordnung dann in Präsenz umgesetzt werden. Ebenso variiert die Beschreibung der entsprechenden Formate im Vergleich. Hier liegt der Fokus zumeist auf dem Grad der Umsetzung in Online beziehungsweise Präsenz (vgl. Abschnitt 2). Insbesondere bei der Gegenüberstellung zum FC zeigt sich jedoch, dass nicht primär der Kontext (Online/Präsenz) als Kriterium herangezogen wird, sondern vielmehr das didaktische Konzept („traditional lecture“, „lecture based learning“). So zeichnet sich die Qualität dieses Formats durch eine didaktische Verknüpfung von selbst- und fremdregulierten Lernepisoden aus, bei der die Möglichkeiten digitaler Lernräume genutzt werden. Hingegen sind andere Studien weniger eingrenzend und setzen den Vergleich allgemeiner zu „traditional teaching“ oder „non-flipped classrooms“. So wird ein spezifisches Format einer Vielzahl anderer Formate mit hoher didaktischer Varianz gegenübergestellt, und es muss überlegt werden, inwiefern dieser Vergleich aussagekräftig ist beziehungsweise wie eine alternative Klassifizierung der Formate vorgenommen werden sollte. Ähnliches gilt es für Vergleiche zu bedenken, bei denen der Einsatz von

(Lern-)technologien im Vordergrund steht (z.B. Bernard et al., 2014; Borokhovski et al., 2016; Tamim et al., 2011). Auch hier steht weniger das Format als vielmehr die Umsetzung innerhalb eines spezifischen Formats im Vordergrund. Darüber hinaus zeigt die Betrachtung der in Studien erfassten abhängigen Variablen als Indikatoren für Lernerfolg, dass primär kognitive Lernergebnisse (z.B. Klausurergebnisse, Kursnoten, Punkte in Leistungstests) berichtet werden. Hingegen sind praktische Tests zur Erhebung von Fähigkeiten in deutlich weniger Studien enthalten (26% bis 43% der Studien), und subjektive Einschätzungen (z.B. Lernerfolg, Zufriedenheit, Evaluationsergebnisse) noch seltener mit 13% bis 26% der Studien. Entsprechend wird auch in der Metaanalyse zweiter Ordnung von Hew et al. (2021) zum FC ausschließlich auf kognitive Lernergebnisse als abhängige Variable verwiesen, da sonst vermutlich nicht genügend Metaanalysen für die Berechnungen verfügbar gewesen wären. Weitere Forschung sollte auf ein breite(re)s Spektrum von Indikatoren für Lernerfolg fokussieren, um so die Effekte der Formate für mehrere unterschiedliche und stärker differenzierte Lernziele (z.B. nach Lernzieltaxonomiestufen) herauszustellen.

5.1 Limitationen

Die berichteten Ergebnisse zur Effektivität der Lehrformate sind maßgeblich durch die in dieses Review einbezogenen Studien, d.h. die gesetzten Inklusionskriterien, determiniert. Folglich müssen die Spezifika des Analysekorpus bei der Interpretation der Ergebnisse berücksichtigt werden. Insbesondere in Bezug auf die untersuchten Stichproben stellt sich die Frage nach der Generalisierbarkeit der Befunde. So beziehen sich einige der einbezogenen Metaanalysen auf ausschließlich ein Land/eine Nationalität und/oder eine Disziplin. Dies wird insbesondere bei den Studien zum FC deutlich. Die chinesischen Studien (Hu et al., 2018; Li et al., 2020; Tan et al., 2017; Xu et al., 2019) sind auf chinesische Stichproben im Gesundheitswesen eingeschränkt. So kritisieren Låg und Sæle (2019, S. 3) zu Recht: „to date, most meta-analyses of flipped classroom studies are limited to the medical and health professions disciplines“. Dadurch können auch die hier gefundenen vergleichsweise hohen Effektivitätsstärken begründet werden. Demgegenüber zeigen sich in den englischsprachigen Metaanalysen aus dem Gesundheitsbereich (*pharmacy education*; Gillette et al., 2018) gar keine Vorteile des FC gegenüber der klassischen Vorlesung. Allerdings schreiben Gillette et al. (2018, S. 435): „All but two reports indicated that the lecture sections consisted of lectures accompanied by active learning“. Darin mag ein zentraler Unterschied zu den Vorlesungen in den asiatischen Studien liegen, da das Ausmaß an aktiver Auseinandersetzung mit den Wissensinhalten in den Vergleichsgruppen nicht vergleichbar ist. Darüber hinaus handelt es sich bei vielen der inkludierten Artikel um gemischte Stichproben, d.h. es werden sowohl der sekundäre als auch der tertiäre Bildungsbereich einbezogen. Diesbezüglich wurde für diesen Beitrag keine weitere Eingrenzung, z.B. in Bezug auf den Anteil an Studien aus dem Hochschulbereich, vorgenommen. Zukünftige Analysen sollten die Rolle stichprobenspezifischer Merkmale weiterführend beachten, da diese Variablen einen Einfluss auf die Wirksamkeit der Formate zeigen (vgl. Abschnitt 4.4). Des Weiteren sind die Befunde auch in Bezug auf methodische Aspekte zu diskutieren. So beziehen sich nicht alle Metaanalysen ausschließlich auf Studien mit einem strikten randomisierten kontrollierten Studiendesign (Randomised-Controlled-Trial), vielmehr werden in manchen Artikeln auch Beobachtungsstudien integriert (z.B. Gillette et al., 2018; hier war nur eine von fünf Studien experimentell ausgerich-

tet). Auch muss bedacht werden, dass ein Teil der Einzelstudien mehrfach in die Metaanalysen eingeschlossen wurde und somit die Aggregation der berichteten Effekte schwer abschätzbare Verzerrungen enthalten könnte. Entsprechend lohnt sich die Hinzunahme von Metaanalysen zweiter Ordnung, die diese potentielle Dopplungen statistisch relativieren und so eindeutige Aussagen in Bezug auf die Wirksamkeit erlauben.

