



MARTIN-LUTHER-UNIVERSITÄT  
HALLE-WITTENBERG

TORSTEN SCHUBERT & MICHAEL GERTH (HRSG.)

# MULTIMEDIALES LEHREN UND LERNEN AN DER MARTIN-LUTHER-UNIVERSITÄT HALLE-WITTENBERG

---

Befunde und Ansätze aus dem  
Forschungsförderprogramm des  
Zentrums für multimediales Lehren und Lernen (LLZ)

Halle (Saale) 2021



## Impressum

### Herausgeber:

Torsten Schubert, Michael Gerth  
Martin-Luther-Universität Halle-Wittenberg  
Zentrum für multimediales Lehren und Lernen  
Hoher Weg 8, 06120 Halle (Saale)  
<https://www.llz.uni-halle.de/>

### Herausgeber und Autoren des Vorworts:

**Torsten Schubert**, Professor für Allgemeine Psychologie an der Martin-Luther-Universität Halle-Wittenberg und Geschäftsführender Direktor des Zentrums für multimediales Lehren und Lernen.

**Michael Gerth**, promovierte in Politologie und ist seit 2012 Geschäftsführer des Zentrums für multimediales Lehren und Lernen an der Martin-Luther-Universität Halle-Wittenberg.

**Wolf Zimmermann**, Prorektor für Studium und Lehre an der Martin-Luther-Universität Halle-Wittenberg, Professor für Software-Engineering und Programmiersprachen.

### Zitation:

Schubert, T., Gerth, M. (2021). Multimediales Lehren und Lernen an der Martin-Luther-Universität Halle-Wittenberg. Befunde und Ansätze aus dem Forschungsförderprogramm des Zentrums für multimediales Lehren und Lernen (LLZ). Halle (Saale). Beiträge zum multimedialen Lehren und Lernen, Bd. 1. ISBN 978-3-96670-087-0, <http://dx.doi.org/10.25673/38465>.

### Über Beiträge zum multimedialen Lehren und Lernen:

Beiträge zum multimedialen Lehren und Lernen ist eine Publikationsreihe des Zentrums für multimediales Lehren und Lernen (LLZ) der Martin-Luther-Universität Halle-Wittenberg, die in regelmäßigen Abständen Forschungsergebnisse zum multimedialen Lehren und Lernen präsentiert. In der Publikationsreihe können Beiträge veröffentlicht werden, die sich mit der Erforschung von Ansätzen und Entwicklungen zum multimedialen Lehren und Lernen oder mit Grundlagen von Lernen und Wissenserwerb wie auch technischen Entwicklungen im Kontext digital geprägten Lehrens und Lernens beschäftigen. Beiträge können zur Veröffentlichung an die Herausgeber gesendet werden, die eine Veröffentlichung prüfen.

## Vorwort

Im aktuellen Sammelband werden Ergebnisse eines Forschungsförderprogrammes für multimediales Lernen und Lehren berichtet, das im Rahmen des Zentrums für multimediales Lehren und Lernen (LLZ) an der Martin-Luther-Universität Halle-Wittenberg durch das Bund-Länder-Programm Qualitätspakt Lehre in der Zeit von 2017 bis 2020 gefördert wurde. Das Forschungsförderprogramm verfolgte das Ziel, Projekte an der Martin-Luther-Universität Halle-Wittenberg zu unterstützen, die sich der Erforschung von Möglichkeiten, Anwendungen, Herausforderungen und auch den Grundlagen beim multimedialen Lehren und Lernen widmen. Das Programm hat dabei in der kurzen Zeit seines Bestehens eine Vielzahl interessanter Forschungsprojekte zum multimedialen Lernen und Lehren angeregt. Die Befunde und Fragestellungen der Projekte zeigen einerseits die Dringlichkeit einer konzertierten Bündelung dieser Forschung an der Universität und stellen andererseits aber auch die Notwendigkeit heraus, den aktuellen Prozess der Digitalisierung von Lehre und Studium im Bildungsbereich durch entsprechende Forschungen zu begleiten und seine Ergebnisse zu evaluieren, um dessen Nachhaltigkeit zu unterstützen und damit verbunden auch neue, wissenschaftlich fundierte Entwicklungen für das digitale Lehren und Lernen anregen zu können.

Das Forschungsförderprogramm ist sehr eng mit dem LLZ verbunden, mit dessen Gründung im Jahre 2012 an der Martin-Luther-Universität Halle-Wittenberg die Grundlagen zur nachhaltigen Entwicklung der digital gestützten Lehre entwickelt wurden. Die anfangs vordringlich auf den E-Service ausgerichtete LLZ-Struktur und -Arbeit ermöglichte in relativ kurzer Zeit eine breite Nutzung der Learning-Management-Systeme (LMS), den breit gestreuten Einsatz von E-Vorlesungen sowie von E-Prüfungen im Rahmen der Lehrveranstaltungen der Universität. Die rasante Entwicklung des guten Know-hows für digital unterstützten und umgesetzten hochschulischen Unterricht fand einen Höhepunkt am Ende der zweiten Förderperiode des LLZ im Jahr 2020, als während der Corona Pandemie die Lehrveranstaltungen an der Martin-Luther-Universität Halle-Wittenberg trotz nicht zu unterschätzender Schwierigkeiten für alle Beteiligten (Lehrende, Studierende, Zentrale Einrichtungen, Universitätsverwaltung und -leitung) und auch trotz einer spürbaren Heterogenität über die Fächer und Individuen hinweg eine gute digitale Umsetzung in der Breite erfahren konnten. Damit konnten eine hohe Qualität von Lehrveranstaltungen und angemessene Studienverläufe bis hin zu den Prüfungen trotz Pandemie aufrechterhalten werden.

Dabei zeigten sich auch wesentliche Vorteile, die mit der Digitalisierung der Lehre verbunden sind (z. B. Zeit- und Ortsunabhängigkeit) und neue Möglichkeiten in Bezug auf Diversität, Heterogenität, Inklusion etc. von Studierenden mit sich bringen. Darüber hinaus ergeben sich durch die Einführung von E-Learning neue didaktische Möglichkeiten und Notwendigkeiten, die zur Erweiterung der traditionellen Lehr-Lern- und Prüfmethode führen.

Gleichzeitig darf freilich nicht übersehen werden, dass die mit dem E-Learning verknüpfte Einführung von digitalen Lehr- und Lernmedien sowie -methoden auch mit erhebli-

chen Anforderungen und Schwierigkeiten verbunden ist, die auf verschiedenen Ebenen wirksam werden. In einer Extremsituation wie der Corona Pandemie können kurzfristig unerwartet aufgetretene Schwierigkeiten und Anforderungen in der Lehre durch zeitlich-befristet erhöhten persönlichen und auch institutionellen Einsatz begegnet und teilweise kompensiert werden. Eine nachhaltig erfolgreiche Entwicklung digitaler Lehr- und Lernmethoden an den Hochschulen bedarf jedoch neben der Unterstützung durch Service- und Unterstützungseinrichtungen auch entsprechende innovative Begleitforschung. Diese Begleitforschung muss eine wissenschaftliche fundierte Evaluierung von didaktischen Maßnahmen und Methoden zum Inhalt haben, die auf entsprechenden Modellen von Lernenden und Lehrenden sowie deren Kompetenzen beruhen, um darauf begründete Entwicklungen digitaler und technologischer Tools vorantreiben können. Dadurch kann diese Forschung zur Entwicklung moderner hochschulischer Lehre beitragen, bei der E-Learning nicht nur als zusätzlicher, sondern integraler Bestandteil der Lehre akzeptiert ist und damit auch künftige Entwicklungsrichtungen für neue Lehr- und Lernformen vorschlagen kann. Gleichzeitig kann sie dabei auch zum Taktgeber für Grundlagenforschung in unterschiedlichen mit Lernen und Lehren sowie auch Technologieentwicklung verbundenen Wissenschaftsbereichen werden.

Vor dem Hintergrund dieser Überlegungen hatten die Verantwortlichen des LLZ in Zusammenarbeit mit dem Prorektorat für Studium und Lehre an der Martin-Luther-Universität Halle-Wittenberg das Forschungsförderprogramm für multimediales Lernen und Lehren initiiert, das Forschung unterstützen sollte, die sich auf jene Herausforderungen bezieht, die durch die verstärkte Anwendung multimedialer Lehr- und Lernszenarios entstehen. Diese Herausforderungen wurden vor allem in Bezug auf drei Schwerpunkte gesehen: (1) anwendungsbezogene Forschung in Bezug auf lehr- und lerndidaktische Konsequenzen multimedialen Lernen und Lehrens, (2) begleitende grundlagenorientierte Forschung über die kognitions- und lernpsychologischen Konsequenzen aus der Anwendung multimedialer Lehre und (3) anwendungsbezogene Aspekte bei der Entwicklung soft- und hardwaretechnischer Lösungen beim multimedialen Lehren und Lernen.

Im Rahmen des Forschungsförderprogrammes konnten dazu insgesamt zehn Forschungsprojekte bis Ende 2020 realisiert werden. Mit bis zu je 12.000 € wurden Vorhaben verschiedenster Fachrichtungen antragsbasiert entsprechend den vorgenannten Schwerpunkten unterstützt. Die Auswahl der Anträge erfolgte durch den Lenkungskreis des LLZ und das Prorektorat für Lehre und Studium, wobei bei der Auswahl der Projekte ein besonderer Schwerpunkt auf eine möglichst breite Fächerpräsenz gelegt wurde. Die administrative Betreuung erfolgte durch den Forschungsbereich am LLZ. Während der Förderperiode fanden regelmäßig Kolloquien mit allen Beteiligten sowie externen Gästen statt, um die doch sehr unterschiedlichen Forschungsziele, -methoden sowie erste Erkenntnisse bereits im Vorfeld zu diskutieren.

Die hier zusammengefassten Arbeiten weisen die große Leistungsfähigkeit der Forschungen an der Universität zum multimedialen Lehren und Lernen aus. Sie sind in einer verhältnismäßig kurzen Zeit entstanden und zeigen eine große Bandbreite an innovativen Ideen der Forschenden an der Martin-Luther-Universität Halle-Wittenberg. Es wird deutlich, dass

sich viele Arbeiten insbesondere mit der Entwicklung von Visualisierungstechniken im Kontext digitalen Lehren und Lernens beschäftigen. Zwei Arbeiten im Bereich der Biologie- und der Geographiedidaktik ordnen sich hier zentral ein und untersuchen dabei aus jeweils unterschiedlichen Gesichtspunkten wie Kriterien für die Gestaltung von Lehrvideos abgeleitet werden können und nach welchen Gesichtspunkten gute Lehrvideos konstruiert werden sollten. **Tellez-Acosta, Acher und Kusitzky** untersuchen dabei zum Beispiel Kriterien, die sich auf die Ebene der epistemischen, fachspezifischen oder didaktischen Komponenten beziehen, um Lernvideos bezüglich ihrer Verwendbarkeit zu charakterisieren, die eine modellbasierte Untersuchung von naturwissenschaftlichen Gesetzen im Biologieunterricht ermöglichen und Studierende dabei unterstützen sollen, Fachwissen in naturwissenschaftliche Modellierungsprozesse zu integrieren. Für die Analyse wurden Kategorien des Professional Vision Frameworks verwendet. Damit sollen Aussagen gewonnen werden, wie Studierende mit der Unterstützung von Multimedia-Tools Erklärungsmodelle konstruieren, die für Vertiefungsprozesse des Wissenserwerbs im Unterricht verwendet werden können. In der Arbeit von **Breuer** werden als Kriterien zur Bewertung von Videos zu geografischen Geländemethoden inhaltlich-methodische, gestalterische, didaktische und zusätzliche Kriterien herangezogen, nach denen die Qualität von sogenannten „Erklärvideos“ eingeordnet und einer systematischen Analyse unterzogen werden kann. Dies erfolgte mithilfe eines Design-Based-Research-Ansatzes. Nach der Auseinandersetzung und Erprobung von Geländemethoden mithilfe von Erklärvideos, wurden die Anforderungen und Bewertungen diesbezüglich mittels begleitender Interviews erfasst. Es wird deutlich, dass eine große Vielfalt an Anforderungen, Erwartungen und Wünschen an Erklärvideos besteht und diese sich auch hin und wieder zu widersprechen scheinen. Durch die Verwendung verschiedener Bewertungsebenen scheint eine Optimierung nach unterschiedlichen Kriterien hin möglich zu sein. Inwiefern dabei die Befunde der beiden Arbeitsgruppen auf andere Wissenschaftsbereiche übertragen werden können, muss späteren Studien überlassen bleiben. **Ritter und Ritter** beschäftigen sich in ihrer Arbeit mit dem Einfluss der Digitalisierung auf den Erwerb des Lesens als generelle Kulturtechnik. Sie fokussieren dabei auf die Frage, wie sich Lesebedingungen als auch Leseverhalten bei Grundschulkindern im Umgang mit dem analogen Papierbuch und dem digital erweiterten Angebot interaktiver Bilderbuch-Apps verändern. Mit ihren Untersuchungen an Kleingruppen von Schülerinnen und Schülern sowie Lehrkräften können sie aufzeigen, welche Konsequenzen auf inhaltlicher, sozialer und technischer Ebene die Einführung von Bilderbuch-Apps im Vergleich zum Lesen von analogen Büchern und damit einhergehenden vermittelnden Gesprächssituationen der handelnden Personen haben. Diese Untersuchungen lassen einen weiten Bogen spannen, der vom Einfluss der Digitalisierung auf die Entwicklung des Lesens im Individuum bis hin zur Entwicklung des Lesens als kulturell vermittelte und vermittelnde Handlung der sozialen Interaktion reicht.

Digitale Lehre gewinnt Attraktivität auch durch die Möglichkeit von filmischer und spielerischer Umsetzung prototypischer Lernsituationen, wobei durch emotionale Involviertheit des/der Lernenden neue motivationale Einflüsse auf das Lernverhalten entstehen können. **Kotschote, Wenzel und Lindner** decken das am Beispiel von Gamification für das Lernen

von ökologischen Nachhaltigkeitsfragen und die Verwendung von Serious Games Scenarios auf. Dabei untersuchen sie, wie das Ausmaß an Immersion der Lernenden in die Spielsituation Voraussetzung für die Effizienz des Lernprozesses sein kann und prüfen dazu neue Parameter wie die Messung von Augenbewegungen neben der Erfassung von subjektiven Ratingwerten. Aus der Kombination beider Parameter können Aussagen über die Aufmerksamkeitsfokussierung von Lernenden und dem Immersionserleben sowie deren direkten Bezug zur Qualität des Lernergebnis getroffen werden.

Erfolgreiche Begleitforschung im Bereich multimediales Lehren und Lernen muss sich auch das Ziel der modellgeleiteten Entwicklung und Evaluierung von didaktischen Lehr- und Lernszenarios stellen. Das zeigen **Thüring und Jäger** durch das Design und die Testung einer Lehrveranstaltung im Bereich Informatik am Beispiel der Lehre zur Objektorientierten Programmierung. Die entsprechende Neuentwicklung berücksichtigt konsequent die Möglichkeiten digitaler Lehr- und Lernmethoden und fußt auf einem elaborierten Blended-Learning-Konzept, wobei ausgewählte Übungen in digitalisierter Form vorliegen und dadurch eine Entzerrung sowie gleichzeitig eine gebündelte Vermittlung von Wissen- und Fertigkeiten im Bereich der objektorientierten Programmierung ermöglichen. Eine Evaluation und Effizienzprüfung erfolgt durch die Analyse der Abschlussnoten im Modul und das Wahlverhalten der Studierenden, die sich verbesserten gegenüber Situationen mit herkömmlicher Didaktik. In der Arbeit von **Eugster, Barth und Ballod** werden generelle Vor- und Nachteile des digitalen Lehrens- und Lernens in Bezug auf Fragen des Kompetenzerwerbs analysiert und eine Erweiterung der didaktischen Möglichkeiten bei E-Learning deutlich gemacht. In dem eher theoretisch orientierten Beitrag werden die Möglichkeiten der Gestaltung des Wissenserwerbs durch Studierende während der Eigenkonstruktion von Quizen eruiert und damit verbundene Änderungen im Lern- und Studierverhalten sowie einem sich daraus begründenden Kompetenzerwerb der Studierenden verdeutlicht. Es ergeben sich dazu vielfältige Möglichkeiten und Notwendigkeiten einer zukünftigen empirischen Prüfung des Ansatzes.

Im Kontext der Effizienz von Tests und Fragenbeantworten können unter anderem auch elaborierte Konzepte aus der Lernpsychologie weiterhelfen, die insbesondere für das Konzept des testenden Lernens entwickelt worden sind. In der Arbeit von **Aslan und Neumann** konnte dabei für den Bereich pharmazeutischen Wissens gezeigt werden, dass testendes Lernen zu Vorteilen gegenüber Lernmethoden, die auf Repetition und Wiederholungslernen ausgerichtet sind, führt. Besonders hervorzuheben ist dabei unter anderem, dass dies für Situationen gezeigt werden konnte, in denen komplexe Sachverhalte wie Lernsätze etc. beim testenden Lernen verwendet wurden, was frühere Befunde erweitert, die das schon für das eher assoziative Lernen im Kontext einfacher Wort-Wortverbindungen zeigte. Damit ergeben sich völlig neue Dimensionen für die Verwendung von digital implementierten Tests, Quizzes und Audience-Response-Systemen, die einen beträchtlichen Anteil digitaler Lehr-Lerntools im didaktischen Repertoire vieler Lehrender ausmachen.

Nicht zuletzt spielen beim multimedialen Lehren und Lernen auch Innovationen im Bereich der Leistungsüberprüfung eine wesentliche Rolle. Hier ist die Studie von **Ranger, Schmidt und Wolgast** interessant, die sich mit der Entwicklung neuartiger, softwarebasierter

Tools zur Detektion von Prüfungshehrlichkeit beschäftigt. Diese basieren auf der Testung verschiedener Parameter des Test- und Antwortverhaltens, die durch Prozeduren des maschinellen Lernens genutzt werden können, um Abweichungen im Antwortverhalten zu detektieren, die auf die Nutzung unerlaubter Hilfsmittel hinweisen können. Während die Untersuchung deutlich macht, dass über den Proof-of-Concept hinaus noch viel Forschung notwendig ist, um die entwickelten Methoden auf die Detektion von Fehlverhalten in neuen Wissensbereichen zu übertragen, zeigt der generelle Ansatz jedoch die Notwendigkeit von verstärkten Forschungsanstrengungen im Bereich Learning Analytics und Methoden der künstlichen Intelligenz zur Weiterentwicklung von multimedialen Lehr- und Lernszenarios und Methoden.

Diesen Arbeiten ist zusätzlich eine Zusammenstellung von Befragungsergebnissen an der Martin-Luther-Universität Halle-Wittenberg vorangestellt, in der die Entwicklung der Erwartungen der Lehrenden an multimediale Lehr-Lernmethoden und deren Anwendungsbedingungen in der hochschulischen Lehre über den Zeitraum von 2012 bis 2018 analysiert werden (*Schubert, Schurer, Gerth und Schrimpf*). Dabei wurde u. a. gefragt, welche Veränderungen des Nutzungsverhaltens multimedialer Lehr- und Lerninstrumente über den genannten Erhebungszeitraum zu verzeichnen sind und welche Faktoren die Nutzung beeinflussen können. Die Ergebnisse der Befragungen geben Aufschluss darüber, welche Vor- und Nachteile sowie welche Schwierigkeiten Lehrende beim Einsatz von multimedialen Elementen erwarten und wie diese mit entsprechenden Maßnahmen auf institutioneller und Serviceebene entsprochen werden kann. Damit können begründete Handlungsempfehlungen für den Ausbau der E-Learning Kapazitäten und entsprechende Service- und Unterstützungsmaßnahmen abgeleitet werden, die auch für andere technologische Transformationsprozesse in Studium und Forschung Gültigkeit haben.

Der aktuelle Sonderband gibt somit einen interessanten Einblick in innovative Entwicklungen auf dem Gebiet der Erforschung multimedialen Lernens und Lehrens im Kontext hochschulischer Lehre und gleichzeitig aber auch in die Reflexionen der Lehrenden, die deren Einschätzungen von Nutzen, Vor- und Nachteilen sowie den erwarteten Schwierigkeiten bei der Durchführung digitaler Lehr- und Lernmethoden betreffen. Damit ist ein Bogen gespannt, durch den sowohl Praktiker/innen als auch Grundlagenforscher/innen wie auch Lehrende im hochschulischen oder anderen Kontexten angesprochen werden.

Die hier vereinten Arbeiten stellen dabei nur einen Ausschnitt der Arbeiten des Forschungsförderprogramms dar. Eine Gesamtübersicht finden Sie auf der FFP-Webseite (<https://www.llz.uni-halle.de/forschung/ffp>). Die Arbeiten weisen auf ein sich noch in der Entwicklung befindliches Gebiet der Forschung hin, das nicht zuletzt durch die breite Kombination und Integration verschiedenster Forschungsansätze und –fragen seinen besonderen Reiz im Kontext von digital gestütztem Lernen, Wissens- und Kompetenzerwerb gewinnt.

Torsten Schubert, Michael Gerth, Wolf Zimmermann

Gefördert durch Bundesministerium für Bildung und Forschung Förderkennzeichen 01PL17065

# Inhalt

<b>1. Erwartungen und Anforderungen bezüglich multimedialer Lehr-Lernmethoden bei Lehrenden an der Martin-Luther-Universität Halle-Wittenberg .....</b>	<b>2</b>
<b>Erwartungen und Anforderungen an die Digitalisierung der Lehre: Befragungsergebnisse von Lehrenden der Martin-Luther-Universität Halle-Wittenberg im Zeitraum 2012 – 2018</b> <i>Torsten Schubert, Teresa Schurer, Michael Gerth, Anne Schrimpf.....</i>	<b>3</b>
<b>2. Lernvideos, Bilderbuch-Apps und Serious Games als digitale Medien in Unterrichtsszenarien .....</b>	<b>24</b>
<b>Unterstützung von Grundschullehramtsstudierenden bei der Integration von naturwissenschaftlichen Fachwissen aus Online-Videos in die Modellierungsprozesse durch „Multimedia-Tools“</b> <i>María Esther Tellez-Acosta, Andres Acher und Leon Kusitzky.....</i>	<b>25</b>
<b>Die Entwicklung und Evaluation von Erklärvideos zu geographischen Geländemethoden</b> <i>Maximilian Breuer .....</i>	<b>41</b>
<b>Gamification im MINT-Unterricht: Eine Untersuchung des Spielverhaltens, Nachhaltigkeits- und Immersionserleben am Beispiel des Serious Game LandYOUs</b> <i>Nils L. Kotschote, Chris Wenzel und Martin Lindner .....</i>	<b>55</b>
<b>Leseproxen im Medienzeitalter: Ergebnisse einer Fallstudie zu digitalen und analogen (Bilderbuch-)Lektüren</b> <i>Alexandra Ritter und Michael Ritter .....</i>	<b>74</b>
<b>3. Entwicklung und Testung ausgewählter didaktischer Szenarien im Kontext von E-Learning.....</b>	<b>91</b>
<b>Das Inverted Classroom Mastery Modell in der Informatik: Untersuchungen zum Kompetenzerwerb in der Objektorientierten Programmierung</b> <i>Annett Thüring und Kathrin Jäger .....</i>	<b>92</b>
<b>Reflektives Peer-Quizzing! Didaktische Potenziale für die nachhaltige Entwicklung adaptiver Lernsettings</b> <i>Benjamin Eugster, René Barth und Matthias Ballod.....</i>	<b>108</b>
<b>Testbasiertes Lernen in der medizinisch-pharmakologischen Lehre</b> <i>Alp Aslan und Joachim Neumann.....</i>	<b>118</b>
<b>4. Innovationen beim digitalen Prüfen .....</b>	<b>129</b>
<b>Detection of Cheating on E-Exams – The Performance of Transferred Detection Rules</b> <i>Jochen Ranger, Nico Schmidt und Anett Wolgast.....</i>	<b>130</b>

# 1 Erwartungen und Anforderungen bezüglich multimedialer Lehr-Lernmethoden bei Lehrenden an der Martin-Luther-Universität Halle-Wittenberg

# Erwartungen und Anforderungen an die Digitalisierung der Lehre: Befragungsergebnisse von Lehrenden der Martin-Luther-Universität Halle-Wittenberg im Zeitraum 2012 – 2018

*Torsten Schubert<sup>1,2</sup>, Teresa Schurer<sup>1</sup>, Michael Gerth<sup>1</sup>, Anne Schrimpf<sup>1</sup>*

*<sup>1</sup> Martin-Luther-Universität Halle-Wittenberg, Zentrum für multimediales Lehren und Lernen*

*<sup>2</sup> Martin-Luther-Universität Halle-Wittenberg, Institut für Psychologie*

Digitale Lehr- und Lernangebote gewinnen in der Hochschullehre immer mehr an Bedeutung. An der Martin-Luther-Universität Halle-Wittenberg unterstützt das Zentrum für multimediales Lehren und Lernen den Einsatz von multimedial gestützten Lehrangeboten durch eine didaktische und technische Betreuung der Lehrenden. In vier Befragungen (2012, 2014, 2016, 2018) wurden  $N = 1017$  Lehrende zu erwarteten Nutzen, Schwierigkeiten, notwendiger Unterstützung, Relevanz multimedialer Elemente in ihrer Lehre und zum Nutzungsverhalten der Lernmanagementsysteme (LMS) befragt. Die Analyse der Befragungsergebnisse über diesen Zeitraum verdeutlicht, dass sich bei den Lehrenden die Einschätzung des erwarteten Nutzens durch elektronische Lehre über die Jahre hinweg erhöht und die Einschätzung der erwarteten Schwierigkeiten durch elektronische Lehre verringert hat. Über die Jahre stiegen auch die Nutzungszahlen der LMS kontinuierlich an. Des Weiteren lassen die Analysen Schlüsse zu, welche Faktoren die Einschätzung der Relevanz von multimedialen Elementen für die Lehre sowie den Wunsch nach zukünftiger Einbindung von multimedialen Elementen in die Lehre bestimmen. Es wird erkennbar, dass die Vorurteile der Lehrenden gegenüber E-Learning sowohl die Nutzung von multimedial gestützten Lehrangeboten als auch die wahrgenommenen Vor- und Nachteile der digital gestützten Tools beeinflussten. Eine geringe Nutzung von E-Learning steht dabei auch häufig in Zusammenhang mit der Wahrnehmung geringer Selbstwirksamkeit und Kompetenz in Bezug auf die Nutzung von E-Learning. Gezielte Unterstützungsmaßnahmen sowie kontinuierliche Beratungen und Weiterbildungen sind deshalb wichtig, um Vorurteile bei Lehrenden abzubauen und entsprechenden Kompetenzen in Bezug auf digitale Tools und entsprechend angepasste Didaktik zu entwickeln.

## Einleitung

Die immense Bedeutung des Einsatzes neuer Technologien und damit verbundener Änderungen didaktischer Methoden und Ansätze für die Hochschullehre zeigten insbesondere die Entwicklungen während der Corona-Pandemie und die damit verbundenen Onlinesemester. Mit diesen Entwicklungen der Hochschullehre, die allerdings schon vor der Pandemie einsetzten, ging auch eine enorme Umstellung der Lehrenden an den Universitäten und Hochschulen einher, da diese sich nun nicht nur auf die klassischen Lehrinhalte und -methoden, sondern auch auf technische Anforderungen im Bereich Informationstechnologie (IT) - und Internetnutzung zur didaktischen Umsetzung der Lehrinhalte einstellen mussten. Trotz steigender Nachfrage und entsprechender Unterstützung erfolgte die Umstellung des Lehrens auf digitale Methoden nicht ohne Hindernisse. Unsicherheiten im Umgang mit neuen Medien oder bezüglich didaktischer Anpassungen, die Sorge vor einem höheren Arbeitsaufwand sowie eine generelle Skepsis bezüglich der Wirksamkeit der neuen Methoden in der Hochschullehre sind häufig auftretende Einwände aufseiten der Lehrenden. Sowohl für die technische als auch die didaktische Komponente des digitalen Lernens benötigen Lehrende und Hochschulmitarbeitende deshalb eine gezielte und breite fachliche und personelle Unterstützung (Martin, Ritzhaupt, Kumar, & Budhrani, 2019; Wingo, Ivankova, & Moss, 2017).

Die vom Bundesministerium für Bildung und Forschung (BMBF) im Jahr 2012 geförderte Gründung des Zentrums für multimediales Lehren und Lernen (LLZ) an der Martin-Luther-Universität (MLU) Halle-Wittenberg diente der institutionellen Konsolidierung sowie der breiten Bereitstellung dieser Unterstützung, vor allem auch mit Hinblick auf die Steigerung multimedial gestützter Lehrangebotsqualität sowie einer technisch und didaktisch fundierten Beratung und Weiterbildung der Lehrenden.

Um die Bedürfnisse und den jeweiligen Stand der Kompetenzen der Lehrenden in dieses Konzept einbinden zu können, fanden in regelmäßigen Abständen Befragungen zur Einschätzung des Bedarfs an technischer und didaktischer Unterstützung bei multimedialer Lehre sowie zur Einschätzung von deren Umsetzung an der MLU statt. Diese systematischen Befragungen dienen nicht nur der Erfassung des Ist-Zustandes, sondern auch als Grundlage für Anpassungen seitens der Servicebereiche der Universität und des LLZ. Dabei wurden die Lehrenden alle zwei Jahre (2012, 2014, 2016, 2018) zu erwarteten Nutzen, Schwierigkeiten, notwendiger Unterstützung und der Relevanz multimedialer Elemente in ihrer Lehre befragt.

Die Ergebnisse dieser Befragungen werden im vorliegenden Aufsatz einer ersten systematischen Analyse mit einem besonderen Fokus auf die Entwicklung der Einschätzungen über den betrachteten Zeitraum unterzogen. Es wurde untersucht, welche Veränderungen des Nutzungsverhaltens multimedialer Lehr- und Lerninstrumente über den genannten Erhebungszeitraum zu verzeichnen sind und mit welchen Faktoren deren Nutzung in Beziehung stehen. Insbesondere wurde untersucht, welchen Nutzen und welche Schwierigkeiten Lehrende der MLU Halle-Wittenberg beim Einsatz von multimedialen Elementen in der Lehre sehen und erwarten und es wurde untersucht, was Lehrende grundsätzlich über den Einsatz multimedialer Elemente des Lehrens und Lernens denken und wie sich das auf die Nutzung

dieser Elemente auswirkt. Im vorliegenden Aufsatz wurde dabei das Nutzungsverhalten von multimedialen Lehr- und Lerninstrumenten durch die Antworten der Lehrenden auf die Frage nach dem Ausmaß der Nutzung von Lernmanagement Systemen (LMS), Stud.IP und ILIAS, operationalisiert, wobei den Autoren bewusst ist, dass das Nutzungsverhalten der Lehrenden in Bezug auf digitale Lehr-Lerninstrumente viel facettenreicher ist als die bloße Erfassung der durch die Lehrenden selbst eingeschätzten LMS-Nutzung. Der Grund für diese Herangehensweise besteht in der durchgängigen Verwendung der entsprechenden Items zur LMS-Nutzung in den Fragebögen über den betrachteten Zeitraum von 2012 bis 2018. Trotz der Bedingtheit der Operationalisierung wird durch diese Herangehensweise eine vorläufige Annäherung an die Frage der zeitlichen Entwicklung der Einschätzungen ermöglicht, die spätere Analysen des realen Nutzungsverhaltens digitaler Lehr- und Lernmittel sowie der damit verbundenen didaktischen Methoden nach sich ziehen soll.

Die vorliegende Untersuchung hat somit explorativen Charakter. Anhand der Untersuchung der aufgeworfenen Fragen sollen Entwicklungstendenzen der Bedarfe der Lehrenden sowie ihre Einstellung zur multimedialen Lehre seit der Gründung des LLZs an der MLU Halle-Wittenberg von 2012 bis 2018 analysiert werden. Dadurch ergibt sich erstmals ein detailliertes Bild über die Entwicklungen im Bereich multimedialer Lehre und deren Reflektion durch die Lehrenden über einen relativ ausgedehnten zeitlichen Umfang. Daneben werden Schlussfolgerungen über mögliche Zusammenhänge zwischen der zeitlichen Entwicklung in den Einstellungen der Lehrenden gegenüber Nutzen und Einsatz multimedialer Lehr- und Lernmethoden sowie der Einschätzung der eigenen digitalen Kompetenz und dem Grad der Umsetzung von Unterstützungsmaßnahmen durch Serviceeinrichtungen ermöglicht.

## **Methoden**

### **Fragebogen**

Der verwendete Fragebogen wurde 2012 von einer Arbeitsgruppe des LLZ entwickelt, die sich aus Mitarbeitenden der Bereiche Qualitätsmanagement und E-Assessment sowie aus technischen Mitarbeitenden zusammensetzte. In den darauffolgenden Jahren wurde der Fragebogen teilweise überarbeitet und verändert, z. B. wurden einige Items entfernt und andere ergänzt. In den hier vorgestellten Analysen wurden nur die Items einbezogen, die in mindestens zwei der vier Erhebungszeiträume konstant verwendet wurden. Trotz des dadurch notwendigen Ausschlusses einiger Items werden durch diese Vorgehensweise Aussagen über Entwicklungstendenzen in den Einschätzungen der Lehrenden hinsichtlich des multimedialen Lehrens und Lernen und der Nutzungszahlen an der MLU Halle-Wittenberg ermöglicht.

Für die web-basierte Evaluierung wurde die Software EvaSys (Electric Paper Evaluationssysteme GmbH, Lüneburg, Germany) verwendet. Die Beantwortung des Fragebogens dauerte durchschnittlich 15 Minuten. Der Fragebogen bestand aus zwei Teilbereichen. Im ersten Bereich wurde die Nutzung der LMS Stud.IP und ILIAS an der MLU durch die Lehrenden erfasst. Die Fragen bezüglich der Nutzung der LMS wurden auf einer Skala von 1 (oft) bis 4 (nie) beantwortet. Auf einer Skala von 1 (nie) bis 6 (> 4) sollten die Lehrenden angeben, in

wie vielen Lehrveranstaltungen sie in den letzten zwei Semestern die Lernplattform ILIAS genutzt haben. Im vorliegenden Aufsatz wird durch diesen Parameter das Nutzungsverhalten im Kontext multimedialer Lehr-Lerninstrumente an der MLU operationalisiert.

Weiterhin wurden subjektive Aussagen zum erwarteten Nutzen, zu erwarteten Schwierigkeiten, zur benötigten Unterstützung und zur wahrgenommenen Relevanz von Multimediaeinsatz in der Lehre erfasst. Die Lehrenden wurden dazu gebeten, die Aussagen mithilfe einer Likert-Skala zu beantworten, bei deren Konstruktion besonderer Wert auf eine symmetrische Konstruktion der Antworten gelegt wurde (Likert, 1932). Da damit eine Annäherung der Daten aus den Likertitems an eine Intervallskalierung ermöglicht wird, ist die Verwendung der Daten aus den Likertantworten analog zu metrischen Daten möglich (Carifio & Perla, 2007). Die Richtung der Kodierung der Antwortmöglichkeiten war allerdings für verschiedenen Items unterschiedlich. Die Fragen bezüglich Nutzen, Schwierigkeiten, benötigter Unterstützung und wahrgenommener Relevanz multimedialer Lehr-Lernmethoden sollten dabei jeweils auf einer Skala von 1 (trifft voll zu) bis 5 (trifft gar nicht zu) beantwortet werden. Bei der Wahrnehmung von Vorteilen hinsichtlich multimedialer Elemente in der Lehre wiesen geringe Werte auf eine positive Entwicklung hin. Bei der Frage nach der Wahrnehmung von Schwierigkeiten hinsichtlich multimedialer Elemente in der Lehre weisen hohe Werte auf eine positive Entwicklung hin. Die Mittelwerte der Item-Beantwortung pro Erhebungszeitraum können in den **Tabellen 2, 4-7** eingesehen werden.

Im zweiten Bereich der Erhebung wurden sozio-demographische Merkmale, z. B. Geschlecht, Fakultät, Position in der universitären Hierarchie und Anzahl der Semesterwochenstunden (siehe **Tabelle 1**) erfasst.

### **Stichprobe**

Bei den Befragungen wurden alle Lehrenden der MLU per E-Mail über ihren universitätsinternen E-Mail-Account gebeten, an der Erhebung teilzunehmen. Diese E-Mail beinhaltete Informationen über den Zweck der Erhebung, Datenschutz sowie einen Link zum Online-Fragebogen. Der Link war TAN-generiert, um sicherzustellen, dass jeder Lehrende nur einmal an der Erhebung teilnimmt. Zweimal wurden jeweils im zweiwöchigen Abstand Erinnerungen an die Lehrenden verschickt. Die Teilnahme war freiwillig und anonym. Insgesamt nahmen  $N = 1017$  Lehrende an den vier Erhebungen von 2012 bis 2018 teil (Informationen zu den Stichproben siehe **Tabelle 1**). Die Befragungen der Lehrenden fanden immer im Wintersemester in folgenden Zeiträumen statt: im Zeitraum von Oktober bis November (2012), Dezember bis Januar (2014), Oktober bis Januar (2016) sowie Dezember bis Januar (2018). Die Erhebungen wurden jeweils mit dem Prorektorat für Studium und Lehre, dem Lenkungskreis des LLZ und dem Personalrat der MLU abgestimmt.

### **Statistische Auswertung**

Für die Auswertung wurden die Antworten der Lehrenden auf die Items der Likertskala jeweils in ihre numerischen Entsprechungen transformiert und dann pro Item die Summenwerte für alle Lehrenden gebildet, die eine Antwort auf das jeweilige Item gegeben haben.

Anschließend wurden die Summenwerte durch die Anzahl der Rater pro Item dividiert und der jeweilige Mittelwert pro Item gebildet. Eine derartige Vorgehensweise zur metrischen Verwendung von Daten aus Likertskalen ist zulässig, wenn die jeweiligen Antwortkategorien auf die jeweiligen Likertitems symmetrisch konstruiert wurden und ein Verständnis von Gleichabständigkeit zwischen den jeweiligen Antwortkategorien der Likertskala vorausgesetzt wird (Carifio & Perla, 2007; Rost, 1996, u. v. a.).

Um zu prüfen, ob sich die Einschätzung der Lehrenden über die Jahre hinweg veränderten, wurde untersucht, ob es einen Einfluss des Faktors Zeit (Jahresscheiben 2012 - 2018) auf die Itemantworten gibt. Obwohl die Verwendung metrischer Verfahren bei Likertdaten debattiert wird (Bortz & Döring, 2014; Carifio & Perla, 2007) wurden die vorliegenden Daten varianzanalytisch ausgewertet, um im Sinne einer explorativen Studie die Chance zur Detektion möglicher Unterschiede zwischen den Antworten in unterschiedlichen Jahren zu erhöhen (siehe Glass et al., 1972). Dazu wurde zunächst mit Hilfe einer multivariaten Varianzanalyse (MANOVA) über die verschiedenen Itemantworten in den unterschiedlichen Themenbereichen (Nutzungshäufigkeiten, Einschätzung des Nutzens, der Schwierigkeiten, erwünschte Rahmenbedingungen, grundsätzliche Einstellung zum multimedialen Lehren und Lernen) getestet, ob es einen Einfluss des Faktors Zeit (Jahresscheiben) auf die kombinierten Variableneffekte der für den jeweiligen Themenbereich betrachteten Items gibt. Anschließend wurden separate univariate Varianzanalysen (ANOVA) mit dem Faktor Zeit über die jeweiligen Items in den einzelnen Bereichen durchgeführt, um detailliertere Aussagen über mögliche Unterschiede in den Itemantworten zu prüfen (siehe **Tabellen 2, 4, 5, 6**).

Mittels geplanter T-Tests wurde dann selektiv geprüft, ob sich Erhöhungen oder Verringerungen zwischen den Ratingwerten am ersten Messzeitpunkt im Jahr 2012 und dem letzten Messzeitpunkt im Jahr 2018 oder zu anderen Messzeiten ergaben, wenn es einen generellen Einfluss des Faktors Zeit über alle Messzeitpunkte auf die Antworten gab (Signifikanz entsprechend Bonferroni-Korrektur). Die berichteten Freiheitsgrade variieren dabei über die untersuchten Items hinweg, in Abhängigkeit von den jeweiligen Anzahlen der Lehrenden, die auf das jeweilige Item antworteten. Da es keinen Zwang zur Beantwortung der Items gab, gaben einige Lehrende bei einigen Items keine Antworten ab.

Zusätzlich wurden Korrelationen nach Pearson (zweiseitig, siehe **Tabelle 3** und **8**) berechnet, um einen Zusammenhang zwischen 1) dem Ausprägungsgrad potentieller Einflussfaktoren auf die Nutzung multimedialer Methoden zu finden, sowie 2) zu ermitteln, inwieweit die Antwort des Items „Der Einsatz von E-Learning/multimedialen Elementen in der Lehre bietet mir keinen Vorteil gegenüber reinen Präsenzveranstaltungen“ im Zusammenhang mit der Beantwortung anderer Items des Fragebogens steht.

**Tabelle 1.** Übersicht über ausgewählte Parameter der Lehrenden, die an den vier Bedarfsanalysen teilgenommen haben.

Stichprobenmerkmale

	2012	2014	2016	2018
Total <i>n</i>	452	184	176	205
Frauen in %	37.4	39.1	40.9	43.4
Anzahl der Semesterwochenstunden	4.00±1.8	4.61±1.7	4.33±1.89	4.16±1.69
Lehrende nach Fakultät in %				
Philosophische Fakultät I	2.7	1.6	18.2	15.6
Philosophische Fakultät II	12.4	9.2	14.2	12.7
Philosophische Fakultät III	11.7	8.7	6.3	8.3
Naturwissenschaftliche Fakultät I	15.3	20.7	9.1	9.3
Naturwissenschaftliche Fakultät II	13.9	16.8	13.1	13.7
Naturwissenschaftliche Fakultät III	5.5	6.5	10.2	13.7
Medizinische Fakultät	12.4	12	8.5	2.9
Juristische und wirtschaftswissenschaftliche Fakultät	13.5	9.2	10.8	16.1
Theologische Fakultät	1.1	0.5	4.5	2.4
Universitätsklinikum	9.5	9.8	0	0
Andere	0.2	3.3	2.3	3.4
Fehlende Angaben	1.8	1.6	2.8	2.0
Lehrende nach Position in %				
Professor/in	22.8	28.3	31.8	26.8
Junior-/Honorar-/Gast-Professor/in	0.8	0	1.7	0.5
Privat-/Universitätsdozent/in	7.1	5.4	3.4	5.9
Wissenschaftliche/r Mitarbeiter/in	59.3	48.3	44.3	57.6
Ärztliches Personal	0.9	2.7	0	0
Lehrbeauftragte	3.5	3.3	6.3	1.5
Lehrkraft für besondere Aufgaben	4.2	10.3	11.4	6.3

**Hinweis:** Die Werte bilden den Mittelwert und die Standardabweichung ab.

## Ergebnisse

### 1. Welche Veränderung der LMS-Nutzung ist über den Erhebungszeitraum seit Gründung des LLZ zu verzeichnen und welche Faktoren korrelieren mit dieser Nutzung?

Die MANOVA ergab einen signifikanten Effekt des Faktors Zeit auf die eingeschätzte Nutzung der LMS ILIAS und Stud.IP ( $F(12,902) = 22.94, p < .001, \text{partielles } \eta^2 = .209, \text{Wilk's } \Lambda = .494$ ). Die Höhe der eingeschätzten Nutzungshäufigkeit sowohl von Stud.IP als auch von ILIAS steigt über den betrachteten Zeitraum stark an, wie zusätzliche Analysen durch weiterführende Tests belegen (siehe entsprechende ANOVA Ergebnisse in **Tabelle 2**). Multiple T-Tests für unabhängige Stichproben zeigen jeweils eine signifikante Veränderung in der Nutzung der LMS zwischen dem ersten Messzeitpunkt im Jahr 2012 und dem letzten Messzeitpunkt im Jahr

2018. In Bezug auf die Einschätzung der Nutzungshäufigkeit von Stud.IP fällt die Nutzungshäufigkeit für das Jahr 2018 ( $M = 1.14$ ;  $SD = .462$ ) signifikant größer aus als im Jahr 2012, ( $M = 1.46$ ;  $SD = .885$ ), ( $t = -4.252$ ,  $p = .000$ ), wobei die Ratingwerte zwischen 1 (oft) und 4 (nie) variieren. Analog verhalten sich die Einschätzungen der Nutzungshäufigkeit bei ILIAS, deren Einschätzungen sich ebenfalls zwischen dem 2018 ( $M = 2.65$ ;  $SD = 1.264$ ) und dem Jahr 2012 ( $M = 3.49$ ;  $SD = .949$ ), ( $t = -6.801$ ,  $p = .000$ ) signifikant unterscheiden.

Zusätzlich sank die Anzahl der Lehrenden stark, die nie eines der LMS nutzten, von 2012 bis 2018 (Stud.IP: 2012: 6.5 %, 2014: 4.4 %, 2016: 2.0 %, 2018: 0.6 %; ILIAS: 2012: 73.0 %, 2014: 58.3 %, 2016: 58.7 %, 2018: 39.6%).

**Tabelle 2.** Antworten der Lehrenden bezüglich der Nutzungshäufigkeit der LMS Stud.IP und ILIAS

Item	2012	2014	2016	2018	F	df	p
Nutzen Sie Stud.IP?	1.46±.89	1.39±.82	1.22±.59	1.14±.46	8.315	3, 935	<.001
Nutzen Sie ILIAS?	3.49±.95	3.21±1.09	3.18±1.12	2.65±1.26	15.420	3, 543	<.001
In wie vielen Ihrer Lehrveranstaltungen wurde ILIAS in diesem Sommersemester eingesetzt?	1.22±.77	2.09±1.44	2.00±1.31	3.37±1.49	118.472	3, 627	<.001
In wie vielen Ihrer Lehrveranstaltungen wurde ILIAS im letzten Wintersemester eingesetzt?	1.25±.85	2.06±1.29	1.90±1.14	3.45±1.64	146.949	3, 695	<.001

**Hinweis:** Die Beantwortung der Fragen bezüglich der Nutzung von Stud.IP und ILIAS reichte auf einer Skala von 1 (oft) bis 4 (nie). Angaben zur Anzahl der Lehrveranstaltungen in denen in den letzten zwei Semestern ILIAS genutzt wurde reichten von 1 (nie) bis 6 (> 4). Die Werte repräsentieren Mittelwert ± Standardabweichung.