5.2 Implikationen und Ausblick

Zusammenfassend lässt sich festhalten, dass sich keine eindeutige Empfehlung zur bevorzugten Nutzung eines der in diesem Beitrag betrachteten Lehrformate ableiten lässt. Dies wirft die generelle Frage auf, wie hoch die Effektstärken sein müssten, um eine solche Empfehlung zur präferierten Nutzung eines der betrachteten Formate zu rechtfertigen. Hier können umfassende Metaanalysen höherer Ordnung für den Bildungsbereich eine gute Orientierung liefern: Z.B. berichtet Hattie (2009, 2018) für den Schulkontext Effektstärken von bis zu 1.57, und für den Hochschulkontext finden Schneider und Preckel (2017) sogar Effektstärken von bis zu 1.91. Diese beziehen sich auf unterschiedliche Variablen, wobei neben lernenden-bezogenen Variablen vor allem die didaktische Komponente und das Verhalten der Lehrenden einen Einfluss haben. Eine entsprechende Übersicht zu den „TOP 5 der wirksamsten Aspekte guter Lehre“ findet sich zudem bei Ulrich (2020, S. 37). Es ist weiterführend zu klären, ob eine lernwirksame Einflussnahme auch (und gegebenenfalls effizienter) über andere Faktoren als das Lehrformat erreicht werden kann. Dies sollte insbesondere bei Formaten mit einem eher hohen Aufwand bei der Erst-Konzeption (wie FC) bedacht werden, da hier eine Kosten-Nutzen-Abwägung besonders relevant ist. Aufbauend auf aktuellen Arbeiten, die diese Fragestellung bereits aufgreifen (Sailer et al., 2021), gilt es, den didaktisch begründeten Einsatz von Technologien in Vergleichsstudien systematisch zu testen. Darüber hinaus erscheinen vertiefende Metaanalysen und Literaturreviews zur Wirksamkeit spezifischer Technologien vielversprechend (z.B. Buchner et al., 2022). Darüber hinaus fließen auch weitere Aspekte in die Entscheidung zur Umsetzung von Hochschullehre online und/oder in Präsenz ein. Aus hochschulpolitischer Sicht könnte beispielsweise auf eine Erhöhung der Studierendenzahlen abgezielt werden, dann wäre die Empfehlung wahrscheinlich hin zu mehr Online-Angeboten. Gleiches gilt, wenn eine Verringerung der Studienabbruchquoten angestrebt wird, weil durch (teil-)digitale Angebote der Heterogenität der Studierendenschaft in der Studieneingangsphase begegnet und so der Lernerfolg in den Grundlagenfächern erhöht werden kann. Wird von den Lernzielen aus gedacht, dann sollten digitale Lernsettings für den Aufbau von Schlüsselqualifikationen für den digitalen Arbeitsmarkt präferiert werden, hingegen bieten Präsenzsettings verstärkt die Möglichkeit zur regelmäßigen (auch informellen) Begegnungen und zum sozialen Austausch und dem Einüben praktischer Fertigkeiten. Insgesamt zeichnet sich gute, innovative Hochschullehre jedoch vor allem durch die Berücksichtigung vielfältiger Faktoren und Ansätze aus – dieser Beitrag liefert dazu eine evidenzbasierte Grundlage und unterstützt so eine systematische und nachhaltige Vorgehensweise.

Literatur

- Adedoyin, O. B. & Soykan, E. (2020). Covid-19 pandemic and online learning: The challenges and opportunities. *Interactive Learning Environments*, 31(2), 863–875. <https://doi.org/10.1080/10494820.2020.1813180>
- Allen, M., Mabry, E., Mattrey, M., Bourhis, J., Titsworth, S. & Burrell, N. (2004). Evaluating the effectiveness of distance learning: A comparison using meta-analysis. *Journal of Communication*, 54(3), 402–420. <https://doi.org/10.1111/j.1460-2466.2004.tb02636.x>
- Amaka, I. H. & Goeman, K. (2017). Selecting media for effective learning in online and blended courses: A review study. *Journal of Educational Multimedia and Hypermedia*, 26(1), 29–59.
- Anthony, B., Kamaludin, A., Romli, A., Raffei, A. F. M., Phon, D. N. A. L. E., Abdullah, A. & Ming, G. L. (2022). Blended learning adoption and implementation in higher education: A theoretical and systematic review. *Technology, Knowledge and Learning*, 27(2), 531–578. <https://doi.org/10.1007/s10758-020-09477-z>
- Aydin, M., Okmen, B., Sahin, S. & Kilic, A. (2021). The meta-analysis of the studies about the effects of flipped learning on students' achievement. *Turkish Online Journal of Distance Education*, 22(1), 33–51. <https://doi.org/10.17718/tojde.849878>
- Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M. & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, 26(1), 87–122. <https://doi.org/10.1007/s12528-013-9077-3>
- Borokhovski, E., Bernard, R. M., Tamim, R. M., Schmid, R. F. & Sokolovskaya, A. (2016). Technology-supported student interaction in post-secondary education: A meta-analysis of designed versus contextual treatments. *Computers & Education*, 96, 15–28. <https://doi.org/10.1016/j.compedu.2015.11.004>
- Bredow, C. A., Roehling, P. V., Knorp, A. J. & Sweet, A. M. (2021). To flip or not to flip? A meta-analysis of the efficacy of flipped learning in higher education. *Review of Educational Research*, 91(6), 878–918. <https://doi.org/10.3102/00346543211019122>
- Buchner, J., Buntins, K. & Kerres, M. (2022). The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning*, 38(1), 285–303.
- Chaudhary, P. & Singh, R. K. (2022). A meta analysis of factors affecting teaching and student learning in higher education [Systematic Review]. *Frontiers in Education*, 6:824504 <https://doi.org/10.3389/educ.2021.824504>
- Chen, J., Wang, M., Kirschner, P. A. & Tsai, C.-C. (2018). The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *Review of Educational Research*, 88(6), 799–843. <https://doi.org/10.3102/0034654318791584>
- Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T. & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research*, 90(4), 499–541. <https://doi.org/10.3102/0034654320933544>
- Chi, M. T. H. & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.
- Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J. & Hamstra, S. J. (2011). Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. *Journal of the American Medical Association*, 306(9), 978–988. <https://doi.org/10.1001/jama.2011.1234>
- Dziuban, C., Graham, C. R., Moskal, P. D., Norberg, A. & Sicilia, N. (2018). Blended learning: The new normal and emerging technologies. *International Journal of Educational Technology in Higher Education*, 15(1), 3. <https://doi.org/10.1186/s41239-017-0087-5>

- Farmus, L., Cribbie, R. A. & Rotondi, M. A. (2020). The flipped classroom in introductory statistics: Early evidence from a systematic review and meta-analysis. *Journal of Statistics Education*, 28(3), 316–325. <https://doi.org/10.1080/10691898.2020.1834475>
- Galvis, Á. H. (2018). Supporting decision-making processes on blended learning in higher education: Literature and good practices review. *International Journal of Educational Technology in Higher Education*, 15(1), 25. <https://doi.org/10.1186/s41239-018-0106-1>
- Garrison, D. R. & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, 7(2), 95–105. <https://doi.org/10.1016/j.iheduc.2004.02.001>
- Gillette, C., Rudolph, M., Kimble, C., Rockich-Winston, N., Smith, L. & Broedel-Zaugg, K. (2018). A meta-analysis of outcomes comparing flipped classroom and lecture. *American Journal of Pharmaceutical Education*, 82(5), 6898. <https://doi.org/10.5688/ajpe6898>
- Graham, C. R. (2005). Blended learning systems: Definition, current trends, and future directions. In C. J. Bonk & C. R. Graham (Eds.), *Handbook of blended learning: Global perspectives, local designs* (pp. 3–21). Pfeiffer.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement* (Reprinted). Routledge.
- Hattie, J. & Clarke, S. (2018). *Visible learning: feedback*. Routledge.
- Heinemann, E. & Funke, J. (2020). Digitalisierung der Hochschullehre: Schluss mit dem betreuten Vorlesen! *Die Zeit*, 32/2020 (30. Juli 2020). <https://www.zeit.de/2020/32/digitalisierung-hochschul-lehre-lehrmethoden-vorlesung-technologie-reform>
- Hew, K. F., Bai, S., Huang, W., Dawson, P., Du, J., Huang, G., Jia, C. & Thankrit, K. (2021). On the use of flipped classroom across various disciplines: Insights from a second-order meta-analysis. *Australasian Journal of Educational Technology*, 37(2), 132–151. <https://doi.org/10.14742/ajet.6475>
- Hodges, C., Moore, S., Lockee, B., Trust, T. & Bond, M. (2020). *The difference between emergency remote teaching and online learning*. <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning>
- Horstmann, K. T., Buecker, S., Krasko, J., Kritzler, S. & Terwiel, S. (2021). Who does or does not use the ‘Corona-Warn-App’ and why? *European Journal of Public Health*, 31(1), 49–51. <https://doi.org/10.1093/eurpub/ckaa239>
- HRK – Hochschulrektorenkonferenz. (2021). *Neue Möglichkeiten schaffen und nutzen: Empfehlungen zur wissenschaftlichen Weiterbildung*. Empfehlung der 32. Mitgliederversammlung der HRK vom 16. November 2021. Stuttgart: HRK.
- Hu, R., Gao, H., Ye, Y., Ni, Z., Jiang, N. & Jiang, X. (2018). Effectiveness of flipped classrooms in Chinese baccalaureate nursing education: A meta-analysis of randomized controlled trials. *International Journal of Nursing Studies*, 79, 94–103. <https://doi.org/10.1016/j.ijnurstu.2017.11.012>
- Jang, H. & Kim, H. J. (2020). A meta-analysis of the cognitive, affective, and interpersonal outcomes of flipped classrooms in higher education. *Education Sciences*, 10, 115. <https://doi.org/10.3390/educsci10040115>
- Karagöl, İ. & Esen, E. (2018). The effect of flipped learning approach on academic achievement: A meta-analysis study. *Hacettepe University Journal of Education*, 1–20. <https://earsiv.ana.dolu.edu.tr/xmlui/handle/11421/14773>
- Kaur, M. (2013). Blended learning – Its challenges and future. *3rd World Conference on Learning, Teaching and Educational Leadership*, 93, 612–617. <https://doi.org/10.1016/j.sbspro.2013.09.248>
- Knaus, T., Merz, O. & Junge, T. (2022). Editorial: Lehre in Zeiten von Corona. Und wir lernen weiter... *Ludwigsburger Beiträge zur Medienpädagogik*, 22, 1–16. <https://doi.org/10.21240/lbzm/22/01>
- Knezek, G. & Christensen, R. (2016). Extending the will, skill, tool model of technology integration: Adding pedagogy as a new model construct. *Journal of Computing in Higher Education*, 28(3), 307–325. <https://doi.org/10.1007/S12528-016-9120-2>

- Låg, T. & Sæle, R. (2019). Does the flipped classroom improve student learning and satisfaction? A systematic review and meta-analysis. *AERA Open*, 5, 233285841987048. <https://doi.org/10.1177/2332858419870489>
- Li, B.-Z., Cao, N.-W., Ren, C.-X., Chu, X.-J., Zhou, H.-Y. & Guo, B. (2020). Flipped classroom improves nursing students' theoretical learning in China: A meta-analysis. *PLOS ONE*, 15(8), e0237926. <https://doi.org/10.1371/journal.pone.0237926>
- Liu, Q., Peng, W., Zhang, F., Hu, R., Li, Y. & Yan, W. (2016). The effectiveness of blended learning in health professions: Systematic review and meta-analysis. *Journal of Medical Internet Research*, 18(1), e2. <https://doi.org/10.2196/jmir.4807>
- Lo, C. K. & Hew, K. F. (2019). The impact of flipped classrooms on student achievement in engineering education: A meta-analysis of 10 years of research. *Journal of Engineering Education*, 108(4), 523–546. <https://doi.org/10.1002/jee.20293>
- Lörz, M., Marczuk, A., Zimmer, L., Multrus, F. & Buchholz, S. (2020). Studieren unter Corona-Bedingungen: Studierende bewerten das erste Digitalsemester. *DZHW Brief*. https://doi.org/10.34878/2020.05.DZHW_BRIEF
- Martin, F., Sun, T. & Westine, C. D. (2020). A systematic review of research on online teaching and learning from 2009 to 2018. *Computers & Education*, 159, 104009. <https://doi.org/10.1016/j.compedu.2020.104009>
- Means, B., Toyama, Y., Murphy, R. & Bakia, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, 115. <https://doi.org/10.1177/016146811311500307>
- Means, B., Toyama, Y., Murphy, R., Bakia, M. & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. US Department of Education. <https://eric.ed.gov/?id=ED505824>
- Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W. & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, 70, 29–40. <https://doi.org/10.1016/j.compedu.2013.07.033>
- Müller, C. & Mildenerger, T. (2021). Facilitating flexible learning by replacing classroom time with an online learning environment: A systematic review of blended learning in higher education. *Educational Research Review*, 34, 100394. <https://doi.org/10.1016/j.edurev.2021.100394>
- Özdemir, A. & Senturk, M. L. (2021). The effect of flipped classroom model on students' academic achievement in science and mathematics education: A meta-analysis study. *Journal of Educational Technology*, 18(3), 22–41.
- Reichert, S. (2019). Chancen und Grenzen der Hochschulevaluation. *fteval Journal for Research an Technology Policy Evaluation*, 49, 42–47.
- Sailer, M., Schultz-Pernice, F. & Fischer, F. (2021). Contextual facilitators for learning activities involving technology in higher education: The Cb-model. *Computers in Human Behavior*, 121. <https://doi.org/https://doi.org/10.1016/j.chb.2021.106794>
- Schley, S. & Buchwald, L. (2022). Wie zufrieden sind Studierende mit der digitalen Lehre? : Eine empirische Studie zum Einfluss von Motivation und Studierverhalten. *Ludwigsburger Beiträge zur Medienpädagogik*, 22, 1–13. <https://doi.org/10.21240/lbzm/22/10>
- Schmidberger, L., Unger, V. & Wacker, A. (2022). Durch das Internet ‚gequetschte‘ Seminare!? Befunde einer explorativen Befragung von Studierenden im ersten Online-Semester. *Ludwigsburger Beiträge zur Medienpädagogik*, 22, 1–15. <https://doi.org/10.21240/lbzm/22/06>
- Schneider, M. & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, 143(6), 565–600. <https://doi.org/10.1037/bul0000098>
- Stegmann, K., Wecker, C., Mandl, H. & Fischer, F. (2016): Lehren und Lernen mit digitalen Medien. Ansätze und Befunde der empirischen Bildungsforschung. In R. Tippelt & B. Schmidt-Hertha (Hrsg.), *Handbuch Bildungsforschung* (S. 1–22). Springer VS.