In früheren Studien konnte gezeigt werden, dass die Entscheidung der Lehrenden, LMS in der Lehre zu nutzen, von verschiedenen Faktoren abhängt, z. B. individuelles Interesse, Zeitmangel, Mangel an technischer Unterstützung sowie wahrgenommener Nutzen oder Benutzerfreundlichkeit (Coleman & Mtshazi, 2017). Zudem konnte gezeigt werden, dass eine Verbesserung des Service für Lehrende, z. B. durch technischen Support oder LMS-Trainingsveranstaltungen, eine stärkere Nutzung der Angebote nach sich zieht (Al-Alwani & Soomro, 2009; Coleman & Mtshazi, 2017). Deshalb wurde in einem zweiten Schritt untersucht, welche der Items in diesen Bedarfsanalysen mit der LMS-Nutzung korrelierten.

Alle Ergebnisse der Korrelationsanalysen können in **Tabelle 3** eingesehen werden. Diese ergaben, dass die subjektiven Anforderungen an die Rahmenbedingungen für multimediale Lehre kaum mit der Nutzung von Stud.IP und ILIAS korrelieren. Vielmehr standen die erwarteten Schwierigkeiten in direktem negativen Zusammenhang mit dem Ausmaß der LMS-Nutzung. Zum Beispiel korrelierte das Item „Meine technischen Kompetenzen reichen dafür nicht aus“ signifikant mit der Nutzung von Stud.IP ( $r = -.131$ ,  $p < .001$ ) und ILIAS ( $r = -.171$ ,  $p < .001$ ). Je mehr die eigenen technischen Kompetenzen demnach als unzureichend eingeschätzt wurden, desto weniger wurden die LMS genutzt. Auch die Sorge, dass

die Kommunikation mit den Studierenden abnehmen könnte korrelierte negativ mit der LMS-Nutzung (Stud.IP:  $r = -.138, p < .001$ ; ILIAS  $r = -.286, p < .001$ ).

Umgekehrt ist zu erkennen, dass das Ausmaß erwarteter Vorteile bei einer Nutzung multimedialer Elemente mit dem Ausmaß der LMS-Nutzung positiv zusammenhängt. Je mehr die Lehrenden zustimmten, dass E-Learning dazu beitragen kann, den Leistungsstand der Studierenden besser einschätzen und Prüfungen besser vorbereiten zu können, sowie den Studienerfolg der Studierenden zu erhöhen, desto häufiger wurden die LMS genutzt. Die Wahrnehmung, dass E-Learning didaktische Möglichkeiten erweitert, korrelierte ebenfalls positiv mit der Stud.IP- und ILIAS-Nutzung, wie in **Tabelle 3** zu sehen ist. Eine positive Korrelation bedeutet dabei, dass je größer der erwartete Nutzen wahrgenommen wurde, desto mehr wurden die LMS genutzt. Eine negative Korrelation bedeutet umgekehrt, dass je größer die erwarteten Schwierigkeiten wahrgenommen wurde, desto weniger wurden die LMS genutzt.

**Tabelle 3.** Korrelationen zwischen den jeweiligen Item-Antworten und den reflektierten Vorteilen von E-Learning

Item	Nutzung von Stud.IP	Nutzung von ILIAS
<b>Erwarteter Nutzen</b>		
Bessere Vereinbarkeit von Studium und Beruf/Privatleben	.107**	.080
Höhere Rücksichtnahme auf unterschiedliche Eingangsvoraussetzungen (Vorwissen, Sprache, Motivation, Lernstrategien)	.114*	.176**
Auffangen von Überlaststudiengängen	.047	.129**
Bessere Einschätzung des Leistungsstandes der Studierenden	.111**	.185**
Mehr Kommunikation mit den Studierenden (bspw. über Chats, Foren etc.)	.125**	-.037
Bessere Prüfungsvorbereitung	.142**	.238**
Steigerung des Studienerfolges der Studierenden	.136**	.210**
Erweiterung meiner didaktischen Möglichkeiten	.169**	.181**
<b>Erwartete Schwierigkeiten</b>		
Meine Lehre wird zunächst arbeitsaufwendiger	-.021	.105*
Meine persönlichen Daten sind nicht hinreichend geschützt (Datenschutz)	-.073*	-.165**
Mein geistiges Eigentum ist nicht hinreichend geschützt (Urheberrecht)	-.051	-.188**
Ich möchte nicht, dass Kollegen eventuell Einblick in meine Lehre erhalten	-.053	-.030
Die Studierenden werden zu sehr allein gelassen	-.090**	-.239**
Die Kommunikation mit den Studierenden nimmt ab	-.138**	-.286**
Die Vorlesungen werden schlechter/weniger besucht	-.067	-.212**
Meine technischen Kompetenzen reichen dafür nicht aus	-.131**	-.171**
Die technischen Kompetenzen der Studierenden reichen dafür nicht aus	.003	.033
<b>Benötigte Rahmenbedingungen</b>		
Didaktische Unterstützung	-.039	.091
Technische Unterstützung	-.049	-.005
Personelle Unterstützung	-.024	.057
Informationsangebote zum Einsatz und Möglichkeiten an der MLU	-.018	-.016
Spezielle Fortbildungsangebote	-.040	-.007
Anerkennung im Sinne der Deputatsanrechnung	-.064	.083
Eine verstärkte Nachfrage von Studierenden	.072*	.042

Ein fächerübergreifender Austausch von Lehrinhalten innerhalb der Universität	.058	.102*
Ein fächerinterner Austausch von Lehrinhalten zwischen den Universitäten	.058	.091*
<b>Grundsätzliche Einstellung</b>		
E-Learning/multimediale Elemente sind nicht relevant für meine Lehre	-.156**	-.373**
Der Einsatz von E-Learning/multimedialen Elementen ermöglicht mir neue didaktische Einsatzszenarien in meiner Lehre	.023	.353**
Der Einsatz von E-Learning/multimedialen Elementen in der Lehre bietet mir keinen Vorteil gegenüber reinen Präsenzveranstaltungen	-.156**	-.360**
Ich fühle mich ausreichend über die Möglichkeiten von E-Learning/multimedial gestützter Lehre informiert	.075*	.133**

**Hinweis:** \*\* = signifikante Korrelation  $p < .01$  (zweiseitig). \* = signifikante Korrelation  $p < .05$  (zweiseitig); Skala von 1 (trifft voll zu) bis 5 (trifft gar nicht zu); positive Korrelation: je größer der erwartete Nutzen, desto mehr werden die LMS genutzt bzw. negative Korrelation: je größer die erwartete Schwierigkeit, desto weniger werden die LMS genutzt.

Auch steht eine skeptische Grundhaltung Lehrender in engem Zusammenhang mit dem Nutzungsverhalten. Das zeigt sich zum Beispiel am Item „Der Einsatz von E-Learning/multimedialen Elementen in der Lehre bietet mir keinen Vorteil gegenüber reinen Präsenzveranstaltungen“, welches stark mit der Nutzung von Stud.IP ( $r = -.156, p < .001$ ) und ILIAS ( $r = -.360, p < .001$ ) dahingehend korrelierte, dass diese LMS von Lehrenden deutlich weniger genutzt wurden, wenn sie darin keinen Vorteil durch E-Learning für ihre Lehre sahen. Weitere Studien müssten jedoch zeigen, ob ein Abbau von Vorurteilen und/oder eine Verstärkung medialer/technischer Kompetenzen der Lehrenden zur Steigerung der LMS-Nutzung in der Hochschullehre führen kann.

## 2. Welchen Nutzen erwarteten Lehrende vom Multimedia-Einsatz in der Lehre und wie veränderte sich diese Sicht über den Erhebungszeitraum?

Die Ergebnisse in **Tabelle 4** zeigen, dass Lehrende nicht unbedingt die Vorteile durch multimediale Lehre darin sehen, um den Leistungsstand der Studierenden besser einschätzen zu können, um Überlaststudiengänge aufzufangen oder um besser mit den Studierenden kommunizieren zu können; das wird durch die durchgängig relativ hohen Werte zu diesen Items in **Tabelle 4** illustriert, durch die die Ablehnung eines Vorteils kodiert wird. Vorteile werden eher darin gesehen, didaktische Möglichkeiten zu erweitern und Prüfungen vorzubereiten. Generell zeigten die MANOVA Ergebnisse, dass sich die Ratingwerte auf die Items bezüglich der Einschätzung von Vorteilen zwar über die Zeit hinweg generell verändern, ( $F(14,530) = 4.224, p < .000$ , partielles  $\eta^2 = .100$ , Wilk's  $\Lambda = .809$ ), allerdings muss der Einfluss des Faktors Zeit für verschiedene Items jedoch unterschiedlich betrachtet werden.

**Tabelle 4.** Antworten der Lehrenden bezüglich des erwarteten Nutzens von Multimedia in der Lehre in Bezug auf verschiedene Nutzenkategorien

Item	2012	2014	2016	2018	F	df	p
Bessere Vereinbarkeit von Studium und Beruf/ Privatleben	2.70±1.20	2.81±1.15	2.86±1.23	2.74±1.29	1.143	3, 822	.510
Höhere Rücksichtnahme auf unterschiedliche Eingangsvoraussetzungen (Vorwissen, Sprache, Motivation, Lernstrategien)	n.a.	3.34±1.19	2.81±1.21	2.55±1.19	17.586	2, 477	<.001
Auffangen von Überlaststudiengängen	3.47±1.29	3.71±1.17	3.69±1.25	3.21±1.42	4.473	3, 709	.004
Bessere Einschätzung des Leistungsstandes der Studierenden	3.56±1.18	3.40±1.20	3.45±1.26	3.21±1.22	3.293	3, 844	.020
Mehr Kommunikation mit den Studierenden (bspw. über Chats, Foren etc.)	3.38±1.31	3.41±1.35	3.28±1.36	3.32±1.30	.358	3, 882	.784
Bessere Prüfungsvorbereitung	2.79±1.20	2.65±1.13	2.43±1.09	2.18±1.02	13.263	3, 893	<.001
Steigerung des Studienerfolges der Studierenden	3.01±1.19	2.86±1.11	2.79±1.25	2.54±1.10	6.216	3, 808	<.001
Erweiterung meiner didaktischen Möglichkeiten	2.78±1.26	2.61±1.17	2.48±1.22	2.20±1.11	9.972	3, 902	<.001

**Hinweis:** Die Beantwortung der Fragen bezüglich des erwarteten Nutzens reichte auf einer Likert Skala von 1 (trifft voll zu) bis 5 (trifft gar nicht zu). Das heißt, je kleiner der Wert, desto größer der erwartete Nutzen. Die Werte repräsentieren Mittelwert ± Standardabweichung. N. a. = not assessed.

Wie die separaten ANOVAs in **Tabelle 4** zeigen, gibt es über die vier Erhebungszeiträume für zwei Items („Bessere Vereinbarkeit von Studium und Beruf/Privatleben“ sowie „Mehr Kommunikation mit den Studierenden (bspw. über Chats, Foren etc.)“) keine Veränderung über die Zeit. Alle anderen Items zeigen von 2012 bis 2018 eine Zunahme der Zustimmung zu einem Nutzen durch multimediale Elemente für die Lehre, z. B. darin, dass diese den Studienerfolg der Studierenden verbessern oder unterschiedliche Eingangsvoraussetzungen der Studierenden besser berücksichtigt werden können. Dabei gilt, je kleiner der Wert, desto größer der erwartete Nutzen. Das wird auch in den Ergebnissen von zusätzlichen T-Tests zu den jeweiligen Items deutlich, für die die ANOVA jeweils einen signifikanten Effekt des Faktors Zeit auf die Ratingurteile der Lehrenden ergab. Für die Items „Bessere Einschätzung des Leistungsstandes der Studierenden“, „Bessere Prüfungsvorbereitungen“, „Steigerung des Studienerfolges der Studierenden“ sowie „Erweiterung meiner didaktischen Möglichkeiten“ zeigten die T-Tests, dass der jeweilige Mittelwert im Jahr 2012 signifikant höher als der jeweilige

Wert im Jahr 2018 ist, kleinstes  $t > 3.15$ ; alle  $ps < .005$ ; wobei ein kleinerer Wert ein größeres Ausmaß an Zustimmung zum erwarteten Nutzen bedeutet.

Für die Items „Höhere Rücksichtnahme auf unterschiedliche Eingangsvoraussetzungen (Vorwissen, Sprache, Motivation, Lernstrategien)“ sowie „Auffangen von Überlaststudienengängen“ zeigten die T-Tests zwar keinen signifikanten Unterschied zwischen den Urteilen im Jahr 2018 im Vergleich zum Jahr 2012, allerdings ergaben sich dann jeweils signifikant höhere Werte im Jahr 2014 bzw. im Jahr 2016 im Vergleich zum Jahr 2018, kleinstes  $t > 1.99$ ; alle  $ps < .005$ .

Die Ergebnisse weisen somit insgesamt auf einen Wandel zu einer positiveren Einschätzung des Nutzens von Multimedia für die Hochschullehre über die Zeit von 2012 bis 2018 hin.

### **3. Welche Schwierigkeiten erwarteten Lehrende vom Multimedia-Einsatz in der Lehre und wie veränderte sich diese Sicht über die Erhebungszeiträume?**

Recht deutlich ist an den Ratingwerten in **Tabelle 5** zu erkennen, dass Lehrende einen großen Nachteil beim Einsatz multimedialer Lehr- und Lernangebote darin sehen, dass die Lehre zunächst arbeitsaufwändiger wird. Die Ratingwerte weisen weiter darauf hin, dass der Einblick durch Kollegen und Kolleginnen in die eigene Lehre sowie möglicherweise fehlende technische Kompetenz seitens der Studierenden als wenig problematisch angesehen werden.

Die Ergebnisse der MANOVA erbrachte einen Einfluss des Faktors Zeit auf die Einschätzungen der Nachteile von E-Learning ( $F(27, 1850) = 2.022, p < .001$ , partielles  $\eta^2 = .029$ , Wilk's  $\Lambda = .917$ ), wobei dieser Effekt für verschiedene Items unterschiedlich ausfällt. So gibt es für die Wahrnehmung des erhöhten Arbeitsaufwandes, den möglichen Einblick durch Kollegen und Kolleginnen in die eigene Lehre, die Bedenken, dass Vorlesungen weniger besucht werden sowie die Einschätzung der eigenen technischen Kompetenzen oder die der Studierenden keine oder wenig Veränderung über die Zeit. Diese Items verhalten sich relativ stabil bzw. es zeigt sich beim Item Wahrnehmung des erhöhten Arbeitsaufwandes ein stetes Auf und Ab der Einschätzungen, wobei generell ein hohes Ausmaß an Zustimmung zum erhöhten Aufwand festzustellen ist (siehe **Tabelle 5**). Alle anderen Items zeigten von 2012 bis 2018 eine signifikante Abnahme der wahrgenommenen Schwierigkeiten, die multimediale Elemente in der Lehre verursachen könnten, z. B. die Sorge um Daten- und Urheberschutz oder die Sorge, dass Studierende zu sehr alleine gelassen werden und die Kommunikation mit ihnen abnimmt (siehe **Tabelle 5**). Je kleiner der Wert, desto größer waren die wahrgenommenen Schwierigkeiten für Multimedia in Bezug auf den dargestellten Bereich. Das wird durch die Ergebnisse von multiplen T-Tests über die Werte in den Items deutlich, für die die ANOVA jeweils einen signifikanten Effekt des Faktors Zeit auf die Ratingurteile zeigte. Für die Items „Meine persönlichen Daten sind nicht hinreichend geschützt“, „Mein geistiges Eigentum ist nicht hinreichend geschützt“, „Die Studierenden werden zu sehr allein gelassen“, „Die Studierenden werden zu sehr allein gelassen“ sowie „Die Kommunikation mit den Studierenden nimmt ab“ zeigten die T-Tests, dass der jeweilige Mittelwert im Jahr 2018 signifikant höher als der jeweilige Wert im Jahr 2012 ausfällt, kleinstes  $t > -4.6$ ; alle  $ps < .001$ . Für das

Item „Meine Lehre wird zunächst arbeitsaufwendiger“ zeigte der T-Test keinen signifikant höheren Wert im Jahr 2018 im Vergleich zum Jahr 2012,  $p > .05$ , allerdings war der Wert im Jahr 2012 höher als im Jahr 2014,  $t = 2.56$ ,  $p < .005$ .

Die Ergebnisse weisen somit auf eine deutliche Tendenz bei den Lehrenden zur Verringerung verschiedener Sorgen bezüglich neuer Technologien in der Lehre und auf einen Wandel zu einer insgesamt positiveren Wahrnehmung von Multimedia in der Hochschullehre über den betrachteten Zeitraum hin.

**Table 5.** Antworten der Lehrenden bezüglich der erwarteten Schwierigkeiten bei der Nutzung von Multimedia in der Lehre in Bezug auf unterschiedliche Schwierigkeitsitems

Item	2012	2014	2016	2018	F	df	p
Meine Lehre wird zunächst arbeitsaufwendiger	2.01±1.04	1.77±.96	1.91±1.00	2.01±.97	2.638	3, 961	.048
Meine persönlichen Daten sind nicht hinreichend geschützt (Datenschutz)	3.17±1.24	3.12±1.28	3.13±1.34	3.63±1.20	6.537	3, 826	<.001
Mein geistiges Eigentum ist nicht hinreichend geschützt (Urheberrecht)	2.89±1.25	2.89±1.25	2.96±1.31	3.39±1.18	7.575	3, 884	<.001
Ich möchte nicht, dass Kollegen eventuell Einblick in meine Lehre erhalten	4.33±.91	4.31±.94	4.40±.97	4.31±.99	.382	3, 950	.766
Die Studierenden werden zu sehr allein gelassen	2.82±1.21	3.04±1.23	3.30±1.23	3.24±1.26	8.590	3, 926	<.001
Die Kommunikation mit den Studierenden nimmt ab	2.88±1.30	3.06±1.32	3.31±1.30	3.28±1.35	6.290	3, 936	<.001
Die Vorlesungen werden schlechter/weniger besucht	2.71±1.28	2.80±1.28	2.81±1.31	2.75±1.30	.306	3, 814	.821
Meine technischen Kompetenzen reichen dafür nicht aus	3.81±1.13	3.68±1.17	3.96±1.06	3.79±1.14	1.821	3, 944	.142
Die technischen Kompetenzen der Studierenden reichen dafür nicht aus	3.83±.94	3.61±.93	4.00±.91	3.81±.97	4.346	3, 848	.005

**Hinweis:** Die Beantwortung der Fragen bezüglich der erwarteten Schwierigkeiten reichte auf einer Likert Skala von 1 (trifft voll zu) bis 5 (trifft gar nicht zu). Die Werte repräsentieren Mittelwert ± Standardabweichung, je kleiner der Wert, desto größer die wahrgenommenen Schwierigkeiten für Multimedia in Bezug auf den dargestellten Bereich, z. B. Kommunikation, Kompetenzen, geistiges Eigentum, etc.)

#### 4. Welche Unterstützung wird von den Lehrenden als wichtig wahrgenommen, um erfolgreich Multimedia-Elemente in die Lehre zu integrieren?

Tendenziell sehen Lehrende die größte Unterstützung für den erfolgreichen Einsatz multimedialer Elemente im technischen und personellen Support. Auch die Anrechnung auf das Lehrdeputat wird mehrheitlich gewünscht. Hingegen wird der fächerübergreifende oder fächerinterne Austausch von Lehrinhalten als weniger notwendig angesehen.

Die Anforderungen an die Rahmenbedingungen verhalten sich über die vier Erhebungszeiträume sehr stabil und zeigen nur wenige Veränderungen auf, was auch durch die Ergebnisse der MANOVA verdeutlicht wird, die keinen signifikanten Effekt der Zeit auf die Einschätzung gewünschter Rahmenbedingungen ergab, ( $F(9, 264) = 0.887, p = .25$ , partielles  $\eta^2 = .029$ , Wilk's  $\Lambda = .971$ ). Zusätzlich durchgeführte separate Analysen, die aufgrund der nicht-signifikanten MANOVA Ergebnisse mit Vorsicht zu interpretieren sind, zeigen jedoch bei zwei Items eine selektive Veränderung über die Zeit: Zum einen scheint sich der Wunsch nach mehr Informationsangeboten seitens der MLU-Lehrenden über die Jahre zu verringern, zum anderen erhöht sich das Ausmaß der wahrgenommenen Nachfrage nach Multimedia durch die Studierenden nach Ansicht der Lehrenden (siehe **Tabelle 6**). Die Ergebnisse weisen insgesamt darauf hin, dass sich der Wunsch nach Vorhaltung bestimmter Rahmenbedingungen für einen erfolgreichen Einsatz von Multimedia in der Hochschullehre größtenteils konstant über die Jahre hinweg auf einem moderaten Niveau verhält.

*Tabelle 6. Antworten der Lehrenden bezüglich der benötigten Rahmenbedingungen für den Einsatz von bzw. der gewünschten Unterstützung beim Einsatz von Multimedia in der Lehre*

Item	2012	2014	2016	2018	F	df	p
Didaktische Unterstützung	n.a.	2.77±1.26	2.85±1.20	2.65±1.28	1.046	2, 502	.352
Technische Unterstützung	2.16±1.16	2.15±1.15	2.26±1.09	2.22±1.20	.396	3, 921	.756
Personelle Unterstützung	2.22±1.22	n.a.	2.31±1.22	2.20±1.24	.409	2, 753	.664
Informationsangebote zum Einsatz und Möglichkeiten an der MLU	2.47±1.26	2.82±1.14	2.84±1.18	2.64±1.21	5.040	3, 883	.002
Spezielle Fortbildungsangebote	2.46±1.25	2.52±1.10	2.70±1.18	2.48±1.25	1.459	3, 892	.224
Anerkennung im Sinne der Deputatsanrechnung	n.a.	2.36±1.38	2.36±1.32	2.39±1.45	.019	2, 472	.982
Eine verstärkte Nachfrage von Studierenden	2.91±1.14	2.70±1.18	2.61±1.07	2.67±1.22	3.449	3, 867	.016
Ein fächerübergreifender Austausch von Lehrinhalten innerhalb der Universität	3.19±1.21	3.19±1.14	3.02±1.20	2.98±1.31	1.636	3, 852	.179
Ein fächerinterner Austausch von Lehrinhalten zwischen den Universitäten	2.89±1.24	2.92±1.22	2.95±1.11	2.84±1.31	.238	3, 863	.870

**Hinweis:** Die Beantwortung der Fragen bezüglich der benötigten Rahmenbedingungen reichte auf einer Likert Skala von 1 (trifft voll zu) bis 5 (trifft gar nicht zu). Je kleiner der Wert, desto wichtiger wird die jeweilige Rahmenbedingung für den Erfolg von Multimedia angesehen. Die Werte repräsentieren Mittelwert  $\pm$  Standardabweichung. N.a. = not assessed.

## 5. Was denken die Lehrenden der MLU grundsätzlich über multimediales Lehren und Lernen und wie wirkt sich dies auf die Nutzung digitaler Lehre aus?

Die Ergebnisse zeigen insgesamt, dass sich die Einstellungen der Lehrenden gegenüber dem Einsatz neuer Technologie und deren Akzeptanz durch die Lehrenden über den betrachteten Zeitraum verbesserten. Das wird durch die Ergebnisse der MANOVA über die Einschätzungen der Lehrenden unter Einschluss der Items belegt, zu denen Daten von mindestens drei der vier Jahresscheiben vorlagen und durch die weiterführenden Analysen. Die entsprechende MANOVA zeigte einen signifikanten Effekt des Faktors Zeit auf die grundsätzlichen Einstellungen der Lehrenden gegenüber digitaler Lehre ( $F(6,1016) = 5,7, p < .001$ ; partielles  $\eta^2 = .029$ , Wilk's  $\Lambda = .936$ ).

Wie **Tabelle 7** zeigt, wird von den Lehrenden die Aussage, dass multimediale Elemente generell nicht relevant für die eigene Lehre sind, mit zunehmender Zeit über die Jahre 2012 bis 2018 immer stärker abgelehnt. So war der Mittelwert für das Item „Der Einsatz von E-Learning/multimedialen Elementen in der Lehre bietet mir keinen Vorteil gegenüber reinen Präsenzveranstaltungen“ im Jahr 2018 signifikant größer als der jeweilige Wert im Jahr 2012 ( $t = -3.076; p < .01$ ), wobei große Werte Ablehnung und kleine Werte Zustimmung zur jeweiligen Aussage bedeuteten. Beim T-Test für das Item „Ich fühle mich ausreichend über die Möglichkeiten von E-Learning/multimedial gestützter Lehre informiert“, war der Mittelwert im Jahr 2018 signifikant kleiner als im Jahr 2012 ( $t = 5.192, p < .001$ ), was unterstreicht, dass sich die Lehrenden mit zunehmenden Jahren besser informiert fühlten als zu Beginn der Erhebungen im Jahr 2012. Interessant ist, dass das Ausmaß der Zustimmung zur Aussage zunimmt, dass der Einsatz von multimedialen Elementen neue didaktische Einsatzszenarien ermöglicht. Allerdings konnte hier nur ein direkter Vergleich von zwei benachbarten Jahresscheiben 2016 ( $M = 2.66$ ) und 2018 ( $M = 2.43$ ),  $t = 2.005, p < .05$ ) vorgenommen werden, da dieses Items nicht in den Jahren 2012 und 2014 erhoben wurde.

**Tabelle 7.** Antworten der Lehrenden bezüglich ihrer grundsätzlichen Einstellung zum Einsatz von Multimedia in der Lehre

Item	2012	2014	2016	2018	F	df	p
E-Learning/ multimediale Elemente sind nicht relevant für meine Lehre	n.a.	3.42 $\pm$ 1.20	3.62 $\pm$ 1.12	3.66 $\pm$ 1.20	2.018	2, 541	.134
Der Einsatz von E-Learning/ multimedialen Elementen ermöglicht mir neue didaktische Einsatzszenarien in meiner Lehre	n.a.	n.a.	2.66 $\pm$ 1.08	2.43 $\pm$ 1.14	4.022	1, 361	.046

Der Einsatz von E-Learning/ multimedialen Elementen in der Lehre bietet mir keinen Vorteil gegenüber reinen Präsenzveranstaltungen	2.93±1.24	3.06±1.20	3.16±1.26	3.27±1.32	3.580	3, 936	.014
Ich fühle mich ausreichend über die Möglichkeiten von E-Learning/ multimedial gestützter Lehre informiert	3.33±1.17	3.30±1.16	2.81±1.13	2.82±1.09	14.989	3, 957	<.001

**Hinweis:** Die Beantwortung der Fragen bezüglich der grundsätzlichen Einstellung reichte auf einer Likert Skala von 1 (trifft voll zu) bis 5 (trifft gar nicht zu). Die Werte repräsentieren Mittelwert ± Standardabweichung. N.a. = not assessed.

Insgesamt weisen diese Befunde darauf hin, dass sich über den Erhebungszeitraum von 2012 bis 2018 sowohl die Akzeptanz multimedialer Elemente in der Lehre als auch die Kenntnisse über die Einsatzmöglichkeiten medial gestützter Lehre verbessert haben.

In einem weiteren Schritt wurde untersucht, in welchem Zusammenhang die generelle Einstellung der Lehrenden zum E-Learning mit dem Antwortverhalten bei der Beantwortung spezieller Items steht, in denen nach spezifischen Vor- bzw. Nachteilen durch E-Learning gefragt wird (siehe **Tabelle 8**). Dazu wurden Korrelationen zwischen den Antwortwerten auf das Item „Der Einsatz von E-Learning/multimedialen Elementen in der Lehre bietet mir keinen Vorteil gegenüber reinen Präsenzveranstaltungen“ und den Werten zum wahrgenommenen Nutzen, zu den Schwierigkeiten sowie zu den notwendigen Rahmenbedingungen korreliert. Für die Berechnung der Korrelationen wurden die Werte in den jeweiligen Items über alle Jahresscheiben zusammengenommen. Die Ergebnisse zeigten, dass das Item „Der Einsatz von E-Learning/ multimedialen Elementen in der Lehre bietet mir keinen Vorteil gegenüber reinen Präsenzveranstaltungen“ mit fast allen anderen Items stark korreliert, was darauf hindeutet, dass die generelle Einstellung der Lehrenden zum E-Learning sehr eng mit Einschätzung des Nutzens, der Schwierigkeiten und des benötigten Supports steht. Mit zunehmender Zustimmung der Lehrenden, dass E-Learning keine Vorteile aufweist, wurden beispielsweise die Items „Bessere Prüfungsvorbereitung“, „Steigerung des Studienerfolgs der Studierenden“ sowie „Erweiterung meiner didaktischen Möglichkeiten“ besonders stark abgelehnt. Zudem empfanden diejenige Lehrenden, für die E-Learning keine Vorteile aufweist, besonders stark, dass die Studierenden zu sehr alleine gelassen werden, die Kommunikation mit ihnen abnimmt und Vorlesungen weniger besucht werden. Damit einher geht dann auch die Ablehnung eines Bedarfs an fächerübergreifenden und fächerinternen Austauschs von Lehrinhalten.

**Tabelle 8.** Korrelationen zwischen den Item-Antworten zu erwartetem Nutzen bzw. erwarteten Schwierigkeiten und den wahrgenommenen Vorteilen von E-Learning

Item	Keine wahrgenommenen Vorteile durch E-Learning
<b>Erwarteter Nutzen</b>	
Bessere Vereinbarkeit von Studium und Beruf/Privatleben	-.293**
Höhere Rücksichtnahme auf unterschiedliche Eingangsvoraussetzungen (Vorwissen, Sprache, Motivation, Lernstrategien)	-.353**
Auffangen von Überlaststudiengängen	-.206**
Bessere Einschätzung des Leistungsstandes der Studierenden	-.384**
Mehr Kommunikation mit den Studierenden (bspw. über Chats, Foren etc.)	-.379**
Bessere Prüfungsvorbereitung	-.461**
Steigerung des Studienerfolges der Studierenden	-.518**
Erweiterung meiner didaktischen Möglichkeiten	-.466**
<b>Erwartete Schwierigkeiten</b>	
Meine Lehre wird zunächst arbeitsaufwendiger	.033
Meine persönlichen Daten sind nicht hinreichend geschützt (Datenschutz)	.307**
Mein geistiges Eigentum ist nicht hinreichend geschützt (Urheberrecht)	.286**
Ich möchte nicht, dass Kollegen eventuell Einblick in meine Lehre erhalten	.109**
Die Studierenden werden zu sehr allein gelassen	.480**
Die Kommunikation mit den Studierenden nimmt ab	.507**
Die Vorlesungen werden schlechter/weniger besucht	.408**
Meine technischen Kompetenzen reichen dafür nicht aus	.117**
Die technischen Kompetenzen der Studierenden reichen dafür nicht aus	.129**
<b>Benötigte Rahmenbedingungen</b>	
Didaktische Unterstützung	-.178**
Technische Unterstützung	-.200**
Personelle Unterstützung	-.111**
Informationsangebote zum Einsatz und Möglichkeiten an der MLU	-.233**
Spezielle Fortbildungsangebote	-.197**
Anerkennung im Sinne der Deputatsanrechnung	-.063
Eine verstärkte Nachfrage von Studierenden	-.128**
Ein fächerübergreifender Austausch von Lehrinhalten innerhalb der Universität	-.214**
Ein fächerinterner Austausch von Lehrinhalten zwischen den Universitäten	-.210**

**Hinweis:** \*\* = signifikante Korrelation  $p < 0.01$  (zweiseitig). Die Items sind für alle Bereiche nach der folgenden Skala kodiert. Skala: 1 = trifft voll zu bis 5 = trifft gar nicht zu). Bei Items zum Bereich erwarteter Nutzen bedeutet eine negative Korrelation: je kleiner der Ratingwert (d. h. mehr Zustimmung zum jeweiligen Nutzen), desto größer werden die Vorteile von E-Learning eingeschätzt. Bei Items zu jeweiligen Schwierigkeiten bedeutet eine positive Korrelation: je kleiner der Ratingwert (d. h. Zustimmung zur jeweiligen Schwierigkeit), desto weniger Vorteile durch E-Learning.

## Diskussion

In der vorliegenden Untersuchung wurden Entwicklungstendenzen in den Einschätzungen Hochschullehrender über den Bedarf an, den Nutzen und die Voraussetzungen zur Anwendung von multimedialen Lehr- und Lernmethoden an der MLU Halle-Wittenberg seit Gründung des LLZ über den Zeitraum von 2012 bis 2018 analysiert. Bei der Betrachtung der Ergebnisse wird deutlich, dass die Lehrenden über die Jahre hinweg ein geändertes Bewusstsein für die Digitalisierung der Lehre entwickelt haben. So hat sich über den betrachteten Zeitraum die Einschätzung des erwarteten Nutzens durch E-Learning, z. B. in Bezug auf die Steigerung des Studienerfolgs und für eine bessere Berücksichtigung von unterschiedlichen Eingangsvoraussetzungen der Studierenden, erhöht. Gleichzeitig haben sich einige Einschätzungen von erwarteten Schwierigkeiten durch E-Learning verringert. Das betrifft zum Beispiel die Sorge um Daten- und Urberschutz oder dass Studierende zu sehr alleine gelassen werden und die Kommunikation mit ihnen abnimmt. Die Ergebnisse weisen auf eine Verringerung der Sorgen bezüglich neuer Technologien in der Lehre hin und gleichzeitig auf eine Verstärkung der wahrgenommenen Vorteile und damit auf einen Wandel hin zu einer insgesamt positiveren Wahrnehmung von E-Learning und multimedialen Lehr-Lernansätzen in der Hochschullehre an der MLU. Die Rahmenbedingungen, die nach Einschätzung der Lehrenden notwendig sind, um erfolgreich multimediale Elemente in ihre Lehre einzubauen, werden über die Jahre hinweg als konstant wichtig angesehen; das betrifft vor allem den technischen und personellen Support, welcher gleichbleibend stark gewünscht wird. Generell kann gesagt werden, dass die Relevanz von multimedialen Elementen in der Lehre von den Lehrenden umso höher eingeschätzt wird, je mehr sie der Ansicht sind, dass multimediale Elemente zur Steigerung des Studienerfolgs der Studierenden beitragen oder sich dadurch bessere Vertiefungsmöglichkeiten bei Prüfungsvorbereitungen ergeben. Es muss an dieser Stelle jedoch erwähnt werden, dass hierbei die subjektive Einschätzung der Bereitschaft der Lehrenden erfasst wurde und diese nicht mit einer objektiven Erfassung des Verhaltens gleichgesetzt werden kann. Nichtsdestotrotz geben die gestiegenen Nutzungszahlen der beiden LMS Stud.IP und ILIAS einen Hinweis darauf, dass über den betrachteten Zeitraum auch verstärkt auf multimediale Elemente in der Lehre zurückgegriffen wurde. Die Nutzungszahlen der LMS Stud.IP und ILIAS stiegen in der Zeit von 2012 bis 2018 kontinuierlich an und die Anzahl der Lehrenden, die nie ein LMS nutzten, reduzierte sich deutlich über den gesamten Erhebungszeitraum.

Seit der Gründung des LLZ gibt es ein vergrößertes Angebot an Weiterbildungsmöglichkeiten, die vermutlich ebenfalls zu einer stärkeren Nutzung beigetragen haben. Verschiedene Studien konnten zeigen, dass eine Verbesserung des technischen Supports und/oder die Einführung von LMS-Trainingsveranstaltungen eine stärkere Nutzung der Angebote nach sich zieht (Al-Alwani & Soomro, 2009; Coleman & Mtshazi, 2017). Ebenso veränderten sich die LMS Stud.IP und ILIAS in dieser Zeit bezüglich einer erhöhten Nutzerfreundlichkeit und Bedienbarkeit. Verschiedene Autoren (Alharbi, & Drew, 2014; Zanjani, Edwards, Nykvist, & Geva, 2017) konnten nahelegen, dass eine einfachere Handhabung der LMS auch zu einer

erhöhten Nutzung dieser Systeme durch die Lehrenden führt. Die Kombination aus verbessertem Support sowie besserer Bedienbarkeit der LMS haben damit möglicherweise ebenfalls zu einer erhöhten Nachfrage beigetragen. Ebenso zeigte sich, dass bei technischen Problemen die Akzeptanz von E-Learning-Maßnahmen negativ beeinflusst wird (vgl. Hunger et al. 2000; Kowalski 2004).

In anderen Studien konnten aber auch weitere Faktoren berichtet werden, die die Entscheidung der Lehrenden beeinflussen, ein LMS in der Lehre zu nutzen; das betrifft z. B. individuelles Interesse, Zeitmangel oder wahrgenommener Nutzen (Coleman & Mtshazi, 2017; Lawrence & Lentle-Keenan, 2013). Wir untersuchten deshalb mit Hilfe einer Korrelationsanalyse die Faktoren, die mit der Nutzung eines LMS an der MLU zusammenhängen. Hervorzuheben ist hierbei die negative Beziehung zwischen der LMS-Nutzung und der Einschätzung der eigenen technischen Kompetenz: je schlechter die technische Kompetenz eingeschätzt wurde, desto weniger wurden die LMS genutzt. Diese Befunde decken sich mit Ergebnissen anderer Studien zur LMS-Nutzung an anderen Hochschulen, wobei die Ursächlichkeit für diesen Zusammenhang unklar ist, d. h., ob wahrgenommene Defizite in der wahrgenommenen Kompetenz zur verringerten Nutzung oder umgekehrt eine verringerte Nutzung der LMS auch damit verbunden ist, dass Lehrende den Umgang mit LMS weniger üben können und somit geringere Möglichkeiten zum Kompetenzerwerb vorhanden sind (Asiri, Mahmud, Abu-Bakar, & Ayub, 2012; Buabeng-Andoh, 2012; Coleman & Mtshazi, 2017; Panda & Mishra, 2007). Lee, Hsiao und Purnomo (2014) konnten in einer elaborierten Studie mit Strukturgleichungsmodellen einen gerichteten Zusammenhang zeigen, der nahelegt, dass die eigene Computer-Selbstwirksamkeit die wahrgenommene Benutzerfreundlichkeit eines Systems beeinflusste und diese wiederum beeinflusste die Absicht, das System zu nutzen. Vor diesem Hintergrund ist es daher ratsam, die Lehrenden in ihren technischen Kompetenzen zu stärken und beispielsweise anzuregen, dass neu eingestellte Lehrende der Universität an den vielfältigen Weiterbildungen teilnehmen, die sowohl die didaktischen als auch die IT-bezogenen Voraussetzungen bei E-Learning umfassen. Die Sorge, dass die Kommunikation mit den Studierenden abnehmen könnte, welche ebenso negativ mit der LMS-Nutzung korrelierte, sollte zudem in diesen Weiterbildungen thematisiert werden, obwohl sich diese Sorgen mit der Zeit verringerten. Auch scheint sich das Wissen um die Vorteile digital gestützter Lehre positiv auf die LMS-Nutzung auszuwirken. Weiterbildungsszenarien für Lehrende sollten daher auch Möglichkeiten aufzeigen, wie der Leistungsstand der Studierenden eingeschätzt, Prüfungen vorbereitet und damit der Studienerfolg erhöht werden können.

Doch nicht nur technisches und didaktisches Vorwissen über E-Learning beeinflusst die Nutzung, sondern auch die übergeordnete Lehranschauung und eventuelle Voreingenommenheit gegenüber neuen Technologien, wie eine Studie von Lawrence und Lentle-Keenan (2013) zeigte. So konnte gezeigt werden, dass Lehrende, die zwischen den technischen Tools und ihrer Lehranschauung eine Übereinstimmung sehen, diese Technologien mit mehr Begeisterung nutzen als Lehrende, die darin keine pädagogische Notwendigkeit sehen

(Lawrence & Lentle-Keenan, 2013). Der Abbau von negativen Vorurteilen gegenüber Multimedia in der Lehre sollte demnach erklärtes Ziel sein, da Fehltritte und –einschätzungen sich auf die Nutzung digitaler Lehr- und Lernmethoden auswirken können.

In der aktuellen Studie wurde u.a. untersucht, wie Vorurteile und Haltungen der Lehrenden mit der Wahrnehmung des Nutzens, der Schwierigkeiten und der Rahmenbedingungen von multimedialer Lehre zusammenhängen. Die Befunde zeigten, dass die Akzeptanz neuer Technologien in der Lehre über die vier Erhebungszeitpunkte gestiegen ist (siehe **Tabelle 4, 7**). Allerdings schätzten Lehrende, die E-Learning als nicht vorteilhaft ansahen ein, dass die Kommunikation mit den Studierenden darunter leidet, der Studienerfolg dadurch nicht gesteigert und die eigenen didaktischen Möglichkeiten nicht erweitert werden können. Kontrollierten Studien der Begleitforschung zum digitalen Lernen, die Erfolg und Voraussetzungen für eine erfolgreiche Anwendung multimedialer Lehr- und Lernmethoden zum Gegenstand haben, kommt demzufolge auch weiterhin eine wichtige Rolle zu, um Fehltritte entgegenzuwirken und durch entsprechende Maßnahmen bei der Gestaltung der Rahmenbedingungen auf universitärer Ebene umzusetzen.

Die aktuelle Studie weist Limitationen auf, die die Aussagefähigkeit der Daten einschränken. So wurden verschiedene Items in den Befragungen zwischen den Jahresscheiben verändert, sodass eine übergreifende Analyse nicht über alle Items in den Fragebögen stattfinden konnte. Diese Veränderungen an den Items wurden vorgenommen, da die Befragungen primär mit dem Ziel der Evaluation des Serviceangebotes des LLZ stattfanden und eine wissenschaftliche Analyse der Lehrendeneinschätzung eher sekundäres Ziel war. Die zusammengefassten Befunde weisen jedoch trotz dieser Einschränkungen auf eine angemessene Analysierbarkeit der Daten hin. Die vorliegenden Likertdaten der Lehrenden wurden mit statistischen Verfahren analysiert, die metrische Abstände zwischen den Antwortkategorien der Items voraussetzen. Auch wenn diese Annahme für die vorliegenden Annahmen intensiv debattiert werden kann, erlauben die vorliegenden Analysen jedoch erste vorläufige Aussagen zu möglichen Tendenzen in den Urteilen und Überlegungen sowie Reflektionen der Lehrenden, die Grundlage für weitergehende systematische Befragungen der Lehrenden mit gezielten empirischen Fragestellungen und für stärker kontrollierte Untersuchungssettings experimenteller Natur sein können (Carifio & Perla, 2007; Glass et al., 1972). Dabei muss auch die Konstruktion der Items optimiert werden, sodass deren Aussagefähigkeit nicht durch Kodierungsprobleme oder Akzeptanzprobleme der Lehrenden beeinträchtigt wird. Der aktuellen Studie kommt somit eher explorative Bedeutung bei, deren Befunde jedoch wichtige Aussagen zur Entwicklung spezifischer Überlegungen und Anforderungen von Lehrenden bezüglich der Notwendigkeit von multimedialer Lehre und deren notwendige Umsetzungsbedingungen über einen längeren Zeitraum der universitären Entwicklung in den letzten Jahren aufzeigen.

Wie die Corona Pandemie in nie zuvor aufgezeigter Deutlichkeit offenbarte, können weite Bereiche der Hochschullehre digital erfolgen oder zumindest von den Möglichkeiten digitalen Lehr- und Lernmethoden profitieren, z. B. in Bezug auf zeitliche und räumliche Flexibilität und auch in Bezug auf die Erweiterung didaktischer Ansätze. Auch wenn derzeit die

didaktischen Potentiale noch nicht ausreichend genutzt werden konnten, bleibt es eine Daueraufgabe, dass diese Systeme von den Lehrenden angenommen und eingesetzt werden. Unterstützt werden kann das vor allem durch Weiterbildungen mit dem Ziel, die Selbstwirksamkeit zu erhöhen, technische Kompetenzen zu entwickeln und mögliche Vorurteile z.B. durch Informationen, praktische Übungen und Selbstreflexionen, etc. zu entkräften.

## Literatur

- Al-Alwani, A. E. S., & Soomro, S. (2009). Integration of information technology in science education at Yanbu, Kingdom of Saudi Arabia. In: Lazinica, A. & Tavares, C., *Technology: Education and Development*, 151-162.
- Alharbi, S., & Drew, S. (2014). Using the technology acceptance model in understanding academics' behavioural intention to use learning management systems. *International Journal of Advanced Computer Science and Applications*, 5(1), 143-155.
- Asiri, M. S., Mahmud, R., Abu-Bakar, K., & Ayub, A. F. (2012). Factors influencing the use of learning management system in Saudi Arabian higher education: A theoretical framework. *Higher Education Studies*, 2(2), 125-137.
- Bortz, J., & Döring, N. (2014). *Forschungsmethoden und Evaluation: in den Sozial- und Humanwissenschaften*. Heidelberg: Springer-Verlag GmbH.
- Buabeng-Andoh, C. (2012). Factors influencing teachers adoption and integration of information and communication technology into teaching: A review of the literature. *International Journal of Education and Development using ICT*, 8(1).
- Carifio, J., & Perla, R. J., (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences*, 3(3), 106-116.
- Coleman, E., & Mtshazi, S. (2017). Factors affecting the use and non-use of Learning Management Systems (LMS) by academic staff. *South African Computer Journal*, 29(3), 31-63.
- Glass, G. V., Peckham, P. D., & Sanders, J.E. (1972). Consequences of failure to meet assumptions underlying the analyses of variance and covariance. *Review of Educational Research*, 42, 237-288.
- Hunger, A., Werner, S., Breges, A., & Schwarz, F. (2000). Einführung neuer Medien in die Hochschulausbildung: sozialkognitive Aspekte der SoftwareEvaluation. In: F. Schermann (Ed.): *Campus 2000 – Lernen in neuen Organisationsformen* (pp. 347–354). Münster: Waxmann.
- Kowalski, S., & Kück, U. (2004). E-Learning – Einsatz internetgestützter Lernmodule. In: T. Brinker, & U. Rössler (Eds.), *Hochschuldidaktik an Fachhochschulen – Neue Ansätze in der Lehre aus den Fachhochschulen des Landes Nordrhein-Westfalen* (pp. 149–156). Bielefeld: wbv.
- Lawrence, B., & Lentle-Keenan, S. (2013). Teaching beliefs and practice, institutional context, and the uptake of Web-based technology. *Distance Education*, 34(1), 4-20.