- Steinmayr, R., Meißner, A., Weidinger, A. F. & Wirthwein, L. (2014). *Academic achievement*. Oxford Bibliographies. <http://www.oxfordbibliographies.com/view/document/obo-9780199756810/obo-9780199756810-0108.xml>
- Talan, T. & Batdi, V. (2020). Evaluating the flipped classroom model through the multi-complementary approach. *Turkish Online Journal of Distance Education*, 21, 31–67. <https://doi.org/10.17718/tojde.803351>
- Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C. & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research*, 81(1), 4–28. <https://doi.org/10.3102/0034654310393361>
- Tan, C., Yue, W.-G. & Fu, Y. (2017). Effectiveness of flipped classrooms in nursing education: Systematic review and meta-analysis. *Chinese Nursing Research*, 4(4), 192–200. <https://doi.org/10.1016/j.cnre.2017.10.006>
- Ulrich, I. (2020). *Gute Lehre in der Hochschule* (2. Aufl.). Springer Wiesbaden. <https://link.springer.com/book/10.1007/978-3-658-31070-7>
- van Alten, D. C. D., Phielix, C., Janssen, J. & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review*, 28, 100281. <https://doi.org/10.1016/j.edurev.2019.05.003>
- Vo, H. M., Zhu, C. & Diep, N. A. (2017). The effect of blended learning on student performance at course-level in higher education: A meta-analysis. *Studies in Educational Evaluation*, 53, 17–28. <https://doi.org/10.1016/j.stueduc.2017.01.002>
- Voutilainen, A., Saaranen, T. & Sormunen, M. (2017). Conventional vs. E-learning in nursing education: A systematic review and meta-analysis. *Nurse Education Today*, 50, 97–103. <https://doi.org/10.1016/j.nedt.2016.12.020>
- Xu, P., Chen, Y., Nie, W., Wang, Y., Song, T., Li, H., Li, J., Yi, J. & Zhao, L. (2019). The effectiveness of a flipped classroom on the development of Chinese nursing students' skill competence: A systematic review and meta-analysis. *Nurse Education Today*, 80, 67–77. <https://doi.org/10.1016/j.nedt.2019.06.005>

Kontakt

Prof. Dr. Elisabeth Mayweg
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin
E-Mail: elisabeth.mayweg@hu-berlin.de

Dr. Natalie Enders
Universität Hildesheim
Universitätsplatz 1
31141 Hildesheim
E-Mail: natalie.enders@uni-hildesheim.de

Prof. Dr. Carla Bohndick
Universität Hamburg
Hamburger Zentrum für Universitäres Lehren und Lernen
Jungiusstr. 9
20355 Hamburg
E-Mail: carla.bohdick@uni-hamburg.de

Dr. Jana Rückmann
Evangelische Hochschule Berlin
Teltower Damm 118–122
14167 Berlin
E-Mail: jana.rueckmann@eh-berlin.de

Anhang

Analysekorpus Onlinelehre

Anmerkung: Fettgedruckte Quellen sind im Analysekorpus Onlinelehre und im Analysekorpus Blended Learning enthalten.

1. Borokhovski, E., Tamim, R., Bernard, R. M., Abrami, P. C. & Sokolovskaya, A. (2012). Are contextual and designed student–student interaction treatments equally effective in distance education? *Distance Education*, 33(3), 311–329. <https://doi.org/10.1080/01587919.2012.723162>
2. He, L., Yang, N., Xu, L., Ping, F., Li, W., Sun, Q., ... & Zhang, H. (2021). Synchronous distance education vs traditional education for health science students: A systematic review and meta-analysis. *Medical Education*, 55(3), 293–308. <https://doi.org/10.1111/medu.14364>
3. Jurewitsch, B. (2012). A meta-analytic and qualitative review of online versus face-to face problem-based learning. *Journal of Distance Education (Online)*, 26(2).
4. **Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers college record*, 115(3), 1–47.**
5. **Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research*, 81(3), 4–28. <https://doi.org/10.3102/0034654310393361>.**
6. Voutilainen, A., Saaranen, T., & Sormunen, M. (2017). Conventional vs. e-learning in nursing education: A systematic review and meta-analysis. *Nurse Education Today*, 50, 97–103. <https://doi.org/10.1016/j.nedt.2016.12.020>
7. **Wandera, S. (2017). *Continuing the conversation about face-to-face, online, and blended learning a meta-analysis of empirical literature 2006-2017*. Wilmington University (Delaware). <https://www.proquest.com/docview/1964254197>**

Analysekorpus Blended Learning

1. Balakrishnan, A., Puthan, S., Sathesh, G. M. K. U, Rashid, M. et al. (2021). Effectiveness of blended learning in pharmacy education: A systematic review and meta-analysis. *PLOS ONE*, 16(6), e0252461. <https://doi.org/10.1371/journal.pone.0252461>
2. Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M. & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, 26(1), 87–122. <https://doi.org/10.1007/s12528-013-9077-3>
3. Borokhovski, E., Bernard, R. M., Tamim, R. M., Schmid, R. F. & Sokolovskaya, A. (2016). Technology-supported student interaction in post-secondary education: A meta-analysis of designed versus contextual treatments. *Computers & Education*, 96, 15–28. <https://doi.org/10.1016/j.compedu.2015.11.004>
4. Car, L. T., Kyaw, B. M., Dunleavy, G., Smart, N. A., Semwal, M., Rotgans, J. I., ... & Campbell, J. (2019). Digital problem-based learning in health professions: systematic review and meta-analysis by the digital health education collaboration. *Journal of Medical Internet Research*. <https://doi.org/10.2196/12945>
5. Dunleavy, G., Nikolaou, C. K., Nifakos, S., Atun, R., Law, G. C. Y. & Car, L. T. (2019). Mobile digital education for health professions: systematic review and meta-analysis by the digital health education collaboration. *Journal of Medical Internet Research*, 21(2), e12937. <https://doi.org/10.2196/12937>

6. Kazu, İ. Y. & Yalçın, C. K. (2022). Investigation of the effectiveness of hybrid learning on academic achievement: A meta-analysis study. *International Journal of Progressive Education*, 18(1). <https://doi.org/10.29329/ijpe.2022.426.14>
7. Kyaw, B. M., Posadzki, P., Paddock, S., Car, J., Campbell, J. & Car, L. T. (2019). Effectiveness of digital education on communication skills among medical students: systematic review and meta-analysis by the digital health education collaboration. *Journal of Medical Internet Rresearch*, 21(8), e12967. <https://doi.org/10.2196/12967>
8. Li, C., He, J., Yuan, C., Chen, B. & Sun, Z. (2019). The effects of blended learning on knowledge, skills, and satisfaction in nursing students: A meta-analysis. *Nurse Education Today*, 82, 51–57. <https://doi.org/10.1016/j.nedt.2019.08.004>
9. Liu, Q., Peng, W., Zhang, F., Hu, R., Li, Y. & Yan, W. (2016). The effectiveness of blended learning in health professions: systematic review and meta-analysis. *Journal of Medical Internet Rresearch*, 18(1), e4807. <https://doi.org/10.2196/jmir.4807>
10. **Means, B., Toyama, Y., Murphy, R. & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, 115(3), 1–47.**
11. Müller, C. & Mildenerger, T. (2021). Facilitating flexible learning by replacing classroom time with an online learning environment: A systematic review of blended learning in higher education. *Educational Research Review*, 34, 100394. <https://doi.org/10.1016/j.edurev.2021.100394>
12. Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R. M., Abrami, P. C., Surkes, M. A. & Woods, J. (2014). The effects of technology use in postsecondary education: A meta-analysis of classroom applications. *Computers & Education*, 72, 271–291. <https://doi.org/10.1016/j.compedu.2013.11.002>
13. **Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C. & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research*, 81(3), 4–28. <https://doi.org/10.3102/0034654310393361>.**
14. Vallée, A., Blacher, J., Cariou, A. & Sorbets, E. (2020). Blended learning compared to traditional learning in medical education: systematic review and meta-analysis. *Journal of Medical Internet Rresearch*, 22(8), e16504. <https://doi.org/10.2196/16504>
15. Vo, H. M., Zhu, C. & Diep, N. A. (2017). The effect of blended learning on student performance at course-level in higher education: A meta-analysis. *Studies in Educational Evaluation*, 53, 17–28. <https://doi.org/10.1016/j.stueduc.2017.01.002>
16. **Wandera, S. (2017). *Continuing the conversation about face-to-face, online, and blended learning a meta-analysis of empirical literature 2006-2017*. Wilmington University (Delaware). <https://www.proquest.com/docview/1964254197>**

Analysekorpus Flipped Classroom

1. Aydin, M., Okmen, B., Sahin, S. & Kilic, A. (2021). The meta-analysis of the studies about the effects of flipped learning on students' achievement. *Turkish Online Journal of Distance Education*, 22(1), 33–51.
2. Bredow, C. A., Roehling, P. V., Knorp, A. J. & Sweet, A. M. (2021). To flip or not to flip? A meta-analysis of the efficacy of flipped learning in higher education. *Review of Educational Research*, 91(6), 878–918. <https://doi.org/10.3102/00346543211019122>
3. Chen, K.S., Monrouxe, L., Lu, Y.-H., Jenq, C.-C., Chang, Y.-J., Chang, Y.-C. et al. (2018). Academic outcomes of flipped classroom learning: a meta-analysis. *Medical Education*, 52, 910–924. <https://doi.org/10.1111/medu.13616>