- Lee, Y. H., Hsiao, C., & Purnomo, S. H. (2014). An empirical examination of individual and system characteristics on enhancing e-learning acceptance. *Australasian Journal of Educational Technology*, 30(5).
- Martin, F., Ritzhaupt, A., Kumar, S., & Budhrani, K. (2019). Award-winning faculty online teaching practices: Course design, assessment and evaluation, and facilitation. *The Internet and Higher Education*, 42, 34-43.
- Panda, S., & Mishra, S. (2007). E-Learning in a Mega Open University: Faculty attitude, barriers and motivators. *Educational Media International*, 44(4), 323-338.
- Rost, J. (1996). *Lehrbuch Testtheorie, Testkonstruktion*. Huber, Bern u. a. 1996, ISBN 3-456-82480-7.
- Wingo, N. P., Ivankova, N. V., & Moss, J. A. (2017). Faculty perceptions about teaching online: Exploring the literature using the technology acceptance model as an organizing framework. *Online Learning*, 21(1), 15-35.
- Zanjani, N., Edwards, S. L., Nykvist, S., & Geva, S. (2017). The important elements of LMS design that affect user engagement with e-learning tools within LMSs in the higher education sector. *Australasian Journal of Educational Technology*, 33(1).

**Torsten Schubert** ist Professor für Allgemeine Psychologie an der Martin-Luther-Universität Halle-Wittenberg und Geschäftsführender Direktor des Zentrums für multimediales Lehren und Lernen. Sein fachliches Spezialgebiet sind Mechanismen der Gedächtnis- und Handlungskontrolle sowie der Aufmerksamkeit und Multitasking. Besonderes Forschungsinteresse kommt digitalen Lernen und digitalen Medien im Kontext von Bildungs- und Lernprozessen zu.

[torsten.schubert@psych.uni-halle.de](mailto:torsten.schubert@psych.uni-halle.de)

**Michael Gerth** promovierte in Politologie und ist seit 2012 Geschäftsführer des Zentrums für multimediales Lehren und Lernen (LLZ) an der Martin-Luther-Universität Halle-Wittenberg.

[michael.gerth@llz.uni-halle.de](mailto:michael.gerth@llz.uni-halle.de)

**Teresa Schurer** ist wissenschaftliche Mitarbeiterin im Zentrum für Multimediales Lehren und Lernen (Forschungsbereich) und befasst sich in ihrer Promotion mit Gedankenabschweifungen beim digitalen Lesen.

[teresa.schurer@llz.uni-halle.de](mailto:teresa.schurer@llz.uni-halle.de)

**Anne Schrimpf** war 2020 Leiterin des Forschungsbereichs des LLZs an der Martin-Luther-Universität Halle-Wittenberg. Derzeit ist sie in der Versorgungsforschung an der Universität Leipzig tätig.

[anne.schrimpf@medizin.uni-leipzig.de](mailto:anne.schrimpf@medizin.uni-leipzig.de)

# 2

Lernvideos, Bilderbuch-Apps und Serious Games  
als digitale Medien in Unterrichtsszenarien

# Unterstützung von Grundschullehramtsstudierenden bei der Integration von naturwissenschaftlichen Fachwissen aus Online-Videos in die Modellierungsprozesse durch „Multimedia-Tools“

*María Esther Tellez-Acosta<sup>1</sup>, Andres Acher<sup>1</sup> und Leon Kusitzky<sup>1</sup>*

*<sup>1</sup> Martin-Luther-Universität Halle-Wittenberg, Institut für Schulpädagogik und Grundschuld-  
daktik*

Am Anfang ihres Studiums suchen Grundschullehramtsstudierende häufig nach Ressourcen, die das Lernen von naturwissenschaftlichem Fachwissen unterstützen. Diese Ressourcen sind zunehmend Online-Videos mit naturwissenschaftlichem Fachwissen. Während diese Videos Studierenden dabei helfen können, das Fachwissen besser zu verstehen, helfen sie nicht automatisch dabei zu verstehen, wie dieses Fachwissen in die naturwissenschaftlichen Modellierungsprozesse integriert werden könnte, die sie unterrichten werden. Wir haben eine Reihe von „Multimedia-Tools“ entworfen, implementiert und untersucht, um Studierende dabei zu unterstützen, Online-Videos für die Entwicklung ihrer eigenen erklärenden Modelle zu nutzen und um zu lernen, wie man damit modellbasierte Untersuchungen plant. Im Rahmen eines Moduls des Sachunterrichts des Studiengangs Grundschullehramt, welches ortsbezogene und online basierte Lerndesigns kombiniert, benutzten wir die drei analytischen Kategorien des „Professional Vision“ Frameworks um zu untersuchen, wie Studierende kollaborativ in einer Online-Umgebung lernen. Die Ergebnisse veranschaulichen eine schrittweise Aushandlung über die Bedeutung naturwissenschaftlichen Fachwissens, die mit den Modellierungsprozessen integriert waren, und führten dazu, dass Gruppen ihre eigenen Erklärungsmodelle entwickelten. Wir diskutieren diese Ergebnisse mit dem Schwerpunkt darauf, wie Studierende die Integration von naturwissenschaftlichem Fachwissen/ Modellierungsverständnissen verbessern und wie sich dies auf ihren zukünftigen Lehrberuf auswirkt.

## Einleitung

Die jüngsten Reformen im naturwissenschaftlichen Unterricht (KMK, 2004; MEN, 2006; NRC, 2012) unterstreichen die Bedeutung der Einbeziehung und Unterstützung von Schülerinnen und Schüler im Hinblick auf ein integratives Lernen von naturwissenschaftlichem Fachwissen und naturwissenschaftlichen Praktiken. Diese Reformen zeigen auch auf, dass naturwissenschaftliches Lehren und Lernen ein sozial verorteter Prozess ist (Wickman, 2012, S. 486). Wenn Grundschullehramtsstudierende ihr Studium beginnen, müssen sie auf ihr Fachwissen und eingebundene Praktiken vertrauen (Zemal-Saul, 2018, S. 235). Sie suchen häufig nach Ressourcen, um mehr über das naturwissenschaftliche Fachwissen zu erfahren, die sie möglicherweise zum Unterrichten benötigen. Eine zunehmend verbreitete Ressource für naturwissenschaftliches Fachwissen sind Online-Videos (Baltaci-Goktalay & Ozdilek, 2010, S. 4740), aus Open-Access-Quellen wie *Youtube.com*, *Vimeo.com* oder wissenschaftsspezifischen Websites wie *kids-and-science.com* oder *bmu.de*. Während diese Online-Videos das Verständnis von Studierenden für naturwissenschaftliches Fachwissen unterstützen können, unterstützen die Online-Videos nicht automatisch ihr Verständnis darüber, wie dieses Fachwissen in wissenschaftliche Praktiken, wie die naturwissenschaftliche Modellierung, integriert werden können. Dies stellt Studierende vor erhebliche Herausforderungen, die bei der Planung einer modellbasierten Untersuchung (MBU) Fachkompetenzen für diese Integration benötigen. **Erstens**, selbst wenn Studierende das naturwissenschaftliche Fachwissen aus Videoquellen verstehen, ist es weniger wahrscheinlich, dass sie dieses Wissen gezielt „sehen“, um Naturphänomene zu modellieren. Zum Beispiel können Studierende YouTube-Videos ansehen, um zu verstehen, was Pflanzen zum Überleben benötigen oder wie Pflanzen wachsen. Diese Videos können Ideen über den Photosynthesevorgang, die Pflanzenernährung oder die verschiedenen Bedingungen enthüllen, die die Pflanze zum Wachstum und zur Entwicklung benötigt. Dies sind jedoch isolierte Konzepte und Beschreibungen, die nicht mit den Prozessen zur Modellierung bestimmter Phänomene zusammenhängen, z. B. wie es möglich ist, dass einige Pflanzen unter Wasser leben können und andere nicht. **Zweitens** verfügen Studierende nicht über die Fachkompetenz, um einen modellbasierten Unterricht zu planen. Es fällt ihnen schwer, diese Modelle als Vermittler zwischen der Fähigkeit der Schülerinnen und Schüler, Naturphänomene zu interpretieren und der vielfältigen und umfassenden Art und Weise, wie sie ihr Verständnis organisieren, zu verwenden (Acher, Arcà, & Sanmartí, 2007, S. 399). Sie könnten lernen, dass Modelle Werkzeuge sind, die zur Theoriebildung und zur Gestaltung von Untersuchungen verwendet werden (Forbes, Lange-Schubert, Böschl, & Vo, 2019, S. 259). Dieses Lernen impliziert eine aktive Beteiligung an soziokulturellen Praxen, um einen charakteristischen Diskurs der professionellen Fachkompetenz zu entwickeln, zu verfeinern und Entscheidungen im interaktiven Austausch zu treffen (Goodwin, 1994, S. 628).

Eine Möglichkeit auf diese Herausforderungen zu reagieren, besteht darin, das Erlernen dieses/dieser Integrationsfachwissens/-modelle bei der Planung einer MBU zu unterstützen, durch gezielt entwickelte Multimedia-Tools. Diese Tools wirken als Katalysator für die

sozial lokalisierten Praxen, an denen die Lernenden beteiligt sind, indem sie IKT-Komponenten einbeziehen, die die Interaktion, Kommunikation, Kooperation und Kreation im Lernprozess beeinflussen (Butrime & Zuzeviciute, 2013, S. 2). Wir untersuchen, wie Studierende durch die Praxis lernen, naturwissenschaftliches Fachwissen in die Modellierungsprozesse zu integrieren, um MBU in einer Online-Umgebung zu planen, die mit modellbasierten Multimedia-Tools unterstützt wird. Wir verwenden die drei analytischen Kategorien des „Professional Vision“ Frameworks (Hervorheben, Codieren und materielle Darstellungen) aus seinen soziokulturellen Grundlagen (Goodwin, 1994, S. 606). Wir konzentrieren uns auf die soziale Interaktion, durch die Studierende eine gemeinsame professionelle Fachkompetenz entwickeln, indem sie auf Online-Videos als Ressourcen für naturwissenschaftliches Fachwissen zurückgreifen und mit den Multimedia-Tools arbeiten. Wir beschäftigen uns mit der folgenden Frage: *Wie lernen Grundschullehramtsstudierende, naturwissenschaftliches Fachwissen aus Online-Videos in ihre eigenen Erklärungsmodelle zu integrieren, um modellbasierte Untersuchungen zu planen, während sie mit Multimedia-Tools unterstützt werden?*

### **Konzeptioneller und theoretischer Rahmen**

Neueste Forschungen im naturwissenschaftlichen Unterricht zeigen, wie wichtig Praxen für Studierende sind, die am Anfang ihres Lehramtsstudiums sind, um für ihren Lehrberuf zu lernen (McDonald et al., 2014; Seidel, Stürmer, Schäfer, & Jahn, 2015). Außerdem betonen einige Studien die Vorteile web-basierter Tools, die in Kombination mit persönlichen Aktivitäten dieses professionelle Lernen unterstützen (Banday, 2013; Butrime & Zuzeviciute, 2013). Studierende lernen naturwissenschaftliches Fachwissen aus Online-Videos in die Modellierungsprozesse zu integrieren und professionelle Fachkompetenz bei der Planung einer MBU zu erweitern, indem sie sich an soziokulturellen Praxen beteiligen, die mit entwickelten modellbasierten Multimedia-Tools unterstützt werden. Zwei konzeptionelle Elemente leiten unsere Überlegungen: Praxen als soziokulturelle Form des Lernens für ihren Lehrberuf und Multimedia-Tools auf der Grundlage wissenschaftlicher Modellierung.

### **Praxen als soziokulturelle Form des Lernens**

Praxisorientierte Ansätze in der naturwissenschaftlichen Lehrerbildung zeigen auf, dass Studierende durch Praxen die Möglichkeit haben, „authentische“ Aktivitäten in der Nähe ihres Lehrberufs und in Zusammenarbeit mit Gleichaltrigen durchzuführen (Davis, 2020, S. 97). Die *Praxis* ist eine Konstruktion, die Studierende während ihres beruflichen Lernens selbst und in einem soziokulturellen Austausch hergestellt haben (Grossman, Schneider, & Pupik, 2018, S. 5). Wir sind der Ansicht, dass Studierende, die mit anderen interagieren, während sie sich an gemeinsamen Praxen beteiligen, lernen, das naturwissenschaftliche Fachwissen zielgerichtet zu „sehen“, in diesem Fall um Modelle mit Erklärungspotential zu entwickeln und Entscheidungen über die Planung einer MBU zu treffen. So führen Studierende mittels Übungen gemeinsam Aktivitäten durch, die sich dem Beruf des naturwissenschaftlichen Lehrers annähern und treffen Entscheidungen zum Erreichen gemeinsamer Ziele. Sie haben die

Möglichkeit, gemeinsam über Bedeutungen auszuhandeln, Erfahrungen, Ideen und Verständnis des anderen aufzubauen und miteinander zu entscheiden.

### **Multimedia-Tools basierend auf naturwissenschaftlicher Modellierung**

In naturwissenschaftlichen Lehramtsstudien wird betont, dass „Tools“ das Lernen von Lehramtsstudierenden durch Praxen positiv beeinflussen (Davis, 2020, S.98). Online-Umgebungen und web-basierte Tools verbessern die elementare Konstruktion von Bedeutungen und die professionelle Kommunikation (Ekici, 2017; Hunt, 2015). Wir sind der Ansicht, dass modellbasierte Multimedia-Tools das professionelle Lernen von Grundschullehramtsstudierenden unterstützen, wie das Fachwissen in erklärende Modelle integriert werden kann, um MBU zu planen. Erstens konzentrieren sich Multimedia-Tools darauf, was Studierende in Zusammenarbeit mit Gleichaltrigen tun (vgl. Cakir, 2013, S. 245), wie sie sich aktiv an Lernaktivitäten beteiligen, diskutieren, einen Konsens erzielen und Produkte entwickeln (z. B. Modelle). Zweitens bieten diese Tools die Möglichkeit des „flexiblen Lernens“ (Garner & Rouse, 2016, S. 31) und reduzieren Kommunikationsbarrieren, wie zeitliche und örtliche Abhängigkeiten, wodurch eine Möglichkeit zur Zusammenarbeit im Online-Fernunterricht entsteht (Weller, 2013, S. 40). Drittens basieren diese Tools auf naturwissenschaftlichen Modellierungsprozessen und implizieren, dass Lehrkräfte zusammenarbeiten, um gemeinsame Ziele in Bezug auf die MBU zu erreichen.

### **Forschungsziel**

Mit dem Schwerpunkt auf den Praxen von Grundschullehramtsstudierenden beim Lernen, MBU in einem onlinebasierten Lehrdesign zu planen, möchten wir charakterisieren, wie sie lernen, naturwissenschaftliches Fachwissen aus Online-Videos in ihre eigenen Erklärungsmodelle zu integrieren, während sie mit Multimedia-Tools unterstützt werden.

## **Methode**

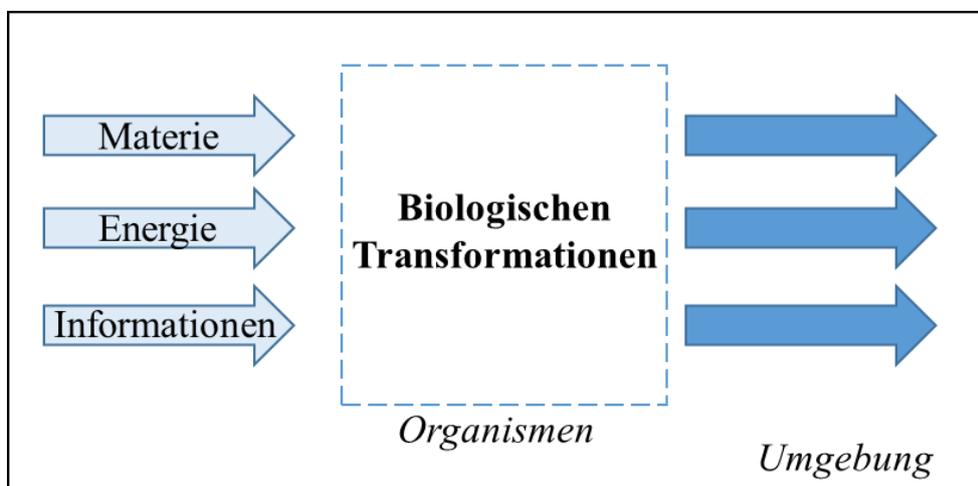
### **Kontext unseres Lehrdesigns**

Wir haben spezifische Lehrdesigns vorbereitet, um Grundschullehramtsstudierende in Bezug auf die Integration von naturwissenschaftlichem Fachwissen aus Online-Videos in ihre eigenen Erklärungsmodelle mit Multimedia-Tools zu unterstützen. Ein solches Lehrdesign wurde als praxisorientierte Planung einer MBU konzipiert. Diese sollte von Studierenden in folgenden Schritten erfolgen: a) Kreieren eines natürlichen phänomenologischen Kontextes; b) Entwickeln einer Leitfrage mit Teilfragen, um diesen Kontext zu untersuchen; c) Vertiefende Auseinandersetzung mit den Kernideen der Naturwissenschaften, um für die Fragen fachwissenschaftlich belastbare Antworten zu finden und d) Artikulieren dieser Kernideen der Naturwissenschaften in Modellen.

## Design von Multimedia-Tools

Wir haben drei Multimedia-Tools mit unterschiedlichen web-basierten Ressourcen (z. B. Foren, Blogs, Datenerfassung, Etherpads usw.) entwickelt. Mit diesen Tools haben wir Gruppen von Studierenden in bestimmte Aktivitäten ihres beruflichen Lernens eingebunden (z. B. Identifizieren was sich in einem Phänomen verändert, Formulierung einer Leitfrage). Vier Komponenten unterstützten die Beteiligung und Kommunikation der Gruppen und die spezifische Arbeit, naturwissenschaftliche Aspekte des Fachwissens (Biologie) in ihre eigenen Erklärungsmodelle bei der Planung einer MBU zu integrieren.

**1- Eine epistemische Komponente: Transformationsboxen (TBs).** Transformationsboxen (Acher & Arcà, 2020), die von einer fachbiologischen Perspektive die Konstruktion von Modellen unterstützt. Es handelt sich um Boxen, mit denen man verschiedene biologische Strukturen darstellen und damit den Austausch von Materie, Energie und Informationen konkretisieren kann, um verschiedene biologischen Transformationen zu erklären (**Abbildung 1**).



**Abbildung 1.** Einfache Darstellungsform von TBs: Ein Organismus, der Materie, Energie und Informationen aus seiner Umgebung transformiert (Acher & Arcà, 2020)

**2- Eine fachspezifische Komponente: Online-Videos als Ressource für biologisches Fachwissen.** Fachspezifische Videos wurden aus zuverlässigen Open Source-Quellen ausgewählt (von Expert\*innen entwickelt und für die vielfältige Benutzergemeinschaft validiert), um sicherzustellen, dass die Informationen relevant und genau sind. Wir haben erwartet, dass die Videos die Studierenden dabei unterstützen, sich mit den Kernideen der Biologie vertraut zu machen und gleichzeitig die Verwendung von TBs zu unterstützen, auf deren Grundlage sie ihre eigene Darstellung erstellen.

**3- Eine didaktische Komponente: a) Tools, die sich auf bestimmte Aspekte der modellbasierten Untersuchung konzentrieren.** Studierende führen Schritt für Schritt bestimmte Aktivitäten durch, die von bestimmten Tools unterstützt werden, um schrittweise ihre modellbasierten Lernergebnisse zu erreichen. *Erstes Multimedia-Tool:* „Veränderungen im Phänomen

wahrnehmen und dabei Aspekte von Stability entdecken“. *Zweites Multimedia-Tool*: „Wahrnehmung mit Transformationsboxen (TBs) um Stability und Changes in dem Naturphänomen zu verstehen und zu ordnen“. *Drittes Multimedia-Tool*: „Unterstützung der Entwicklung modellbasierter Untersuchungen“.

**b) Interaktive web-basierte Ressourcen**, die die Beteiligung der Studierenden unterstützten, die gemeinsamen Lernziele der Modellierung zu erreichen. Wir haben verwendet: Foren um die Studierenden, durch „Diskussionsthreads“ an Aktivitäten zu beteiligen; Etherpads zur Unterstützung der Teilnehmer\*innen beim Austausch verschiedener Ideen und bei gemeinsamen Entscheidungen; Datenerfassungstabellen, die ihnen helfen, ihre Auswahl zusammenzufassen und unterschiedliche Ideen gegenüberzustellen; und Blogs, um die Kommunikation und Zusammenarbeit zwischen den Teilnehmer\*innen aufrechtzuerhalten. Wir haben erwartet, dass diese web-basierten Ressourcen die Studierenden dabei unterstützen, professionelle Fachkompetenz bei der Integration des naturwissenschaftlichen Fachwissens aus Videoressourcen in Erklärungsmodelle für die Planung einer MBU zu erreichen.

### **Die beteiligten Studierenden**

An der Studie haben 48 Studierende aus dem Modul des Sachunterrichts des Studiengangs Grundschullehramt aus dem zweiten Fachsemester (95 % weiblich; 5 % männlich) teilgenommen. Sie haben selbst zwölf Gruppen mit jeweils vier bis sechs Studierenden (männlich und weiblich gemischt) gebildet. Alle Gruppen haben dieses Lehrdesign und die Multimedia-Tools ausprobiert. Wir haben eine dieser Gruppen als Beispiel ausgewählt. Wir haben uns für die Gruppe entschieden, die sich auf die Entwicklung der biologischen Kernidee Informationsverarbeitung bei Organismen fokussierte, weil diese Gruppe verschiedene Aspekte dieser Idee aus den Videos für die Entwicklung des Modells ausgehandelt hat, während sie durch die Multimedia-Tools an den gemeinsamen Praxen beteiligt waren. Diese Kernidee der folgenden Frage umfasst: *Wie erkennen, verarbeiten und nutzen Organismen Informationen über die Umwelt?* Um eine MBU über diese Kernidee zu planen, haben die Studierenden einen phänomenologischen Kontext gestaltet, in dem manche Schülerinnen und Schüler auf dem Schulhof an einem Herbsttag Gänsehaut bekommen und manche nicht. Den Ausgangspunkt dieses Kontextes zu hinterfragen war eine Leitfrage -*Wieso hat nur Maria eine Gänsehaut gekriegt, wenn wir alle draußen im Schulhof waren?* Sie haben mehrere Teilfragen wie z. B. *Welche Reize können Gänsehaut auslösen?* Und *Welche Wege nehmen die Signale im Körper?* entwickelt, um ein Modell schrittweise zu konstruieren.

Die Multimedia-Tools mit den fachspezifischen Videos wurden hier genutzt, um die Konstruktion dieses Modells anhand der Transformationsboxen zu unterstützen. Zwei Open Access Videos, die für die Studierenden bekannt sind, wurden benutzt (Giesecke & Schork, 2014, 2015). In den Videos wurden die wichtigsten Begriffe der Haut (z. B. Funktionen wie Schutz vor Austrocknung/Wärmeverlust) und des Nervensystems (z. B. Nerv, Neuron, Nervenzelle und Synapse) erklärt. Die Studierenden haben in jedem Multimedia-Tool nacheinander an verschiedenen Aufgaben gearbeitet. Sie haben synchron und asynchron gearbeitet, um an den von uns entworfenen Aktivitäten teilzunehmen und die Eingabeaufforderungen

zu diskutieren, die wir als Unterstützung für die Tools bereitgestellt hatten. Aus diesen Aufgaben haben wir uns unsere Daten genommen. Die Gruppe hat an sieben von acht vorgeschlagenen Aktivitäten teilgenommen. Wir haben drei Momente dieser Aktivitäten ausgewählt, weil diese die Zusammenarbeit der Gruppe mit den unterstützenden, interaktiven web-basierten Ressourcen, mit denen wir die Tools entworfen haben, repräsentierten.

## Analyse

Wir haben die drei diskursiven Praktiken der „Professional Vision“ (Goodwin, 1994, S. 606) als analytische Kategorien verwendet, um das Lernen der Gruppen in den drei ausgewählten Momenten zu interpretieren: *Hervorheben*, *Codieren* und *materielle Darstellungen*. Das *Hervorheben* bezieht sich darauf, was für die Gruppe über die Kernideen der Biologie (aus den Online-Videos) in den Modellierungsprozessen relevant war. Das *Codieren* bezieht sich darauf, wie sie Bedeutungen ausgehandelt oder interpretiert haben und was sie bei der Entwicklung von Erklärungsmodellen über TBs hervorheben. *Materielle Darstellungen* beziehen sich auf die Erklärungsmodelle biologischer Phänomene, die die Gruppe konstruiert oder verwendet haben.

In der ersten Analyserunde haben wir Schlüsselereignisse identifiziert (Kelly & Chen, 1999, S. 890). Diese entsprechen Aktivitätsmomenten, in denen die Gruppe Aushandlungen über die Integration der biologischen Kernidee in Erklärungsmodelle über TBs zur Planung der MBU führt. In der zweiten Runde haben wir die diskursive Arbeit und Konstruktion in einem soziokulturellen Austausch detailliert analysiert. Zeile für Zeile interpretierten wir die Aushandlungsprozesse (Hervorheben, Codieren und Konstruieren/Verwenden von materiellen Darstellungen) in jedem Schritt der Planungspraxis rückwirkend, da sie an den web-basierten Ressourcen teilgenommen haben. Im Folgenden veranschaulichen wir die Beteiligung der Gruppe bei der Aushandlung von Bedeutungen rund um die Informationsverarbeitung der biologischen Kernidee im Kontext des spezifischen Phänomens „Gänsehaut“. Diese Aushandlung war gekennzeichnet durch die Integration der Aspekte dieser biologischen Kernidee in die Modellierungsprozesse: Konstruieren, Beurteilen, Überarbeiten und Anwenden von Modellen als Erklärungsinstrumente.

## Ergebnisse

In der folgenden **Tabelle 1** stellen wir die Aushandlungsprozesse von Studierenden während der drei ausgewählten Aktivitätsmomente ihrer Arbeit mit den Multimedia-Tools dar. Diese Momente entsprechen: a) Forum für Identifikation und Charakterisierung von Veränderungen in Naturphänomenen; b) Datenerfassungstabellen für das Aufstellen von Stability/Change (gegensätzliche Paare) und Einschätzen ihrer Relevanz für das Verstehen des Phänomens in Bezug auf die Kernidee; und c) Forum für Konstruktion von erklärenden Modellen mit Hilfe von TBs. Wir haben einige Beschreibungen unserer Interpretation hinzugefügt, aus denen wir das Lernen der Gruppe als Entwicklung ihrer professionellen Fachkompetenz charakterisiert haben.

Tabelle 1. Beispiele für die Arbeit einer Gruppe von Studierenden mit Multimedia-Tools

Erste Multimedia-Tool		Beschreibung																		
Aushandlung der Bedeutung im Forum: „Identifikation und Charakterisierung von Veränderungen in Naturphänomenen“																				
		<p><b>Hervorheben</b> von "Schauer", "Zittern" oder "Gänsehaut im Gesicht" als weitere Prozesse, die durch extreme Bedingungen in Bezug auf die Energiedimension (z. B. Kälte) ausgelöst werden können. Emotionen im Zusammenhang mit dem „Gänsehaut“</p> <p><b>Codieren</b> der Beziehung zwischen der Eingabe und den intern-externen Prozessen.</p>																		
		<p><b>Materielle Darstellungen:</b> Während die Armhaut als TB hervorgehoben wird. Energiedimension der Changes (z. B. Wärmeisolation oder Körperwärme wird abgegeben)</p> <p><b>Codieren</b> der biologischen Umwandlung des "kalter Wind" in der Armhaut-TB in "Wärme" und in eine Energieform (die die Armhaut mit anderen Körperteilen verbinden könnte).</p>																		
<p><b>Abbildung 2. Materielle Darstellung Teilnehmerin 1</b>      <b>Abbildung 3. Materielle Darstellung Teilnehmerin 2</b></p>																				
<table border="1"> <thead> <tr> <th>Changes</th> <th>Stability</th> <th>Relevanz für das Verstehen des Phänomens in Bezug auf die Kernidee ↓</th> </tr> </thead> <tbody> <tr> <td>Energieniveau</td> <td>Nervenzelle</td> <td>Weiterleiten des Reizes</td> </tr> <tr> <td>Kontraktion der Haarbalgmuskeln</td> <td>Haarlänge bleibt gleich, keine Oberflächenveränderung an den Lippen und Hand/Fußflächen, Ohren</td> <td>Sichtbares Element des Vorganges, Entstehen einer Isolierungsschicht</td> </tr> <tr> <td>Wärmempfindung</td> <td>Innere Körpertemperatur bleibt gleich</td> <td>Reaktion auf den Umweltreiz</td> </tr> <tr> <td>Umweltreize (Wind/ Berührung)</td> <td>Körper an sich bleibt gleich bis zur Verarbeitung des Reizes</td> <td>Auslöser für Körperreaktion</td> </tr> <tr> <td>Emotionen/ Empfindungen</td> <td>(Wahrnehmung)/ Reizverarbeitung</td> <td>Auslöser für Gänsehaut</td> </tr> </tbody> </table>		Changes	Stability	Relevanz für das Verstehen des Phänomens in Bezug auf die Kernidee ↓	Energieniveau	Nervenzelle	Weiterleiten des Reizes	Kontraktion der Haarbalgmuskeln	Haarlänge bleibt gleich, keine Oberflächenveränderung an den Lippen und Hand/Fußflächen, Ohren	Sichtbares Element des Vorganges, Entstehen einer Isolierungsschicht	Wärmempfindung	Innere Körpertemperatur bleibt gleich	Reaktion auf den Umweltreiz	Umweltreize (Wind/ Berührung)	Körper an sich bleibt gleich bis zur Verarbeitung des Reizes	Auslöser für Körperreaktion	Emotionen/ Empfindungen	(Wahrnehmung)/ Reizverarbeitung	Auslöser für Gänsehaut	<p><b>Hervorheben</b> von Energie-, Struktur-, Material- und Informationsdimensionen der Changes.</p> <p><b>Codieren</b> z. B. Der Körper nimmt die Inputs (Informationen) von außen (z. B. Wind) durch die Haut wahr. Der Körper überträgt die Informationen durch die Nervenzellen und reagiert nach der Verarbeitung der Informationen (z. B. Oberflächenveränderung der Haut).</p>
Changes	Stability	Relevanz für das Verstehen des Phänomens in Bezug auf die Kernidee ↓																		
Energieniveau	Nervenzelle	Weiterleiten des Reizes																		
Kontraktion der Haarbalgmuskeln	Haarlänge bleibt gleich, keine Oberflächenveränderung an den Lippen und Hand/Fußflächen, Ohren	Sichtbares Element des Vorganges, Entstehen einer Isolierungsschicht																		
Wärmempfindung	Innere Körpertemperatur bleibt gleich	Reaktion auf den Umweltreiz																		
Umweltreize (Wind/ Berührung)	Körper an sich bleibt gleich bis zur Verarbeitung des Reizes	Auslöser für Körperreaktion																		
Emotionen/ Empfindungen	(Wahrnehmung)/ Reizverarbeitung	Auslöser für Gänsehaut																		

## Zweite Multimedia-Tool

Aushandlung der Bedeutung im Forum: "Konstruktion von erklärenden Modellen mit Hilfe von TBs"

Beschreibung

The screenshot shows a forum thread with two posts. The first post, by Franziska, discusses the concept of a 'Transformationsbox Nervenzelle' and includes an attachment 'Screenshot\_20190603-171943.jpg'. The second post, by Johannes, explains the biological process at a synapse: 'In der Synapse wird ein elektrisches Signal zu einem chemischen Signal (synaptischer Spalt) und anschließend wieder zu einem elektrischen Signal. D.h. die Information muss nach dem synaptischen Spalt anders aussehen. Ich baue das mal in die verbundenen Transformationsboxen ein.' It includes an attachment 'Note\_03.06.2019\_18\_21\_04.pdf'.

**Hervorheben** der Nervenzelle als TBs. Informationsdimension der Veränderungen. **Codieren** der „Synapse“ als biologische Transformation. Information → Elektrisches Signal → Chemisches Signal und dann wieder → Elektrisches Signal.

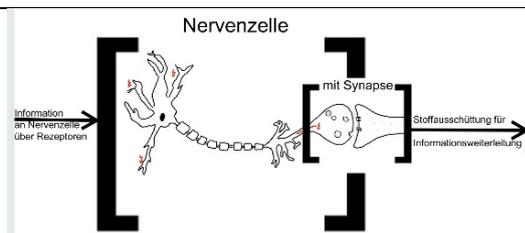


Abbildung 4. Konstruktion von Robots-Modell.

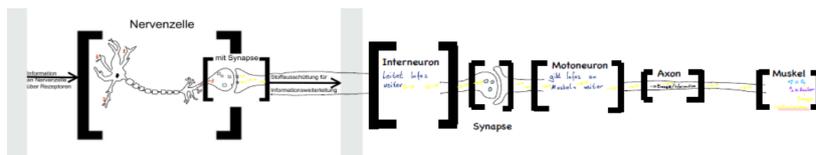


Abbildung 5. Weitere Erarbeitung des Modells.

### Materielle Darstellungen:

- a) Nervenzelle-TBs (Synapsen): Rezeptoren erhalten die Informationen, die dann weitergeleitet werden.
- b) Interneuron (Synapse): von der Nervenzelle zur motorischen Nervenzelle
- c) Motorische Nervenzelle (Axon): Signal von einer Zelle zur anderen übertragen.
- d) Muskel: wandelt Sauerstoff und Zucker in Energie um, die zur Kontraktion verwendet wird.

Im ersten Forum haben die Teilnehmer\*innen in einer synchronen Arbeit (am selben Tag 21.05 zur selben Zeit, von 11:42-13:17) ihr Verständnis darüber integriert, wie der Körper Temperaturunterschiede wahrnimmt (Inputs) und darauf reagiert (Outputs). Die Aspekte wurden zur Konstruktion von ersten Darstellungen von Veränderungen im Phänomen „Gänsehaut“ angewendet. Wie die erste Zeile des Forums zeigt, haben die Teilnehmer\*innen in jedem Beitrag verschiedene Aspekte der biologischen Kernidee Informationsverarbeitung und des Phänomens (z. B. Kälte im Zusammenhang mit der Energiedimension) hervorgehoben und diese Aspekte mit Hilfe der Aufgaben a-e herausgearbeitet (z. B. Beziehung zwischen dem Input und den intern-externen Prozessen). Diese Diskussion im Forum hat die Gruppe dazu geführt, erste Modelle von Veränderungen zu konstruieren (siehe **Abbildung 2** und **3**), die sie später verwendeten, um Stability-Aspekte im Zusammenhang mit diesen Veränderungen zu identifizieren und über die relevantesten zu untersuchenden Stabilities/Changes zu entscheiden. Diese Modelle haben die meisten Ideen aller Teilnehmer des Forums gesammelt und als Brücke für die Arbeit mit der Datenerfassungsressource gedient (dritte Zeile der Tabelle).

In der Tat hängen die Stabilities/Changes in der Tabelle mit den Aspekten der biologischen Kernidee zusammen, die in diesen Darstellungen hervorgehoben und codiert sind. Zum Beispiel beziehen sich die Changes „Wärmeempfindung“ und „Umweltreize (Wind / Berührung)“ auf das Modell in **Abbildung 3** (a) und auf die Modelle in **Abbildung 2** (a-b). Eine Teilnehmerin hat drei verschiedene Momente der Veränderung (vorher, während und nachher) der Armhaut-TB dargestellt und Aspekte der „Wärmeisolation“ des Körpers hervorgehoben und sie als „Reaktion auf den Umweltreiz“ codiert. Eine andere Teilnehmerin war spezifischer in Bezug auf die Energiedimension der Veränderung. Sie hat den „kalten Wind“ als Input hervorgehoben, und die biologische Umwandlung dieses „kalten Windes“ in der Armhaut-TB in „Körperwärme“ und eine andere Form von „Energie“ codiert. In diesem ersten Teil der Arbeit hat sich die Gruppe auf die Körperwahrnehmung und die Reaktion auf externe Informationen (Inputs und Outputs) konzentriert. Erst im letzten Teil des Modells in **Abbildung 2** (c) hat sie den Weg eröffnet, die biologischen Transformationen jenseits der Armhaut-TB zu verstehen, indem sie die Idee der Nervenbahnen hervorgehoben haben. Dieser Aspekt stand im Zusammenhang mit dem Change „Energieniveau“ in der Datenerfassungstabelle.

Im zweiten Forum hat sich die Gruppe erneut mit einer synchronen Arbeit (03.06, von 16:41-19:19) beschäftigt, in der sie mit Online-Videos gearbeitet haben. Sie haben detaillierte Aspekte des Videos in ein robustes Erklärungsmodell integriert, um die internen Mechanismen des Körpers für die Verarbeitung der externen Informationen zu interpretieren. Sie haben begonnen, verschiedene Aspekte der Informationsverarbeitung hervorzuheben, und diese Aspekte dann mit Unterstützung der auf TBs basierenden Inputs zu codieren. Obwohl in **Tabelle 1** nicht die gesamten Modellierungsprozesse im zweiten Forum dargestellt sind, dienen die Beispiele zur Veranschaulichung der Arbeit der Gruppe. Zum Beispiel hat eine Teilnehmerin in ihrem Beitrag die Nervenzelle als TBs hervorgehoben.

Im Modell, das diesen Beitrag begleitete (**Abbildung 4**), hat die Teilnehmerin dann auch die Informationsdimension (Information an Nervenzelle über Rezeptoren) und die Stoffausschüttung hervorgehoben. Sie hat den Diskussionsthread zur Interpretation der biologischen Transformation „Synapse“ zum Output „Informationsweiterleitung“ geleitet. Sie hat die Gruppe zum Codieren geführt, welches auf den strukturellen und funktionalen Aspekten der Transformation basiert, nachdem der Körper den externen Input erhalten hat, sowie basierend auf dem Fluss von Energie und Information. Ein Teilnehmer hat dieses Codieren durch Integration des Aspekts der biologischen Kernidee gemacht: „Die von den Rezeptoren empfangenen Informationen wandeln sich in ein elektrisches Signal um, dieses elektrische Signal wandelt sich in ein chemisches Signal und dann wieder in ein elektrisches Signal um, das weitergeleitet wird.“ In seinem Modell (**Abbildung 5**) hat er die Arbeit seiner Begleiter\*innen beurteilt und überarbeitet unter Verwendung zusätzlicher Aspekte aus den Videos. Er hat einige Aspekte in Bezug auf den Unterschied zwischen den Nervenzellen gemäß den Funktionen aufgebracht, die in die Darstellung integriert wurden: a) Motorische Zellen (Rückenmark) übertragen niemals Signale vom Gehirn oder Rückenmark an die Muskeln, b) Interneurone sind Vermittler, welche Signale zwischen verschiedenen Körperteilen vermitteln.

Zusammenfassend hat die Gruppendiskussion in den Foren gezeigt, wie die Studierenden in vier sich ergänzenden Bereichen der MBU-Planung, die von Multimedia-Tools unterstützt werden, professionelle Fachkompetenz erweitert haben. **Erstens** haben sie Fachkompetenzen in ihrem Verständnis der biologischen Kernidee der Informationsverarbeitung erlangt, indem sie ihr Verständnis geteilt haben. Wir haben herausgefunden, wie sie zusammengearbeitet haben, um die einzelnen Ideen in den Videos zu erarbeiten und einen Konsens mit gemeinsamer Bedeutung zu erzielen. **Zweitens** haben sie in gemeinsamer Entwicklung mit der biologischen Kernidee Fachkompetenz in ihrem Verständnis der naturwissenschaftlichen Modellierung erlangt. Sie haben biologische Modelle konstruiert und als Werkzeuge verwendet, um das Phänomen zu definieren und zu interpretieren sowie ihre eigenen Modelle anzuwenden und zu überarbeiten. Sie haben professionelle Fachkompetenz erlangt, um die Ideen für naturwissenschaftliches Fachwissen, insbesondere die von Online-Videos, für Modellierungszwecke zu „modifizieren“, insbesondere basierend auf den Unterstützungen, die von den in den Multimedia-Tools eingebetteten TBs angeboten werden. **Drittens** haben sie Fachkompetenz in pädagogischen Aspekten der Planung einer MBU erlangt. Sie haben Entscheidungen darüber getroffen, wie die Untersuchung parallel zu ihrem Verständnis der Bedeutung der biologischen Kernideen für die Modellierung geführt werden soll. Zum Beispiel entschied sich die Gruppe aus ihren anfänglichen Darstellungen und der Relevanz der Stabilities/Changes die Veränderungen von „Gänsehaut“ zu wählen, um das Phänomen zu erklären und daraus eine angemessene Leitfrage zu formulieren. Bei solchen Entscheidungen überlegte die Gruppe, wie der modellbasierte Unterricht so geführt werden sollte, wie sie die Schülerinnen und Schüler von Anfang an in die Untersuchung einbeziehen und sie dabei unterstützen, biologische Kernideen strategisch in Erklärungsmodelle zu integrieren. **Viertens** hat die Gruppe Autonomie beim Lernen erlangt, die MBU zu planen. Sie haben die Themen des Forums synchron verfolgt, ohne dass externe Interventionen erforderlich waren. Sie vertrauen auch auf den Austausch mit Gleichaltrigen, um Entscheidungen über ihr Modell zu treffen. Diese Autonomie erweitert ihre professionelle Fachkompetenz, weil sie lernen, wie sie auf den Ideen und Erfahrungen des jeweils anderen aufbauen können und dabei die Bedeutung naturwissenschaftlicher Ideen und Praktiken auf ähnliche Weise konstruieren, wie sie es in ihren zukünftigen Klassenzimmern tun werden.

## Diskussion

Die Relevanz unserer Ergebnisse für die Entwicklung der professionellen Fachkompetenz von Grundschullehramtsstudierenden hat zwei Gründe. **Der erste Grund** dafür liegt in den epistemischen (TBs) und fachspezifischen Komponenten (Videos) der Multimedia-Tools, die die Integration von Aspekten einer biologischen Kernidee aus Videos in Erklärungsmodelle unterstützen. Studien wie Acher et al. (2007, S. 41) und Manz (2012, S. 1101) haben betont, wie wichtig es ist, dass Grundschullehrerinnen und -lehrer die Schülerinnen und Schüler dazu bewegen, wissenschaftliche Ideen gemeinsam mit den Modellierungsprozessen zu integrieren,

um Naturphänomene besser zu erklären. Dieser Aufruf erfordert die Vorbereitung von Lehramtsstudierenden, um zu lernen, wie diese Integration selbst durchgeführt werden kann. Unsere Studie legt nahe, dass Gruppen von Studierenden, die ihre eigenen Erklärungsmodelle konstruieren, diese Integration von den Kernideen der Naturwissenschaften (Biologie) und Modellierungsprozesse durchführen, die durch epistemische Unterstützung (TBs) vermittelt werden. **Der zweite Grund**, aus dem wir die Relevanz unserer Ergebnisse argumentieren, hängt von den didaktischen Komponenten der Multimedia-Tools ab, die die Aushandlung von Bedeutungen bei der Planung einer MBU unterstützen. Im Bereich der naturwissenschaftlichen Grundschullehrerbildung hat die Forschung das Potenzial von entworfenen Praxen und Tools aufgezeigt, um Studierende beim professionellen Lernen zu unterstützen, Wissen für modellbasierten Unterricht zu entwickeln (Davis, 2020, S. 110). Diese Studien konzentrierten sich hauptsächlich auf Praxen in ortsbezogene Lerndesign (Kenyon et al., 2011; Téllez-Acosta, Acher, McDonald, 2020). In der vorliegenden Studie haben die Ergebnisse gezeigt, dass ähnliche Erfolge möglich sind, wenn eine Kombination von Webressourcen (z. B. Foren, Datenerfassung, Etherpads) implementiert werden. Einerseits stellten die Foren eine sinnvolle Ressource dar, um Ideen auszutauschen und Bedeutungen auszuhandeln. Diese erhöhten die Zusammenarbeit und Kommunikation der Studierende. Unsere Gruppe teilte schriftliche Ideen und Darstellungen mit, die als Auslöser für die Aushandlung von Bedeutungen über die Integration von biologischen Kernideen in Erklärungsmodelle verwendet wurden. Forschungen, die mit Lehramtsstudierenden geführt wurden, haben das Potenzial web-basierter Ressourcen untersucht, um ihre Diskussion, ihre Beteiligung oder ihre Motivation zum Lernen für ihre professionelle Fachkompetenz zu erweitern (Ekici, 2017; Hunt, 2015). Der Schwerpunkt lag jedoch auf der Verwendung von Umfrageansätzen zur Untersuchung des Engagements im Internet (Cakir, 2013, S. 246) oder in gemischten Lernumgebungen (Weller, 2013; Garner & Rouse, 2016). In unserem Fall haben wir uns auf die Diskussion und die Beteiligung der Studierenden an ihrem eigenen Prozess konzentriert und gezeigt, dass sowohl das synchrone Engagement als auch die soziokulturellen Mechanismen ihrer Arbeit erfolgreich sind. Das Forum ist kein einfacher Informationsspeicher, sondern eine Gelegenheit, für Studierende ihre Ideen gemeinsam auszuhandeln und zu gestalten. Andererseits erleichtert *die Konzentration auf bestimmte Aspekte der modellbasierten Untersuchung die schrittweise Entscheidungsfindung*. Die Konzentration auf einen bestimmten Aspekt der damaligen Planungspraxis (Forzani, 2014, S. 360), unterstützte die Gruppe, Schritt für Schritt authentische Aktivitäten unter reduzierter Komplexität durchzuführen und schrittweise Entscheidungen zu treffen (Grossman et al., 2018, S. 12).

Unsere Erkenntnisse, die wir auf Grund unseres ausgewählten Beispiels gewonnen haben, erlauben es uns die Relevanz folgendes Verständnisses zu klären: mit der Unterstützung von Multimedia-Tools haben die Studierenden naturwissenschaftliches Fachwissen aus Online Videos genutzt, um eine MBU durchzuführen. Dies eröffnet den Horizont für das Design von Online Lernumgebungen, die mehr Möglichkeiten zur Entwicklung „praxisorientierter“ professioneller Lernprozesse bieten und Studierende zusätzlich zu den Foren für Teamarbeit in anderen Webressourcen motivieren.

## Literatur

- Acher, & Arcà (2020). Transformation Boxes: epistemic supports for teaching and learning scientific modeling for biological core ideas. (Unveröffentl. Forschungsarbeit). Universität Bielefeld, Bielefeld
- Acher, A., Arcà, M., & Sanmartí, N. (2007). Modeling as a Teaching Learning Process for Understanding Materials: A Case Study in Primary Education. *Science Education*, 91(1), 398–418.
- Baltaci-Goktalay, S., & Ozdilek, Z. (2010). Pre-service teachers' perceptions about web 2.0 technologies. In *Procedia - Social and Behavioral Sciences*, 2 (2), 4737- 4741. doi:10.1016/j.sbspro.2010.03.760
- Banday, M. T. (2013). Web 2.0 in e-Learning. *ICST Transactions on E-Education and e-Learning*, 1(3), 1-11. doi: 10.4108/el.1.3.e2
- Butrime, E., & Zuzeviciute, V. (2013). E-learning as a socio-cultural system (elements to be influenced and influencing elements). IMSCI 2013 - 7th International Multi-Conference on Society, Cybernetics and Informatics, Proceedings. Abgerufen am 15.07.2020 von [https://www.vdu.lt/wp-content/uploads/2015/05/SMF\\_Butrime\\_Zuzeviciute\\_38.pdf](https://www.vdu.lt/wp-content/uploads/2015/05/SMF_Butrime_Zuzeviciute_38.pdf)
- Cakir, H. (2013). Use of blogs in pre-service teacher education to improve student engagement. *Computers and Education*, 68, 244–252. doi: 10.1016/j.compedu.2013.05.013
- Davis, E. (2020). Approximations of Practice: Scaffolding for preservice teachers. In E. Davis, C. Zembal-Saul, & S. M. Kademian (Hrsg.), *Sensemaking in Elementary Science: Supporting Teacher Learning* (S. 86-112). Abingdon, Oxon: Routledge, Taylor & Francis Group. doi: 10.4324/9780429426513
- Ekici, D. I. (2017). The effects of online communities of practice on pre-service teachers' critical thinking dispositions. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(7), 3801–3827. doi: 10.12973/eurasia.2017.00759a
- Forbes, C. T., Lange-Schubert, K., Böschl, F., & Vo, T. (2019). Supporting Primary Students' Developing Modeling Competence for Water Systems. In A. Upmeyer zu Belzen, D. Krüger, & J. van Driel (Hrsg.), *Towards a Competence-Bases View on Models and Modeling in Science Education* (S. 257–273). Cham: Springer. doi: 10.1007/978-3-030-30255-9\_15
- Forzani, F. M. (2014). Understanding “Core Practices” and “Practice-Based” Teacher Education: Learning From the Past. *Journal of Teacher Education*, 65(4), 357–368. doi: 10.1177/0022487114533800
- Garner, R., & Rouse, E. (2016). Social presence – connecting pre-service teachers as learners using a blended learning model. *Student Success*, 7(1), 25–36. doi: 10.5204/ssj.v7i1.299
- Giesecke, A., & Schork, N. (28.12.2014). Nervenzelle, Nerv, Axon, Synapsen – Grundbegriffe des Nervensystems [Video]. Abgerufen am 03.06.2019 von <https://www.youtube.com/watch?v=20m6fhh-G7U>.

- Giesecke, A., & Schork, N. (06.04.2015). Die Haut – Organe des Menschen [Video]. Abgerufen am 03.06.2019 von <https://www.youtube.com/watch?v=OicZrm0wM4Y>
- Goodwin, C. (1994). Professional Vision. *American Anthropologist*, 96(3), 606–633. doi: 10.1525/aa.1994.96.3.02a00100
- Grossman, P., Schneider, S., & Pupik, C. G. (2018). The Turn Towards Practice In Teacher Education. In P. Grossman (Hrsg.), *Teaching Core Practices in Teacher Education*. Cambridge, MA 02138: Harvard Education Press.
- Hunt, A. M. (2015). Blended Online Learning in Initial Teacher Education: A Professional Inquiry into Pre-service Teachers' Inquiry Projects. *Journal of Open, Flexible, and Distance Learning*, 19(2), 48–60. Abgerufen am 13.08.2020 von <https://www.learntechlib.org/p/161942/>
- Kelly, G. J., & Chen, C. (1999). The sound of music: Constructing science as sociocultural practices through oral and written discourse. *Journal of Research in Science Teaching*. 36(8), 883–915. doi: 10.1002/(SICI)1098-2736(199910)36:8<883::AID-TEA1>3.0.CO;2-I
- Kultusministerkonferenz (KMK) (2004). *Bildungsstandards der Kultusministerkonferenz. Erläuterungen zur Konzeption und Entwicklung*. Berlin, Germany: Luchterhand.
- Manz, E. (2012). Understanding the codevelopment of modeling practice and ecological knowledge. *Science Education*, 96(6), 1071–1105. doi: 10.1002/sce.21030
- McDonald, M., Kazemi, E., Kelley-Petersen, M., Mikolasy, K., Thompson, J., Valencia, S. W., & Windschitl, M. (2014). Practice Makes Practice: Learning to Teach in Teacher Education. *Peabody Journal of Education*, 89(4), 500–515. doi: 10.1080/0161956X.2014.938997
- Ministerio de Educación Nacional (MEN) (2006). *Estándares Básicos de Competencias en lenguaje, matemáticas, ciencias y ciudadanas*. *Revolución Educativa*. Abgerufen am 20.08.2020 von [https://www.mineducacion.gov.co/1759/articles-81033\\_archivo\\_pdf.pdf](https://www.mineducacion.gov.co/1759/articles-81033_archivo_pdf.pdf)
- National Research Council (NRC) (2012). *A Framework for K-12 Science Education : Practices , Crosscutting Concepts , and Core Ideas*. Social Sciences. Washington, D.C.: The National Academies Press. doi: 10.17226/13165
- Seidel, T., Stürmer, K., Schäfer, S., & Jahn, G. (2015). How preservice teachers perform in teaching events regarding generic teaching and learning components. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, 47 (2), 84-96. doi: 10.1026/0049-8637/a000125
- Téllez-Acosta, M., Acher, A., & McDonald, S. (2020). *Pedagogies and tools to support elementary preservice teachers in learning to plan modeling-based investigations while utilizing science content resources*. (Unveröffentl. Doktorarbeit). Martin-Luther-Universität, Halle-Wittenberg
- Weller, A. (2013). The use of Web 2.0 technology for pre-service teacher learning in science education. *Research in Teacher Education*, 3(2), 40–46. Abgerufen am 26.09.2020 von

[https://repository.uel.ac.uk/download/06baa967023fc233c891c33adcce7ad02d96e61db0c17672dea516c6b6655322/1436052/RiTE\\_3\\_2\\_Weller.pdf](https://repository.uel.ac.uk/download/06baa967023fc233c891c33adcce7ad02d96e61db0c17672dea516c6b6655322/1436052/RiTE_3_2_Weller.pdf)

Wickman, P. O. (2012). Using pragmatism to develop didactics in Sweden. *Zeitschrift Fur Erziehungswissenschaft*, 15(3), 483–501. doi: 10.1007/s11618-012-0287-7

Zembal-Saul, C. (2018). The Role of Teacher Education in Advancing Reform in Primary Science Education. In Y. Lee., J. Tan (Hrsg.), *Primary Science Education in East Asia. Contemporary Trends and Issues in Science Education* (229–241). Cham: Springer. doi: 10.1007/978-3-319-97167-4\_12

**María Esther Téllez-Acosta** ist Doktorandin an der Martin-Luther-Universität Halle-Wittenberg in Partnerschaft mit der Pennsylvania State Universität, USA. Sie beschäftigt sich mit dem Lernen und Lehren von Naturwissenschaft aus einer soziokulturellen und praktischen Perspektive. Ihr Forschungsschwerpunkt liegt auf dem Lernen von Grund- und Sekundarschullehrer\*innen naturwissenschaftlicher Fächer in praxisorientierten und authentischen Kontexten.

[maria.tellez-acosta@student.uni-halle.de](mailto:maria.tellez-acosta@student.uni-halle.de)

**Andres Acher** leitete die Arbeiten an diesem Projekt während seiner Zeit als wissenschaftlicher Mitarbeiter im Arbeitsbereich Sachunterricht, Institut für Schulpädagogik und Grundschuldidaktik an der Martin-Luther-Universität Halle-Wittenberg. Seine Forschung konzentriert sich auf verschiedene Aspekte des Lehrens und Lernens wissenschaftlicher Modellierung als wissenschaftliche Praxis aus soziokulturellen Perspektiven.

[andres.acher@uni-bielefeld.de](mailto:andres.acher@uni-bielefeld.de)

**Leon Kusitzky** ist Studierender des Lehramts an Grundschulen mit dem Drittfach Sachunterricht an der Martin-Luther-Universität Halle-Wittenberg. Als studentische Hilfskraft der philosophischen Fakultät III im Bereich Grundschullehramt Sachunterricht (Naturwissenschaftlich-technisches Lehren und Lernen im Sachunterricht) beschäftigte er sich mit der Optimierung und Online-Integration von Hilfestellungen für Studierende, um den Prozess der Modellierung zu strukturieren.

[leon.kusitzky@student.uni-halle.de](mailto:leon.kusitzky@student.uni-halle.de)

# Die Entwicklung und Evaluation von Erklärvideos zu geographischen Geländemethoden

*Maximilian Breuer<sup>1</sup>*

*<sup>1</sup> Martin-Luther-Universität Halle-Wittenberg, Institut für Geowissenschaften und Geographie, Didaktik der Geographie*

Geographische Geländemethoden sind trotz ihrer Erwünschtheit und Einforderung durch curriculare Vorgaben in der schulischen Praxis unterrepräsentiert. Ein Hauptgrund dafür ist die unzureichende Qualifikation angehender Lehrerinnen und Lehrer durch die hochschulische Ausbildung. Im Kontext multimedialen Lehrens und Lernens spielen auch Erklärvideos eine zunehmend größere Rolle, da ihre Verwendung sowie ihr Angebot quantitativ steigen und sie für individualisierte und selbstständige Lernprozesse eingesetzt werden können. Für Lehramtsstudierende können Erklärvideos also potentiell zur Aneignung geographischer Geländemethoden dienen. Die unzureichende Verfügbarkeit von Erklärvideos zu geographischen Geländemethoden erfordert deren Entwicklung und Evaluation, die im Projekt mithilfe eines Design Based Research-Ansatzes beforscht werden sollen. Die vorliegende Arbeit fokussiert dabei die Erprobung und Weiterentwicklung von Erklärvideos. Zur Erprobung wird ein Quasi-Experiment mit 30 Lehramtsstudierenden des Fachs Geographie durchgeführt. In begleitenden Interviews werden die Bewertungen und Anforderungen von den Lehramtsstudierenden bezüglich der untersuchten Videos erfasst. Alle Aktivitäten und Äußerungen der Probandinnen und Probanden werden videographiert. Die qualitative Auswertung dieser Daten erfolgt mithilfe der inhaltlich-strukturierenden Inhaltsanalyse. Die Analyse zeigt ein breites Spektrum unterschiedlicher Einschätzungen und Implikationen für die Erklärvideos, die durch Übersichten des Kategoriensystems und detaillierter durch drei Fälle kontrastierend dargestellt werden. Die Unvereinbarkeit bestimmter Implikationen führt zu der Erkenntnis, dass eine Diversifizierung und Adaption der Erklärvideos an verschiedene Zielgruppen hilfreich, aber nur schwer zu realisieren ist.

## Einleitung

Auswirkungen der Windstärke beobachten, den Bodennitratgehalt bestimmen, den Flusslauf kartieren, die Temperatur messen, Schritte zählen, um eine Entfernung zu ermitteln – zahlreiche Merkmale unserer Umwelt, sogenannte Geofaktoren, lassen sich durch geographische Geländemethoden auch von Nicht-Fachwissenschaftlern wie Lehrerinnen und Lehrern sowie Schülerinnen und Schülern unkompliziert durchführen. Damit leisten Geländemethoden einen wesentlichen Beitrag zum wissenschaftlichen Arbeiten im Fach Geographie, da sie die systematische Erfassung, Dokumentation und Analyse beobachtbarer Prozesse und Strukturen auf der Erde zum Ziel haben (Falk, 2015). Im schulischen Kontext werden geographische Geländemethoden durch die Bildungsstandards (DGfG, 2020) und die verschiedenen Curricula der Bundesländer eingefordert. Dementsprechend sind auch in der Lehrerbildung unter anderem die Informationsbeschaffung im Gelände sowie durch Medien und die adressatengemäße Präsentation als fachliche Kompetenzen im Bereich Methoden vorgesehen (KMK, 2008). In der Praxis werden geographische Geländemethoden im Rahmen von Exkursionen durchgeführt (Rinschede, 2019), die trotz ihrer Beliebtheit bei Lehrenden und Lernenden nur selten realisiert werden (Lößner, 2011). Daher sind auch die Geländemethoden von geringer Einsatzhäufigkeit betroffen. Neben organisatorischen und systembedingten Hindernissen der Schule (ebd.), ist nach Selbsteinschätzung von Lehrkräften die unzureichende Qualifikation im Rahmen der universitären Ausbildung ursächlich dafür (Höhnle & Schubert, 2016).

Ähnlich problembehaftet erscheint laut der Meta-Analyse von Findeisen, Horn und Seifried (2019) die Fähigkeit des Erklärens angehender Lehrpersonen, was in der Regel auf Defizite im Fachwissen und fachdidaktischen Wissen zurückzuführen ist. Bestimmte Bestandteile der Erklärfähigkeit – beispielsweise die Verwendung eines logischen Aufbaus, eines passenden Sprachniveaus, von Zusammenfassungen, Visualisierungen und Beispielen, die Auswahl und Fokussierung relevanter Inhalte – gelten aber als erlernbar (Brown & Daines, 1981; Miltz, 1972). Dies ist von besonderer Bedeutung, da das Fachinhalte-erklären-Können als bedeutende Qualifikation einer Lehrperson gilt (Aff & Schopf, 2017; Vogt, 2009). Darüber hinaus konnte ein Zusammenhang zwischen der Qualität der Erklärung einer Lehrperson sowie dem Lernerfolg und der Zufriedenheit der Schülerinnen und Schüler nachgewiesen werden (Hattie, 2009).

Erklärvideos sind bei Kindern, Jugendlichen und Erwachsenen zu einem beliebten Werkzeug der Informationsbeschaffung geworden (Bitkom, 2017; Rat für kulturelle Bildung e. V., 2019) und in großer Zahl zu den verschiedensten Themen der Bildung und des Alltags online verfügbar (Wolf 2015). Definitionsgemäß sind es „Filme, in denen erläutert wird, wie man etwas macht oder wie etwas funktioniert, bzw. in denen abstrakte Konzepte erklärt werden“ (ebd., S. 123). Als Spezialfall der Erklärvideos sind sogenannte How-To-Videos oder Tutorials zu verstehen, „in denen eine Fertigkeit oder Fähigkeit im Sinne einer vollständigen Handlung explizit zum Nachmachen durch die Zuschauer vorgemacht wird“ (ebd., S. 123). Im Gegensatz zu Wolfs Unterscheidung von Erklärvideos als von Amateuren produzierten Videos

einerseits und Lehrfilmen als Videos mit „zunehmende[r] didaktische[r] und mediengestalterische[r] Planung und Professionalisierung“ (ebd., S. 122) andererseits wird in dieser Arbeit – unabhängig von der Ausprägung dieser Merkmale – Erklärvideo als Oberbegriff für alle Formen verwendet, da die Übergänge fließend sind und eine trennscharfe Begriffsverwendung somit nicht möglich ist.

Erklärvideos visualisieren Inhalte und können den Erwerb von Fähigkeiten, Fertigkeiten und Kenntnissen unterstützen (Zander, Behrens, & Mehlhorn, 2018). Dies ist erklärbar mit der sozialkognitiven Lerntheorie von Bandura (1977; 1986), laut der ein kognitives Schema des Verhaltens eines Modells erworben wird, das es den Beobachtenden erlaubt, das beobachtete Verhalten zukünftig selbst auszuführen. Unter Annahme dieses Modells ist der Lernerfolg beim Einsatz von Erklärvideos unter anderem von der Komplexität und Sichtbarkeit der Handlung, dem Vorwissen der Betrachtenden und interaktiven Gestaltungselementen des Videos abhängig (Findeisen, Horn, & Seifried 2019).

Können sich demnach Lehramtsstudierende mithilfe von Erklärvideos geographische Geländemethoden selbstständig aneignen? Können Lehramtsstudierende außerdem in die Lage versetzt werden, geographische Geländemethoden für Schülerinnen und Schüler fach- und adressatengerecht zu erklären? Und wie sollten diese Erklärvideos dann gestaltet sein?

Da nur wenige Videos zu geographischen Geländemethoden existieren, die zudem fast ausschließlich einen fachwissenschaftlichen Personenkreis adressieren, müssen diese von Grund auf neu entworfen werden. Ziel dieses Projekts ist daher die Konzeption, Produktion, Erprobung und Weiterentwicklung von Erklärvideos, die Lehramtsstudierenden zur Aneignung geographischer Geländemethoden dienen sollen. Da die Darstellung aller Teilaufgaben mit den damit verbundenen Forschungsfragen den Rahmen des vorliegenden Berichts sprengen würde, liegt der Fokus auf der Weiterentwicklung und damit der Bearbeitung und Beantwortung der folgenden Teilforschungsfrage: Wie sollten Erklärvideos zu geographischen Geländemethoden aus Sicht von Lehramtsstudierenden konzipiert sein?

Darf ein Video länger als die von Marketingexperten oder Psychologen empfohlene Dauer sein? Braucht es aufwendige Animationen? Welche Wissensbestände und Fähigkeiten können vorausgesetzt werden? Über solche und ähnliche Fragen soll die folgende Untersuchung Aufschluss geben. Da die Lehramtsstudierenden als Adressaten der Erklärvideos zugleich auch als Experten ihres individuellen Lernprozesses angesehen werden, erscheint die Identifikation der individuellen Anforderungen an Erklärvideos zu geographischen Geländemethoden auch jenseits allgemeiner Kriterienkataloge lohnenswert.

## **Methodischer Ansatz**

Im Folgenden soll das methodische Vorgehen des gesamten Projekts kurz erläutert werden, um die behandelte Teilforschungsfrage einordnen zu können. Zur Bearbeitung der oben vorgestellten Zielstellungen folgt das Projekt einem Design Based Research-Ansatz, der in iterativen Zyklen die Aufgaben der Entwicklung, d. h. das Gestalten, Verändern und Optimieren und die Aufgaben der Forschung, d. h. das Analysieren, Verstehen und Erklären, durch die

beobachteten bzw. initiierten Lehr- und Lernprozesse miteinander verknüpft (Hußmann, Thiele, Hinz, Prediger, & Ralle, 2013). Zentraler Bezugspunkt dieser Aufgaben sind die Erklärvideos zu geographischen Geländemethoden. Für die Untersuchung wurden unter Berücksichtigung von Kriterien für die Erklärqualität (Kulgemeyer, 2020) Erklärvideos zu zwei geographischen Geländemethoden, die sich in ihrer Komplexität und inhaltlichen Ausrichtung unterscheiden, zusammen mit Expertinnen und Experten der Fachdidaktik und Fachwissenschaft der Geographie konzipiert und mit Unterstützung des Zentrums für multimediales Lehren und Lernen (LLZ) der Martin-Luther-Universität produziert. Bei den vorgestellten Geländemethoden handelt es sich um die Bestimmung der Fließgeschwindigkeit eines Gewässers und die Bestimmung des Calciumcarbonatgehalts im Boden. Die zwei Erklärvideos sind unter folgenden Links erreichbar:

[https://filer.itz.uni-halle.de/dl/1170/pub/Ev1\\_Fliessgeschwindigkeit.mp4](https://filer.itz.uni-halle.de/dl/1170/pub/Ev1_Fliessgeschwindigkeit.mp4)

[https://filer.itz.uni-halle.de/dl/1170/pub/Ev2\\_Carbonatgehalt.mp4](https://filer.itz.uni-halle.de/dl/1170/pub/Ev2_Carbonatgehalt.mp4)

### **Untersuchungsdesign**

Im Sommersemester 2019 wurde ein Quasi-Experiment mit 30 Lehramtsstudierenden des Fachs Geographie durchgeführt, die am Seminar „Außerschulische Lernorte/Exkursionsdidaktik“ teilnahmen. Die Probandinnen und Probanden hatten jeweils die Aufgabe, sich mit den Erklärvideos vorzubereiten, die zwei Geländemethoden durchzuführen und für eine 9. Klasse fach- und adressatengerecht zu erklären. Ergänzt wurden diese Phasen um Interviews (s. **Abbildung 1**), die unmittelbar vor und nach der Vorbereitung, nach der Performanz, d. h. nach der Durchführung und Erklärung, sowie etwa zwei Wochen nach dem Untersuchungszeitpunkt geführt wurden. Mithilfe halbstrukturierter Interviewleitfäden wurden die Probandinnen und Probanden zu ihrem Vorwissen, ihrer Bewertung der Erklärvideos sowie der Reflexion ihrer Vorbereitung und Performanz befragt. Die Aktivitäten und Äußerungen der Probandinnen und Probanden wurden sowohl während der Vorbereitung mit den Erklärvideos als auch während der Durchführung und Erklärung der Geländemethoden videografiert. Die Fragen zur Bewertung der Erklärvideos zielten vor allem auf die Einschätzung gelungener oder nicht gelungener Aspekte sowie Änderungs- und Ergänzungsvorschläge. Mithilfe der Auswertung dieser Aussagen soll die Forschungsfrage beantwortet werden.

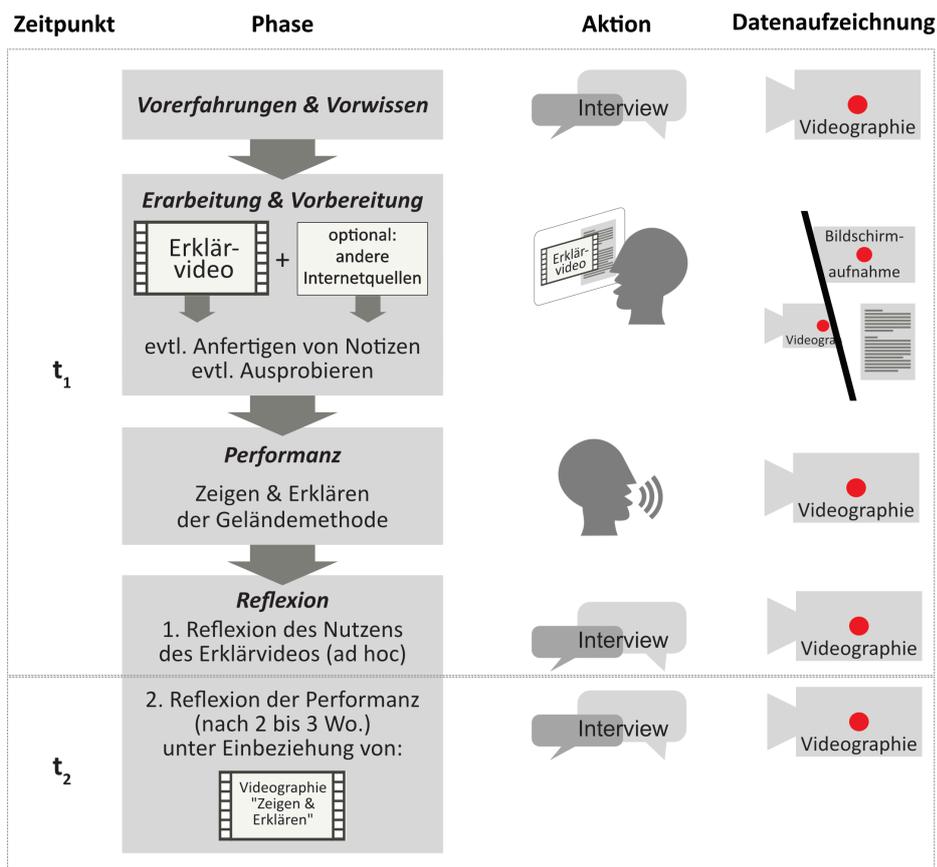


Abbildung 1. Untersuchungsdesign (eigene Darstellung)

## Stichprobe

Von den insgesamt 30 Fällen wurden vier aufgrund unvollständiger Daten von der gesamten Auswertung ausgeschlossen, sodass sich eine Stichprobengröße von 26 Lehramtsstudierenden ergibt, von denen 13 männlich und 13 weiblich sind. 15 Probandinnen und Probanden waren für das Lehramt an Gymnasien und elf für das Lehramt an Sekundarschulen eingeschrieben und befanden sich während des Untersuchungszeitraums hauptsächlich im 6. oder 8. Fachsemester.

## Datenaufbereitung und -auswertung

Für die Erfassung der Lernendenperspektive auf die Gestaltung der Erklärvideos wurden die zugehörigen Aussagen zunächst mittels Erweiterter inhaltlich-semantischer Transkription des Gesagten nach Dresing und Pehl (2018) verschriftlicht. Im Anschluss erfolgte die Auswertung mithilfe der inhaltlich-strukturierenden Inhaltsanalyse nach Kuckartz (2016). Zur systematischen Analyse wurde ein Kategoriensystem auf Basis deduktiver und induktiver Kategorien gebildet. Die deduktiven Kategorien wurden aus den 14 Gütekriterien für Erklärvideos von Kulgemeyer (2020) übernommen, die schon zur Konzeption und Produktion der untersuchten Videos dienten. Alle anderen Kategorien wurden induktiv am Material gebildet, indem für inhaltlich übereinstimmende Kodierungen allgemeine (induktive) Kategorienbezeichnungen und -definitionen formuliert wurden. Kodierungen, die sich keiner Kategorie

zuordnen ließen, wurden hingegen als in-vivo-Codes belassen. Äußerungen, die aus einer eingenommenen Schülerinnen- und Schülerperspektive vorgenommen wurden (z. B. „Ein Schüler/eine Schülerin der neunten Klasse kann sich, glaube ich, nicht so lange darauf konzentrieren.“), wurden nicht kodiert, da nur die Perspektive der Lehramtsstudierenden als Zielgruppe der Erklärvideos für die Untersuchung relevant ist. Aus allen Kategoriebezeichnungen, Kategoriedefinitionen und dafür passenden Ankerbeispielen wurde ein Kodierleitfaden erstellt. Für eine das gesamte Spektrum abdeckende und kontrastierende Darstellung der Ergebnisse wurden nach Kriterien der maximalen strukturellen Variation (Hensel & Kreuz 2018) neun Fälle ausgewählt und von einer weiteren Person mithilfe des Kodierleitfadens kodiert, um die Ergebnisse der Auswertung konsensuell zu validieren. Wiederum drei dieser neun Fälle werden in den Ergebnissen kurz vorgestellt, da sie durch ihre Ausführlichkeit und Fokussierung auf eine bestimmte Einschätzung der Erklärvideos jeweils ein Extrem darstellen. Die Fallbezeichnungen 23m, 26w und 29m ergeben sich aus der fortlaufenden Nummer der Probandinnen und Probanden im Untersuchungszeitraum und dem Geschlecht. Weiterhin werden alle in den neun Fällen zu beiden Erklärvideos gefundenen Kategorien aus dem Kodierleitfaden in einer Tabelle systematisiert aufgeführt.

## Ergebnisse

Der Fall 29m hebt sich durch seine zahlreichen Anmerkungen zu inhaltlichen Aspekten des zweiten Erklärvideos im Vergleich zu den anderen Fällen hervor. Sein Wissen in diesem Fachgebiet ist dabei wohl auf sein Chemiestudium als zweiten Lehramtsfach zurückzuführen. Bei der Durchführung der Geländemethode kritisiert 29m die aus seiner Sicht eher zweckmäßige Bestimmung des Calciumcarbonatgehalts direkt aus dem Boden, die nicht so präzise ist wie die in der Chemie übliche Bestimmung aus einer zuvor hergestellten wässrigen Bodenlösung. Eine Präzisierung fordert er zudem für die Arbeitsaufträge bei der Zugabe der Salzsäure mit konkreten Volumenangaben, anstatt Salzsäure „auf Gutedünken rein[zumachen], bis alles bedeckt ist“.

29m hätte es lieber und, wenn er keine Ahnung gehabt hätte, wäre es für ihn logischer, wenn die Reaktion von Calciumcarbonat mit Salzsäure als Reaktionsgleichung vor der Beobachtung besprochen werden würde. Ein großes Defizit sieht er im Fehlen von Hinweisen zu der von ätzender Salzsäure ausgehenden Gefahr und dem korrekten Umgang durch Schutzmaßnahmen damit. Auch die am Ende des Erklärvideos angesprochene Entsorgung wird seiner Einschätzung nach nicht begründet. Die Notwendigkeit dieser Inhalte betont 29m vor dem Hintergrund seiner persönlichen Erfahrungen in den schulpraktischen Übungen mit Kommilitonen anderer Fächerkombinationen, „[die viel unsicherer waren, auch nicht mit vielen Sachen umzugehen wussten, [...] die nie Experimente durchführen]“. Weiterhin fehlt ihm beim ermittelten Carbonatgehalt der „direkte Bezug mit einem Vergleichswert“, um diesen Stoffgehalt auch einordnen und interpretieren zu können.

Während 23m bei der zweiten Geländemethode, die Bestimmung des Calciumcarbonatgehalts im Boden, ebenfalls vorwiegend inhaltliche Änderungswünsche hat, bezieht er sich in

seinen Aussagen zur Methode ‚Bestimmung der Fließgeschwindigkeit eines Gewässers‘ vor allem auf fachdidaktische und pädagogische Aspekte. Zur Ergänzung der Erklärvideos wünscht er sich Hinweise, „um einfach [s]einer Planung das ein bisschen klarer zu machen“. Dies kann seiner Meinung nach in Stichwörtern erfolgen. Damit soll beantwortet werden, „welche Probleme auftreten können“ oder worauf man achten sollte, wie zum Beispiel die Auswahl des Ortes. Im weiteren Verlauf des Interviews führt 23m das anhand der ersten Geländemethode aus, indem er auf die Gefahren eines steilen Uferbereichs in Verbindung mit Schülerinnen und Schülern, „die vielleicht ein bisschen, ja, nervöser sind, schwieriger“, hinweist und die Wahl eines flachen und einsehbaren Uferbereichs vorschlägt. Er führt seine Überlegungen sogar noch weiter: Eine zusätzliche Absicherung könnte durch die Bestätigung der Eltern über die Schwimmfähigkeit der Kinder und die Mitnahme von Schwimmwesten erfolgen.

Auch die Beschreibung einer anderen Methode, womit er vermutlich nicht Geländemethode, sondern Aktionsform meint, oder einer anderen Sozialform ist für 23m wünschenswert, um einerseits „aus diesem Lehrervortrag [herauszutreten] und [...] die Schüler selbst in die Interaktion [zu] bekomme[n]“ und andererseits „selbst dann frei entscheiden [zu können]“ zwischen den methodischen Vorgehensweisen. Diesbezüglich sollten für 23m die Vorschläge aber „schon einmal evaluiert wurde[n]“ sein bzw. „[funktionieren]“.

Anhand ihrer Aussage, dass „[kein Erklärvideo da reicht]“, „um mir das so wirklich erklärbar zu machen“, „wenn man da wirklich so komplett ohne Ahnung rangeht“ wird deutlich, dass die Einschätzungen von 26w gegenüber den beiden Fällen zuvor in eine völlig andere Richtung gehen. Sie ist unsicher, „ob [sie] noch mehr Informationen gebraucht hätte oder wenn nicht sogar ein bisschen weniger“. „[D]as war einfach zu viel“ für sie, sodass sie ab der Einblendung der chemischen „Gleichung nicht hinterhergekommen“ ist und sie das Video „ein bisschen verloren [hat]“. Mit Blick auf die in der Untersuchung folgende Performanz, insbesondere die Erklärung dieser Geländemethode, stellt 26w fest: „[D]as kann ich nicht nacherzählen. Da müsste ich, glaube ich, nach jedem Satz das Video erstmal anmachen und gucken, ob es stimmt“. Diese Aussagen von 26w wurden bei der Inhaltsanalyse unter anderem mit der Kategorie „Hoher Cognitive Load: Wiederholen notwendig“ kodiert. Um dem entgegenzuwirken, empfiehlt sie Reaktionsgleichungen und Berechnungsformeln vor „eine[m] neutralen Hintergrund, dass man wirklich nur die Gleichung sieht“, und „dann hört man [die Sprecherin] vielleicht aus dem Hintergrund“. Außerdem „kann man ja auch wirklich so eine Gleichung einmal animiert umstellen; nicht von einem Bild ins nächste, sondern das richtig animieren, wie man was umstellt und macht“. Dennoch empfindet 26w das Erklärvideo als „sehr schön anschaulich“ und „auch gut gezeigt“, welche einzelnen Schritte zu machen sind. Ihrer Meinung nach „soll ein bisschen die Methode im Vordergrund stehen und nicht die ganzen Gleichungen da dahinter“. Daher schlägt sie vor, die im Erklärvideo bereits getroffene Unterscheidung in eine qualitative und eine quantitative Bestimmung auch in zwei Videos umzusetzen. „Das erste wäre quasi so ein Überblick und einfacher durchzuführen. Und das andere geht ja dann nochmal tiefer rein.“

Diese drei Fälle heben sich von den übrigen vor allem durch ihre Ausführlichkeit und Fokussierung auf eine bestimmte Einschätzung der Erklärvideos ab. Das heißt, dass die Erklärvideos inhaltlich sowie didaktisch erweitert und weiterentwickelt, gleichzeitig aber weniger komplex und kürzer werden sollten. Ähnliche Aussagen lassen sich verteilt auch bei anderen Fällen finden. Die etwas ausführlicher formulierten Kategorienbezeichnungen, die diese Aussagen inhaltlich strukturiert widerspiegeln, sind den beiden Tabellen im Folgenden zu entnehmen. Dabei sind weniger die Häufigkeiten, die durch die Zahlen in Klammern dokumentiert werden, sondern das inhaltliche Spektrum der Kategorien von Interesse. In der ersten Tabelle (s. **Tabelle 1**) werden Aspekte zur Verbesserung der Erklärvideos, in der zweiten (s. **Tabelle 2**) bereits gelungene Aspekte der Erklärvideos aus Sicht der Lehramtsstudierenden aufgeführt. Die Kategorien werden nochmals einer inhaltlich-methodischen, gestalterischen, didaktischen oder offenen Oberkategorie und den beiden Geländemethoden zugeordnet, über die die Aussagen getroffen wurden. Des Weiteren wird eine Zuordnung der Kategorien vorgenommen, die bei beiden bzw. nur bei einer von beiden Geländemethoden identifiziert werden konnten.

**Tabelle 1.** Kategoriensystem – Aspekte zur Verbesserung der Erklärvideos aus Sicht von Lehramtsstudierenden

	Erklärvideo zur Bestimmung der Fließgeschwindigkeit	Erklärvideo zur Bestimmung des Calciumcarbonatgehalts
<b>inhaltlich-methodisch</b>	<ul style="list-style-type: none"> <li>Vergleichswerte angeben</li> </ul>	
	<ul style="list-style-type: none"> <li>Ungenauigkeit beim Messen im Video (2)</li> <li>Inkonsequenz: Gummiente als Praxisbeispiel genannt, aber Holzstock im Versuch verwendet (3)</li> <li>Hinweis zum genauen Messen der Strecken (3)</li> </ul>	<ul style="list-style-type: none"> <li>Inhaltliche Ergänzung (3)</li> <li>Wunsch nach mehr Hintergrundinformationen (1)</li> <li>Teilung in jeweils ein Video zur qualitativen bzw. quantitativen Bestimmung (3)</li> <li>Hinweis zum Notieren der Werte (1)</li> <li>Unklare Herkunft von Werten (3)</li> <li>Fehlen weiterer Einflussfaktoren im Video (4)</li> <li>Anforderungen der Bodenprobe (1)</li> <li>Angabe zur Wartedauer auf Reaktion (3)</li> </ul>
<b>gestalterisch</b>	<ul style="list-style-type: none"> <li>Potential zur Raffung (5)</li> <li>Zusätzliche Einblendungen wünschenswert (9)</li> <li>Unzureichende Ästhetik/optische Abwechslung (7)</li> <li>Mangelnde Anschaulichkeit/Sichtbarkeit von Details (7)</li> </ul>	<ul style="list-style-type: none"> <li>Wahl der Kameraperspektive (2)</li> <li>Monotone Sprechweise (10)</li> <li>Langsame Sprechweise (4)</li> <li>Starrheit in Mimik und Gestik (6)</li> </ul>
	<ul style="list-style-type: none"> <li>Kameraperspektive zum Start-/Endpunkt und Stock im Video (3)</li> </ul>	<ul style="list-style-type: none"> <li>Gesamtlänge des Videos (5)</li> <li>Zeitraffer ohne genaue Angabe der realen Zeit (2)</li> </ul>
<b>didaktisch</b>	<ul style="list-style-type: none"> <li>Unterrichtliche Einbindung: Aufteilung in Phasen, die outdoor bzw. indoor erfolgen können (2)</li> <li>Hinweise zu Sicherheitsmaßnahmen (5)</li> <li>Hinweise zu örtlichen Voraussetzungen (2)</li> <li>Hinweise zu Materialien (4)</li> <li>Problemorientierung nicht relevant für Schülerinnen und Schüler (6)</li> </ul>	
	<ul style="list-style-type: none"> <li>Fehlende fachwissenschaftliche Fragestellung (1)</li> </ul>	<ul style="list-style-type: none"> <li>Hinweis zum Einsatz der qualitativen bzw. quantitativen Bestimmung (1)</li> <li>Umfang der Problemorientierung zu groß (5)</li> </ul>

<b>offen</b>	<ul style="list-style-type: none"> <li>• Komplexität / Schnelligkeit / Umfang der Formeln (12)</li> <li>• Hoher Cognitive Load: Stoppen notwendig (6)</li> <li>• Hoher Cognitive Load: Wiederholen notwendig (6)</li> </ul>
--------------	---

*Table 2. Kategoriensystem – gelungene Aspekte der Erklärvideos aus Sicht von Lehramtsstudierenden*

	Erklärvideo zur Bestimmung der Fließgeschwindigkeit	Erklärvideo zur Bestimmung des Calciumcarbonatgehalts
<b>inhaltlich-methodisch</b>	<ul style="list-style-type: none"> <li>• Struktur (7)</li> <li>• Vorgehensweise/Schrittfolge/-haftigkeit (8)</li> <li>• Kohärenz des Gesagten (2)</li> <li>• Detailliertheit der Anweisungen (4)</li> </ul>	
	<ul style="list-style-type: none"> <li>• Kürze/Prägnanz/keine Exkurse (5)</li> </ul>	
<b>gestalterisch</b>	<ul style="list-style-type: none"> <li>• Deutliche Aussprache (2)</li> <li>• Sprechweise/-tempo (4)</li> <li>• Anschaulichkeit/Sichtbarkeit von Details (3)</li> <li>• Einblendungen im Video (6)</li> <li>• Passung von Erklärung und Sichtbarem (5)</li> </ul>	
	<ul style="list-style-type: none"> <li>• angenehme Länge (1)</li> <li>• Zeitraffer/Vermeidung von Längen (1)</li> </ul>	
<b>didaktisch</b>	<ul style="list-style-type: none"> <li>• Praxisbezug/Problemorientierung (13)</li> </ul>	<ul style="list-style-type: none"> <li>• Verdeutlichung der Relevanz (1)</li> </ul>
<b>offen</b>	<ul style="list-style-type: none"> <li>• sehr informativ (7)</li> <li>• Lob/keine Beanstandungen (4)</li> </ul>	

## Diskussion und Ausblick

Wie den Ergebnissen zu entnehmen ist, existiert ein großes Spektrum von Anforderungen, Erwartungen und Wünschen an Erklärvideos zu geographischen Geländemethoden. Das führt in Hinblick auf deren Weiterentwicklung zu der Frage, ob und wie man diese zahlreichen, sich teilweise sogar widersprechenden Aspekte miteinander vereinbaren kann. Ohne größere Schwierigkeiten oder Widerspruch lassen sich die meisten gestalterischen Verbesserungsvorschläge realisieren. Viele Lehramtsstudierende merkten beispielsweise die monotone und langsame Sprechweise der erklärenden Person im Video an, die aber an sechs anderen Stellen positiv beurteilt wurde. Allerdings ist wohl nicht davon auszugehen, dass eine lebhaftere und schnellere Sprechweise auf Ablehnung bei den Probandinnen und Probanden stößt, die sich bereits positiv äußerten. Mit der Interpretation der Sprechweise als zu monoton und zu langsam sowie der Konsequenz, diese in einer Weiterentwicklung lebhafter und schneller zu gestalten, verbindet sich aber auch die Frage nach der Wahl eines richtigen Maßes, bevor es manch einer Adressatin oder manch einem Adressaten wiederum zu schnell sein könnte.

Das richtige Maß bzw. genaue Vorgaben für die Produktion von Erklärvideos, auch zu anderen Themen oder im Bereich anderer Fachwissenschaften, abzuleiten, kann durch die vorliegende Untersuchung nur bedingt erfolgen. Vielmehr bieten die beiden Erklärvideos zu geographischen Geländemethoden bezüglich der sogenannten gelungenen Aspekte (s. **Tabelle 2**) Beispiele und Orientierungen dafür, wie z. B. der Inhalt eines Videos strukturiert, Einblendungen gestaltet oder ein Praxisbezug konzipiert sein kann. Auf der anderen Seite (s. **Tabelle 1**) kann eine solche Orientierungsfunktion dafür, wie etwas verbessert oder nicht umgesetzt werden sollte, nicht unterstellt werden, da überwiegend themenspezifische Aspekte von den Probandinnen und Probanden geäußert wurden, die keine Verallgemeinerung zulassen. Dazu bedarf es einer weiteren Iteration der Entwicklung und fokussierten Evaluation im Sinne des Design Based Research-Ansatzes.

Die Vereinbarung der verschiedenen inhaltlichen und didaktischen Aspekte zur Verbesserung der zwei untersuchten Erklärvideos, deren Extreme in den Ergebnissen dargestellt wurden, scheint jedenfalls nicht möglich und wenig sinnvoll, wenn beispielsweise mehrere inhaltliche Ergänzungen und weitere Hinweise in ein noch kürzeres oder weniger komplexes Erklärvideo integriert werden sollen. Somit kann ‚ein Erklärvideo für alle‘ nicht die Lösung dieses Problems darstellen. Um die individuellen Anforderungen zu erfüllen, müsste wiederum mit für den jeweiligen Einzelfall adaptierten Formaten gearbeitet werden. Da eine solche Adaption bei nicht live produzierten Medien wie Erklärvideos in der Regel unmöglich ist, schlagen Wolf und Kulgemeyer (2016) vor, den Lernenden eine Auswahl von mehreren Erklärvideos zum selben Thema zur Verfügung zu stellen, sodass „[nicht der Erklärer den Erklärungsansatz, sondern der Rezipient den Erklärer bzw. Erkläransatz wechselt]“ (S. 39). Bei der großen Zahl bereits verfügbarer Erklärvideos zu obligatorischen Themen des schulischen

Lehrplans ist das unproblematisch, für das eher spezielle Thema der geographischen Geländemethoden, die an Lehrende gerichtet sind, ist es hingegen eine große Herausforderung.

In der Untersuchung konnte aber auch festgestellt werden, dass mehrere Lehramtsstudierende keine Beanstandungen an den Videos hatten und die Geländemethoden für sich zufriedenstellend durchführen und erklären konnten. Grundsätzlich kann man auch mit den Worten von 26w gegenargumentieren, dass „die Methode im Vordergrund stehen [soll]“ und sich die Lehrenden bei Bedarf weitere für sie relevante Informationen beschaffen und selbst auswählen können. Da dieses Vorgehen der üblichen Praxis entspricht und eine Weiterentwicklung und Evaluation von verschiedenen Varianten der Erklärvideos noch mehr Zeit und Ressourcen in Anspruch nimmt, wurde die Weiterführung des Design Based Research-Ansatzes in diesem Projekt verworfen. Einzelne Phasen dieses Prozesses werden aber gesondert fortgesetzt. So fokussiert eine künftige Arbeit des Autors den Einfluss von Erklärvideos auf den Prozess des Erklärens bei Lehramtsstudierenden.

## Literatur

Aff, J., & Schopf, C. (2017). Verständlich, problemorientiert und interessant Erklären – eine Selbstverständlichkeit? *wissenplus*, 1(16/17), 14–15.

Bandura, A. (1977). *Social learning theory*. Eaglewood Cliffs: Prentice-Hall.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs: Prentice-Hall.

Bitkom (2017). Nachhilfe im Netz: Steigende Nutzerzahlen für Video-Tutorials. Abgerufen am 04.03.2020 von <https://www.bitkom.org/Presse/Presseinformation/Nachhilfe-im-Netz-Steigende-Nutzerzahlen-fuer-Video-Tutorials.html>

Brown, G. A., & Daines, J. M. (1981). Can explaining be learnt? Some lecturers' views. *Higher Education*, 10(5), 573–580.

DGfG (Hrsg.) (2020). *Bildungsstandards im Fach Geographie für den Mittleren Schulabschluss – mit Aufgabenbeispielen*. Bonn: Eigenverlag.

Dresing, T., & Pehl, T. (2018). *Praxisbuch Interview, Transkription & Analyse. Anleitungen und Regelsysteme für qualitativ Forschende*. Marburg: Eigenverlag.

Falk, G. (2015). Exkursionen. In S. Reinfried, & H. Haubrich (Hrsg.), *Geographie unterrichten lernen* (S. 150–153). Berlin: Cornelsen.

Findeisen, S., Horn, S., & Seifried, J. (2019). Lernen durch Videos – Empirische Befunde zur Gestaltung von Erklärvideos. *Medienpädagogik: Zeitschrift für Theorie Und Praxis Der Medienbildung*, 2019(Occasional Papers), 16-36. doi: 10.21240/mpaed/00/2019.10.01.X

Hattie, J. A. C. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hensel T., & Kreuz S. (2018). (Um-)Wege im Feld: qualitative Fallauswahl zwischen Gegenstandskonstituierung und Feldbeschaffenheit. In M. Maier, C. Keßler, U. Deppe, A.