4. Cheng, L., Ritzhaupt, A. D. & Antonenko, P. (2019). Effects of the flipped classroom instructional strategy on students' learning outcomes: a meta-analysis. *Educational Technology Research and Development*, 67(4), 793–824. <https://doi.org/10.1007/s11423-018-9633-7>
5. Farmus, L., Cribbie, R. A. & Rotondi, M. A. (2020). The flipped classroom in introductory statistics: Early evidence from a systematic review and meta-analysis. *Journal of Statistics Education*, 28(3), 316–325. <https://doi.org/10.1080/10691898.2020.1834475>
6. Gillette, C., Rudolph, M. Kimble, C. Rockich-Winston, N., Smith, L. & Broedel-Zaugg, K. (2018). A meta-analysis of outcomes comparing flipped classroom and lecture. *American Journal of Pharmaceutical Education*, 82(5), 433–440.
7. Hew, K. F. & Lo, C. K. (2018). Flipped classroom improves student learning in health professions education: a meta-analysis. *BMC Medical Education*, 18(1), 1–12. <https://doi.org/10.1186/s12909-018-1144-z>
8. Hew, K. F., Bai, S., Huang, W., Dawson, P., Du, J., Huang, G., Jia, C. & Thankrit, K. (2021). On the use of flipped classroom across various disciplines: Insights from a second-order meta-analysis. *Australasian Journal of Educational Technology*, 37(2), 132–151. <https://doi.org/10.14742/ajet.6475>
9. Hu, R., Gao, H., Ye, Y., Ni, Z., Jiang, N. & Jiang, X. (2018). Effectiveness of flipped classrooms in Chinese baccalaureate nursing education: A meta-analysis of randomized controlled trials. *International Journal of Nursing Studies*, 79, 94–103. <https://doi.org/10.1016/j.ijnurstu.2017.11.012>
10. Jang, H. Y. & Kim, H. J. (2020). A meta-analysis of the cognitive, affective, and interpersonal outcomes of flipped classrooms in higher education. *Education Sciences*, 10(4), 115. <https://doi.org/10.3390/educsci10040115>
11. Karagöl, İ. & Esen, E. (2018). The effect of flipped learning approach on academic achievement: A meta-analysis study. *Hacettepe University Journal of Education*, 34(3), 708–727. <https://earxiv.anadolu.edu.tr/xmlui/handle/11421/14773>
12. Låg, T. & Sæle, R. G. (2019). Does the flipped classroom improve student learning and satisfaction? A systematic review and meta-analysis. *AERA Open*, 5(3), 1–17. <https://doi.org/10.1177/2332858419870489>
13. Li, B.Z., Cao, N.W., Ren, C.X., Chu, X.J., Zhou, H.Y. & Guo, B. (2020). Flipped classroom improves nursing students' theoretical learning in China: A meta-analysis. *PloS One*, 15(8), e0237926. <https://doi.org/10.1371/journal.pone.0237926>
14. Lo, C. K., Hew, K. F. & Chen, G. (2017). Toward a set of design principles for mathematics flipped classrooms: A synthesis of research in mathematics education. *Educational Research Review*, 22, 50–73. <https://doi.org/10.1016/j.edurev.2017.08.002>
15. Lo, C. K. & Hew, K. F. (2019). The impact of flipped classrooms on student achievement in engineering education: A meta-analysis of 10 years of research. *Journal of Engineering Education*, 108(4), 523–546. <https://doi.org/10.1002/jee.20293>
16. Özdemir, A. & Senturk, M. L. (2021). The effect of flipped classroom model on students' academic achievement in science and mathematics education: A meta-analysis study. *Journal of Educational Technology*, 18(3), 22–41.
17. Shi, Y., Ma, Y., MacLeod, J. & Yang, H. H. (2020). College students' cognitive learning outcomes in flipped classroom instruction: a meta-analysis of the empirical literature. *Journal of Computers in Education*, 7(1), 79–103. <https://doi.org/10.1007/s40692-019-00142-8>
18. Strelan, P., Osborn, A. & Palmer, E. (2020). The flipped classroom: A meta-analysis of effects on student performance across disciplines and education levels. *Educational Research Review*, 30, 100314. <https://doi.org/10.1016/j.edurev.2020.100314>
19. Talan, T. & Batdi, V. (2020). Evaluating the flipped classroom model through the multi-complementary approach. *Turkish Online Journal of Distance Education*, 21(4), 31–67.
20. Tan, C., Yue, W.-G. & Fu, Y. (2017). Effectiveness of flipped classrooms in nursing education: Systematic review and meta-analysis. *Chinese Nursing Research*, 4(4), 192–200. <https://doi.org/10.1016/j.cnre.2017.10.006>

21. Van Alten, D. C., Phielix, C., Janssen, J. & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review*, 28, 100281. <https://doi.org/10.1016/j.edurev.2019.05.003>
22. Xu, P., Chen, Y., Nie, W., Wang, Y., Song, T., Li, H. et al. (2019). The effectiveness of a flipped classroom on the development of Chinese nursing students' skill competence: A systematic review and meta-analysis. *Nurse Education Today*, 80, 67–77. <https://doi.org/10.1016/j.nedt.2019.06.005>
23. Zhang, Q., Cheung, E. S. T. & Cheung, C. S. T. (2021). The impact of flipped classroom on college students' academic performance: A meta-analysis based on 20 experimental studies. *Science Insights Education Frontiers*, 8(2), 1059–1080. <https://doi.org/10.15354/sief.21.re019>
24. Zheng, L., Bhagat, K. K., Zhen, Y. & Zhang, X. (2020). The effectiveness of the flipped classroom on students' learning achievement and learning motivation: A meta-analysis. *Educational Technology & Society*, 23(1), 1–15.

Tabelle A.1: Moderatoren und Effektgrößen in der Praxiskategorie bezogen auf das Cb-Modell (Sailer et al., 2021)

Kategorie	Moderator	Effektgröße	Artikel	Kontraste	
Cb Model – Higher Education Teachers' Digital Technology Use					
Didaktische Elemente mithilfe digitaler Technologie	Technologie-Einsatz	k.A.	Bernard et al. (2014)		
	Medienfeature	.20 - .21	Means et al. (2013)	Nur Text vs. Text plus zusätzliche Medien	
	Hilfsmittel		.07 - .30	He et al. (2020)	Ja vs. Nein
		Pädagogisch-didaktisches Design	-.02 - .16	Müller & Mildenberger (2021)	Flipped Classroom (FC) vs. Andere Formate
	FC-Modell		.68 - .57	Zheng et al. (2020)	Klassisches FC-Modell vs. Innovatives FC-Modell
		FC-Design	.45 - .52	Strelan et al. (2020)	Klassisches FC vs. Minimal FC (kein Signifikanztest)
	One-Way Video oder Audio	.09 - .25	Means et al. (2013)	Vorhanden vs. Nicht vorhanden (nicht angegeben)	
	Computer-basierte Instruktionselemente	.18 - .23	Means et al. (2013)	Vorhanden vs. Nicht vorhanden (nicht angegeben)	
	Vorbereitende Aktivitäten	.25 - .46	Hew & Lo (2018)	Ja vs. Nein	
	Verfügbarkeit Übungen/Literatur/Textmaterial vor dem Unterricht	.30 - .35	Hew & Lo (2018)	Ja vs. Nein	
	Verfügbarkeit bestimmter Aktivitäten vor dem Unterricht	.16 - .42*	Lo & Hew (2019)	Lehrvideos (LV) vs. Texte (T) vs. Online-Quizzes (OQ) vs. LV+T+OQ vs. LV+T vs. LV+OQ vs. T+OQ	
	Vorbereitende Aktivitäten vor dem Unterricht	.35 - .78	Strelan et al. (2020)	Nicht berichtet vs. Diskussion vs. Reflexion vs. andere Bewertungsform; kein Signifikanztest	
	Interaktionen vor dem Unterricht	.50 - .77	Zheng et al. (2020)	Lesen von Lernmaterial vs. Anschauen von Lehrvideos vs. zwei Arten von Interaktion vs. drei oder mehr Arten von Interaktion	
	Aktivitäten vor dem Unterricht	.66 - .75	Zheng et al. (2020)	Online Learning Plattform vs. Andere	
Aktivitäten vor dem Unterricht		.58 - .84	Zheng et al. (2020)	Lehrvideos vs. Zwei Arten von Interaktion vs. Drei oder mehr Arten von Interaktion	
	Quiz vor dem Unterricht	.26 - .56*	Hew & Lo (2018)	Ja vs. Nein	
Kursaktivitäten	-.31 - .68*	Lo & Hew (2019)	Quiz zu Beginn des Kurses vs. Wiederholung vs. Vorträge vs. Einzelaufgaben vs. Kleingruppenaktivitäten vs. Verschiedene Kombinationen		
Kursaktivitäten	.06 - .52	Strelan et al. (2020)	Studierenden-zentriert vs. Dozierenden-zentriert vs. Nicht berichtet; kein Signifikanztest		
Zugang zu Archiven	.03 - .24	He et al. (2020)	Ja vs. Nein		
Zeitliche Aufteilung der digitalen Lerngelegenheiten	Reduktion der Unterrichtszeit	-.01 - .19	Müller & Mildenberger (2021)	30-40% vs. 41-59% vs. 60-79%	
	Grad der Verblendung	k.A.	Borokhovski et al. (2016)		
	Zeitlicher Anteil von Blended Learning Unterrichtszeit (Online vs. Face-to-Face (f-2-f)) Unterrichtszeit		k.A.	Bernard et al. (2014)	
			.18 - .45	Means et al. (2013)	Online > f-2-f vs. f-2-f > Online
	.15 - .47	Bredow et al. (2021)	Äquivalent vs. Weniger bei FC		