- Leuthold-Wergin, & S. Sandring (Hrsg.), *Qualitative Bildungsforschung. Studien zur Schul- und Bildungsforschung* (Bd. 68). (S. 75–92). Wiesbaden: Springer VS. doi: 10.1007/978-3-658-18597-8\_5
- Höhnle, S., & Schubert, J. C. (2016). Hindernisse für den Einsatz naturwissenschaftlicher Arbeitsweisen im Geographieunterricht aus Studierendenperspektive – Ausgewählte Ergebnisse einer empirischen Studie mit Lehramtsstudierenden. *GW Unterricht, 142/143* (2-3), 153–161. doi: 10.1553/gw-unterricht142/143s153
- Hußmann, S., Thiele, J., Hinz, R., Prediger, S., & Ralle, B. (2013). Gegenstandsorientierte Unterrichtsdesigns entwickeln und erforschen. In M. Komorek, & S. Prediger (Hrsg.), *Der lange Weg zum Unterrichtsdesign. Zur Begründung und Umsetzung fachdidaktischer Forschungs- und Entwicklungsprogramme* (S. 25–42). Münster: Waxmann.
- KMK (2008): *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung*. Abgerufen am 24.04.2020 von [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2008/2008\\_10\\_16-Fachprofile-Lehrerbildung.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf)
- Kuckartz, U. (2016). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Weinheim: Beltz Juventa.
- Kulgemeyer, C. (2020). A framework of effective science explanation videos informed by criteria for instructional explanations. *Res Sci Educ 50*, 2441–2462. doi: 10.1007/s11165-018-9787-7
- Lößner, M. (2011). *Exkursionen im Erdkundeunterricht: didaktisch gewünscht und in der Realität verschmäht? Ergebnisse einer empirischen Untersuchung an mittelhessischen Gymnasien*. Geographiedidaktische Forschungen, Vol. 48, Münster: Monsenstein und Vannerdat.
- Miltz, R. (1972). *Development and evaluation of a manual for improving teachers' explanations*. Technical Report No. 26, Stanford University.
- Rat für kulturelle Bildung e.V. (2019). *Eine repräsentative Umfrage unter 12- bis 19-Jährigen zur Nutzung kultureller Bildungsangebote an digitalen Kulturorten*. Abgerufen am 19.06.2020 von [https://www.rat-kulturelle-bildung.de/fileadmin/user\\_upload/pdf/Studie\\_YouTube\\_Webversion\\_final\\_2.pdf](https://www.rat-kulturelle-bildung.de/fileadmin/user_upload/pdf/Studie_YouTube_Webversion_final_2.pdf)
- Rinschede, G. & Siegmund, A. (2019). *Geographiedidaktik*. Paderborn: Schöningh.
- Vogt, R. (2009). Die Organisation von Erklärprozessen im Unterricht. In R. Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven* (S. 201–225). Tübingen: Stauffenburg.
- Wolf, K. (2015). Bildungspotentiale von Erklärvideos und Tutorials auf YouTube. *merz - medien + erziehung – zeitschrift für medienpädagogik, 1*(59), 30-36.
- Wolf, K., & Kulgemeyer, C. (2016). Lernen mit Videos? Erklärvideos im Physikunterricht. *Naturwissenschaften im Unterricht Physik, 27*(152), 36-41.

Zander, S., Behrens, A., & Mehlhorn, S. (2018). Erklärvideos als Format des E-Learnings. In H. M. Niegemann, & A. Weinberger (Hrsg.), *Lernen mit Bildungstechnologien: Praxisorientiertes Handbuch zum intelligenten Umgang mit digitalen Medien* (S. 1–12). Berlin und Heidelberg: Springer.

**Maximilian Breuer** ist wissenschaftlicher Mitarbeiter und Lehrkraft für besondere Aufgaben in der Didaktik der Geographie an der Martin-Luther-Universität Halle-Wittenberg. Er beschäftigt sich mit dem Einsatz digitaler Medien in der geographiedidaktischen Lehre und forscht zu Erklärprozessen und Erklärvideos im Kontext von Outdoor Education.

[maximilian.breuer@geo.uni-halle.de](mailto:maximilian.breuer@geo.uni-halle.de)

# Gamification im MINT-Unterricht: Eine Untersuchung des Spielverhaltens, Nachhaltigkeits- und Immersionserleben am Beispiel des Serious Game LandYOUs

*Nils L. Kotschote<sup>1</sup>, Chris Wenzel<sup>1</sup> und Martin Lindner<sup>1</sup>*

*<sup>1</sup> Martin-Luther-Universität Halle-Wittenberg, Institut für Biologie, Didaktik der Biologie*

Mit Ausweitung der Digitalisierung der Lehre in Bildungseinrichtungen erweitert sich stets das Angebot an digitalen Medien, Anwendungen und Apps. Diese Vielzahl an Inhalten verringert die Übersichtlichkeit für Lehrende, wodurch viele den Nutzen der neuen Angebote zunächst kritisch betrachten. Auf dieser Basis baut die vorliegende Arbeit auf. Die Grundlage ist eine eigenständig durchgeführte Studie über das digitale Serious Games LandYOUs. Dabei war das Ziel die Untersuchung des Lerneffekts und des Immersionserlebens für den Einsatz im Bereich der Fächer des MINT- und gesellschaftswissenschaftlichen Unterrichts an Schulen. Insgesamt haben 46 Personen aus zwei Zielgruppen, Schülerinnen und Schüler / junge Erwachsene teilgenommen. Im Rahmen der Datenerhebung wurden Fragebögen und Eye-Tracking eingesetzt. In den folgenden dargestellten Ergebnissen werden Lernerfolge und Änderungen von Sichtweisen bei den Teilnehmenden durch Nutzung des Spiels nachgewiesen. Weiterhin ist ein hoher Grad an Immersion messbar, was auf eine hohe Akzeptanz von LandYOUs bei den Teilnehmenden hinweist. Damit ist diese Publikation für Personen in der digitalen Lehre, besonders Lehrende an Schulen, interessant und zeigt neue Chancen bei der Gestaltung von Unterricht auf.

## Einleitung

Was ist guter Unterricht? Auf Basis dieser Frage untersuchten zahlreiche Studien die Rahmenbedingungen von Wissensvermittlung und –erarbeitung im Kontext des Schulunterrichts. Beispielsweise wurden bei der 2009 publizierten ausführlichen Metauntersuchung zum Thema Lernerfolg von John Hattie ca. 800 Metauntersuchungen mit 80.000 Einzeluntersuchungen, welche sich dem Thema Unterricht widmen, ausgewertet (Hattie, 2009). Ein Faktor für einen guten Unterricht stellt unter anderem der zweckmäßige und adressatengerechte Einsatz von Medien dar. Einhergehend mit der zunehmenden Digitalisierung der Gesellschaft rücken ebenso die digitalen Medien mehr und mehr in den Fokus der Bildung.

Eine Möglichkeit zur Nutzung dieser Medien und des daraus generierten Potenzials für den Unterricht ist Gamification. Der Begriff stammt aus dem Jahr 2002 von dem Spielentwickler Pelling (2011) und umfasst neben dem Einsatz von Computerspielen auch die generelle Nutzung aller digitalen Anwendungen mit spieltypischen Designelementen (Deterding, Dixon, Khaled, & Nacke, 2011). Untersuchungen zeigten, dass demnach die Lernenden als Adressaten motivierter sind, sich mit dem Lerngegenstand auseinanderzusetzen (Lefers & Birkenkrahe, 2016).

Als Beispiel für die Gamification im Bereich des MINT-Unterrichts soll die Nutzung des Spiels LandYOUs zu Grunde gelegt werden, welches den zentralen Gegenstand der Studie darstellt. LandYOUs ist frei zugänglich und wird online gespielt. Hinsichtlich des Genres wird es den Serious Games zugeordnet. Das Hauptziel dieser Art von Spielen ist nicht der Unterhaltungsaspekt, sondern die spielerische Wissensvermittlung in simulierten Szenarios. Dadurch tritt eine Integration von Elementen aus Simulationen, Unterhaltungsspielen und Lernanwendungen auf (Belotti, Berta & De Gloria, 2010). Entwickelt wurde LandYOUs 2014 vom Helmholtz-Zentrum für Umweltforschung. Im Jahr 2017 wurde das Spiel in Kooperation mit dem Klett-Verlag durch eine Handreichung für den Unterricht ergänzt (Berning et al., 2016). Das Spiel ist frei zugänglich und wird online gespielt. Da die Bedeutung für die Motivation aus den Untersuchungen der Biologiedidaktik der vergangenen Jahre belegt ist, soll in der vorliegenden Studie der Lerneffekt weiter eruiert werden.

Thematisch ist LandYOUs im Bereich der Bildung für nachhaltige Entwicklung, mit Schwerpunkt auf nachhaltiges Landmanagement, zu verorten (Schulze et al., 2014). Die Spielenden übernehmen die Rolle des Präsidenten von Ecotania, dem fiktiven Land, das es zu verwalten gilt. Sie müssen durch Investitionen in verschiedenen Sektoren ihre Nation gleichzeitig zu ökonomischen, ökologischen und sozialen Erfolgen führen. Durch Rückkopplungen ihrer Entscheidungen und verschiedene, im Algorithmus integrierte, Zufallsmechanismen ist jeder Spieldurchgang individuell.

Seit 2015 wird der Einsatz des Serious Game LandYOUs in der Biologiedidaktik durch didaktische Forschungsmethoden begleitet. Es wurde zunächst in Deutschland, ab 2016 auch in verschiedenen Ländern (Slowakei, Tschechien, Finnland, Türkei, Nepal), in studentischen Gruppen eingesetzt, später dann auch mit Schülerinnen und Schülern erprobt. Dazu wurde es in verschiedene Sprachen übersetzt. Die Resultate wurden in Workshops für Lehrkräfte

eingebettet, die auch im Rahmen eines ERASMUS+ Programms (SciVis, Laufzeit 2014-2016) international veranstaltet sowie in die Veröffentlichung des Klett-Verlags eingebracht wurden. Die bisherigen Ergebnisse zeigen, dass LandYOUs sehr hohe motivationale Elemente besitzt und von Schülerinnen und Schülern sehr gut als Medium in der Lehre angenommen wird. Von dieser Basis ausgehend widmet sich die vorliegende Studie der Untersuchung von Lern-effekten, dem Nachhaltigkeits-, sowie dem Immersionserleben.

Der Begriff Immersionserleben bezieht sich auf den Grad des Versinkens der Probandinnen und Probanden im Spiel. Nach Brown und Cairns (2004) gibt es drei Stufen der Immersion: die Teilnahme, Vertiefung und die totale Immersion. Die Merkmale der Stufe Teilnahme sind das Erlernen der Bedienung und das Sich-Bewusst-Werden als Spielender. Die Vertiefung ist über einen emotionalen Bezug zum Spiel und einem Verlust des Zeitgefühls definiert. In der letzten Stufe ist die Person vollkommen im Spiel versunken. Der Spielende nimmt das Spiel als real wahr und verliert das Bewusstsein für die eigene Umwelt. Diese Immersionsstufe besitzt mit dem 1990 von Csikszentmihalyi beschriebenen Flowleben Gemeinsamkeiten. Für Serious Games ist dieses Spielerleben von Hoblitz (2014) erforscht. Der Vergleich beider zeigt, dass Immersion im Gegensatz zu Flow bei jedem Spielvorgang auftritt. Dadurch ist die Erfassung von Immersion vorteilhafter.

Ziele der Studie sind die Erörterung des Spielverhaltens, der Immersion und der Veränderung des Nachhaltigkeitserlebens, woraus sich folgende Forschungsfragen ergeben:

1. Welche Muster oder Strategien können im Spielverhalten der Probandinnen und Probanden beim Spielen von LandYOUs in Hinblick auf den erfolgreichen Abschluss nachgewiesen werden?
2. Inwiefern kann bei dem Spielen von LandYOUs ein Immersionserleben in Abhängigkeit von den Methoden des Fragebogens und des Eye-Trackings nachgewiesen werden?
3. Welche Veränderungen ergeben sich im Nachhaltigkeitserleben bei den Probandinnen und Probanden aufgrund des Spielens von LandYOUs? Welchen Einfluss haben die Faktoren Immersion und Spielerfolg auf die Veränderung des Nachhaltigkeitserlebens?

Die Auseinandersetzung mit diesen Forschungsfragen erfolgt über einen experimentellen Ansatz. Neben einer Erhebung von Ratingwerten ist die Messung von Augenbewegungen ein zentraler Bestandteil der Studie. Damit soll eine Beantwortung der Forschungsfragen mit mehreren, möglichst objektiven Parametern gewährleistet sein. Zusätzlich wird in dieser Studie ein neuer Ansatz untersucht. Durch eine Kombination beider Methoden sollen erstmals Aussagen über die Aufmerksamkeitsfokussierung von Probandinnen und Probanden und ihrem Immersionserleben getroffen werden.

## Methodischer Ansatz

Die vorliegende Studie untersuchte in jeweils einer eigenen Teiluntersuchung drei verschiedene Probandengruppen. In der ersten Gruppe wurden 33 Studierende im Alter zwischen 18 bis 35 Jahren erfasst (21 W; 12 M). Innerhalb der mittels Eye-Tracking erhaltenen Ergebnisse verringert sich die Anzahl der Stichprobenstärke auf 32. Dies ist auf technische Schwierigkeiten bei der Aufzeichnung aufgrund einer sehr hohen Brillenstärke der Versuchsperson zurückzuführen. Demnach gibt es eine Variation der Stichprobengröße von 32 oder 33 in Abhängigkeit der Verwendung von Daten aus dem Eye-Tracking. Die zweite Gruppe beinhaltet eine Untersuchung von 13 Probandinnen und Probanden im Alter zwischen 14 und 18 Jahren (8 W; 5 M). Diese Probandinnen und Probanden sind Schülerinnen und Schüler aus zwei Gymnasien der Stadt Halle (Saale). In der letzten Gruppe sind 20 Personen in einem Alter von 35 bis 60 Jahren erfasst worden (8 W; 12 M). Die Teilnahme an der Studie erfolgte freiwillig und ohne Vergütung. Die Ergebnisse der dritten Gruppe wurden bereits in einer wissenschaftlichen Hausarbeit ausgewertet und veröffentlicht.

Für die Datenerhebung werden zwei Fragebögen und Eye-Tracking als quantitative Untersuchungsmethode verwendet. Die Durchführung der Studie erfolgte in einem Computerlabor des Zentrums für multimediales Lehren und Lernen (LLZ) der Martin-Luther-Universität Halle-Wittenberg. Dabei erfolgte die Datenerhebung für jede Versuchsperson einzeln.

Zeitlich untergliederte sich der Ablauf folgendermaßen: Zunächst füllte die Versuchsperson einen ersten Fragebogen aus. Danach erfolgte eine kurze Beschreibung der Benutzeroberfläche von LandYOUs durch den Versuchsleiter. Nach der Kalibrierung des Eye-Trackers wurden 25 Minuten gespielt. Während der praktischen Durchführung ist nach einem realen Spielabschluss, also einem Game Over oder einem Sieg, der Vorgang beendet worden. Dadurch entstanden Abweichungen von ca. fünf Minuten. Zuletzt füllte der Proband einen weiteren Fragebogen aus.

Im Fragebogen werden, neben den allgemeinen Angaben zur Person, in einer Prä-Post-Evaluation 27 Items zum Thema Nachhaltigkeit in sieben verschiedenen Dimensionen erfasst. Diese Dimensionen sind der *eigene individuelle Informationsgrad zur Thematik*, die *Bewertung von Nachhaltigkeit*, die *Bedeutung von Bildung*, der *Zukunftsaspekt* und die *soziale Komponente von Nachhaltigkeit*, die *Einschätzungen der eigenen Lebensweise hinsichtlich Nachhaltigkeit* und die *Beurteilung der Lebensweise der allgemeinen Bevölkerung*.

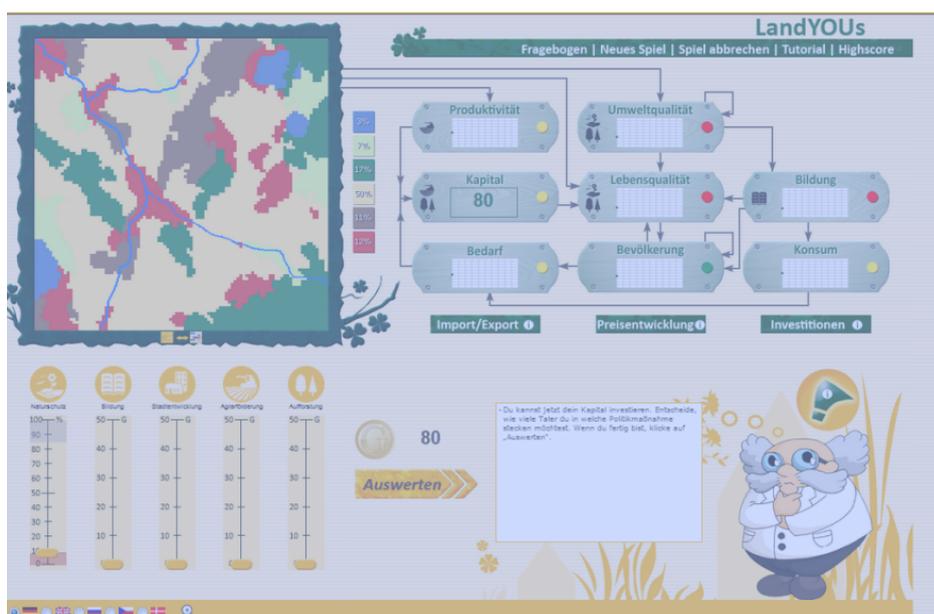
Bei den späteren Stichproben der Teiluntersuchungen in Phase zwei und drei sind die Dimensionen um die Faktoren *Einfluss der Politik*, *ökonomische Nachhaltigkeit* und *Erweiterung der ökologischen Nachhaltigkeit* ergänzt. Die dafür genutzten Items basieren auf einer Weiterentwicklung des Science-Camps Fragebogen der Biologiedidaktik der Martin-Luther-Universität und auf explizit im Rahmen dieser Studie neu entwickelte Items.

Im zweiten Fragebogen sollen die Probandinnen und Probanden, zusätzlich zu den im ersten Fragebogen erhobenen Daten, Angaben über das Spiel- und Immersionserleben tätigen. Die Grundlage der Immersionsuntersuchung dieser Arbeit bildet die Studie von Jennett (2008). Dazu sind 25 Items aus dem Englischen übersetzt und auf das Spiel LandYOUs

angepasst worden. Die Items werden in der Auswertung den drei Immersionsstufen zugeordnet. Dabei unterteilen sich diese Stufen nochmals in verschiedene Dimensionen. Die erste Immersionsstufe umfasst die Dimensionen *Einschätzung der Grafik*, der *Bedienung des Spiels*, der *eigenen Leistung* und der *Herausforderung* durch das Spiel. Zu der zweiten Stufe zählen die *Bereitschaft weiter zu spielen*, der *emotionale Bezug zum Spiel* und *das Zeitempfinden*. Bei der letzten Stufe werden die Probandinnen und Probanden nach ihrem wahrgenommenen Umweltbezug und der Empfindung des Spiels als „real“ befragt. Zusätzlich sollen die Probandinnen und Probanden ihre Versunkenheit im Spiel selbst beurteilen.

Allgemein werden zu jeder abgefragten Dimension 2 bis 4 Items zugeordnet. Für die Einschätzung der Items findet die fünfstufige Likert-Skala Verwendung. Die Erfassung der persönlich erlebten Immersion erfolgt auf einer zehnstufigen Skala. Die vorherige Pilotierung des Fragebogens ist in zwei Phasen mit 31 und 11 Probandinnen und Probanden erfolgt.

Mit dem Eye-Tracking wurden während des Spielens die Augenbewegungen der Testpersonen auf dem Bildschirm aufgezeichnet und nach sogenannten Areas Of Interest (kurz AOI) ausgewertet. Areas Of Interest sind durch die auswertende Person definierte und anschließend markierte Flächen in der aufgezeichneten Anwendung, beispielsweise in der **Abbildung 1** blau unterlegt. Mit Hilfe der AOI kann das Auswertungsprogramm die Gesamtheit der Daten zeit- und bereichsspezifisch filtern. Dadurch werden hinsichtlich der Fragestellung nur relevante Augenbewegungen ausgewertet. Verwendet wurden drei große Overlays von AOI. Die Immersionsuntersuchung baute auf zwei Overlays auf. Im ersten wurden Fixationen per Zeiteinheit gemessen. Dafür wurde für eine Zeitspanne von je zwei Minuten ein einzelnes AOI über den gesamten Spielbereich gelegt (**Abbildung 1**). Die Ränder des Bildschirms wurden nicht berücksichtigt, da nur die Anzahl der Fixationen im aktiven Spielbereich relevant ist.



**Abbildung 1.** Schematische Darstellung für eine Area Of Interest über den gesamten Spielbereich

Im zweiten Immersionsoverlay erfolgte eine Aufmerksamkeitsmessung. Bestimmt wurde diese über den Zeitanteil, in welchem der Proband auf das Spielfeld blickte. Das AOI der Aufmerksamkeit ist in Bezug auf die Position der **Abbildung 1** ähnlich. Es ist ein wenig vergrößert und umfasste zeitlich die gesamte Spieldauer. Ziel des leicht vergrößerten AOI ist, auch kurze Überlegungen zum Spiel mit einzubeziehen, was durch einen leicht abschweifenden Blick charakterisiert ist, obwohl sich die Person gedanklich noch im Spiel befand.

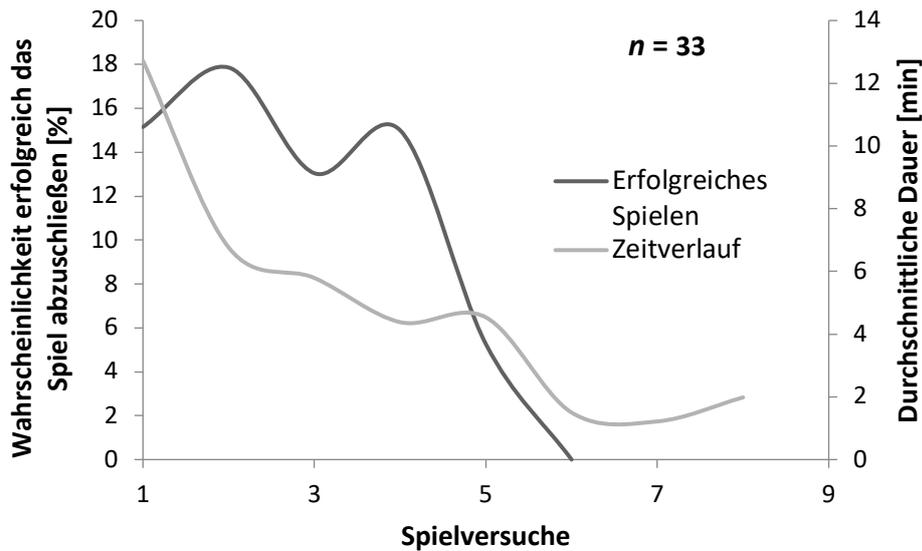
Das dritte Overlay wurde für die Erfassung des Spielverhaltens genutzt. Es umfasste sieben verschiedene AOI, welche für das Spiel relevante Bereiche abdecken. Durch den prozentualen Anteil der einzelnen AOI im Verhältnis zur gesamten Betrachtungszeit konnten Rückschlüsse auf Spielstrategien geschlossen werden.

## Ergebnisse

Die Auswertung der Daten, die mithilfe des Eye-Trackings erhoben wurden, geschah mit dem Programm Tobii Pro Lab. Die statistischen Berechnungen wurden mit den Programmen SPSS und PSCP umgesetzt. Verwendet wurden die Korrelationsrechnung nach Spearman, die lineare Regressionsanalyse, der Kolmogorov-Smirnov-Test und t-Tests für abhängige Stichproben.

Zuerst soll die Forschungsfrage zum Spielverhalten beantwortet werden. Die meisten dargestellten Daten sind explorativ erhoben. In der **Abbildung 2** sind die Maxima an durchschnittlich erreichten sowie die durchschnittliche Anzahl an beendeten Spielrunden abgebildet. 17 Probandinnen und Probanden haben das Spiel erfolgreich abgeschlossen, während 16 bereits vor Ende der Spielzeit Game Over gegangen sind. Bei genauerer Betrachtung der Verteilung ist zu erkennen, dass sich verschiedene Gruppierungen bilden. Unter Anwendung der k-Means-Clusteranalyse mit sich unterscheidenden Zahlen sind drei verschiedene, realistische Cluster gefunden. Das erste umfasst das nicht erfolgreiche Spielen, mit einer Wertespanne von 4 bis 6 bei den maximal erreichten Runden. Das zweite Cluster stellt das durchschnittlich erfolgreiche Spielen dar. Es beinhaltet die durchschnittliche Rundenzahl von 5 bis 8, bei der maximal erreichten Runde 10. Das dritte Cluster liegt bei dem Wert von ca. 10 auf beiden Achsen. Aus der Abbildung geht hervor, dass Probandinnen und Probanden, die die Runde 5 oder 6 schaffen, nicht mehr in den kommenden Runden Game Over gehen werden, sondern die Runde 10 erreichen. Dadurch ist der Unterschied zwischen dem Cluster 1 und den anderen beiden am auffälligsten.





**Abbildung 3.** Kombinierte Darstellung der Wahrscheinlichkeit das Spiel erfolgreich abzuschließen und dem durchschnittlichen zeitlichen Verlauf der Spielversuche der Probandinnen und Probanden

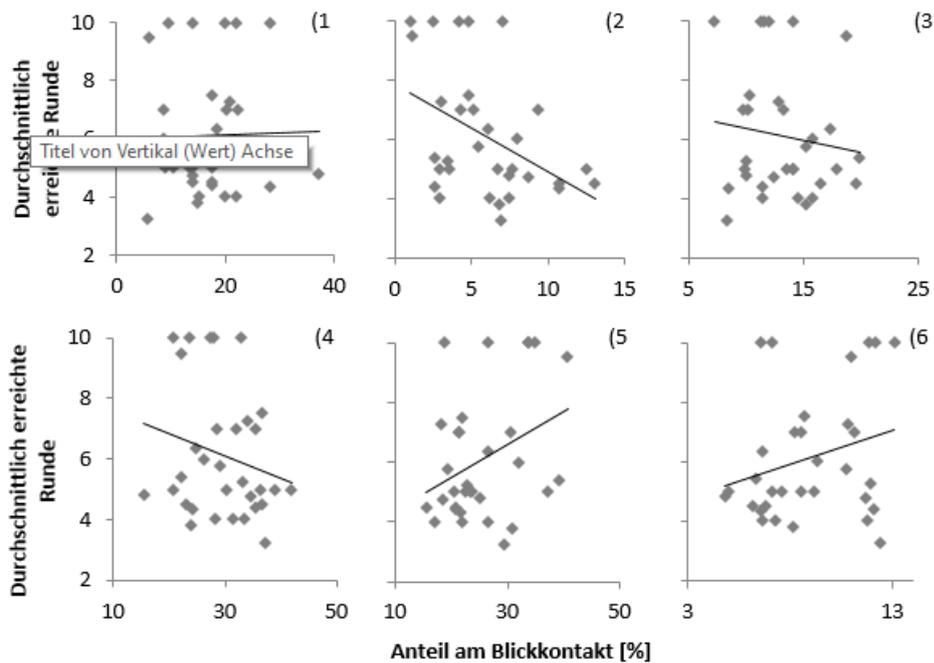
Zuletzt erfolgt die Darstellung der Ergebnisse der Spielbereich-AOI. Hierfür ist die Dauer des Blickkontakts mit jedem AOI bestimmt worden. Aus der Summe aller AOI-Zeiten wird der Prozentsatz des Blickkontakts mit dem Bereich ermittelt. In der **Abbildung 4** werden die erhaltenen Ergebnisse mit linearer Regression abgebildet. Der Startbildschirm ist als AOI nicht aufgelistet, er wird in der Berechnung nicht berücksichtigt. Der Test auf Regression zeigt lediglich bei dem AOI *Karte* einen signifikanten Zusammenhang (**Tabelle 1**).

**Tabelle 1.** Ergebnisse des Tests der AOI auf lineare Regression

AOI	$p$	$R^2$
Diagramme	.675	.01
Karte	.016*	.18
Unterer Rechter Bildschirm mit Geld	.418	.02
Regler	.172	.06
Auswertungsbildschirm	.061	.11
Pressespiegel	.172	.06

**Anmerkungen.**  $n = 32$ . Korrelation nach Spearman. \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

Bei Betrachtung des Diagramms ist zu erkennen, dass das AOI einen negativen Effekt auf den Spielerfolg hat. Trotz der fehlenden weiteren Regressionssignifikanzen kann eine graphische Einschätzung erfolgen, die negative Zusammenhänge der AOI *Regler* und dem unteren rechten AOI mit dem Spielerfolg erkennen lässt. Positive Effekte treten bei dem *Auswertungsbildschirm* und dem *Pressespiegel* auf.

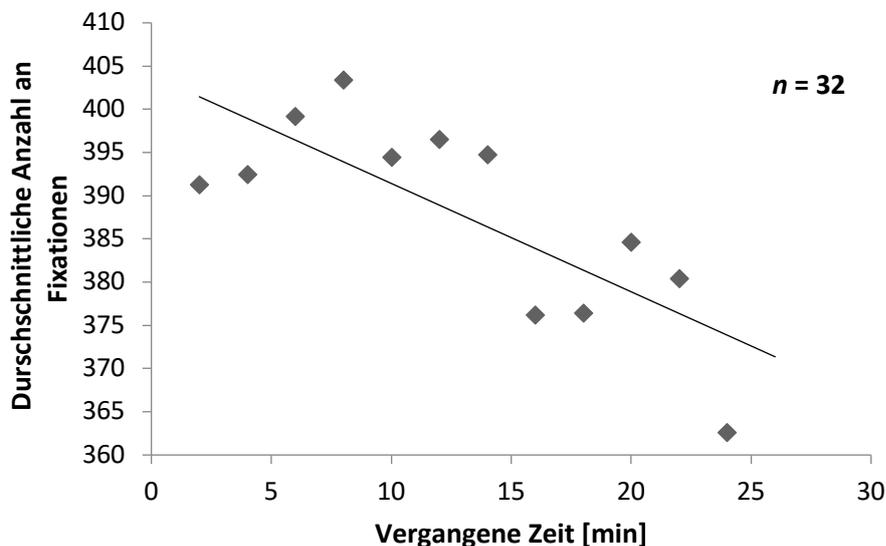


**Abbildung 4.** Zeitanteil der AOI im Bezug zur durchschnittlich erreichten Runde der Probandinnen und Probanden mit linearer Regression. Reihenfolge der AOI: Diagramme (1, Karte (2, unterer Rechter Teil (Geld und Landstein) (3, Regler (4, Auswertungsbildschirm (5 und Pressespiegel 6)

Zur Beantwortung der Forschungsfrage 1) Welche Muster oder Strategien können im Spielverhalten der Probandinnen und Probanden beim Spielen von LandYOUs in Hinblick auf den erfolgreichen Abschluss nachgewiesen werden? können folgende Aussagen getroffen werden:

- Es gibt Auffälligkeiten beim Spielen in Bezug zur Runde 5 oder 6.
- Mit zunehmenden Versuchen sinkt die benötigte Zeit des Spielversuchs.
- Die Wahrscheinlichkeit, im Spiel in Runde 10 zu kommen, deszendiert ab dem 5. Versuch stark.
- In Abhängigkeit zu der Betrachtungsdauer einzelner AOI variiert die Erfolgswahrscheinlichkeit im Spiel.

Nach der Untersuchung des Spielerverhaltens folgen die Ergebnisse im Hinblick auf das Immersionserleben. Zuerst wird die durchschnittliche Anzahl der Fixationen je Zeiteinheit bestimmt, um auf das Vorhandensein von Immersion zu prüfen (**Abbildung 5**).



**Abbildung 5.** Bestimmung des Immersionsauftretens nach Jennett [2008] über die durchschnittliche Anzahl an Fixation je Zeiteinheit mit linearer Regression

Die lineare Regressionsanalyse ergibt ein Signifikanzniveau von  $p = .004$  bei einem  $R^2$ -Wert von  $.58$ . Anhand der Ergebnisse ist von einem fallenden linearen Graphen auszugehen, mit einer Verringerung der Fixationen pro Zeiteinheit.

Die nachfolgende **Tabelle 2** zeigt die Ergebnisse der Korrelationsanalyse der Immersionsstufen, berechnet und selbst angegeben, sowie den Zeitanteil der Aufmerksamkeit. Die berechnete Immersion wird aus den Werten der drei Immersionsstufen gebildet.

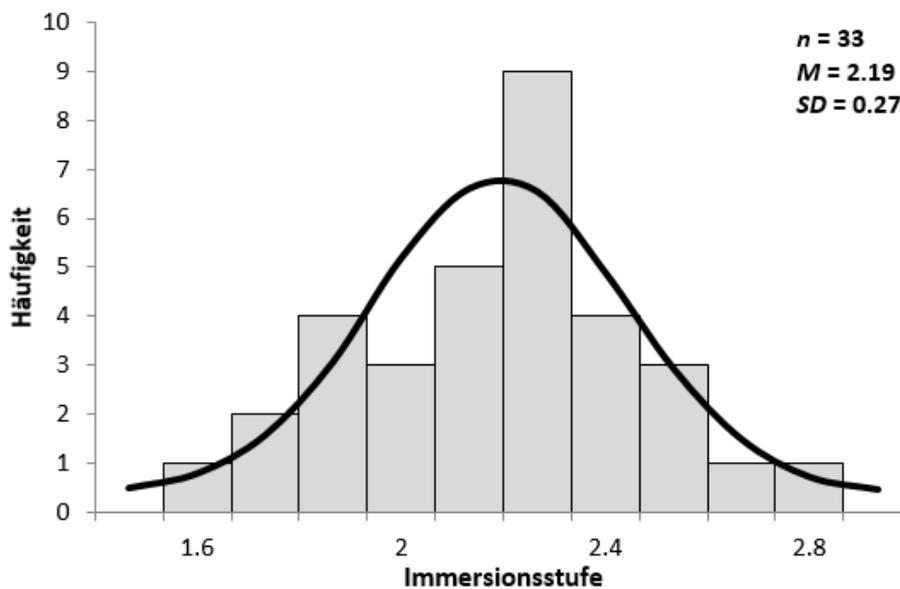
**Tabelle 2.** Immersion des Serious Game LandYOUs

	DR	Immersion	
		Berechnet	Selbst angegeben
Berechnete Immersionsstufen	.19		
Selbst angegebene Immersion	.11	.72***	
Zeitanteil Aufmerksamkeit	.10	.38*	.53**

**Anmerkungen.**  $N = 33$ . bei der prozentualen Aufmerksamkeit  $n = 32$ . Korrelation nach Spearman. DR = Durchschnitt der erreichten Runden aller Spielversuche der einzelnen Probandinnen und Probanden. \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

Die Tabelle zeigt eine höchst signifikante Korrelation zwischen der selbst eingeschätzten und der berechneten Immersion, sowie ein Signifikanzverhalten zwischen der Immersion und dem Zeitanteil der gemessenen Aufmerksamkeit. Der Spielerfolg, dargestellt in erreichten Runden, hat keinen messbaren Einfluss auf das Immersionserleben der Probandinnen und Probanden.

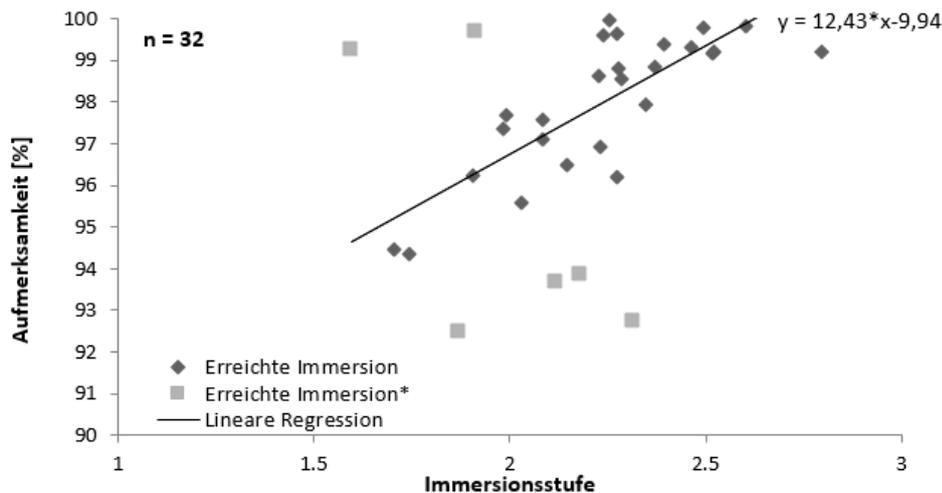
Resultierend aus dem hohen Signifikanzniveau wird für die weitere Bestimmung der erreichten Immersionsstufen der Mittelwert aus der berechneten und der selbst eingeschätzten Immersion gebildet. Nachfolgend werden die erreichten Immersionsstufen auf eine mögliche Normalverteilung untersucht. Dazu ist ein Kolmogorov-Smirnov-Test durchgeführt worden. Die extremste Differenz beträgt für die Stichprobe 0,11. Dieser Wert ist kleiner als der Literaturwert von 0,231, womit die theoretische Annahme auf eine Standardverteilung nicht abgelehnt wird. Die asymptotische Signifikanz liegt mit einem Wert von 0,845 ebenfalls über dem Grenzwert von 0,05. Die grafische Auswertung der **Abbildung 6** bestätigt ebenso eine Normalverteilung der Immersionsstufen.



**Abbildung 6.** Häufigkeitsverteilung der Immersionsstufen der Probandinnen und Probanden mit Normalverteilung

**Anmerkungen.** Die Spannweite der Immersion geht von dem Wert 1 bis 3. Sie ist analog zu den drei Stufen der Immersion.

Die letzten Ergebnisse der Immersionsuntersuchung umfassen die gemessene Aufmerksamkeit mithilfe des Eye-Trackings. In der **Abbildung 7** sind die Aufmerksamkeitsanteile der Probandinnen und Probanden zu den jeweils erreichten Immersionsstufen abgebildet.

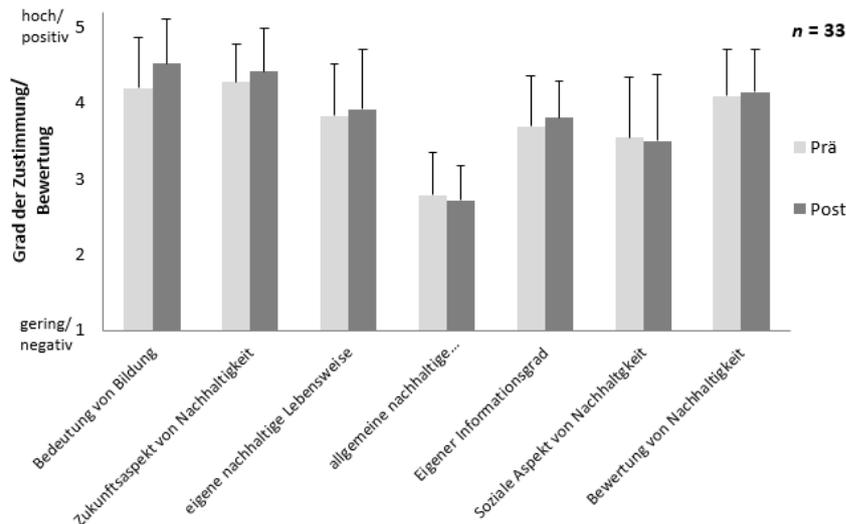


**Abbildung 7.** Anteil der Aufmerksamkeit am Spiel im Bezug zu der erreichten Immersionsstufe  
**Anmerkungen.** \* Die grau unterlegten Punkte sind bei der linearen Regression nicht berücksichtigt.

Aus der Abbildung geht eine gleichverteilte Häufung der erreichten Immersionsstufen zur gemessenen Aufmerksamkeit hervor. Für die Untersuchung der Häufung ist ein Test auf lineare Regression durchgeführt worden. Die extremsten Werte wurden hellgrau unterlegt und sind ausgeschlossen. Das Ergebnis des Regressionstests ergibt mit einem  $R^2$  von .65, einer Signifikanz  $\alpha < .001$  und  $t \neq 0$  eine Annahme des linearen Zusammenhangs (ohne Korrektur  $R^2 = .22$  und  $\alpha = .007$ ).

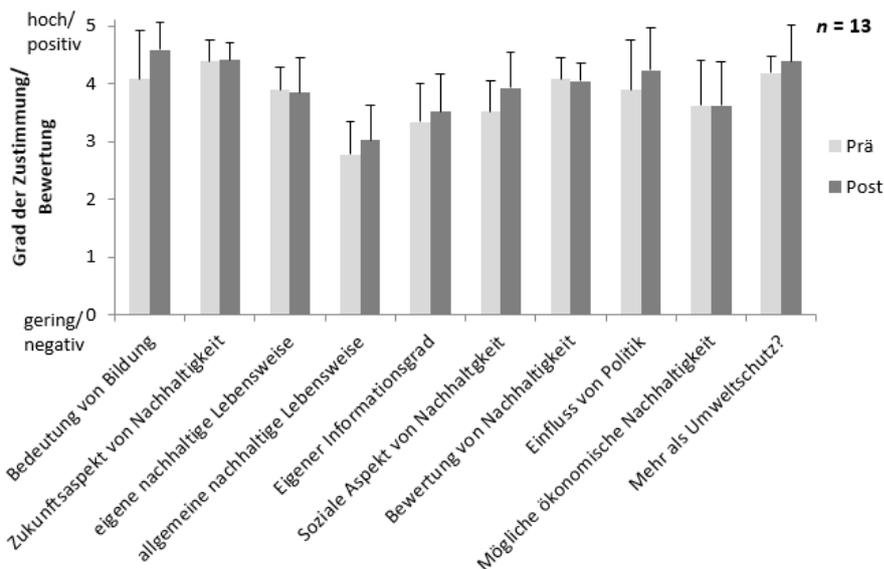
Im Hinblick auf die Beantwortung der zweiten Forschungsfrage Inwiefern kann bei dem Spielen von LandYOUs ein Immersionserleben in Abhängigkeit von den Methoden des Fragebogens und des Eye-Trackings nachgewiesen werden? ergibt sich, dass es nachgewiesen zu einer messbaren Immersion während des Spielens des Serious Game LandYOUs kommt. Die durchschnittlich erreichte zweite Immersionsstufe deutet auf ein Vorhandensein eines emotionalen Bezugs zum Spiel hin. Tendenziell trifft auf die Probandinnen und Probanden weiterhin ein Verlust des Realitätsgefühls zu. Die im Rahmen der Studie genutzte Methode des Eye-Trackings zeigt für die gemessene Aufmerksamkeit signifikante Korrelationen mit der durch den Fragebogen bestimmten Immersionsstufe.

Die Ergebnisse zur Beantwortung der letzten Forschungsfrage mit Fokus auf das Nachhaltigkeitserleben basieren auf der Prä-Post-Erhebung mittels der Fragebögen. Für die Stichprobe der ersten Phase, welche die Altersgruppe zwischen 18 und 35 Jahren erfasst, werden die Mittelwerte der Dimensionen vor und nach dem Spielen in der nachfolgenden **Abbildung 8** dargestellt.



**Abbildung 8.** Prä-Post-Vergleich der Mittelwerte der einzelnen Dimensionen des Nachhaltigkeitserlebens der Probandinnen und Probanden im Alter von 18-35 Jahren. Die Fehlerbalken repräsentieren die Standardabweichung.

Ersichtlich ist ein starker Anstieg der Mittelwerte im Bereich Bildung ( $\Delta M = 0,32$ ) und ein Anstieg in den Bereichen Zukunftsaspekt von Nachhaltigkeit und individueller Informationsgrad ( $\Delta M > 0,1$ ). Bezüglich der vier weiteren Dimensionen sind nur geringfügige Unterschiede erkennbar. Für eine Antwort auf den ersten Teil der letzten Forschungsfrage *Welche Veränderungen ergeben sich im Nachhaltigkeitserleben bei den Probandinnen und Probanden aufgrund des Spielens von LandYOU?* ist ein t-Test für abhängige Stichproben durchgeführt worden. Bei der Bildungsdimension ergibt sich eine signifikante Veränderung durch das Spielen,  $t(33) = -3,46, p < 0,002$ . In den anderen sechs Dimensionen liegt kein repräsentatives Signifikanzniveau vor.



**Abbildung 9.** Prä-Post-Vergleich der Mittelwerte der einzelnen Dimensionen des Nachhaltigkeitserlebens der Probandinnen und Probanden im Alter von 14-18 Jahren. Die Fehlerbalken repräsentieren die Standardabweichung.

In der **Abbildung 9** sind die Ergebnisse der Stichprobe aus der zweiten Phase, welche Testpersonen im Alter zwischen 14 und 18 Jahren erfasst, dargestellt. Bei den Mittelwerten sind bei Betrachtung der bildungsbezogenen ( $\Delta M = 0,5$ ), der politischen ( $\Delta M = 0,35$ ) und der sozialen Dimension von Nachhaltigkeit ( $\Delta M = 0,41$ ) starke Zuwächse wiederzufinden. Weiterhin sind Anstiege bei den Mittelwerten der Dimensionen Einschätzung der allgemeinen nachhaltigen Lebensweise, des eigenen Informationsgrads und der Erweiterung des Nachhaltigkeitsbegriffs erkenntlich ( $\Delta M > 0,15$ ). Eine Signifikanzüberprüfung erfolgte erneut mit t-Tests für abhängige Stichproben. Die Bildungsdimension zeigt eine sehr signifikante Veränderung durch das Spielen,  $t(13) = -2,55$ ,  $p < 0,025$  und eine höchst Signifikante bei dem sozialen Aspekt der Nachhaltigkeit,  $t(13) = -4,79$ ,  $p < 0,000$ .

Um eine Aussage über den zweiten Teil der letzten Forschungsfrage *Welchen Einfluss haben die Faktoren Immersion und Spielerfolg auf die Veränderung des Nachhaltigkeitserlebens?* abzuleiten wird in der **Tabelle 3** die Korrelationsanalysen der Faktoren Spielstrategien und Immersionserleben mit den Dimensionen der Nachhaltigkeit abgebildet. Die Grundlage der Veränderung im Nachhaltigkeitserleben bildet die Differenz der Mittelwerte aus der Prä-Post-Untersuchung.

**Tabelle 3.** Wechselbeziehungen der Änderung von Nachhaltigkeitsdimensionen im Prä- und Posttest mit den durchschnittlich erreichten Runden und der Immersion der Probandinnen und Probanden im Alter von 18 bis 35 Jahren

Dimensionen	DR	Nachhaltigkeitsdimensionen						Bewertung
		Bildung	Eigene Lebensweise	Zukunft	Allgemeine Lebensweise	Selbstinformation	Soziale Nachhaltigkeit	
Bildung	-.17							
Eigene Lebensweise	-.11	.00						
Zukunft	.10	-.00	-.16					
Allgemeine Lebensweise	-.14	.02	-.27	.03				
Selbstinformation	.38*	.07	-.03	-.08	-.08			
Soziale Nachhaltigkeit	.25	-.11	-.45**	.30	.34	-.08		
Bewertung	.25	.26	.21	.19	-.20	.28	-.00	
Immersion	.16	.04	-.21	.17	-.05	.02	.12	.19

**Anmerkungen.**  $n = 33$ . Korrelation nach Spearman. DR = Durchschnitt der erreichten Runden aller Spielversuche der einzelnen Probandinnen und Probanden. \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

Die Ergebnisse zeigen, dass zwischen dem Faktor Spielerfolg und der Selbstinformation ein signifikant positiver Zusammenhang besteht. Die zweite untersuchte Variable, die Immersion, weist keine signifikanten Einflüsse in Bezug auf die Veränderung des Nachhaltigkeitserlebens auf. Zusätzlich wird in **Tabelle 3** eine Korrelation innerhalb der Veränderungen der

Nachhaltigkeitsdimensionen ersichtlich. Dieser sehr signifikant negative Zusammenhang besteht zwischen der Einschätzung der sozialen Nachhaltigkeit und der eigenen Lebensweise.

Bei Betrachtung der Ergebnisse der letzten Forschungsfrage *Welche Veränderungen ergeben sich im Nachhaltigkeitserleben bei den Probandinnen und Probanden aufgrund des Spielens von LandYOUs? Welchen Einfluss haben die Faktoren Immersion und Spielerfolg auf die Veränderung des Nachhaltigkeitserlebens?* kann generell festgehalten werden, dass das Spielen von LandYOUs signifikante Änderungen im Nachhaltigkeitserleben bewirkt. Die wichtigsten Resultate dabei sind:

- Die Probandinnen und Probanden schätzen altersunabhängig nach dem Spiel den Faktor Bildung für die Thematik der Nachhaltigkeit wichtiger ein.
- Jüngere Spielerinnen und Spieler erkennen nach dem Spiel verstärkt einen sozialen Aspekt in der Nachhaltigkeit.
- Probandinnen und Probanden, welche größere Spielfortschritte gemacht haben, fühlen sich nach dem Spiel besser über die Thematik der Nachhaltigkeit informiert.

## Diskussion und Ausblick

Nach der ausführlichen Untersuchung des Serious Games LandYOUs im Rahmen der Studie können verschiedene Ableitungen anhand der Ergebnisse getroffen werden. Zuerst wird Bezug auf die Forschungsfragen zum Spielverhalten genommen. Aus der Clusteranalyse geht eine sehr hohe Diskrepanz zwischen dem ersten und den anderen beiden Clustern hervor. Basierend auf diesen kann angenommen werden, dass der Ausgang des Spiels womöglich in den ersten Runden festgelegt wird. Den Entscheidungen zu Spielbeginn entsprechend schaffen die Spielenden entweder den erfolgreichen Spielabschluss mit Runde 10 oder das vorzeitige Game Over in der Runde 5/6. Mit dem daraus resultierenden Lerneffekt kann den Spielerinnen und Spielern verdeutlicht werden, dass für ein nachhaltiges Landmanagement bereits frühzeitig die Entwicklungsrichtung durch die selbst gewählten Handlungen durch Definition mittels der Regler festgelegt wird. Entstehende Spätfolgen, wie das Game Over, gründen sich auf die eingangs getroffenen Entscheidungen.

Weiterhin zeigt der zeitliche Verlauf der Spiele die Veränderungen im Spielverhalten auf (vgl. **Abbildung 3**). Der deutlich erhöhte Zeitaufwand der ersten Spielrunde deutet auf die Orientierung im Spiel hin. Nach dem Überblick über die wichtigsten Spielelemente sinkt die investierte Zeit ab dem zweiten Spielversuch. Daraus kann abgeleitet werden, dass sich die Spielenden auf spielrelevante Objekte fokussieren. Die Abnahme der Fixationen je Zeiteinheit aus der Immersionsuntersuchung in der **Abbildung 5** weist ebenso auf diese Annahme hin. Die sehr geringe Versuchszeit ab dem Spielversuch 5 macht eine erste ersichtliche Spielstrategie ersichtlich, dem Trial-and-Error-Prinzip. Einerseits ist mit ca. 2-3 Minuten der aufgewandte Zeitumfang sehr gering und andererseits deutet die Erfolgswahrscheinlichkeit nicht auf einen Sieg im Spiel hin. Durch die geringe Zeit sollte ein Lesen der Texte bei diesen

Spielerinnen und Spielern unwahrscheinlich sein. Es ist somit anzunehmen, dass diese Probandinnen und Probanden durch Ausprobieren versuchen, den Spielealgorithmus zu verstehen.

Mit Hilfe der **Abbildung 4** können Spielstrategien der Testpersonen abgeleitet werden. Das Auseinandersetzen mit Auswertungsbildschirm und Pressespiegel als Feedbackelemente hat positive Auswirkungen auf das Spiel. Die Probandinnen und Probanden bekommen durch diese AOI ein besseres Verständnis für die Auswirkungen ihrer Entscheidungen. Daher gelingt Spielenden mit Fokus auf diese Bereiche eine positive Veränderung des Wirkungsgefüges. Hingegen ist die Karte nachweislich ein für das Spiel irrelevanter Bestandteil. Der Fokus auf diesen Bereich steht in Korrelation mit einer Spielniederlage und sollte weniger Aufmerksamkeit erhalten. In den Daten der AOI Regler und Geld (mit Prof. Landstein) kann auch das Vorhandensein des Trial-and-Error-Prinzips interpretiert werden. Personen, welche sich vorrangig auf diese AOI konzentrieren, legen den Fokus auf das Ausprobieren. Für sie haben die Erklärungen anderer AOI weniger eine Relevanz. Ein Scheitern wird mit dieser Spielart wahrscheinlicher. Bei dem Einsatz von LandYOUs im Lehrkontext ist meist nur eine begrenzte Zeit vorhanden. Mit Hilfe besserer Markierungen im Spiel selbst oder durch gezieltere Verweise auf diese „relevanten“ AOI könnte Spielenden schneller zu Spielerfolgen verholfen werden.

Hinsichtlich des Immersionserlebens kann anhand der linearen Regression der durchschnittlichen Fixationen per Zeiteinheit ein Vorhandensein von Immersion angenommen werden. Die weitere Betrachtung begründet sich auf diesem Ergebnis. Die erhaltene höchst signifikante Korrelation zwischen der berechneten und der eingeschätzten Immersion zeigt eine enge Verknüpfung beider Messvariablen. Daher können die Selbsteinschätzung und die berechneten Immersionsstufen für eine objektivere Immersionsbestimmung zusammen genutzt werden. Auch deutet diese Korrelation auf die Bestätigung hin, dass die Berechnung der Immersion aus den Items in dieser Form möglich ist. Der Einsatz des Eye-Trackings für die Immersionsbestimmung weist weiterhin signifikante Korrelationen mit den angegebenen Stufen auf, sodass die Methode des Eye-Tracking in dieser Art auch Immersionsstufen von Probandinnen und Probanden über den Blickkontakt mit dem Spielbildschirm bestimmen kann. Über eine größere Stichprobe könnte eine weitere Untersuchung einen repräsentativen Zusammenhang mit einer Gleichung erfassen. Die Verbindung der drei Varianten, dem Fragebogen, der Selbsteinschätzung und dem Eye-Tracking würde in Zukunft eine Evaluation der Immersion deutlich genauer ermöglichen.

Der letzte Punkt innerhalb der Immersionsuntersuchung betrifft die erreichten Stufen selbst (**Abbildung 6**). Aus den Ergebnissen geht eine anzunehmende Normalverteilung für die aufkommende Immersion im Spiel LandYOUs hervor. Demzufolge könnte jedes Spiel ein individuelles Immersionspotential für die Spielenden besitzen. Bezugnehmend auf die durchschnittliche Immersionsstufe von 2,2 ist das Ergebnis überraschend. Der durchschnittliche Wert zeigt nicht nur eine Teilnahme an LandYOUs, sondern dass auch individuell ein emotionales Verhältnis zu dem Spiel aufgebaut wird. Eine Zeit von 25 Spielminuten reicht

demnach bereits für das Erleben eines hohen Grades an Immersion. Bei dem Einsatz in Schulen bietet LandYOUs dahingehend zwei Vorteile. Es kann in einem kleinen zeitlichen Rahmen eingesetzt werden und es aktiviert die Emotionen von den Lernenden. Damit sich diese in die Lehre einbringen, wird häufig versucht, einen lebensweltlichen oder emotionalen Bezug herzustellen. Durch das Spielen von LandYOUs erfolgt bereits der zweite Punkt. Auf Basis der gemachten Spielerfahrungen kann die Lehrperson aufbauen und weitere Bezüge zu Emotionen, dem Leben und der Thematik herstellen.

Die Ergebnisse der Untersuchung des Nachhaltigkeitserlebens weisen auch auf Vorteile beim Einsatz des Spiels im Lehrkontext hin. Die **Abbildungen 8** und **9** haben gezeigt, dass bei Probandinnen und Probanden durch das Spielen von LandYOUs eine Einstellungsänderung beziehungsweise ein Lernfortschritt in der Thematik Nachhaltigkeit auftritt, bei den jugendlichen Probandinnen und Probanden sogar in einem größeren Umfang. Dadurch ist eine Wissensvermittlung durch das Serious Game LandYOUs nachgewiesen.

Weiterhin besitzt auch der Spielfortschritt einen Einfluss auf die Änderung des Nachhaltigkeitserlebens. Im Spiel erfolgreiche Probandinnen und Probanden fühlen sich im Post-Test besser über die Thematik informiert als weniger erfolgreiche Testpersonen. Für den Einsatz im Lehrkontext bedeutet das, dass möglichst alle Spielenden erfolgreich im Spiel sein sollten. Methoden wie das gemeinsame Spielen oder das Spielen zu zweit wären für das Erreichen des Ziels denkbar. Diese Möglichkeiten bieten den Teilnehmenden neben den unterschiedlichen Spielerlebnissen die Chance auf Förderung der Kommunikationskompetenz.

Die auffällige negative Korrelation zwischen der Änderung in den Dimensionen der sozialen Nachhaltigkeit und der eigenen nachhaltigen Lebensweise kann an dieser Stelle nicht begründet werden. Über Ursachen für diesen Zusammenhang sind nur Spekulationen möglich. Für eine weiterführende Forschung bietet sich an, den langfristigen Lernerfolg nach dem Einsatz von LandYOUs zu ermitteln.

Allgemein ist bei der gesamten Studie festzuhalten, dass durch eine größere Stichprobe möglicherweise mehr statistisch signifikante Ergebnisse erhalten worden wären. Die Methode des Eye-Trackings liefert sehr genaue und vielseitige Daten, doch der hohe Zeitaufwand der Einzelaufnahmen limitiert die Stichprobengröße auf eine verhältnismäßig geringe Anzahl an Probandinnen und Probanden. Für zukünftige Untersuchungen im Nachhaltigkeitserleben sind auch Erhebungen ohne Eye-Tracking, beispielsweise im Schulkontext, denkbar. Dabei sind größere Stichprobenzahlen deutlich praktikabler.

Abschließend kann man Lehrerinnen und Lehrern sehr empfehlen, Serious Games in ihrem Unterricht einzusetzen. Bereits eine Spielzeit von 20-30 Minuten reicht für eine Vermittlung nachhaltigkeitsrelevanter Inhalte. Damit besitzt LandYOUs nicht nur für den Unterricht motivierende Elemente, sondern eignet sich auch im Rahmen des Wissenserwerbs.

## Literatur

- Abdellatif, A., Mccollum, B., & MacMullan, P. (2018). Serious Games: Quality Characteristics Evaluation Framework and Case Study. (S. 112-119). 2018 IEEE Integrated STEM Conference (ISEC).
- Belotti, F., Berta, R., & De Gloria, A. (2010). Designing Effective Serious Games: Opportunities and Challenges for Research. *International Journal of Emerging Technologies in Learning*, 5.
- Berning, M., Schmidt, T., & Scholz, P. (2016). *Das Onlinespiel LandYOUs in der Schule - Unterrichtsimpuls zur Bildung für nachhaltige Entwicklung*. Stuttgart: Klett MINT.
- Brown, E., & Cairns, P. (2004). A Grounded Investigation of Game Immersion. *Conference on Human Factors in Computing Systems - Proceeding*, 10, S. 1297-1300.
- Csikszentmihalyi, M. (1990). *Flow: the psychology of optimal experience*. New York: Harper & Row.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From Game Design Elements to Gamefulness: Defining Gamification. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek*, 11, S. 9-15.
- Hattie, J. (2009). *Visible Learning*. London, New York: Routledge.
- Hoblitz, A. (2014). *Spielend Lernen im Flow*. Paderborn: Universität Paderborn.
- Jennett, C. (2008). Measuring and defining the experience of immersion in games. In *International Journal of Human-Computer Studies* 66 (S. 641-661).
- Lefers, B., & Birkenkrahe, M. (2016). How can educators with limited resources gamify their classes? A Design-based approach. *10th European Conference on Games Based Learning*. Paisley.
- Pelling, N. (2011). *Funding Startups (& other impossibilities)*. Abgerufen am 15. September 2020 von <https://nanodome.wordpress.com/2011/08/09/the-short-prehistory-of-gamification/>
- Schulze, J., Martin, R., Finger, A., Lindner, M., Pietzsch, K., Werntze, A., et al. (2014). Design, implementation and test of a serious online game for exploring complex relationships of sustainable land management and human well-being. *Environmental Modelling & Software* 65 (20115), S. 58-66.

**Nils Ludwig Kotschote** hat einen Abschluss in den Fächern Biologie, Chemie, Geographie und Astronomie für das Lehramt an Gymnasien. Er war studentischer Mitarbeiter in der Biologiedidaktik der Martin-Luther-Universität Halle-Wittenberg. Neben der Mitarbeit an diversen Projekten forscht er vor allem zu digitalen Medien und der Digitalisierung im Bereich MINT. Weiterhin ist er auch am Institut für Chemie in der Festkörperchemie tätig.

[nils.kotschote@gmail.com](mailto:nils.kotschote@gmail.com)

**Chris Wenzel** studiert Biologie und Sozialkunde für das Lehramt an Sekundarschulen. Seit einigen Jahren arbeitet er als studentischer Mitarbeiter in der Biologiedidaktik der Martin-Luther-Universität Halle-Wittenberg. Er wirkt bei verschiedensten Projekten mit und leitet jene auch in den Gebieten Aquaponic und Social & Science Camps auch im Ausland selbst. Weitere Forschungsfelder von ihm sind digitale Medien und MINT im Biologieunterricht.

[chris.wenzel@biodidaktik.uni-halle.de](mailto:chris.wenzel@biodidaktik.uni-halle.de)

**Martin Lindner** ist Professor der Didaktik für Biologie und Geographie an der Martin-Luther-Universität Halle-Wittenberg. Sein Hauptbeschäftigungsfeld sind die Interessen und Motivationsforschung im Bereich MINT. Dabei untersucht er vorrangig Science Camps und digitale Medien.

[martin.lindner@biodidaktik.uni-halle.de](mailto:martin.lindner@biodidaktik.uni-halle.de)

# Lesepaxen im Medienzeitalter: Ergebnisse einer Fallstudie zu digitalen und analogen (Bilderbuch-)Lektüren

*Alexandra Ritter<sup>1</sup> und Michael Ritter<sup>1</sup>*

*<sup>1</sup> Martin-Luther-Universität Halle-Wittenberg, Institut für Schulpädagogik und Grundschuld-  
daktik, Arbeitsbereich Deutsch*

Präsentiert werden Ergebnisse des Teilprojektes *Lesepaxen im Medienzeitalter. Digitales Lesen von Bilderbuch-Apps zwischen Leseförderung und Literarischem Lernen*. Im Mittelpunkt stehen Bilderbuchlektüren im Grundschulkontext auf der Basis von analogen und digitalen Bilderbuchmedien. Konkret werden sozial-interaktive literarische Vermittlungspraktiken von Lehrkräften untersucht, die in kleinen Gruppen mit Kindern analoge Bilderbücher und Bilderbuch-Apps rezipieren. Fokussiert wird dabei auf die Frage, wie sich die strukturellen Lektürebedingungen im Umgang mit dem analogen Papierbuch und dem digital erweiterten Angebot der interaktiven und multimodalen Apps verändern (Innovation vs. Remediation). Wie beeinflussen die digitalen Medienangebote der Bilderbuch-Apps die Lesepaxis und ihr Interaktionsprofil in Vorlesegesprächen? Wie werden Zugänge zur literarischen Substanz durch die ausgewählten Medien ermöglicht? Findet die Rezeption unabhängig oder in konkreter Abhängigkeit von der medialen Verfasstheit der literarischen Werke statt? Kann von einer spezifischen „digitalen Literalität“ gesprochen werden? Bei den hier vorgestellten Ergebnissen handelt es sich um kasuistische Interaktionsprofilanalysen, die komparativ interpretiert und didaktisch kontextualisiert werden. Es wird herausgearbeitet, wie die mediale Substanz der literarischen Werke Einfluss auf die Gesprächsstruktur des Vermittlungshandelns nimmt, dass pauschalisierende Gegenüberstellungen von analogen und digitalen Lesepaxen jedoch in dieser Perspektive nicht tragfähig sind. Der Aufsatz leistet damit einen Beitrag zur Erforschung des professionellen literarischen Vermittlungshandelns im Kontext sich verändernder Medien(erfahrungs)welten.

## **Veränderte Medienwelt – verändertes Leseverhalten?!**

Gesamtgesellschaftlich befinden wir uns schon länger in einem Wandlungsprozess in Bezug auf die Digitalisierung. So verändern sich auch die informellen Bildungsbereiche und Enkulturationsprozesse in einer zunehmend digitalisierten Welt (Dawidowski, 2009; 2013). Damit einher gehen sich verändernde Lektüreangebote. Beispielsweise gibt es in vielen Familien neben einem analogen Bücherregal mitunter noch ein digitales, in dem aktuelle E-Books gesammelt werden.

Dawidowski (2013) und Krommer (2019) sprechen deutlich von einem Leitmedienswechsel und zeigen, dass solche Prozesse immer wieder die Menschheitsgeschichte geprägt und verändert haben; denn die Leitmedien stellen nicht nur den Zugang zu Informationen dar, sie beeinflussen durchaus auch unser Denken und die Art des Umgangs mit dem Wissen.<sup>1</sup>

Die Digitalität stellt nun andere Formate der Wissensstrukturierung zur Verfügung, die ebenfalls auf die veränderten technischen Möglichkeiten zurückzuführen sind. Es finden sich weniger lineare Ordnungssysteme, sondern Wissensnetze, die von vielen genutzt und selbst bearbeitet werden können. Viele Informationen sind über das Internet jedem zugänglich, allerdings gilt es – als neue Anforderung in der digitalen Welt – zwischen wesentlichen und unwesentlichen Informationen zu unterscheiden und die Qualität der Informationen prüfen zu können (vgl. Krommer, 2019). Dawidowski konstatiert eine Veränderung auch beim literarischen Lesen. Diese Lesehaltung wird von einem kursorischen, stark auf (schnelle) Informationsentnahme orientierten Lesen bestimmt, dessen Ursache Dawidowski in der Omnipräsenz von Virtualität und Vernetzung sieht (vgl. Dawidowski, 2013, S. 11).

Von Seiten der Zielgruppe der Kinder- und Jugendlichen bestätigt die KIM-Studie von 2018 eine veränderte Mediennutzung und spiegelt den gesellschaftlichen Wandlungsprozess wider: „Kinder wachsen heute mit einem sehr vielfältigen Medienrepertoire auf. In den Haushalten, in denen Kinder im Alter von sechs bis 13 Jahren leben, besteht Vollausrüstung bei Fernseher, Internetzugang sowie Handy/Smartphone“ (MPFS, 2019, S. 8). 97 % dieser Haushalte verfügten über ein Handy bzw. Smartphone, 81 % über einen Computer/Laptop. Die verhältnismäßige niedrige Verbreitung von Tablets mit 38 % dürfte seit 2018 weiter gestiegen sein (vgl. MPFS, 2019, S. 9). Die miniKIM Studien von 2012 zeigt zudem deutlich, dass das Buch als tradiertes Lesemedium bereits deutlich vor dem Schuleintritt seinen Stellenwert als Leitmedium der kindlichen Lesesozialisation einbüßt und Text- und Literaturbegegnung in anderen Medien an Bedeutung gewinnt (vgl. MPFS, 2013, S. 11).

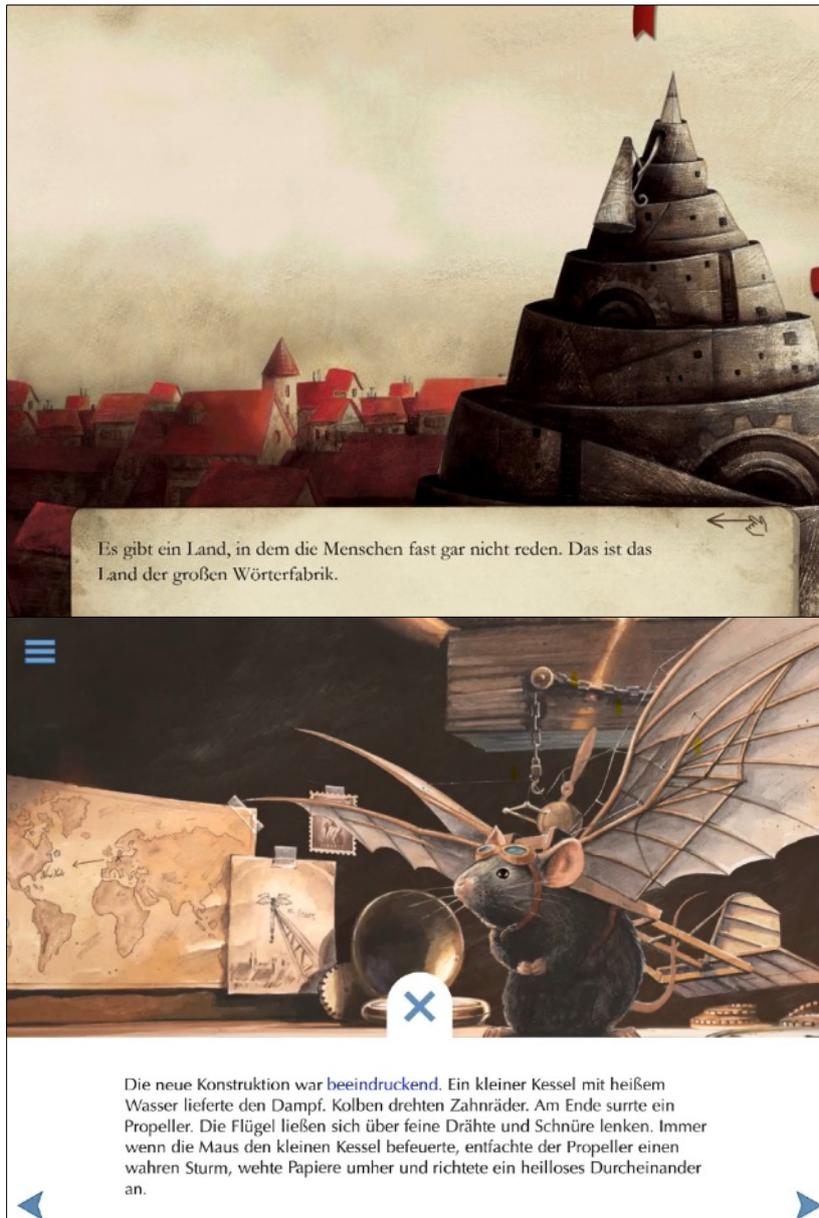
### **Bilderbuch – digital und analog**

Auch die Literatur selbst wird von Adaptions- und Transformationsprozessen beeinflusst (vgl. Orland & Ritter, 2019, S. 73), was man am Beispiel von Bilderbuch und Bilderbuch-App gut herausarbeiten kann. Literatur in digitalen Medien erscheint als symmedialer Text, „insofern Text, Bild und Ton in immer neue medial-semiotische Verbindungen treten“ (Frederking,

---

<sup>1</sup> Krommer stellt der Digitalisierung die kulturanthropologischen Phasen der Oralität, der Skriptografie und der Typografie voran, die hier aus Platzgründen nicht ausgeführt werden können (vgl. Krommer, 2019; Ritter & Ritter, 2020).

2013, S. 543). Solche Texte „erfordern symmediale Leseweisen, d. h. das Decodieren und Verstehen literaler, piktoraler, auditiver, audiovisueller und hypermedialer Texte bzw. ihrer medialen und semiotischen Verknüpfung“ (ebd.). Das Zitat deutet an, dass digitale Wissenskulturen einerseits andere Anforderungen an die Verarbeitung von Sinnangeboten stellen, weil die Buchstabenschrift viel stärker als bisher in Kombination mit anderen Modalitäten erschlossen werden muss, andererseits werden auch bei der Erschließung und Beurteilung von Sinnangeboten andere (Meta-)Modi der Verarbeitung nötig, die die Vernetzung von Wissen und die Einbettung in bestimmte Diskurse und deren implizite Intentionalität stärker in den Blick nehmen. Dies soll an zwei, für die Studie relevanten Beispielen aus dem Bereich der narrativen Bilderbuch-Apps gezeigt werden.



**Abbildung 1.** Lestrade & Docampo: *Die große Wörterfabrik* ([App] © München: mixtvision, 2013) und Kuhlmann & Salhi: *Lindbergh* ([App] © Hamburg: Friedrich Oetinger, 2015)

*Die große Wörterfabrik* (2013) ist eine App-Sekundärproduktion zum gleichnamigen Buch von Agnes de Lestrade und Valeria Docampo und beschreibt ein Land, in dem man Wörter kaufen und schlucken muss, um sie aussprechen zu können. Diese werden in der titelgebenden Fabrik hergestellt. Die App-Adaption zeigt sich als typisches realisiertes App-Angebot (vgl. Ritter, 2019), das die lineare Hard-Rail-Struktur übernimmt und die Handlung der Bild-Text-Erzählung durch ergänzende interaktive Wort-Spiel-Angebote aufwertet (**Abbildung 1**). Dennoch ist eine Tendenz zur Remediation, also zur Rückbindung an das (Papier-)Bilderbuch, erkennbar.

Die App *Lindbergh* ist ebenfalls eine Sekundärproduktion aus dem Jahr 2015, die das gleichnamige Bilderbuch von Torben Kuhlmann adaptiert. Sie erzählt von einer Maus, die es sich zur Aufgabe macht, ein Flugobjekt zu bauen, um den Alleinflug von Hamburg über den Atlantik nach New York zu meistern. Auch diese App ist in Analogie zum Bilderbuch linear konzipiert. Die Geschichte präsentiert sich in Seitensequenzen, der Erzähltext ist mit einer Highlight-Funktion ausgestattet und kann ein- oder ausgeblendet werden (**Abbildung 1**). Dazu können lektürebegleitend Animationen (Geräusche oder Bewegungen) aktiviert werden. Diese sind durch einen aufblinkenden Kreis leicht auffindbar.

Beide Apps sind linear aufgebaut mit fakultativen Animationen der Bildelemente bei *Lindbergh* und interaktiven Spielangeboten bei der *Wörterfabrik*. Die medienspezifischen Erweiterungen treiben aber die Handlung nicht voran, sondern dienen eher illustrativ der Vorstellungsbildung. Aufgrund des ähnlichen Aufbaus und der narrativ angelegten, literar-ästhetisch innovativen Bild-Text-Erzählung in Bilderbuch und App, wurden beide als Gegenstand für die vorliegende Studie ausgewählt.

## **Forschungsreview – Studien zu digitalen Lesepraxen**

Bezüglich der Lektüre im analogen und digitalen Medium gibt es bereits zahlreiche Studien und Forschungsreviews. Einige Ergebnisse sollen an dieser Stelle exemplarisch angeschnitten werden:

Aufenanger (2018) zeigt in seinem Forschungsüberblick zu Tablets im Unterricht, dass die Akzeptanz derselben bei Schüler\*innen und Lehrkräften bezogen auf die Nutzung von Tablets für schulische Lernangebote hoch sei (vgl. Aufenanger, 2018, S. 50; Anders, 2018, S. 238). Gleichzeitig konnten positive Effekte auf die Lernmotivation beobachtet werden, die allerdings auf einen Neuigkeitseffekt zurückgeführt werden könnten (vgl. Karsenti & Fievez, 2013, S. 6, zit. n. Aufenanger, 2018, S. 50 f.). Dawidowski stellt fest, dass das Lesen im Digitalen deutlich fragmentarischer und cursorischer abläuft (Dawidowski, 2013, S. 12). Bezogen auf das Leseverständnis kommt Maik Philipp in seiner Synopse unterschiedlicher Studien zu einem ähnlichen Ergebnis, allerdings zeigt sich der Mediumseffekt vor allem bei Sachtexten. „Gibt es ein zeitliches Limit, verstehen die Testpersonen die digitalen Texte schlechter. Ohne Zeitlimit fällt der Mediumseffekt erheblich geringer aus.“ (Philipp, 2020, S. 6) Ursachen sieht Philipp vor allen Dingen auch im Scrollen von Texten. Bei Texten, die digital ohne Scrollen gelesen wurden, waren die Mediumseffekte weniger stark. Zudem konnte der Mediumseffekt bei narrativen Texten nicht nachgewiesen werden (vgl. ebd., 7f.).

Heike Schaumburg betont, dass „der Einsatz digitaler Medien im Kontext des personalisierten Lernens kein Selbstläufer“ ist (Schaumburg, 2020a). Die Einbindung in den Unterricht und die Unterstützung durch die Lehrperson seien entscheidend für die Lernwirksamkeit digitaler Lernumgebungen (vgl. ebd.). Des Weiteren kommt Schaumburg in einer Meta-Analyse zu Effekten von explorativen computerbasierten Lernumgebungen bezüglich E-Books zu dem Ergebnis, dass multimediale Effekte wie Ton oder Animationen, wie sie auch in den Bilderbuch-Apps zu finden sind, deutlich lernwirksamer seien als interaktive Elemente, besonders für schwächere Schüler\*innen (ES=.66) (vgl. Schaumburg, 2020b, Folie 17).

Wieler, Brandt, Naujok, Petzold und Hoffmann (2008) machten bei der Untersuchung von familien- und schulbezogenen Lektürekulturen die Beobachtung, dass digitale literarische Medien<sup>2</sup> in der Familie und in der Schule unterschiedlich eingebunden werden. Insgesamt beobachten die Forschenden „Schwierigkeiten der Integration der neuen Medien in die kommunikative Alltagskultur von Familie und Unterricht“ (Wieler, Brandt, Naujok, Petzold, & Hoffmann, 2008, S. 266). Pointiert stellen sie zwei Befunde gegenüber:

„Durchweg erfolgte die Rezeption (zumindest) dieses einen Buches im Beisein der Lehrerin, sie wurde bereits durch das Vorlesen gesteuert und darüber hinaus durch Gespräche und Arbeitsaufträge intensiv begleitet“ (ebd., S. 267). Demgegenüber bildete „die Rezeption der CD-ROMs gewissermaßen den Gegenpol zu dieser ‚geteilten‘ Buchrezeption; der Umgang mit den Computermedien wurde den Schülerinnen und Schülern vorwiegend zur freien Beschäftigung überlassen“ (ebd.).

Das mag einerseits daran liegen, dass die digitalen Lektüren durch Vorlesefunktionen, Animationen und eingebundene Interaktionsangebote die Steuerung der Lektüre technisch anbieten und vorstrukturieren, gleichzeitig ist aber auch zu vermuten, dass bezüglich der kaum geläufigen digitalen Angebote Lehrer\*innen weniger konventionelle Routinen im Umgang mit solchen Artefakten abrufen können. Zudem sind die didaktischen Erwartungen an die digitalen Angebote deutlich geringer (vgl. Wieler et al., 2008, S. 269).

Bettina Muratovic, die Vorlesegespräche zu analogen und digitalen Bilderbüchern im familiären Kontext vergleicht, kommt hingegen zu dem Ergebnis, dass die Vorlesegespräche zu den digitalen Angeboten keine prinzipielle Strukturänderung bezüglich der interaktiven Realisierung aufweisen (vgl. Muratovic, 2014, S. 157).

„Demnach können *keine grundlegenden Veränderungen* hinsichtlich der *Vorlesedialoge mit Printmedien* gegenüber denen *mit digitaler [sic] Medien* festgestellt werden.“ (Hervorhebungen im Original, d. Verf.; Muratović, 2014, S. 137)

Auf Studien zu (Vor-)Lesegesprächen kann hier nur verwiesen werden (vgl. z. B. Schneider, 1995; Wieler, 1997; Härle & Steinbrenner, 2004; Kruse, 2013; Ritter, 2014; Fuhrmann & Merklinger, 2015; Kruse, 2016; Wieler, 2019).

---

<sup>2</sup> Hier waren es sog. Living Books auf CD-ROM.

## **Forschungsdesign zum Projekt *Lesepaxen im Medienzeitalter***

Im Mittelpunkt der hier dargestellten Studie stehen Lektüregespräche im Grundschulkontext auf der Basis von analogen und digitalen literarischen Medien. Konkret werden literarische Vermittlungspraktiken von Lehrkräften, die in kleinen Gruppen von Schüler\*innen Bilderbücher und Bilderbuch-Apps rezipieren, und deren Anregungspotenziale untersucht. Fokussiert wird dabei auf die Frage, wie sich die strukturellen Lektürebedingungen im Umgang mit dem analogen Papierbuch und dem digital erweiterten Angebot der interaktiven und multimodalen Apps vergleichen lassen (Innovation vs. Remediation). Wie beeinflussen die digitalen Medienangebote der Bilderbuch-Apps den Leseakt und sein Interaktionsprofil in Lesegesprächen? Wie werden Zugänge zur literarischen Substanz durch die ausgewählten Medien ermöglicht? Findet die Rezeption unabhängig oder in konkreter Abhängigkeit von der medialen Verfasstheit der literarischen Werke statt? Kann von einer spezifischen digitalen Literalität (Harrison, 2009; kulturanthropologisch auch Krommer, 2019) gesprochen werden? Auch die Perspektiven der Lehrer\*innen auf den Gegenstand sollen im Rahmen leitfadengestützter Interviews einbezogen und berücksichtigt werden.

Für die Erhebung der untersuchten Vorlesegespräche wurden vier Lektüregruppen zusammengestellt. Die vier zufällig ausgewählten Lehrkräfte wurden gebeten, je eine analoge und eine digitale Bilderbuchlektüre mit einer selbst zusammengestellten Gruppen von Kindern (A-D) durchzuführen. Vorgegebene Rahmenbedingungen waren:

1. *Format:* Die Lehrkräfte wurden gebeten, in Kleingruppen zwei Bilderbücher in verschiedenen Medienformen zu lesen und zu betrachten.
2. *Medienauswahl:* Gelesen werden sollten zu je einem Termin die Bilderbücher *Lindbergh* (Kuhlmann, 2014; Kuhlmann & Salhi, 2015) und *Die große Wörterfabrik* (de Lestrade & Docampo 2010; 2013). Die beiden Lektüren fanden im Abstand von ca. einer Woche statt. Die Lehrkräfte hatten im Vorfeld Buch und App zur Eigenlektüre und zur Vorbereitung des Lernsettings zur Verfügung.
3. *Medienverteilung:* In jeder Gruppe wurde ein Werk als analoges Bilderbuch und eins als Bilderbuch-App rezipiert. Die Verteilung wurde von der Versuchsleitung vorgenommen, so dass je zwei Gruppen *Lindbergh*-Bilderbuch/*Die große Wörterfabrik*-App und zwei Gruppen *Lindbergh*-App/*Die große Wörterfabrik*-Bilderbuch bearbeiteten. So sollte gewährleistet werden, dass systematische Vergleiche zwischen den Lektürepraxen möglich werden. Sowohl konnte so im Rahmen der Datenauswertung fallübergreifend die Lektüre gleicher Werke als auch unterschiedlicher Werke je gleicher oder unterschiedlicher Medienform in den Blick genommen werden.
4. *Didaktisch-methodische Entscheidungen:* Für die Gestaltung der Vorlesegespräche wurden keine Vorgaben gemacht. Die Lehrer\*innen sollten vielmehr die Prozesse so gestalten, wie es ihren Vorstellungen und Routinen im Hinblick auf die Bilderbuch-Lektüre entspricht, um möglichst alltagsnahe Formen der Vorlesegespräche

mit entsprechend routinierten Vermittlungs- und Erschließungspraktiken in den Blick nehmen zu können.

5. *Gruppengröße:* Es sollten drei bis vier Kinder aus dem dritten oder vierten Schuljahr (Alter: 8-10 Jahre) an der Erhebung teilnehmen. Die teilnehmenden Kinder sollten bei beiden Lektüren möglichst dieselben sein. Es sollten jeweils Jungen und Mädchen teilnehmen. Konkretere Vorgaben hinsichtlich der Auswahl der Kinder wurden nicht gemacht. Die Auswahl wurde von den Lehrkräften vorgenommen.
6. *Erhebungsort:* Alle Lektüren fanden in Schulgebäuden statt, wobei eher Funktionsräume (Schulbibliothek, Mensa) als Klassenräume genutzt wurden.
7. *Interview:* Im Anschluss an das zweite Lektüretreffen wurde mit der Lehrkraft ein kurzes leitfragengestütztes Interview zu den beiden Lektüren durchgeführt, in dem der Gesamteindruck der Lehrer\*innen und ihre medienbezogenen Orientierungen thematisiert werden sollten. Da die Interviewergebnisse hier keine Rolle spielen, wird dieser Studienbaustein nicht eingehender dargestellt.
8. *Datenaufzeichnung:* Die Lektüren wurden videographiert und sicherheitshalber auch noch mit Diktiergeräten mitgeschnitten. Die visuelle Ebene schien auch deshalb wichtig, weil so nonverbale mimische, gestische und deiktische Performanzphänomene gesichert und in die Auswertung einbezogen werden konnten. Die Interviews wurden lediglich als Audio-Mitschnitte aufgenommen.
9. *Aufbereitung:* Im Anschluss an die Datenerhebung wurden sämtliche erhobenen Daten transkribiert. Die Transkriptionsvorschrift orientierte sich am gesprächsanalytisch konventionellen GAT-Verfahren (vgl. Heller & Morek, 2018). So sollten neben der verbalen Dimension der Interaktionen auch non- und paraverbale Interaktionsdimensionen miterfasst werden.
10. *Datenauswertung:* Es wurde zunächst eine kasuistische Interaktionsprofilanalyse angelegt, deskriptiv-statistisch ausgewertet, komparativ interpretiert und didaktisch kontextualisiert. Weiterhin wurden ausgewählte Fälle der Vorleselektüre hermeneutisch-rekonstruktiv erschlossen. Das Vorgehen orientierte sich dabei an der interpretativen Unterrichtsforschung (Krummheuer & Naujok, 1999) und am linguistischen Verfahren der Gesprächsanalyse (Deppermann, 2008; Heller & Morek, 2018; ausführlich bei Ritter & Ritter, 2020).

Dieser Beitrag beschränkt sich aus Platzgründen auf Einblicke in die vergleichende Fallrekonstruktion am Beispiel der Gruppe C, die exemplarisch Einblicke in die Ergebnisse der Gesamtstudie zu geben vermögen.

## Einblick in die komparative Fallanalyse Gruppe C

Anhand der drei Lektüresituationen sollen vergleichend Ergebnisse der Studie vorgestellt werden. Einführend werden einige Aspekte der Interaktionsprofilanalyse und der daraus folgenden deskriptiven Statistik für eine erste Orientierung präsentiert. Danach werden exemplarisch ausgewählte Szenen der Lektüren analysiert. Gruppe C bestand aus einer Lehrerin und 2 Schülerinnen bei der Bilderbuch-Lektüre und einem weiteren Schüler bei der App-Lektüre, der bei der ersten Erhebungssituation aus Krankheitsgründen nicht teilnehmen konnte.

Insgesamt liegt die **Dauer beider Lektüren** deutlich unter den Durchschnittswerten: die Bilderbuch-Lektüre von *Die große Wörterfabrik* dauert 14.03 min ( $\emptyset$  Bilderbuch-Lektüren: 26.01 min), die App-Lektüre von *Lindbergh* dauert 32.22 min ( $\emptyset$  App-Lektüren: 40.09 min).

Zudem ergibt sich bei diesen Werten eine starke Differenz zwischen der Dauer der beiden Lektüren. Diese korrespondiert einerseits mit dem größeren Werkumfang von Lindbergh, andererseits aber auch mit einer Tendenz zur ausführlicheren App-Lektüre, die bei allen anderen Lektüregruppen ebenfalls nachgewiesen werden konnte (vgl. Ritter & Ritter, 2020). Allerdings sind die durchschnittliche **Dauer der Sequenzen**<sup>3</sup> und die durchschnittliche **Anzahl der Interaktionseinheiten pro Sequenz** in beiden Lektüren annähernd ähnlich. Die Lektüren scheinen sich auf der Mikroebene daher zunächst wenig zu unterscheiden.

Die Verteilung der **Sprechakte** lässt allerdings schnell erkennen, dass sich das Interaktionsverhältnis in den beiden Lektüresituationen verändert. Während bei der Bilderbuch-Lektüre die Lehrerin häufiger spricht (62 % aller Sprechakte,  $n=108$ ), nimmt sie sich bei der App-Lektüre stärker zurück (46,4 % aller Sprechakte,  $n=181$ ). Es zeigt sich also, dass die Schüler\*innen während der App-Lektüre häufiger zu Wort kommen und sich mit Gesprächsbeiträgen in die Lektüresituation einbringen.

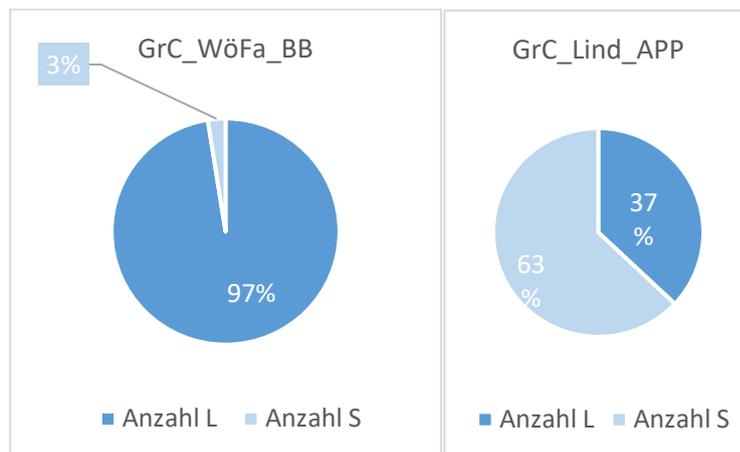


Abbildung 2. Interaktionseröffnende Aktivitäten von Lehrerin und Schüler\*innen

<sup>3</sup> Die Lektüren wurden grob in Sequenzen gegliedert, die größere Sinnabschnitte markieren und alle Interaktionen zu einer Buch- bzw. App-Seite erfassen. Sequenzwechsel war jeweils das Umblättern. Die Sequenzen wurden weiterhin in Interaktionseinheiten gegliedert, die aus semantisch aufeinander bezogenen Interakten bestehen, also kleinere Sinneinheiten innerhalb der Sequenzen darstellen.

Blickt man nur auf die **interaktionseröffnenden Aktivitäten**<sup>4</sup>, so werden die Unterschiede noch deutlicher (**Abbildung 2**). Bei der Bilderbuch-Lektüre eröffnet fast ausschließlich die Lehrerin die Interaktionseinheiten (97,5 %,  $n = 43$ ), wobei es vorrangig um Steuerungsaktivitäten geht, wie z. B. die Vergabe des Vorleserechts und die Initiierung des Sequenzwechsels. Außerdem ist die Lehrerin um die Verständnissicherung bemüht. Sie setzt hingegen nur wenige produktive Impulse. Anders ist dies bei der App-Lektüre: Hier verschiebt sich das dargestellte Verhältnis in zwei Richtungen. Die beteiligten Kinder wirken dominanter, sie eröffnen nun 63,1 % der Interaktionseinheiten ( $n = 131$ ). Sie übernehmen z. T. die Steuerung und erkunden zudem selbstständig die Bildschirmoberfläche. Außerdem sind deutlich mehr produktive Interaktionseinheitseröffnungen bei der App-Lektüre zu beobachten, wie z. B. das Formulieren von Entdeckungen und Irritationen oder das Aufstellen von Hypothesen bzw. das Antizipieren des Fortgangs der Geschichte.

Das reine Vorkommen dieser Ereignisse zeigt bestimmte Lektüretendenzen, kann aber noch keine Auskünfte über die Qualität der Vermittlungspraktiken und der Interaktionen geben. Das sollen die nun folgenden ausgewählten, sequenz-analytisch rekonstruierten Szenen herausstellen:

### *Szene 1 der Bilderbuch-Lektüre*

- 171 [Bilderbuchseite: das Mädchen und der dunkle Junge stehen sich gegenüber, daneben sind ganz groß Wörter auf Zetteln dargestellt]  
 L: ich LIEBE DICH VON GANZEM HERZEN MEINE MARIE (.) EINES TAGES DASS WEIß ICH (.) WERDEN WIR (.) HEIRATEN (3) wer sagt das
- 175 M1: oskar  
 L: mhh (.) was denkt ihr wie viel geld hat das gekostet  
 M1: hunderttausend euro  
 [L lacht]  
 L: ok (.) wie viele wör- wörter musste denn er kaufen
- 180 (6)  
 M2: sechzehn  
 L: hmh (.) genau  
 [L blättert Seite um]

(GrC\_WöFa\_BB, Z. 171-183)

Diese Szene zeigt exemplarisch den stabil-formalisierten Ablauf der Bilderbuchlektüre, der sich in Gruppe C etabliert hat. Die Lehrerin eröffnet die Sequenzen, indem sie selbst vorliest oder das Vorleserecht vergibt. Dann stellt sie verständnissichernde, in der Regel geschlossene Fragen, die seitens der Schüler\*innen mit einem Wort oder einzelnen Phrasen beantwortet werden. Die Antworten scheinen der Lehrerin zu genügen, denn es wird nicht nach weiteren Aussagen zur Elaborierung oder Begründung, z. B. bei der Antwort von M1 in Zeile 177, gefragt. Die Antwort wird weder bewertet noch vertieft. Die Lehrerin ist am Ende der Szene diejenige, die das evaluierende Schlusswort hat und das Umblättern initiiert.