Kategorie	Moderator	Effektgröße	Artikel	Kontraste
Cb Model – Higher Education Teachers' Digital Technology Use				
Gesamtzeitraum der Intervention	Dauer der Intervention	.14 - .23	Means et al. (2013)	Weniger als ein Monat vs. Mehr als ein Monat
	Dauer der Intervention	.46 - .56	Shi et al. (2020)	Mehr als ein Semester vs. Weniger als ein Semester
	Dauer der Intervention	.24 - .74	Talan & Batdi (2020)	2-4 Wochen vs. 5-6 Wochen vs. 7-8 Wochen vs. 9-18 Wochen vs. Unspezifisch
	Dauer der Intervention	.46 - 1.11*	Zheng et al. (2020)	2-4 Wochen vs. 5-8 Wochen vs. 9-24 Wochen vs. Mehr als 24 Wochen
	Dauer der Intervention	.41 - .71	Karagöl & Esen (2019)	1-4 Wochen vs. 5-8 Wochen vs. 9 oder mehr Wochen vs. Unspezifisch
	Häufigkeit der Kurse Ausmaß des Unterschieds zwischen Intervention und Kontrolle Flip-Anteil	.04 - .25 .19 - .24* .17 - .55	He et al. (2020) Schmid et al. (2014) Bredow et al. (2021)	1-10 Niedrig vs. Mittel vs. Hoch Flip vs. Teil-Flip
Cb Model – Students' Learning Activities Involving Digital Technology				
Funktionen der eingesetzten Technologien	Funktionen der Technologie	.14 - .63*	Borokhovski et al. (2016)	Kognitive Unterstützung vs. Kommunikationsunterstützung vs. Gemischt
	Funktionen der Technologie	.22 - .59*	Bernard et al. (2014)	Kommunikationsunterstützung vs. Such- und Abrufunterstützung vs. Kognitive Unterstützung vs. Inhaltliche/präsentative Unterstützung vs. Kognitive plus inh./präs. Unterstützung
	Funktionen der Technologie Funktionen der Technologie	.31 - .42* -.06 - .42*	Tamim et al. (2011) Schmid et al. (2014)	Direkter Unterricht vs. Unterstützung des Unterrichts Kommunikationsunterstützung vs. Such- und Abrufunterstützung vs. Kognitive Unterstützung vs. Inhaltliche/präsentative Unterstützung vs. Kognitive plus inh./präs. Unterstützung
Interaktionen und Kommunikation	Gelegenheit zu persönlichen Gesprächen mit Dozierenden	.05 - .30	Means et al. (2013)	Während des Unterrichts vs. Vor oder nach dem Unterricht vs. Abwesend/nicht angegeben
	Gelegenheit zu persönlichen Gesprächen mit Kommiliton*innen	.01 - .30	Means et al. (2013)	Während des Unterrichts vs. Vor oder nach dem Unterricht vs. Abwesend/nicht angegeben
	Interaktion in F2F-Kurs	.65 - .70	Zheng et al. (2020)	Zwei Arten von Interaktionen vs. Drei oder mehr Arten von Interaktionen
Lernmöglichkeiten nach dem Unterricht	Übungsmöglichkeiten	.16 - .21	Means et al. (2013)	Vorhanden vs. Nicht vorhanden (nicht angegeben)
	Wöchentliche Quizzes Aktivitäten nach dem Unterricht	k.A.* .11 - .66	Farmus et al. (2020) Lo & Hew (2019)	Ja vs. Nein Online Quizze vs. Übungen vs. Beides vs. Nichts
	Interaktion	.22 - .50*	Borokhovski et al. (2012)	Natürliche vs. Angeleitete Interaktion

Kategorie	Moderator	Effektgröße	Artikel	Kontraste
Cb Model – Students' Learning Activities Involving Digital Technology				
Instruktions- ansatz	Pädagogisch- didaktischer Ansatz	.05 - .39*	Means et al. (2013)	Dozierendengeleitet/exemplarisch vs. Selbständig/aktiv vs. Kooperativ/interaktiv
	Pädagogisch- didaktischer Ansatz	.36 - 1.05*	Shi et al. (2020)	Individualisiert vs. Kollaborativ vs. Teambasiert
	Computervermittelte Kommunikation mit Dozierenden	.04 - .24	Means et al. (2013)	Nur asynchron vs. Synchron & asynchron
	Computervermittelte Kommunikation mit Dozierenden	.17 - .27	Means et al. (2013)	Nur asynchron vs. Synchron & asynchron
Feedback und Assessment	Bewertungsmethoden am Ende des Kurses	.36 - .41	Vo et al. (2017)	Eine vs. Mehrere
	Leistungsbeurteilung	-.01 - .17	Müller & Mildenberg (2021)	Prüfung/Posttest vs. Kursnote
	Leistungsbeurteilung	.47 - .76	Strelan et al. (2020)	Reliabel (min. 20 Items) vs. Weniger reliabel; kein Signifikanztest
	Feedback	.20 - .21	Means et al. (2013)	Vorhanden vs. Nicht vorhanden (nicht angegeben)
Leistungs- prüfung	Tests zur Vorbereitung	.31 - .40	Lağ & Sæle (2019)	Ja vs. Nein
	Übungen zur Vorbereitung	.05 - .63	Bredow et al. (2021)	Keine vs. Aufgabe vor dem Unterricht vs. Quiz vor dem Unterricht vs. Quiz im Unterricht
Sonstige				
	Kursgröße	k. A.	Farmus et al. (2020)	Groß (mehr als 100 Studierende) vs. Klein
	Gruppengröße bei Gruppenarbeiten	.26 - .48	Strelan et al. (2020)	2-5 vs. 6-10 vs. >11 Studierende; kein Signifikanztest
	Unterrichtsmethoden im F2F-Unterricht	.57 - .89	Zheng et al. (2020)	Eine Unterrichtsmethode vs. Zwei vs. Drei oder mehr Methoden

Anmerkung: Signifikante Moderatoren mit * markiert. Effektstärken auf zwei Nachkommastellen gekürzt.