<sup>4</sup> Jeweils der erste Interakt einer Interaktionseinheit.

In Gruppe C hat sich demnach eine analoge Lektürepraxis etabliert, die durch ein stabiles Muster aus Sequenzeröffnung inkl. Vorlesen und meta-literarischer Auseinandersetzung auf Basis von durch die Lehrkraft gestellten Fragen geprägt ist. Die Schüler\*innen reagieren auf die Impulse der Lehrerin, lesen vor und beantworten Fragen. Daran zeigt sich eine deutliche strukturelle Dominanz und inhaltliche Deutungshoheit, die der Lehrkraft zukommt. Ein Austausch zu den Inhalten des Buches, ggf. sogar auf deutungsbezogener Augenhöhe, scheint hier nicht denkbar zu sein. Die Lehrerin leitet und lenkt das Gespräch und gibt, wenn nötig, auch noch zusätzliche Informationen und Zusammenfassungen der Sequenzen. So haben die Kinder nur selten eine Möglichkeit, eigene Gedanken einzubringen. Werden diese genannt, werden sie kaum hinterfragt, sondern nur als „korrekte“ Antworten bestätigend evaluiert. Ein Austausch über individuelle Lesarten und eine bewusste Konfrontation divergierender Interpretationen findet nicht statt.

Bezogen auf die App-Lektüre sollen zwei Stellen zum Tragen kommen, die unterschiedliche Aspekte der Lektürepraxis aufzeigen:

### *Szene 1 der App-Lektüre*

- 125 *[Vorlesestimme (V): botschaften aus einem fernen land ...]*  
M1 [flüstert]: (das gibt es doch auch als buch?)  
*[... die wochen vergingen (.) weit und breit war keine andere maus mehr aufzuspüren (.) die kleine maus hatte mittlerweile in den zeitungsen viel über die neuen mausefallen gelesen ...]*  
*[J1 tippt auf Bildschirm] [V: eine alpträumerhafte erfingung; Rezoom; ... sicher hatten diese fallen*  
130 *alle ihre freunde vertrieben (.) nur (.) wohin waren sie verschwunden (.) vielleicht nach amerika (.) alle mäuse kannten geschichten über diesen fernen ort ...]* [J1 tippt auf das Foto auf der Zeitungsseite] [V: es klappt; ... es hieß dass dort eine große statue alle neuankömmlinge begrüße ...] [M2 tippt  
Zeitung] [V: ganz egal ob mensch oder maus]  
[J1 und M1 tippen auf Bild, J1 kommt dabei auf Weiter-Pfeil] [Bootshupe ertönt; Bild  
135 blättert um, Blick auf den Hafen scheinbar hinter einem Pfeiler hervor vom Fußboden aus]
- 
- J1: m tz och  
138 L: wenn du auf den pfeil drückst dann geht's immer weiter  
(GrC\_Lind\_App, Z. 125-138)

Auffallend ist, dass die Schüler\*innen während der Lektüre kaum sprechen, dafür aber intensiv mit dem Tippen auf dem Tablet zur Aktivierung von Animationen und Geräuschen in der App beschäftigt sind. Die Gruppe hat sich dafür entschieden, den Text elektronisch vorlesen zu lassen, und so kann während des Vorlesens die App erkundet werden. Kritisch kann man jetzt fragen, wo der spezifische „Mehrwert“ der App für die literarische Bildung und für die Lesekompetenz liegt, der den Einsatz eines solchen Mediums auch im Rahmen des Deutschunterrichts legitimieren kann. Ein Lektüregespräch kommt an dieser Stelle zumindest nicht zustande und auch Lesefertigkeiten werden nicht gefordert. Wie viel die Schüler\*innen an Vorstellungsbildung und Wahrnehmung von Bild und Text bzw. der Narration der Geschichte im Rahmen von literarischen Lernprozessen entwickelt haben, kann anhand dieser Situation nicht beurteilt werden. Und doch scheint ein deutlicher Fokus auf dem Tippen, den haptisch-

kinästhetisch erfahrbaren und dann auch akustisch oder optisch veränderbaren Animationen zu liegen, das die Bedeutung des literarischen Gegenstands zu marginalisieren scheint.

### Szene 2 der App-Lektüre

[V: die neue konstruktion war beeindruckend ein kleiner kessel ...] [...]  
410 [...mit heißem wasser lieferte den dampf kolben drehten zahnräder am ende surrte ein propeller ...]  
J1: (wies aussieht wird er es jetzt schaffen?)  
[V: ... die flügel ließen sich über feinde drähte und schnüre lenken ...]  
L: mhm  
[V: immer wenn die maus den kleinen kessel befeuerte entfachte der propeller einen wahren sturm ...]  
[ein Kreis  
415 blinkt auf dem bild auf]  
[J1 tippt auf den Kreis] [Bildperspektive verschiebt sich mehr auf die Karte]  
[V: wehte papiere umher und richtete ein heillosos durcheinander an ...] [ein neuer Kreis blinkt auf]  
[J1 tippt auf den Kreis] [die Flügel der Flugmaschine bewegen sich]  
[M1 tippt auf die Maus]  
420 L: die sieht aus als würd sie schAFFen  
J1: mhm (.) JA  
[M1 und M2 nicken]  
L: woran siehst du das  
J1: hier an den flügeln sehn so fest aus (.) und hier mit den stöckeln (is das irgendwie so?)  
[zeigt auf die  
425 Flügel und das Gestänge daran]  
L: ok (2) woran erlNnern euch die flügel [zeigt auf die Flügel]  
M1: an die fledermaus  
J1: an die fledermausflügel  
L: mhm (.) sehn n bisschen aus wie knochen oder so ne diese stÖCKER [zeigt auf Gestänge]  
430 M1: hm  
(2)  
L: genau (.) und die MAUS guckt die entschLOSSEN ode:r wie guckt die ÄNGStlich  
[J1 blättert um] [blättert um, es sind Dächer und Schornsteine zu sehen]  
[L blättert wieder zurück]  
435 M2: n bisschen (.) verZWEIfel:t  
J1: eigentlich guckt die so wie immer  
[blättert zurück auf die Maus mit ihren Flügel und vor der Landkarte]  
L: verZWEIFELT  
[Kreis blinkt auf Bild. J1 tippt auf den Kreis, Bild verschiebt sich Richtung Karte, Vorleser beginnt  
440 erneut den bereits gehörten Text zu sprechen]  
L: ok [tippt auf Weiterpfeil, flüstert] das hattmer schon  
M2: [flüstert unverständlich] hm  
[blättert um auf Hausdächer] (GrC\_Lind\_App, Z. 408-443)

Nach einer anfänglichen, stark auf die Animationen und Funktionen der App ausgerichteten Lektüre geraten nach dem Erkunden der Animationen mehr und mehr einzelne Aspekte der Geschichte in den Blick; auch durch Impulse der Lehrerin. In Zeile 420 greift die Lehrerin eine Aussage von J1 fast wortwörtlich auf. Durch die Reformulierung wird sie relevant gesetzt und als Thema für den folgenden Diskurs etabliert. Die Aussage wird im Folgenden von allen Kin-

dern bestätigt. An dieser Stelle vertieft nun die Lehrerin das Gespräch, indem sie nach Begründungen für die These fragt (Z. 423). J1 antwortet und verweist dabei auf seine Bildbeobachtungen. Die Ausführungen machen deutlich, dass die Kinder – auch ohne es verbalisiert zu haben – intensiv die Geschichte mitverfolgt und Vorstellungen zu den handlungstragenden Elementen – in diesem Falle zur Technik der Flugmaschine – aufgebaut haben.

Strukturell markant an dieser Szene ist, dass die einsetzende, meta-literarische Auseinandersetzung ihren Ausgangspunkt bei einer Aussage von J1 nimmt. Diese wird von Frau C aufgenommen und anhand verschiedener Fragen elaboriert und angereichert. Dabei bringt Frau C nicht nur Fragen, sondern auch eigene Erläuterungen – teils als Antwortvarianten, teils als Antwortergänzungen – ins Gespräch ein. Die Schüler\*innen sind an diese Deutungen jedoch nicht gebunden. Zwar bestätigen sie zumeist die eingebrachten Deutungen, aber es gibt trotz der geschlossenen Frageform eine alternative Deutung von M1 und einen offenen Widerspruch durch J1. Dieser seltene Moment diskrepanter Lektüreeindrücke wird jedoch nicht zum Thema gemacht, sondern einseitig im Sinne von M1 bestätigt. Frau C beharrt gleichzeitig aber nicht auf ihrer Deutung, es findet allerdings auch keine Klärung zu den verschiedenen Interpretationsansätzen statt. Dennoch wird deutlich, dass die klare Hierarchie- und Dominanzstruktur, wie sie bei der Bilderbuch-Lektüre prägend war, zugunsten einer stärkeren Beteiligung der Schüler\*innen modifiziert wurde. Sie bringen diese selbstbewusste, abweichende Deutung ins Spiel und übernehmen selbstständig die Lektüreorganisation.

## **Ergebnisse im Überblick**

Im Folgenden sollen zentrale Befunde des gruppeninternen Fallvergleichs in Gruppe C zusammengefasst werden.

1) Die Sequenzen der App-Lektüre weisen – inhaltlich betrachtet – im Vergleich zur Bilderbuch-Lektüre eine eigene Charakteristik auf. Nach der Einführung in die App und den Lektüreprozess erkunden die Kinder die App anfangs weitgehend kommentarlos. Im Mittelpunkt steht die Inszenierung der App als performatives Angebot, das sich aus Bild, Vorlesetext und Animationen zusammensetzt und durch die geschickte Aktivierung der Animationen als künstlerisches Artefakt zur Aufführung gebracht werden muss. Das leisten insbesondere die Schüler\*innen, wozu sie anfangs von der Lehrkraft explizit herausgefordert werden. Besonders dominant tritt J1 auf. Das meta-literarische Sprechen über die narrativ-literarästhetische Dimension der App tritt hier in den Hintergrund. Das kann passieren, weil die Lehrkraft sich selbst als Gesprächsleiterin zurücknimmt. Sie delegiert die Aufgabe der Organisation und Strukturierung des Lektürevorgangs weitgehend an die Kinder und interveniert vor allen Dingen dann, wenn organisatorisch-technische Fragen auftreten. Diese Beobachtung steht im Kontrast zur Bilderbuch-Lektüre, bei der die Lehrerin organisatorisch und inhaltlich dominant agiert und durch enge, geschlossene Fragen den Lektüreprozess bestimmt. Mit fortschreitender App-Lektüre tritt die Lehrerin verstärkt als Impulsgeberin auf. Dabei nimmt sie aber – anders als bei der Bilderbuch-Lektüre, bei der die Lehrerin ihre eigenen Fragen zur Sicherung des Textverständnisses nutzt – tendenziell stärker Kommentare,

Beobachtungen und Denkanstöße der Schüler\*innen auf und macht diese zum Ausgangspunkt einer vertiefenden Auseinandersetzung. Die Kinder reagieren darauf weniger mit geschlossenen Antworten als mit Thesen, Vermutungen und Begründungen. Divergierende Interpretationen bleiben jedoch vielfach unkommentiert nebeneinander stehen.

2) Auch interaktions-strukturell sind deutliche Unterschiede erkennbar. Während sich bei der analogen Bilderbuchlektüre ein festes Muster von Frage-Antwort-Evaluierung (IRE<sup>5</sup>) zu den Einzelseiten etabliert hat (vgl. Mehan, 1979; Wenzl, 2014), das von einer eindeutigen Hierarchie und Dominanz der Lehrerin geprägt war, finden sich bei der App-Lektüre unterschiedliche Strukturmuster der kollaborativen Sinngestaltung. Die Rolle der Lehrkraft ist dabei wesentlich weniger dominant; mitunter tritt sie ganz zurück.

3) In diesen beiden Phasen der App-Lektüre gibt es deutliche Unterschiede zur Bilderbuch-Lektüre, die jedoch weniger in der Logik des Mediums als im Zugriff begründet liegen. Die Inszenierungspraxis bei der App-Lektüre ist tatsächlich einem multimodalen Mehrwert der App geschuldet, das offene Nachdenken und der Rückzug der Lehrkraft von ihrer Rolle als Gesprächsleiterin hingegen sind programmatisch wichtige Momente der Vorlesegespräche. Markant hier ist, dass diese Differenz zur Bilderbuch-Lektüre auftritt und durchaus für Gruppe C als medienspezifische Differenz zu deuten ist. Der dabei beobachtbare Rollenwechsel der Lehrkraft im Verlauf der App-Lektüre deutet auf eine gewisse Unsicherheit hin, die sich – im Gegensatz zur konsistenteren und sicherlich auch erfahrungsgesättigteren und routinierteren Bilderbuch-Lektüre – dem neuen Medium gegenüber zeigt. Frau C reagiert mit vorsichtiger Zurückhaltung, bevor sie sich wieder stärker in die Lektüre einbringt. Entlastet von der Rolle der organisatorischen Leitung und performativen Inszenierung der Lektüre, ist sie jedoch auch offener und interessierter an den eigenständigen Beobachtungen und Kommentierungen der Schüler\*innen, die sie durch Rückfragen und Reformulierungen relevant setzt und vertiefend ins Gespräch bringt.

4) Grundsätzlich kann nicht mit Sicherheit Auskunft über die Zielsetzungen der beiden Lektüren gegeben werden. Deutlich wird aber, dass beim Bilderbuch der Fokus mehr auf dem Inhalt liegt, bei der App die materielle und technische Seite stark in den Blick rückt. Das hat auch zur Folge, dass Bilder bei der App-Lektüre genauer besprochen und Vorstellungen zu ihren Inhalten entwickelt werden. Auch die Struktur der Gespräche wird offener, die Fragen sind stärker produktiv ausgerichtet und der Diskursverlauf wirkt variabler. Das bestätigen auch die Befunde, die sich bereits im Rahmen der Analyse der deskriptiven Statistiken zeigten.

## **Diskussion und Ausblick**

Die vergleichende Interpretation der beiden Lektüren von Gruppe C offenbart erhebliche Differenzen in der Ausgestaltung der Lektüresituation. Während beim Lesen des analogen Bil-

---

<sup>5</sup> Initiation – reply – evaluation sequence.

derbuchs eine starre, geschlossene und lehrer\*innendominierte Form der Frage-Antwort-Interaktion vorherrscht, findet sich für die Bilderbuch-App eine offenere Gesprächsstruktur, die gerade der Lehrer\*in die Möglichkeit eröffnet, sich in einer eher beobachtenden und impulsaufnehmenden Rolle wiederzufinden.

Auch wenn der Fall von Gruppe C im Sample der Untersuchung einen Extremfall markiert, finden sich ähnliche Tendenzen auch in den anderen Fällen (vollständige Ergebnisdarstellung vgl. Ritter & Ritter, 2020). Die App-Lektüren sind nicht nur ausführlicher und interaktiv dichter, sie geben auch den schüler\*innendominierten Aktionsformen (z. B. nennen einer Beobachtung und Irritation, Auffinden und Erproben der Bildschirmanimationen) einen größeren Stellenwert und bieten damit den Lehrkräften die Möglichkeit, stärker auf die spontanen Lektüreaktionen der Schüler\*innen einzugehen. Dies ist im Sinne der Programmatik der Vorlesegespräche (vgl. z. B. Wieler, 2019) als Vorteil zu werten, da so die Rolle der Schüler\*innen als aktiv sinnkonstruierende Lektüresubjekte gestärkt wird (ausführlich vgl. Ritter & Ritter, 2020).<sup>6</sup> Damit zeigt sich, dass es sich wenigstens lohnen sollte, Bilderbuch-Apps als spezifische Formate der Lektüre intensiver in den Fokus systematischer Forschungen zum Lesen(-lernen) mit digitalen Medium einzubeziehen.

## Primärliteratur

De Lestrade, A., & Docampo, V. (2013). *Die große Wörterfabrik*. München: mixtvision. [App]

Kuhlmann, T., & Salhi, D. (2015). *Lindbergh. Die abenteuerliche Reise einer fliegenden Maus*. Hamburg: Friedrich Oetinger. [App]

## Sekundärliteratur

Anders, P. (2018). Visuelle und digitale Medien. In J.M. Boelmann (Hrsg.), *Empirische Forschung in der Deutschdidaktik. Band 3: Forschungsfelder* (S. 231–250). Baltmannsweiler: Schneider Verlag Hohengehren.

Aufenanger, S. (2018). Forum 5: Lehren und Lernen mit digitalen Medien. Ein kurzer Forschungsüberblick. In: GEW (Hrsg.), *Die digitale R\*Evolution? Herausforderungen für Berufliche Bildung und Weiterbildung. Dokumentation der Bundestagung Berufliche Bildung und Weiterbildung. 23./24.11.2017* (S. 50–52). Frankfurt a.M.: GEW.

Dawidowski, Chr. (2009). *Literarische Bildung in der heutigen Mediengesellschaft. Eine empirische Studie zur kultursoziologischen Leseforschung*. Frankfurt a.M.: Lang.

Dawidowski, Chr. (2013). Lesen Digital Natives anders? *Julit*, 39(2), 7–15.

Deppermann, A. (2008). *Gespräche analysieren. Eine Einführung*. Wiesbaden: VS.

---

<sup>6</sup> Ausführliche Ergebnisdarstellungen und Begründungen finden sich im Projektbericht (Ritter & Ritter, 2020). An dieser Stelle kann nur exemplarische auf den Lektüervergleich in Gruppe C eingegangen werden.

- Frederking, V. (2013). Symmedialer Literaturunterricht. In V. Frederking, H.-W. Huneke, A. Krommer, & Ch. Meier (Hgg.). *Literatur und Mediendidaktik. Bd. 2*, 2. erw. Aufl. (S. 535–567). Baltmannsweiler: Schneider Verlag Hohengehre.
- Fuhrmann, C., & Merklinger, D. (2015). Literarisches Lernen in Vorlesegesprächen. In H. De Boer (Hrsg.). *Gespräche über lernen – Lernen im Gespräch* (S. 251–266). Wiesbaden: Springer.
- Härle, G., & Steinbrenner, M. (Hgg.) (2004). *Kein endgültiges Wort. Die Wiederentdeckung des Gesprächs im Literaturunterricht*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Harrison, C. (2009). Digitale Literalität. Lesen, Schreiben, Wissen und Kommunikation durch Computer und Internet unterstützen. In A. Bertschi-Kaufmann, & C. Rosebrock (Hgg.). *Literalität. Bildungsaufgabe und Forschungsfeld* (S. 73–90). München, Weinheim: Juventa Verlag.
- Heller, V., & Miriam M. (2018). Gesprächsanalyse. Mikroanalytische Beschreibung sprachlicher Interaktion in Bildungs- und Lernzusammenhängen. In J.M. Boelmann (Hrsg.). *Empirische Forschung in der Deutschdidaktik. Band 2: Erhebungs- und Auswertungsverfahren* (S. 221–244). Baltmannsweiler: Schneider Verlag Hohengehren.
- Krommer, A. (2019). *Paradigmen und palliative Didaktik. Oder: Wie Medien Wissen und Lernen prägen*. URL: <https://axelkrommer.com/2019/04/12/paradigmen-und-palliative-didaktik-oder-wie-medien-wissen-und-lernen-praegen/>Kim 2918 (Abgerufen am 15.09.2020)
- Krummheuer, G., & Naujok, N. (1999). *Grundlagen und Beispiele Interpretativer Unterrichtsforschung*. Wiesbaden: VS.
- Kruse, I. (2016). Ästhetisches Lernen, Sprachbildung und Teilhabe. *Grundschulunterricht Deutsch*, 62(1), 4–8.
- Kruse, I. (2013). Hoch inferentes Rating: Qualität angeleiteter Bilderbuchrezeption. In F. Lipowsky, & G. Faust (Hgg.). *Dokumentation der Erhebungsinstrumente des Projekts "Persönlichkeits- und Lernentwicklung von Grundschulkindern" (PERLE) (Materialien zur Bildungsforschung, 23/3)* (S. 219–253). Münster, München, Berlin: Waxmann.
- Medienpädagogischer Forschungsverbund Südwest (2013). *miniKIM-Studie 2012. Kleinkinder und Medien. Basisuntersuchung zum Medienumgang 2- bis 5-Jähriger*. Stuttgart: MPFS.
- Medienpädagogischer Forschungsverbund Südwest (2019). *KIM-Studie 2018. Kindheit, Internet, Medien. Basisuntersuchung zum Medienumgang 6- bis 13-Jähriger*. Stuttgart: MPFS.
- Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge: Harvard University Press.
- Muratović, B. (2014). *Vorlesen digital. Interaktionsstrukturierung beim Vorlesen gedruckter und digitaler Bilderbücher*. Berlin: De Gruyter.

- Orland, C. C., & Ritter, M. (2019). Lesetradition und Medienwandel. Überlegungen zur Erforschung von Lesepraxen und Leseerwerb unter sich ändernden schriftkulturellen Bedingungen. *kj&m*, 71(4), 70–74.
- Philipp, M. (2020). Analoges versus digitales Lesen – 1:0? *Julit*, 46(1), 3–10.
- Ritter, A. (2014). *Bilderbuchlesarten von Kindern*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Ritter, A., & Ritter, M. (2016). Papierdenken im Datenstrom. Kindliche Lesarten von Bilderbüchern und ihren Apps. In G. Scherer, & St. Volz (Hgg.). *Im Bildungsfokus: Bilderbuchrezeptionsforschung* (S. 339–352). Trier: WVT.
- Ritter, M. (2019). *Bilderbuch-Apps*. Lexikoneintrag Universität Duisburg-Essen. URL: <http://kinderundjugendmedien.de/index.php/152-fachlexikon/fachdidaktik/2735-bilderbuch-apps> (Abgerufen am 30.03.2020)
- Ritter, A., & Ritter M. (2020). *Lesepraxen im Medienzeitalter. Vorlesegespräche zu Bilderbuch-Apps. Ein Projektbericht*. München: kopaed.
- Schaumburg, H. (2020a). Personalisierung mit digitalen Medien. In G. Brägger, & H.-G. Rolff (Hgg.). *Kompetenzorientiert Unterrichten und Lernen mit digitalen Medien. Handbuch für kooperative Unterrichtsentwicklung und Förderung personalisierten Lernens*. Weinheim: Beltz.
- Schaumburg, H. (2020b). *Individuelle Förderung mit digitalen Medien aus didaktischer Perspektive. Vortragfolien zum Vortrag auf der Mercator-Tagung: Hauptsache digital?!*, 18.02.2020.
- Schneider, S. (1995). *Entwicklung sozialer Handlungsfähigkeit. Formen der Interaktionsstrukturierung beim gemeinsamen Bilderbuchlesen von Eltern und Kind*. Frankfurt/M.: Peter Lang.
- Wenzl, T. (2014). *Elementarstrukturen unterrichtlicher Interaktion. Zum Vermittlungszusammenhang von Sozialisation und Bildung im schulischen Unterricht*. Wiesbaden: Springer.
- Wieler, P. (1997). *Vorlesen in der Familie. Fallstudien zur literarisch-kulturellen Sozialisation von Vierjährigen in der Familie*. Weinheim, München: Juventa Verlag.
- Wieler, P. (2019). Reden, Zuhören, Bedeutung konstruieren bei der gemeinsamen Bilderbuchrezeption. In J. Knopf, & U. Abraham (Hgg.). *Bilderbücher Band 1: Theorie* (S. 176–188). Baltmannsweiler: Schneider Verlag Hohengehren.
- Wieler, P., Brandt, B., Naujok, N., Petzold, J., & Hoffmann, J. (2008). *Medienrezeption und Narration: Gespräche und Erzählungen zur Medienrezeption von Grundschulkindern*. Freiburg i. Br.: Fillibach.

**Alexandra Ritter**, Dr. phil., ist wissenschaftliche Mitarbeiterin am Arbeitsbereich Deutsch des Instituts für Schulpädagogik und Grundschuldidaktik der Martin-Luther-Universität Halle-Wittenberg. Im Schwerpunkt beschäftigt sie sich mit Fragen der Kinder- und Jugendliteraturentwicklung und ihrer Rezeption und Didaktik, insbesondere bezogen auf das Bilderbuch und seine Digitalisierung. Sie leitet die bundesweite Arbeitsgemeinschaft Jugendliteratur und Medien (AJuM in der GEW).

[alexandra.ritter@paedagogik.uni-halle.de](mailto:alexandra.ritter@paedagogik.uni-halle.de)

**Michael Ritter**, Dr. phil., ist Professor für Grundschuldidaktik Deutsch/Ästhetische Bildung am Institut für Schulpädagogik und Grundschuldidaktik der Martin-Luther-Universität Halle-Wittenberg. Neben Fragen der Bilderbuchforschung (analog und digital) interessiert er sich besonders für die Konzeption des Deutschunterrichts in der inklusiven Schule und die fachbezogene Professionalisierung von Lehrkräften.

[michael.ritter@paedagogik.uni-halle.de](mailto:michael.ritter@paedagogik.uni-halle.de)

# 3

Entwicklung und Testung ausgewählter didaktischer Szenarien im Kontext von E-Learning

# Das Inverted Classroom Mastery Modell in der Informatik: Untersuchungen zum Kompetenzerwerb in der Objektorientierten Programmierung

*Annett Thüring<sup>1</sup> und Kathrin Jäger<sup>2</sup>*

<sup>1</sup> *Martin-Luther-Universität Halle-Wittenberg, Institut für Informatik*

<sup>2</sup> *Martin-Luther-Universität Halle-Wittenberg, Zentrum für multimediales Lehren und Lernen*

Objektorientierte Programmierung (OOP) ist eine Grundlagenveranstaltung der Informatik. An der Martin-Luther-Universität Halle-Wittenberg wird diese für ca. 300 Studierende im ersten Semester als klassische Präsenzvorlesung mit Übung angeboten. In den letzten Jahren waren hohe Abbruchquoten seitens der Studierenden zu verzeichnen. Dieser Aspekt sowie die Bedürfnisse und Wünsche der Studierenden nach individuelleren und flexibleren Studienmodellen rückten die Neukonzeption der Veranstaltung in den Fokus. Nach einer detaillierten wissenschaftlichen Analyse der Veranstaltung wurde in einer umfassenden didaktischen Designplanung ein lernerorientiertes Blended Learning-Konzept entwickelt. Dieses greift die Idee der Kompetenzorientierung im Inverted Classroom Mastery Model (ICMM) auf. Für einen engen Praxisbezug und den Erwerb von Fähigkeiten wurde als Grundmodell für Problemlöseaufgaben das Cognitive Apprenticeship-Model (CAM) herangezogen. Ziel ist, den Lernerfolg schrittweise zu ermöglichen, zu messen und transparent, sowohl für den Lehrenden als auch den Studierenden, sichtbar zu machen. Um die Integration des Wissens in die individuelle Wissensstruktur zu ermöglichen und zu fördern, nutzen Onlinephasen unterschiedlich aufbereitete digitale Elemente, unterstützen verstärkt selbstreguliertes Lernen und fokussieren Mastery Learning zum Erwerb von Fachkompetenz. Zusätzlich orientieren Präsenzphasen auf intensiven Diskurs sowie Anwendung und Vertiefung. Nach einer ersten Testphase können praktische Erfahrungen positiv diskutiert werden.

## Einleitung

Das Modul OOP wird vom Institut für Informatik seit dem WiSe 2012/13 am Beispiel der Programmiersprache Java durchgeführt. Die Veranstaltung wurde bisher im klassischen Präsenz-Lehrformat Vorlesung und Übung über ein Semester angeboten. Sie ist obligatorisch für Studierende der Informatik im ersten Semester und wahlobligatorisch für weitere Studienrichtungen. In den letzten Jahren sind steigende Studierendenzahlen, aber auch eine zunehmende Heterogenität hinsichtlich der Vorkenntnisse und im Lehrveranstaltungsverlauf häufige Kursabbrüche seitens der Studierenden zu verzeichnen. Vor diesem Hintergrund ist eine zielführende Gestaltung der Präsenzveranstaltungen, die den Anforderungen aber auch Bedürfnissen aller Studierenden gerecht wird, erschwert. Einer Neukonzeption der Veranstaltung geht eine Analyse der Lehrveranstaltung ab WiSe 2012/13 voraus, um im Ergebnis aus den Daten Rückschlüsse für Veränderungen und Maßnahmen auf der Ebene verschiedener Handlungsfelder abzuleiten.

## Analyse der Lehrveranstaltung

Im Fokus der Analyse stehen die Studierenden der Lehrveranstaltungen, die Vermittlungsart und Verfügbarkeit der Fachinhalte, die Studienleistungen und die Ergebnisse der summativen Leistungsmessung sowie die durchgeführten Evaluationen. Zur Analyse liegen sieben Kohorten vor, zuzuordnen den OOP-Lehrveranstaltungen der WiSe 2012/13 bis WiSe 2018/19. Zur Datenerhebung werden zusätzlich Stud.IP<sup>1</sup>, YAPEX<sup>2</sup> und das Löwenportal<sup>3</sup> herangezogen (**Tabellen 1-3**).

---

<sup>1</sup> Stud.IP – (Studienbegleitender Internetsupport von Präsenzlehre) ist die an der Martin-Luther-Universität Halle-Wittenberg eingesetzte webbasierte Plattform zur Organisation der Studierenden in den Lehrveranstaltungen.

<sup>2</sup> YAPEX – (Yet Another Practical EXercise Platform) ist eine von der Martin-Luther-Universität Halle-Wittenberg am Institut für Informatik entwickelte webbasierte Arbeitsumgebung zum Bearbeiten von Programmieraufgaben. Das Verwaltungsmanagement von YAPEX liefert statistische Übersichten über alle bearbeiteten Aufgaben und Bewertungen (Dähne, 2016).

<sup>3</sup> Das Löwenportal registriert die Anmeldung der Studierenden zum Modul und den zugehörigen Prüfungen und dokumentiert die erreichten Studien- und Prüfungsleistungen.

**Tabelle 1.** Kohorten Modul OOP WiSe 2012/13 bis WiSe 2018/19 (klassisches Format: Vorlesung/Übung)

Semester	Anzahl Studierende des Moduls	Anzahl Studienrichtungen	Anzahl Studierende angemeldet	Klausur bestanden	Klausurergebnis Durchschnitt
2012/13	166	26	127	103	2,5
2013/14	144	28	103	79	2,6
2014/15	130	26	102	64	3,2
2015/16	157	37	102	75	3,7
2016/17	182	40	104	85	3,0
2017/18	237	49	140	88	3,8
2018/19	279	51	174	131	3,0

In den letzten Jahren ist neben einer steigenden Anzahl Studierender des Moduls OOP auch eine zunehmende Heterogenität hinsichtlich der Studienrichtungen zu verzeichnen. Seit dem WiSe 2014/15 schließen weniger als die Hälfte der Studierenden die summative Prüfung erfolgreich ab. Zur Analyse möglicher Ursachen werden im Folgenden die Veranstaltungsführung und die zu erbringenden Studienleistungen herangezogen (**Tabellen 2 und 3**).

**Tabelle 2.** Analyse der Lehrveranstaltung OOP WiSe 2012/13 bis WiSe 2018/19

Semester	Präsenz VL mit Skript/ Übung	Bereitstellung von Online-Material (nicht verpflichtend)				Studienleistung und Klausur im gleichen Semester	Klausur klassisch bzw. E-Klausur (*)
		Aufzeichnung Vorlesung	Aufzeichnung Übung	ILIAS-Modul Erweiterung ab 2015/16 Aufgaben: Video/Text	YAPEX Schnittstelle ILIAS/ YAPEX ab 2018/19		
2012/13	+/+					+	+
2013/14	+/+	+	+				+
2014/15	+/+	+	+	+			+
2015/16	+/+		+	+			+
2016/17	+/+		+	+	+		+
2017/18	+/+		+	+	+		+
2018/19	+/+	+	+	+	+		+*

**Tabelle 3.** Studienleistungen zum Modul OOP WiSe 2012/13 bis WiSe 2018/19

Semester	Studienleistung
2012/13	als Klausurvorleistung – 50 % der Punkte aller Home-Aufgaben der Übungen und 2 Praxistests, einer bestanden
2013/14	50 % der Punkte aller Home-Aufgaben der Übungen und 2 Praxistests, einer bestanden
2014/15	50 % der Punkte aller Home-Aufgaben der Übungen und 50 % der Punkte praktischer in den Übungen gelöster Aufgaben
2015/16 2016/17 2017/18	50 % der Punkte aller Home-Aufgaben der Übungen und mindestens drei in den Übungen erfolgreich gelöste Programmieraufgaben oder Praxistest am Ende des Semesters bestanden
2018/19	50 % der Punkte aller Home-Aufgaben der Übungen und 6 Praxispunkte aus Programmiertests maximal 2 Punkte pro Programmiertest möglich

Der Vergleich der Durchführung des Moduls von WiSe 2012/13 bis WiSe 2018/19 zeigt eine Vielzahl veränderter Bedingungen. Diese beziehen sich auf folgende Maßnahmen und Aspekte:

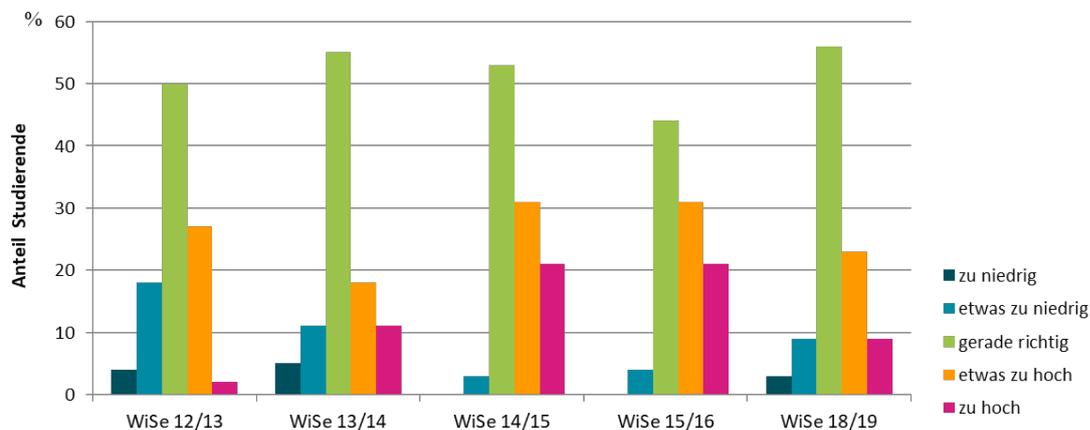
- zusätzliche Bereitstellung von Aufzeichnungen der Vorlesungen und Übungen (unterschiedlich über die Semester)
- Bereitstellung eines nicht verpflichtend zu bearbeitenden Online-Begleitmoduls im LMS ILIAS und Integration der YAPEX-Übungsplattform über eine Schnittstelle im LMS (ab WiSe 2014/15)
- Terminierung der Abgabe der Studienleistungen (nur WiSe 2012/13 sind die Studienleistungen zeitgleich im Semester und als Klausurvorleistung zu erbringen)
- Art der Leistungsmessung (Papierklausur bis WiSe 2017/18, danach E-Klausur)
- Art der Studienleistungen (zu erbringende Studienleistungen wechseln über den Verlauf, **Tabelle 3**)

Zusätzlich trat ein Wechsel der Dozenten für die Vorlesungseinheit auf. Dieser fand in den Semestern 2015/16 und 2018/19 statt. Fachdidaktisch nutzten die Dozenten die gleiche Vorgehensweise (objects-later). Folgende Unterschiede zeigten sich:

- Wissensvermittlung mit Verweis auf eigene Webseiten zur Thematik und begleitende Fachinhalte, ohne Aufzeichnung (WiSe 2012/13), zusätzliche Aufzeichnung (WiSe 2013/14, WiSe 2014/15, WiSe 2018/19)
- Wissensvermittlung ohne Aufzeichnung (WiSe 2015/16 bis WiSe 2017/18)

Eine Auswertung wird durch die veränderten Durchführungen erschwert. Im WiSe 2012/13 konnten 62 % der Studierenden des Moduls die Klausur erfolgreich abschließen. In den folgenden sechs Jahren bestanden prozentual weniger Studierende die Klausur. Seit WiSe 2016/17 meldeten sich weniger Studierende zur Prüfung an. Es wird zudem vermutet, dass auch die Studienleistung, die nur im WiSe 2012/13 als Klausurvorleistung zu erbringen war, das Klausurergebnis positiv beeinflusste. Für alle darauffolgenden Semester war eine Abgabe der Studienleistung auch im späteren Studienverlauf möglich. Die einzubringenden Leistungen für WiSe 2012/13 bis WiSe 2018/19 waren in Folge unterschiedlich (**Tabelle 3**).

Die Lehrveranstaltungsevaluation erfasste, dass die Studierenden die Praxistests als starke Belastung wahrnahmen. Mit der als Erleichterung angedachten Änderung wurde scheinbar bei den Studierenden ein gegenteiliger Effekt erreicht. Die gefühlte Wahrnehmung der Studierenden hinsichtlich der Anforderungen wird den Lehrveranstaltungsevaluationen WiSe 2012/13 bis WiSe 2015/16 und des WiSe 2018/19 entnommen (**Abbildung 1**). Die Bewertung des Items „Die Anforderungen zur Erreichung der Leistungspunkte sind...“ erfolgte in Form der Auswahl auf einer fünfstufigen Likert Skala. Die Einstufung der Anforderungen im Bereich „etwas zu hoch“ oder „zu hoch“ kann im Zusammenhang mit dem Lernerfolg diskutiert werden. Mehr als die Hälfte der Studierenden nimmt die Anforderungen im WiSe 2014/15 und WiSe 2015/16 als „etwas zu hoch“ oder „zu hoch“ wahr, der Klausurdurchschnitt verschob sich im Vergleich zu den Vorsemestern zu schlechteren Noten. Ob auch veränderte Vermittlungsarten ursächlich sind oder aber die Bereitstellung von Vorlesungsaufzeichnungen einen Effekt hat, kann aufgrund der zur Verfügung stehenden Daten jedoch nicht vollständig und isoliert diskutiert werden.



**Abbildung 1.** Einschätzung der Anforderungen – gefühlte Wahrnehmung der Studierenden

Da die auf den Evaluationen beruhenden jährlichen Anpassungen keinen höheren Lernerfolg erzielten, geriet eine Neukonzeption der Gesamtveranstaltung in den Fokus. Grundlage dazu war neben der Lehrveranstaltungsanalyse zusätzlich eine Analyse universitärer Lehr-Lern-Konzepte und -Methoden für OOP sowie die Entwicklung eines Modells zur Messung des Kompetenzerwerbs über das OOP-Modul.

## Forschungsbasis

Die nachfolgende Literatur zeigt unterschiedliche didaktische und methodische Ansätze sowie Lehrveranstaltungskonzepte mit der Intention, den Erwerb von deklarativem und prozeduralem Wissen und die Fähigkeit des Programmierens zu erlernen, zu unterstützen. Die hier vorgestellte Recherche stützt sich auf drei Aspekte:

- Fachdidaktische Konzepte zur Vermittlung der Fachinhalte von OOP
- didaktische Modelle zur kompetenzorientierten Ausrichtung der Lehrveranstaltung
- Kompetenzentwicklungsmodelle für OOP

Fachdidaktische Konzepte verfolgen bei gleichen Fachinhalten des Moduls zwei unterschiedliche Vorgehensweisen: objects-first und objects-later. Objects-first beginnt mit der objektorientierten Sichtweise, objects-later stellt die prozedurale Sichtweise an den Anfang. Der Unterschied liegt somit in der Abfolge der Fachinhalte. Für beide Konzepte werden in Fachkreisen spezifische Vor- und Nachteile diskutiert. Trotz einer facettenreichen fachdidaktischen Diskussion besteht nach wie vor kein Konsens über ein favorisiertes Modell zur Vermittlung von OOP (Bennedsen & Schulte, 2007; Ehlert & Schulte, 2009; Ehlert, 2012; Uysal, 2012).

Zur Ausrichtung der Lehrveranstaltung OOP rücken fähigkeitsorientierte Modelle in den Fokus. Basis liefert Weinert (2001) mit seiner Kompetenzdefinition, in der Kompetenzen als die „bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen“ verstanden werden, aber auch „die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften“, die für einen erfolgreichen und verantwortungsvollen Transfer notwendig sind, in die Betrachtungen eingehen (Weinert, 2001, S. 27). Zur Vermittlung von Kompetenzen greift Bligh (2000) in seinen Ausführungen als Basis der didaktischen Führung die Passung von Lehr-Lern-Form und Zielgruppe bzw. Heterogenität der Zielgruppe auf. Bligh (2000) argumentiert, dass eine passive Vorlesung zur Vermittlung nachhaltigen Wissens für alle Zielgruppen schwierig ist und stellt heraus, dass eine Vorlesung für die Entwicklung von Fähigkeiten ungeeignet ist. Ins Blickfeld für OOP geraten somit prozessorientierte Ansätze, die auf soziokonstruktivistische und situierte didaktische Konzepte orientieren. Diese setzen sich insbesondere mit den individuellen Lern- und Lösungsprozessen als aktiven und konstruktiven Prozess der Lernenden auseinander und erfordern eine Eingebundenheit in eine konkrete Situation (Mandl, Gruber, & Renkl, 1995; Brown, Collins, & Duguid, 1989). Die Ansätze solcher Modelle wie Cognitive Apprenticeship (Collins, Brown, & Newman, 1989; Collins, Brown, & Holum, 1991) oder Anchored Instruction (The Cognition and Technology Group at Vanderbilt, 1990) finden sich zudem in speziellen didaktischen Ansätzen für OOP wieder. So fokussiert das 4-Schritte-Modell von Fässler (2007) die Kompetenzorientierung mit dem Ziel der Aneignung aktiven transferfähigen Wissens durch intensives Einüben der vermittelten Konzepte und Fertigkeiten bei zunehmender Selbstständigkeit der Studierenden. Die Vermittlung der Lerninhalte erfolgt analog

der Kognitiven Meisterlehre (Collins et al., 1989; 1991). Das nachgestellte Anwendungsebenenmodell zielt im Fortgang auf die flexible Anwendung der erlernten Grundkonzepte in komplexen Programmieraufgaben (Fässler, Hinterberger, Dahinden, & Wyss, 2006; Fässler, 2007). Dahm et al. (Dahm, Barnjak, & Heilemann, 2015; Dahm, Rose, & Köhler, 2020) hingegen verfolgen den Ansatz einer integrierten Entwicklungsumgebung. Das didaktische Konzept basiert auf den fünf Schritten auf dem Weg vom Problem zum Programm - Lesen, Verstehen, Überlegen, Aufschreiben, Codieren. Das Ziel ist die kognitive Belastung, die bei der Erstellung eines Programms insbesondere für Anfängerinnen und Anfänger sehr hoch ist, zu reduzieren (Williams, 2014).

Ein weiteres Ziel ist, die individuell erworbenen Fähigkeiten und Fertigkeiten messbar zu machen. Ansätze liefern Kompetenzentwicklungsmodelle. Empfehlungen für Kompetenzmodelle für das Studienfach Informatik werden von der Gesellschaft für Informatik (GI) herausgegeben (Zukunft, 2016). Diese beinhalten kognitive und nicht kognitive Dimensionen, erstere basieren auf der Bloom'schen Taxonomie (vgl. Anderson & Krathwohl, 2001; Bloom, Engelhart, Fürst, Hill, & Krathwohl, 1956). Voraussetzung für Kompetenzentwicklungsmodelle sind sowohl Niveaustufung und adäquate Aufgabenentwicklung sowie Bewertungsmaßstäbe der Einzelniveaus. Nelles et al. (Nelles, Rhode, & Stechert, 2010) gehen dazu von vier Kompetenzdimensionen aus. Rohde (2013) stellt ein Konzept zur Messung informatischer Modellierungskompetenz vor und zeigt, dass es möglich ist, einen statistisch signifikanten Kompetenzzuwachs im Vergleich von Vor- und Nachtest zu messen. Auch Kramer et al. (Kramer, Hubweiser, & Brinda, 2016) identifizierten die Teilkompetenzen für OOP und ordneten diese einem vierstufigen Modell zu. Die Zuordnung und Messbarkeit der Kompetenzen erfordert jedoch empirische Studien und liegt in vielen Bereichen noch in der Entwicklung. Ein „Prototyp eines Messwerkzeuges zur empirischen Bestimmung solcher Kompetenzen“ wird von Kramer et al. (Kramer, Samimi, & Brinda, 2017) vorgestellt.