Kathrin Henrich

Studium als biographischer Übergang

Eine qualitative Studie zu den Perspektiven der Studierenden

Reflexive Übergangsforschung – Doing Transitions, Band 12

2023 • ca. 240 S. • kart. • ca. 34,00 € (D) • ca. 35,00 € (A)

ISBN 978-3-8474-2754-4 • eISBN 978-3-8474-1924-2

Der Übergang ins Studium ist für viele eine der wichtigsten Veränderungen im Lebensweg. Welche Erfahrungen machen Studierende? Wie gehen sie mit diesen Erfahrungen um? Die Studie wertet biographisch-narrative Interviews aus und bietet eine subjektorientierten reflexiven Blick auf das Studium als Übergang. Es wird gezeigt, dass die Themen, die im Studium für die Menschen relevant werden, ähnlich sind, sie diese jedoch in unterschiedlichen Modi des Doing-Transition gestalten.

Die Autorin:

Dipl.-Päd. Kathrin Henrich, Universität Tübingen



Marco Rieckmann, Bror Giesenbauer,
Benjamin Nölting, Thomas Potthast,
Claudia T. Schmitt

Nachhaltige Entwicklung von Hochschulen

Erkenntnisse und Perspektiven zur
gesamtinstitutionellen Transformation

*Schriftenreihe **Ökologie und Erziehungswissenschaft** der Kommission Bildung für nachhaltige Entwicklung der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE)*

2023 • ca. 160 S. • kart. • ca. 40,00 € (D) • ca. 41,20 € (A)

ISBN 978-3-8474-2551-9 • eISBN 978-3-8474-1698-2

Welchen Beitrag zur Nachhaltigen Entwicklung können Hochschulen in den Bereichen Lehre, Forschung, Betrieb und Transfer leisten? Wie gelingt die gesamtinstitutionelle Nachhaltige Entwicklung von Hochschulen? Diese Fragen standen im Fokus des Verbundprojekts „Nachhaltigkeit an Hochschulen (HOCH^N)“ mit elf deutschen Hochschulen. Der Sammelband präsentiert und diskutiert die Ergebnisse.

Die Herausgeber*innen:

Prof. Dr. Marco Rieckmann (Universität Vechta), Dr. Bror Giesenbauer (Universität Bremen), Prof. Dr. Benjamin Nölting, Hochschule für nachhaltige Entwicklung Eberswalde), Prof. Dr. Thomas Potthast (Universität Tübingen), Dr. Claudia T. Schmitt (Universität Hamburg)



Mathias Barth, Daniel Fischer,
Gerd Michelsen (Hrsg.)

Bildung für nachhaltige Entwicklung in der Hochschule

**Wege und Wirkungen am Beispiel der
Leuphana Universität Lüneburg**

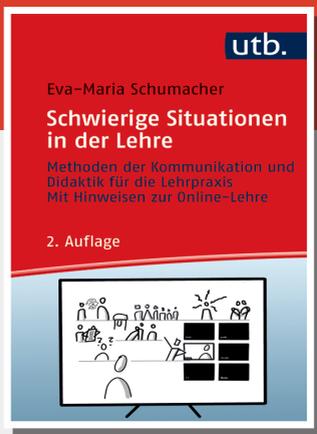
*Schriftenreihe **Ökologie und Erziehungswissenschaft** der Kommission Bildung für nachhaltige Entwicklung der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE)*

2023 • ca. 220 S. • kart. • ca. 29,90 € (D) • ca. 30,80 € (A)

ISBN 978-3-8474-2154-2 • eISBN 978-3-8474-1175-8

Bildung gilt als einer der Schlüssel zur Erreichung einer weltweiten nachhaltigen Entwicklung. Dabei kommt insbesondere der Hochschulbildung eine besondere Bedeutung zu, denn sie soll zukünftige Entscheidungsträger für die Herausforderungen einer nachhaltigen Entwicklung sensibilisieren und zur Übernahme einer aktiven und verantwortungsvollen Gestaltungsrolle befähigen. Die Leuphana Universität Lüneburg hat sich als eine der ersten Universitäten in Deutschland auf den Weg gemacht, sich systematisch mit den damit zusammenhängenden Herausforderungen auseinanderzusetzen. Mit dem Band werden die dabei erprobten Ansätze und Wege sowie Erkenntnisse über die Wirksamkeit eines nachhaltigkeitsorientierten Studiums vorgestellt. Er gibt damit praxiserprobte und wissenschaftlich fundierte Impulse zu Möglichkeiten, Bildung für eine nachhaltige Entwicklung in der Hochschullehre zu integrieren.

www.shop.budrich.de



Eva-Maria Schumacher

Schwierige Situationen in der Lehre

Methoden der Kommunikation und
Didaktik für die Lehrpraxis.
Mit Hinweisen zur Online-Lehre

2. Auflage

Kompetent lehren, Band 2

2022 • 188 S. • kart. • 16,90 € (D) • 17,40 € (A) • utb S

ISBN 978-3-8252-5635-7 • eISBN 978-3-8385-5635-2

Sind Sie genervt von Störungen und Widerständen in Ihren Veranstaltungen? Mit Hilfe dieses Praxisleitfadens verstehen Sie, wodurch Störungen und Widerstände entstehen und wie Sie lösungsorientiert damit umgehen können. Didaktische Prävention hilft Ihnen ebenso wie der supervisorische Blick und andere Ansätze in diesem Buch. Die 2. Auflage widmet sich auch der digitalen Lehre und ihren Herausforderungen und nimmt neue Modelle aus dem Bereich des Coachings in den Blick.

Die Autorin:

Eva-Maria Schumacher, Dipl.-Päd., Supervisorin, Rollenspielleiterin, Lehrtrainerin und Lehrcoach (DVNLP), Leiterin von constructif – Institut für konstruktives Arbeiten, Leben, Lernen

www.shop.budrich.de