## **Vorbetrachtungen zur Entwicklung eines Modells zur Messung des Kompetenzerwerbs für das Modul OOP**

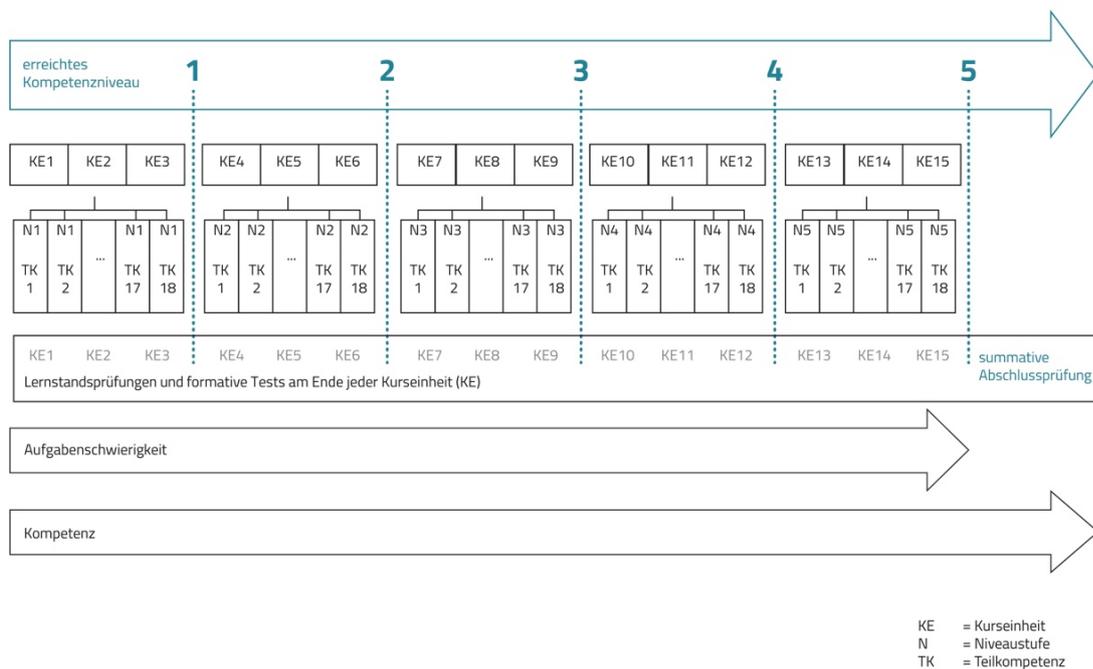
In diesem Beitrag soll basierend auf dem aufgezeigten Wissensstand zunächst der Fokus auf die Analysen und Entscheidungen hinsichtlich des Kompetenzerwerbs gelegt und daraus ein Modell zur Messung des Kompetenzerwerbs, speziell für das hier durchzuführende OOP-Modul, entwickelt werden. Vor diesem Hintergrund werden folgende Aspekte vorangestellt:

- Lern- und Kompetenzziele des Moduls
- Stufen der Kompetenzentwicklung
- favorisiertes Modell zum Kompetenzerwerb
- Lehr-Lern-Format
- formative und summative Kompetenzmessung

Ziel des Moduls OOP ist, unter Verwendung objektorientierter Konzepte, eigene Programme in der Programmiersprache Java zu entwickeln. Zunächst wurden alle Lernziele nach der

Bloom'schen Taxonomie jeder Kurseinheit formuliert (vgl. Anderson & Krathwohl, 2001; Bloom et al, 1956), die insbesondere die von Weicker et al. (Weicker, Draskoczy, & Weicker, 2006) ermittelten Schlüsselkompetenzen für Informatikerinnen und Informatiker berücksichtigen. Die ausgewählten 18 Schlüsselkompetenzen werden jedem Kompetenzniveau zugeordnet. Aus den Lern- und Kompetenzziele lassen sich direkt Prüfungsaufgaben ableiten (constructive alignment) (Biggs, 1996). Obwohl constructive alignment als Konzept für die Hochschullehre kontrovers diskutiert wird (Reinmann, 2018), wird es hier unterstützend eingesetzt, da Studierende im Erstsemester zumeist wenig Erfahrung mit selbstorganisiertem Lernen besitzen. Die Lern- und Kompetenzziele verbunden mit den Prüfungszielen begünstigen selbstreguliertes Lernen und geben Studierenden im Modulablauf Strukturierung, Motivation und Verbindlichkeit.

Im Modulablauf über 15 Semesterwochen werden zusätzlich Stufen der Kompetenzentwicklung favorisiert, die sich aus Teilaspekten zusammensetzen. Für jede Teilkompetenz werden aufeinander aufbauende Konzepte und Handlungsschritte benötigt. Damit scheint eine Unterteilung in Kompetenzniveaus schlüssig. Basierend auf der Taxonomie und auf dem Kompetenzentwicklungsmodell von Dreyfus et al. (Dreyfus & Dreyfus, 1980) wurden für alle Teilkompetenzen fünf Niveaustufen einer kontinuierlichen Kompetenzskala über jeweils drei Kurseinheiten formuliert (**Abbildung 2**). Zur Abgrenzung der Niveaus wurden die konkreten Inhalte und die zur Leistungsmessung relevanten Aufgaben herangezogen.



**Abbildung 2.** Kompetenzerwerb im Modul OOP

Favorisiertes didaktisches Modell zum Kompetenzerwerb ist das erweiterte Cognitive Apprenticeship-Modell (CAM) mit der Abfolge *Content, Situated Learning, Modeling and Explaining, Scaffolding, Fading, Coaching, Exploration, Articulation, Reflection* (Collins et. al., 1989; 1991). Dieses prozessbegleitende Konzept intendiert die individuelle Unterstützung während

des Lösungsprozesses und die Weiterentwicklung des Lernenden und regt dessen Auseinandersetzung mit der eigenen und der Vorgehensweise anderer an. Die hierbei sehr viel höhere und auch konkretere Intensität instruktorischer Unterstützung im Vergleich zu anderen Methoden, steht im universitären Lehrkontext zumeist nicht im Fokus. Für Studierende im Erstsemester wird dieses Modell favorisiert, da es auch Vorteile zur Entwicklung eigener Lernstrategien bietet.

Zur Umsetzung des Moduls OOP wird das Inverted Classroom-Format favorisiert (Bergmann & Sams, 2012). Das Konzept zielt darauf ab, die Wissensvermittlung online vorzulagern und die Präsenzzeit zur Übung und zum intensiven Diskurs zu nutzen (vgl. Handke, Loviscach, Schäfer, & Spannagel, 2012). Gleichermaßen bietet das Format Raum für individualisierte Lernwege, die sowohl der Motivation zuträglich sind, aber auch die Kreativität der Studierenden fördert und das Problemlösen unterstützt. Basierend auf dem o.g. Kompetenzmodell wurden schrittweise Tests entwickelt, die sich jeweils einzelnen Kompetenzfacetten zuordnen lassen. Für eine Auswertung werden die Wahrscheinlichkeiten der Antwortkategorien über Schwellen definiert. Eine Schwelle bildet dabei die Grenze zwischen zwei Niveaustufen (vgl. Rasch, 1960). Umfangreiche Testszenarien werden umgangen, indem in der Umsetzung kürzere Tests in alle Kurskomponenten integriert werden. Positiv wirkt sich zudem die „Rückholpraxis“ beim Lernen durch Testen aus (vgl. Adesope, Trevisan, & Sundararajan, 2017).

### **Kursumsetzung**

Zur Neugestaltung der Veranstaltung wird ein umfassendes Instruktionsdesign mit Planungs- und Entscheidungsparametern zur Entwicklung, Umsetzung und Evaluation genutzt (Reinmann, 2015). Gestaltungsleitend zur Kursumsetzung waren die o.g. Analysen. Fachdidaktisch wird das objects-first-Modell genutzt. Das Voranstellen der objektorientierten Konzepte bietet den Studierenden über das Semester ausreichend Zeit diese durch den Einsatz in verschiedenen Anwendungskontexten zu vertiefen. Unter Beachtung des Workloads wurden die 15 Kurseinheiten in ein interaktives Lernmodul mit formativem Assessment ins Lernmanagementsystem ILIAS<sup>4</sup> überführt und praktische Aufgaben zusätzlich über die YAPEX-Schnittstelle integriert<sup>5</sup> (zur Strukturierung Modul und Kurseinheiten zum Kompetenzerwerb OOP, sowie Praxisansatz, Motivations- und Betreuungskonzept siehe Thüning & Jäger, 2020a). Online wird der Wissens- und Kompetenzerwerb im ICMM durch Mastery Learning und formatives Assessment mit Feedback unterstützt (vgl. Bernitzke, 1987). Die Kurskomponenten wurden entsprechend dem erweiterten ACM kombiniert. Aufgaben zu erworbenem Wissen über mehrere Kurseinheiten greifen auf das Anwendungsebenenmodell zurück (Fässler, 2007). Formative Tests aber auch Umfragen dienen der Reflexion und zielen zudem auf metakognitive Strategien (Thüning & Jäger, 2020b).

---

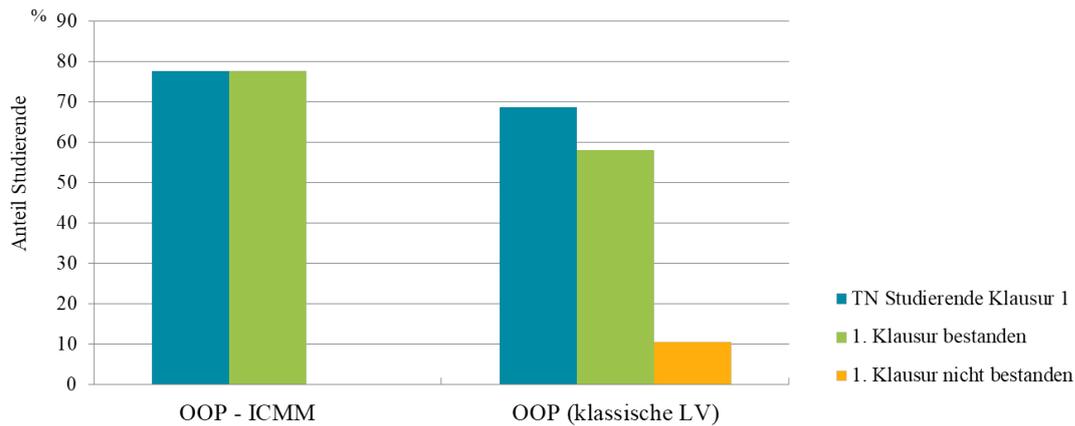
<sup>4</sup> ILIAS – Integriertes Lern-, Informations- und Arbeitskooperations-System ist ein Open-Source Learning Management System. [<https://www.ilias.de/>]

<sup>5</sup> Demo-Kurs abrufbar unter: <https://bit.ly/2UJfSud>. (Nutzer: ICMMDemo, Passwort: ICMM\_ilias2020)

## Ergebnisse und Diskussion

Im WiSe 2019/20 wurde das Modul in Form des Blended Learning-Kurses in einer ersten Testphase für 20 Studierende angeboten. Die Teilnahme der Studierenden am Kurs im ICMM-Format beruhte auf Freiwilligkeit. Die klassische Veranstaltung besuchten im gleichen Semester 217 Studierende. Im Semesterverlauf wurden für den Kurs im ICMM sowohl über Beobachtungen und Online-Umfragen (anonym nach jeder Kurseinheit) als auch über eine abschließende Evaluation unterschiedliche lernrelevante Aspekte untersucht. Zusätzlich wurde der Lernfortschritt über das Semester gemessen (Lernfortschrittskontrolle im ILIAS, anonymisierte Zeit- und Seitenerfassung sowie die Ergebnisse aus formativen Assessments ausgewertet). Die Motivation und hohe Aktivität in Form partizipativer Mitgestaltung der Online-Lernumgebung durch die Studierenden äußerte sich darin, dass Studierende eigenständig ergänzende Testfälle für YAPEX-Aufgaben erstellten und die Kursgestaltung durch Erweiterung der Online-Kurseinheiten mit praxisnahen Beispielen bereicherten. Zur Bewertung des Kompetenzzuwachses und Lernerfolgs wurden die Ergebnisse der Online-Tests und der wöchentlichen Online-Programmieraufgaben herangezogen. Anhand der gezeigten Leistungen konnten im neuen Veranstaltungsformat abschließend positive Rückschlüsse auf erworbene Kompetenzen gemessen werden. Eine vergleichende quantitative Analyse ergab, dass 10 % der Studierenden den Kurs im ICMM abbrachen, im klassischen Kurs hingegen liegt der Kursabbruch bei 39 %. Die summative Prüfung am Ende des Semesters misst inwieweit die Studierenden objektorientierte Systeme modellieren und Algorithmen in eine Programmiersprache umsetzen, sowie Testfälle formulieren können. In die Auswertung wurde aufgrund der veränderten Bedingungen nur die erste Prüfungsmöglichkeit einbezogen (aufgrund der durch COVID-19 bedingten Einschränkungen wurde die zweite Prüfung als Online-Prüfung mit nachträglichem Rücktrittsrecht angeboten, nicht bestandene Prüfungen wirken sich nicht auf die Anzahl der Prüfungsversuche aus). Die Ergebnisse (**Abbildung 3**) zeigen:

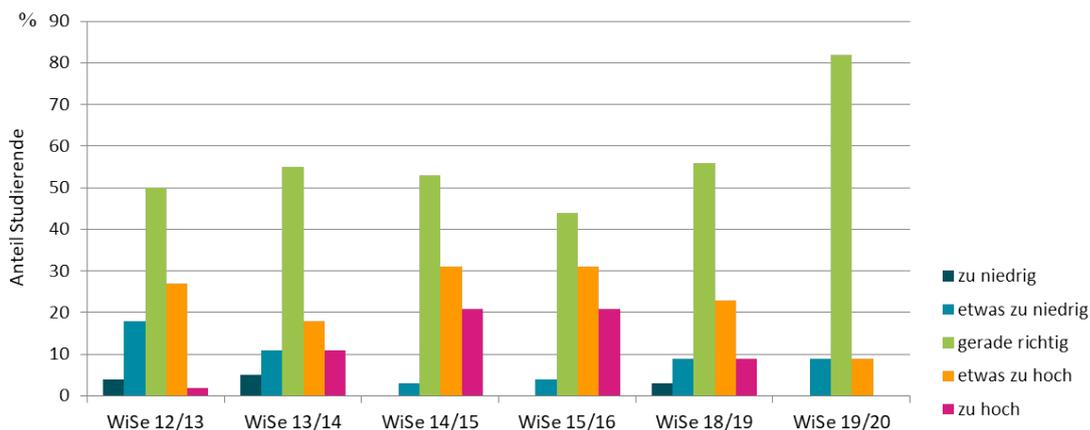
- Es meldeten sich anteilig mehr Studierende aus dem Kurs im ICMM zur ersten Klausur an.
- Alle angemeldeten Studierenden des Kurses im ICMM beendeten die Klausur erfolgreich (schlechteste Note 3,3 - ein Studierender). Das Ergebnis für Studierende im dem klassischen Kurs zeigte, dass 10,68 % der Studierenden die Klausur nicht bestanden, sieben Studierende erhielten zudem die Note 3,7 oder 4,0.



**Abbildung 3.** Klausurteilnahme und summativer Lernerfolg (Klausur 1)

Erste Untersuchungen müssen jedoch unter Vorbehalt betrachtet werden, da die Freiwilligkeit der Kurswahl Alternativerklärungen zulässt. Die erhobenen Daten wurden zusätzlich mit in internationalen Studien ermittelten Ergebnissen von Bennedsen und Caspersen (2007; 2019) verglichen. Bennedsen und Caspersen (2007; 2019) untersuchten das Verhältnis zwischen erfolgreichem Bestehen und Misserfolg in Programmierungseinführungskursen auf Universitätsebene. Der Lernerfolg liegt hier bei 67 %, eine Bilanz, die auch mit dem Ergebnis der summativen Prüfung des Kurses im ICMM gezogen werden kann. Der hohe Misserfolg einführender Programmierkurse wird von Bennedsen und Caspersen (2007; 2019) zudem, wie auch in dieser Arbeit, hinsichtlich der facettenreichen Faktoren des Studieneinstiegs und der individuellen Bedürfnisse der Studierenden in dieser Phase diskutiert.

Vergleichend zur ausgehenden Analyse wird die Bewertung des Items „Die Anforderungen zur Erreichung der Leistungspunkte sind...“ auch in der Evaluation des Kurses im ICMM-Format einbezogen.



**Abbildung 4.** Einschätzung der Anforderungen – gefühlte Wahrnehmung der Studierenden in den Semestern WiSe 2012/13-WiSe 2018/19 mit Vergleich ICMM WiSe 2019/20

Studierende, die den OOP-Kurs im ICMM-Format nutzten, nahmen die Anforderungen im für sie passenden Bereich wahr. Die in diesem Zusammenhang eingangs diskutierten Auswirkungen

gen auf den Lernerfolg bei einer gefühlten stimmigen Wahrnehmung, können mit den Ergebnissen der Testgruppe bestätigt werden. Da die Fachinhalte aller Kurse identisch sind, bleibt jedoch zu prüfen, ob die Freiwilligkeit der Kurswahl einen starken Einfluss auf den gefühlten Einfluss der Anforderungen hat oder welche Aspekte der Umgestaltung den Studierenden am zuträglichsten sind.

## **Fazit und Ausblick**

Die Neukonzeption der Veranstaltung OOP zur Kompetenzvermittlung im ICMM-Format fokussiert auf die verstärkte Lernerzentrierung ausgerichtet an den differenzierten Bedürfnissen der Studierenden. Beobachtungen und praktische Erfahrungen bei der Nutzung der Online-Lernumgebung im ICMM weisen darauf hin, dass Studierende die Kursinhalte mit hoher Motivation ergebnisorientiert bearbeiteten. Formativer und summativer Lernerfolg lassen zudem vermuten, dass die didaktische und mediale Gestaltung der Lernumgebung des ICMM, einschließlich der objects-first Vorgehensweise und der Cognitive Apprenticeship-Methode, für den Erwerb der Kompetenz der Programmierfähigkeit einen Mehrwert zu bieten scheint. Für Studierende im Erstsemester bieten möglicherweise die höhere und auch konkretere Intensität instruktionaler Unterstützung verbunden mit Rückmeldung zum Lernstand und Feedback auch Vorteile zur Entwicklung von Problemlösestrategien. Hervorzuheben sind die geringe Abbruchquote und der höhere Lernerfolg im Vergleich zur klassischen Veranstaltung. Vor dem Hintergrund der positiven Bilanz der Testphase wird die Veranstaltung im ICMM-Format weiter ausgebaut.

Aufgrund der aktuellen COVID-19 bedingten Situation (Hybridsemester) werden erst im WiSe 2021/22 sowohl das klassische Format als auch der Kurs im ICMM parallel angeboten. Studierende können dann selbst wählen, welches Lernformat sie nutzen. Studierende, die sich nicht aktiv entscheiden, werden randomisiert einem der Kurse zugeteilt. Angestrebt wird eine empirische Studie zum Vergleich beider Veranstaltungen in einem 4-Gruppen-Vor-Nachtest-Design. In beiden Formaten soll der Kompetenzzuwachs im Verlauf gemessen werden und das vorgestellte Kompetenzmodell empirisch untersucht werden. Über eine Analyse der Daten soll zudem die Frage beantwortet werden, ob ein vollständiger Überblick über den aktuellen persönlichen Kompetenzstand im Bezug zur Kurskompetenz während einer Veranstaltung zu einer intensiveren Auseinandersetzung mit der eigenen Leistung führt und sich in Folge positiv auf das Lernverhalten auswirken kann.

## **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.

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Boston: Pearson Education Group.

- Bennedsen, J., & Caspersen, M. E. (2007). Failure rates in introductory programming. *SIGCSE Bulletin*, 39(2), 32-36.
- Bennedsen, J., & Caspersen, M. E. (2019). Failure Rates in Introductory Programming: 12 Years Later. *ACM Inroads*, 10(2), 30-36.
- Bennedsen, J., & Schulte, C. (2007). What does “Objects-First” Mean? An International Study of Teachers’ Perceptions of Objects-First. In R. Lister, & B. Simon (Hrsg.), *Proceedings of the Seventh Baltic Sea Conference on Computing Education Research*, CRPIT Vol. 88 (S. 21-29). Koli Calling, Finland.
- Bergmann, J., & Sams, A. (2012). *Flip your classroom: reach every student in every class every day*. Washington DC: International Society for Technology in Education.
- Bernitzke, F. H. (1987). Mastery-Learning-Strategie als Unterrichtsalternative – Empirische Studie zur Effektivität der Mastery-Learning-Strategie und zu Interdependenzen mit Schülermerkmalen. *Europäische Hochschulschriften*, (Reihe 11) Pädagogik/Education/Pédagogie 305. Frankfurt/M.: Peter Lang, Internationaler Verlag der Wissenschaften.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347–364.
- Bligh, D. (2000). *What’s the Use of Lectures?* San Francisco: Jossey-Bass Publishers.
- Bloom, B. S., Engelhardt, M. D., First, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: Handbook I. Cognitive domain*. New York: David McKay.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In L. B. Resnick (Hrsg.), *Knowing, learning and instruction: Essays in honor of Robert Glaser* (S. 453-495). Hillsdale, NJ: Erlbaum.
- Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, 15(3), 6-11.
- Dähne, J. (2016). Sprachunabhängiges automatisiertes Testen und Bewerten von Programmieraufgaben. (Unveröffentl. Bachelorarbeit). Martin-Luther-Universität Halle-Wittenberg, Halle.
- Dahm, M., Barnjak, F., & Heilemann, M., (2015). 5Code - Eine integrierte Entwicklungsumgebung für Programmieranfänger. In H. Pongratz, & R. Keil (Hrsg.), *DELFI 2015 - Die 13. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V.* (S. 119-130). Bonn: Gesellschaft für Informatik e.V.
- Dahm, M., Rose, J., & Köhler, M., (2020). Programmier-Praktikum für Erstsemester – Erfahrungen aus mehreren Iterationen. In R. Zender, D. Ifenthaler, T. Leonhardt, & C. Schumacher (Hrsg.), *DELFI 2020 – Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.* (S. 259-264). Bonn: Gesellschaft für Informatik e.V.

- Dreyfus, H. L., & Dreyfus, S. E. (1980). *A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition*. ORC-80-2, Operations Research Center, University of California: Berkeley.
- Ehlert, A. (2012). *Empirische Studie: Unterschiede im Lernerfolg und Unterschiede im subjektiven Erleben des Unterrichts von Schülerinnen und Schülern im Informatik- Anfangsunterricht (11. Klasse Berufliches Gymnasium) in Abhängigkeit von der zeitlichen Reihenfolge der Themen (OOP-First und OOP-Later)*. Dissertation. Freie Universität Berlin, Berlin. Online verfügbar unter <http://dx.doi.org/10.17169/refubium-14069>.
- Ehlert, A., & Schulte, C. (2009). Empirical comparison of objects first and objects later. In M. Clancy, M. Casperse & R. Lister (Hrsg.), *ICER '09: Proceedings of the fifth international workshop on Computing education research workshop* (S. 15-26). New York: Berkeley.
- Fässler, L. E., Hinterberger, H., Dahinden, M., & Wyss, M. (2006). Evaluating student motivation in constructivist, problem-based introductory computer science courses. In T: Reeves & S. Yamashita (Hrsg.), *Proceedings of E-Learn 2006--World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (S. 1178-1185). Honolulu, Hawaii, USA: Association for the Advancement of Computing in Education (AACE).
- Fässler, L. E. (2007). *Das 4-Schritte-Modell – Grundlage für ein kompetenzorientiertes E-Learning*. Dissertation Nr. 17521, Eidgenössische Technische Hochschule Zürich, Zürich.
- Handke, J., Loviscach, J., Schäfer, A. M., & Spannagel, C. (2012). Inverted Classroom in der Praxis. In B. Berendt, B. Szczyrba, & J. Wildt (Hrsg.), *Neues Handbuch Hochschullehre* (E 2.11, 1-18), Ergänzungslieferung 57, Dezember 2012. Berlin: Raabe.
- Kramer, M., Hubwieser, P., & Brinda, T. (2016): A Competency Structure Model of Object-Oriented Programming. In J. E. Guerrero (Hrsg.), *2016 International Conference on Learning and Teaching in Computing and Engineering: LaTiCE* (S. 1-8). Piscataway, NJ: IEEE.
- Kramer, M., Samimi, V., & Brinda, T. (2017). Entwicklung eines Online-Tools zur Bestimmung objektorientierter Programmierkompetenzen. In I. Diethelm (Hrsg.), *Informatische Bildung zum Verstehen und Gestalten der digitalen Welt* (S. 425-426). Bonn: Gesellschaft für Informatik e.V.
- Mandl, H., Gruber, H., & Renkl, A. (1995). Situiertes Lernen in multimedialen Lernumgebungen. In L. J. Issing, & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia* (S. 167-178). Weinheim: Beltz.
- Nelles, W., Rhode, T., & Stechert, P. (2010). Entwicklung eines Kompetenzrahmenmodells – Informatisches Modellieren und Systemverständnis. *Informatik-Spektrum*, 33(1), 45-53.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Reinmann, G. (2015). *Studientext Didaktisches Design*. Abgerufen am 22.11.2020 von [https://gabi-reinmann.de/wp-content/uploads/2013/05/Studientext\\_DD\\_Sept2015.pdf](https://gabi-reinmann.de/wp-content/uploads/2013/05/Studientext_DD_Sept2015.pdf)

- Reinmann, G. (2018). Shift from Teaching to Learning und Constructive Alignment: Zwei hochschuldidaktische Prinzipien auf dem Prüfstand. *Impact Free*, 14.
- Rohde, T. (2013). *Entwicklung und Erprobung eines Instruments zur Messung informatischer Modellierungskompetenz im fachdidaktischen Kontext*. Dissertation, Universität Paderborn, Paderborn.
- The Cognition and Technology Group at Vanderbilt (1990). Anchored Instruction and Its Relationship to Situated Cognition. *Educational Researcher*, 19(6), 2-10.
- Thüring A., & Jäger K. (2020a). Objektorientierte Programmierung im Inverted Classroom. In S. Kahmann & S. Ludwigs (Hrsg.), *So gelingt E-Learning, Reader zum Higher Education Summit* (S. 162-169). München: Pearson.
- Thüring, A., & Jäger, K., (2020b). Objektorientierte Programmierung - Kompetenzerwerb im Mastery Model des Inverted Classroom. In R. Zender, D. Ifenthaler, T. Leonhardt, & C. Schumacher (Hrsg.), *DELFI 2020 – Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.* (S. 205-210). Bonn: Gesellschaft für Informatik e.V.
- Uysal, M. P. (2012). The Effects of Objects-First and Objects-Late Methods on Achievements of OOP Learners. *Journal of Software Engineering and Applications*, 5(10), 816-822.
- Weicker, N., Draskoczy, B., & Weicker, K. (2006). *Fachintegrierte Vermittlung von Schlüsselkompetenzen der Informatik*. In P. Forbrig, G. Siegel, & M. Schneider (Hrsg.), *HDI 2006: Hochschuldidaktik der Informatik – Organisation, Curricula, Erfahrungen* (S. 51-62). Bonn: Gesellschaft für Informatik e. V.
- Weinert, F. E. (2001). *Leistungsmessungen in Schulen*. Weinheim: Beltz.
- Williams, J. S. (2014). *A Computer Learning Environment for Novice Java Programmers That Supports Cognitive Load Reducing Adaptations and Dynamic Visualizations of Computer Memory*. Dissertation, Paper 574, Univ. of Wisconsin, Milwaukee.
- Zukunft, O. (2016). *Empfehlungen für Bachelor- und Masterprogramme im Studienfach Informatik an Hochschulen*. Bonn: Gesellschaft für Informatik e.V.

**Annett Thüring** ist Dipl. Informatikerin und als Leiterin der IT-Abteilung am Institut für Informatik tätig. Sie verfügt über langjährige Erfahrung im Bereich Digitalisierung von Lehre. Ihre aktuellen Forschungsinteressen fokussieren die Verbesserung des Lernerfolgs in der Studieneingangsphase durch den gezielten Einsatz lernerzentrierter Blended Learning-Konzepte, speziell im Bereich der Programmierung.

[annett.thuering@informatik.uni-halle.de](mailto:annett.thuering@informatik.uni-halle.de)

**Kathrin Jäger** studierte Synthesechemie und promovierte als Ingenieurin in der Thermischen Verfahrenstechnik. Am Zentrum für multimediales Lehren und Lernen unterstützt sie alle Phasen der Integration von Blended Learning in den Naturwissenschaftlichen Fakultäten der Universität Halle. Schwerpunkte ihrer wissenschaftlichen Arbeiten liegen im Instruktionsdesign und im elektronischen Assessment.

[kathrin.jaeger@llz.uni-halle.de](mailto:kathrin.jaeger@llz.uni-halle.de)

# Reflektives Peer-Quizzing! Didaktische Potenziale für die nachhaltige Entwicklung adaptiver Lernsettings

*Benjamin Eugster<sup>1</sup>, René Barth<sup>2</sup> und Matthias Ballod<sup>3</sup>*

*<sup>1</sup> Martin-Luther-Universität Halle-Wittenberg, Institut für Germanistik, Projekt [D-3] Deutsch Didaktik Digital*

*<sup>2</sup> Martin-Luther-Universität Halle-Wittenberg, Zentrum für Lehrer\*innenbildung (ZLB), Projekt DikoLa*

*<sup>3</sup> Martin-Luther-Universität Halle-Wittenberg, Professur für Didaktik der deutschen Sprache und Literatur*

Das Projekt „Wissenskonstruktion und Wissensdiagnostik. Das Potenzial von Quizze für die Erstellung adaptiver Lernszenarien zur Selbstreflexion und Selbstbewertung“ ging der Frage nach, wie formative Formen des elektronischen Assessments für eine nachhaltige Wissensakquise und adäquate Kompetenzprüfung eingesetzt werden können. Aufbauend auf einem ILIAS-Fragenpool wurden im Verlauf von zwei Jahren neue Methoden zur Entwicklung und dem Einsatz von formativen Fragen im Lehramtsstudium Deutsch pilotiert und evaluiert. Als Potenziale haben sich dabei einerseits eine integrative und nachhaltige Nutzung und Erweiterung von Fragensammlungen herausgestellt, die sowohl in summativen als auch formativen Kontexten genutzt werden können. Andererseits hat sich die Arbeit mit unterschiedlichen methodischen Ansätzen im Umgang mit formativen Testverfahren (Kompetenzorientierung durch studentisch erstellte Quizze, Qualitätssicherung durch Lehrpersonen, bis zur Nutzung von Teststatistiken) als besonders erfolgsversprechend abgezeichnet. Basierend auf Ansätzen des ‚Inquiry Based Learning‘ wird das Konzept des ‚Peer-Quizzing‘ als das Einbinden von Studierenden in den nachhaltigen Aufbau und die Nutzung digitaler Fragebestände theoretisch kontextualisiert. Schlaglichter wirft der Beitrag auf zukünftige Entwicklungen integrativ adaptiver Lernsettings mit Assessment-Elementen. Skizziert wird dazu ein Konzept, das zentrale Elemente klassischer Lehr-Lernorganisation (Selbstlernen, Lehren, Prüfen...) konsequent didaktisch und methodisch aufeinander bezieht. Im Beitrag sollen diese Potenziale nachgezeichnet und eine curriculare sowie didaktische und methodische Einbindung in die (hoch)schulische Lehre aufgezeigt werden.

## Einleitung

Dieser Beitrag basiert auf der Erfahrung mit der langjährigen Nutzung von ILIAS-Fragenpools in formativen und summativen Testsettings. Im Zentrum stehen Erkenntnisse, die im Kontext des FFP-Projektes „Wissenskonstruktion und Wissensdiagnostik. Das Potenzial von Quizze für die Erstellung adaptiver Lernszenarien zur Selbstreflexion und Selbstbewertung“ in unterschiedlichen Lehr-/Lernsettings mit der Erstellung von Quiz-Fragen durch Studierende des Lehramts Deutsch gewonnen wurden.

Dieser Beitrag reflektiert das didaktische Potenzial integrierter Prozesse zur Fragen-erstellung und -auswertung zum einen vor dem Hintergrund der didaktischen Möglichkeiten des ‚Inquiry Based Learning‘ (Bell, Schanze, & Ploetzner, 2010) in der Form von studentisch erstellten Test- und Quizfragen und zum anderen der nachhaltigen Verwaltung und Qualitätsentwicklung von größeren Beständen an digitalen Testfragen. Mit den studentisch generierten Fragen erhalten Aspekte wie die Fragequalität und die Kompetenzorientierung eine höhere Dringlichkeit. Gerade deshalb stellen diese Methoden ein geeignetes Mittel dar, die Fragengestaltung und -verwaltung von Lehrenden zu ergänzen und die Anwendungsfelder jenseits formativer und summativer Settings weiterzuentwickeln.

## E-Assessment in integrierten Lernsettings

Der Begriff E-Assessment bezieht sich auf unterschiedliche technologische Umsetzungen, aber auch unterschiedliche Phasen der Wissenskonstruktion und -diagnostik mit digitalen Hilfsmitteln. Eine übliche definitorische Orientierung gibt hier die Aufteilung in unterschiedliche Phasen, die den Lernprozess rahmen oder begleiten (vgl. Schmees, Krüger, & Schaper, 2013, S. 21). Während diagnostische Assessments zu Beginn eines fokussierten Lernprozesses die Eingangskompetenzen überprüfen und formative Assessment-Formen im Verlauf des Lernprozesses durch Feedback und Wiederholungen maßgeblich zur Wissenskonstruktion und Kompetenzentwicklung beitragen, setzen summative Formen des Assessments klar auf eine abschließende Kontrolle der erworbenen oder nicht erworbenen Wissensbestände oder Kompetenzen.

Mit diagnostischem, formativem und summativem Testen werden daher nicht nur Einsatzgebiete unterschieden, zumeist sind auch unterschiedliche technische Realisierung damit verbunden. Während summative Formen der Wissensdiagnostik, wie z. B. klassische Klausuren oder geschlossene digitale Prüfungsformate, besondere Anforderungen an die Absicherung und technische Beständigkeit stellen (Ruedel, Schiefner, Noetzli, Seiler, & Schiedt, 2007, S. 187), ist die Nutzung formativer Formen der Wissenskonstruktion in Gestalt von Quizze oder Übungsumgebungen weitaus offener und flexibler gestaltet (vgl. Derr, 2021). So ist es im Gegensatz zum geschlossenen Setting erwünscht, dass die Fragen gut zugänglich sind und mehrfach wiederholt werden können. Auch im Falle eines technischen Zwischenfalls kann das Quiz einfach von neuem gestartet werden. Gleichzeitig bieten diagnostische Testformate gerade im Bereich der Sprachdiagnostik und Sprachförderung, z. B. in der Form digitaler Einstufungstests, neue Möglichkeiten (Meister & Oevel, 2017, S. 58 ff.).

Aufgrund dieser technologischen Ausdifferenzierung ergibt sich für eine nachhaltige Nutzung dieser Systeme eine grundlegende Herausforderung: Wie können die Fragebestände der unterschiedlichen Plattformen, technischen Tools und Organisationen miteinander kombiniert werden? Die vielfältigen Fragen, die hier thematisiert werden, sowie die variable Einbindung in konkrete sich verändernde Seminarkonzepte zeigen auf, dass es sich um keine bloß technische Herausforderung handelt.

### **E-Assessment im Wandel: Von Selbstlernmodulen zu mobilen Quizzing-Tools**

Die Entwicklung von Selbstlerneinheiten für formative Test- und Lernsettings stellt bei Weitem keine neue Erscheinung dar. Erste Erfahrungen mit online-gestützten Selbstlerneinheiten wurden in der Linguistik bereits in den frühen 2000er Jahren im Rahmen des hochschulübergreifenden E-Learning-Projektes PortalLingua gesammelt. Damit verbunden war auch schon die Ernüchterung ob der eingeschränkten didaktischen Möglichkeiten von digitalen Test-Fragen der damals verwendeten Software *Lernplus ASP*: „Die didaktischen Funktionen des Test-Centers beschränken sich ebenfalls auf einfache, an behavioristischen Vorstellungen vom Lernen ausgerichteten Testverfahren (vor allem *multiple choice*).“ (Wagner & Antos, 2003, S. 14)

Dennoch ist im Laufe der Jahre durch die Nutzung der Lernplattform ILIAS und die Durchführung digitaler Prüfungen eine beträchtliche Menge von Testfragen entstanden. Während einige dieser Fragenkomplexe verwaist sind, hatten andere an der MLU Halle-Wittenberg einen festen Platz. Die über die Jahre zustande gekommene Sammlung von rund 2000 Fragen im Bereich der linguistischen Grundlagen wird nach wie vor in Selbstlerneinheiten und elektronischen Prüfungen verwendet. Diese waren vor dem Hintergrund der Prüfungsvorbereitung notgedrungen von genau diesem Fokus auf formatives Lernen im Sinne des Self-Assessments geprägt. Derartige Formen des computergestützten Testens und multimedialen Lernens – die insbesondere auch aus dem Fremdsprachenlernen (z. B. Vokabeltrainer, Fremdsprachen-Apps) bekannt sind – basieren jedoch meistens auf der behavioristischen Lerntheorie, die durch Drill und Practice gekennzeichnet ist. Gerade mit dem Aufkommen von Mobilgeräten wurde nicht nur die Verbreitung, sondern auch die Akzeptanz entsprechender Ansätze wieder erhöht. Indem durch Sprachlern-Apps wie Duolingo, Babbel u. a. der Fokus auf dem direkten Feedback, individualisierten Hilfestellungen und einem adaptiven Aufbau der Lerneinheiten liegt, rückt der Aufwand für die Erstellung einzelner Frage-Items in den Hintergrund resp. wird bei Applikationen von Dritten komplett unsichtbar.

Beim Mutterspracherwerb und in der Muttersprach-Didaktik spielen Varianten der Wissensdiagnostik, die in vielen Wissensdomänen eine Selbstverständlichkeit darstellen (Multiple-Choice-Abfragen), hingegen kaum bis gar keine Rolle. Insbesondere auf höheren Kompetenzstufen wird dabei aus guten Gründen auf offene und reflexive Formen des Assessments zurückgegriffen. Eine Konsequenz ist jedoch, dass die didaktischen Kompetenzen der digitalen Medien mitunter unterentwickelt sind respektive nicht deren volles Potenzial ausgeschöpft wird. Es fehlt eine für die digitalen Medien angepasste Muttersprach-Didaktik, die

eine produktive Wissenskonstruktion und Wissensdiagnostik, wie forschendes Lehren oder Lernen durch Fragenentwicklung, ermöglicht.

Zwischenzeitlich haben sich Formate des formativen Prüfens und des Lernens mit Quizze insbesondere außerhalb des Hochschulkontexts entwickelt. Von der interaktiven Umsetzung von Quiz-Shows im Netz (z. B. Quiz-Duell) bis zu mobilen Apps des Sprachlernens (z. B. Duolingo, Babbel, etc.) stellen formative digitale Testmethoden einen festen Bestandteil der Populärkultur von Wissensgesellschaften dar. Vom Wandel der Internetnutzung von einem vermeintlich passiven Medienkonsum hin zur aktiven Medienproduktion im Sinne des Web 2.0 und der sozialen Medien waren diese Ansätze jedoch nur geringfügig betroffen. Mit Plattformen wie LearningSnacks, LearningApps, Quizlet etc. erreichte das niederschwellige Erstellen von Testfragen durch Lernende oder durch Lehrpersonen auch die Klassenzimmer. Anreize zu einer integrierten oder systematischen Einbindung von Lernenden in den Prozess der Fragenerstellung liefern diese Programme jedoch kaum.

### **Studentische Fragenerstellung als Form des ‚Inquiry Based Learning‘**

Der bewusste Einbezug von Lernenden in die Fragenerstellung für Studien- und Prüfungszwecke wird seit den 1980er-Jahren in unterschiedlichen Kontexten als Methode diskutiert und angewandt (Song, 2016, S. 58). Sichtbar wurden diese Verfahren vor allem durch spezifisch dafür erstellte Lernsoftware wie z. B. PeerWise (Hakulinen, 2010) oder das 2015 für die Lernplattform Moodle entwickelte Plugin StudentQuiz (Filip, Pudło, & Marchewka, 2018, S. 187), das die Erstellung von Test- und Quizfragen durch Studierende mittels Rating- und Ranking-Mechanismen spielerisch gestaltet.

Der Ansatz des forschenden Lernens oder des ‚Inquiry Based Learning‘ (Bell et al., 2010) baut auf der Annahme auf, dass eine valide Kompetenzprüfung und -entwicklung auf offene und produktionsorientierte Lernformen angewiesen ist. Gerade in der Ausbildung zukünftiger Lehrpersonen ist dieser produktionsorientierte Ansatz besonders bedeutend. Durch den Fokus auf die Erzeugung eigener Medienprodukte oder Fragensettings kann nicht nur der Inhalt auf eine andere Art und Weise durchdrungen werden, sondern es werden digitale Kompetenzen erlernt, die sich auf die eigene Vermittlung und Überprüfung von Leistungen beziehen. Mit diesem Shift geht auch eine Verschiebung von instruktionalen Settings zu einer stärkeren Selbsterschließung von Inhalten einher. Das Konzept des ‚Inquiry Based Learning‘ stellt damit die gesamte Logik von rein instruktionalen Settings komplett auf den Kopf. Um die Lernwege von Lernenden systematisch gestalten zu können, muss das Finden und Mitbestimmen eigener Lernwege im Sinne des ‚Inquiry Based Learning‘ bereits früh ins Studium integriert werden (vgl. Reitlinger, 2013, S. 25).

Mit Blick auf die stärkere Integration digitaler Kompetenzen wie die Bedienung lernspezifischer Software oder die multimediale Gestaltung von Lernsettings stellen studentisch generierte Fragen in zweierlei Hinsicht eine besondere Chance dar. Einerseits kann die Aktivierung durch die eigene Fragenformulierung für eine nachhaltige und reflektierte Wissenskonstruktion genutzt werden, indem analog zum veränderten Medienproduktionsverhalten auch mehr Partizipation durch Lernende eingefordert wird. Bereits in analogen Settings

konnte aufgezeigt werden, dass das Verständnis von Lernenden durch das eigene Formulieren von Wissensfragen gesteigert werden kann (vgl. King, 1992). Andererseits können damit aber auch mediendidaktische Kompetenzen praxisorientiert in die Lehre eingebunden werden. Gerade in der ersten Phase des Lehramtsstudiums kann diese Produktionsorientierung wertvolle Kompetenzen in der Konzeption, Gestaltung und Nutzung von digitalen Lernmedien vermitteln.

Für die produktionsorientierte Lehre bedeutet dies letztlich auch, den Fokus vom Beantworten von Fragen auf die Formulierung und Kontextualisierung von Fragen zu verlagern. Studierende in die Erstellung von Quiz-Fragen einzubeziehen, stellt dann eine unmittelbare Methode zur Diversifizierung des Lernerfolgs der Studierenden dar, indem sie den Perspektivwechsel vollziehen, von der Rolle der Geprüften zur Instanz der Prüfenden. Sie erproben damit zugleich ein Hauptanforderungsprofil ihrer späteren Profession.

Trotz der didaktischen Einschränkung durch automatisierte und teilautomatisierte Frageformate erfordert diese Sicht vom Lernenden vernetztes Denken, bei dem theoretische Grundlagen auf Anwendungsfelder zu übertragen sind. Zur didaktischen und technischen Umsetzung derartiger Settings gibt es unterschiedliche Möglichkeiten. Während sich beispielsweise die kollaborative Arbeit in Wikis dazu nutzen lässt, gemeinsame Aufgaben- und Fragensets zu erarbeiten (vgl. Reinmann, 2009), existieren auch elaborierte technische Lösungsansätze, die regelrechte Spiele mit dem Fokus auf ‚Peer-Quizzing‘ (vgl. Kiron, Adaji, Long, & Vassileva, 2019) erarbeitet haben. Der am stärksten integrierte Ansatz liegt hier wohl im Plugin „StudentQuiz“ für die Lernplattform Moodle vor (vgl. Filip et al., 2018, S. 187), das dank der Kompatibilität der Frageformate in die Lernplattform integriert werden kann und dennoch mit unterschiedlichen Interaktionsmöglichkeiten aufwartet, welche die Fragenerstellung, -verwaltung und -beantwortung auf spielerische Weise abzubilden versuchen. Durch die Möglichkeit einer Bewertung von Fragen können Qualität und Komplexität durch die Lernenden gegenseitig beurteilt werden und auf einer gemeinsamen Rangliste wird laufend angezeigt, welche Lernenden bei der Lösung von anderen Quizfragen bereits die meisten Punkte gesammelt haben. Doch auch mit anderen Lernplattformen lassen sich vergleichbare Settings erstellen.

### **Didaktische Umsetzung im Seminarkontext**

Für die Umsetzung im Rahmen der Lehramtsstudiengänge im Fach Deutsch an der MLU (außer im Lehramt Grundschule) werden die Grundfunktionalitäten der Lernplattform ILIAS so genutzt, dass die Erstellung von Quizfragen in bestimmten Bereichen für Studierende freigeschaltet ist. Nach der Erprobung des didaktischen Designs im Wintersemester 2018/2019, in dem Studierende bereits prüfungsrelevante Fragen erstellten, diese aber zunächst in einem separaten Wiki gesammelt wurden, fand die ILIAS-integrierte Durchführung erstmals in den fachdidaktischen Einführungen des Wintersemesters 2019/2020 statt.

Hierfür wurden die Studierenden zu Semesterbeginn einerseits in die technischen Besonderheiten der Lernplattform ILIAS eingeführt. Andererseits erhielten die Seminarteilnehmenden auch wichtige methodische Hinweise zum Erstellen von Quizfragen, d. h. von

Fragen im Antwort-Wahl-Format, vor allem dazu, welche üblichen Fehler bei der Formulierung von Distraktoren zu vermeiden sind und wie letztere das Niveau der Frage bestimmen.

Dabei übernahmen die Mitglieder der in Zweier- oder Dreier-Teams zusammengeschlossenen Studierendengruppen im Seminarverlauf einmal die Rolle von Expert\*innen. Als solche bereiteten sie nicht nur den thematischen Input zu einer Sitzung vor, sondern zeichneten in der darauffolgenden Sitzung auch für die niederschwellige Ermittlung des Wissensstandes ihrer Kommiliton\*innen verantwortlich. Im Verlauf des Semesters erstellten alle Studierenden eine Frage. Beginnend mit Single- und Multiple-Choice-Fragen wurden im Semesterverlauf zunehmend komplexere Frage-Typen wie Zuordnungs-, Lückentext- mit Auswahl- und Textlücken, Kprim-Choice- oder Fehlertextfragen verwendet. Diese Fragen mussten die Teilnehmenden in ihrer Gruppe in kritischer Auseinandersetzung mit dem Stoff nach dem Ende ihrer „Expert\*innen-Sitzung“ erstellen.

In der darauffolgenden Sitzung wurden diese Fragen zur Repetition und Reflexion des Themas aufgegriffen. Bei dieser Gelegenheit zeigten sich vor allem die Herausforderungen in der Formulierung valider Fragen. So stand hier die reflexive Diskussion von Schwächen und Stärken der Fragen im Vordergrund, nachdem die drei Fragen der letzten Sitzung jeweils durch alle Studierenden beantwortet wurden. Die daraus resultierenden Ergebnisse dienten dabei als Indiz für allfällige Missverständnisse in der Fragestellung, Qualitätsmängel in der Formulierung der Distraktoren oder technische Fehler. Insbesondere bei einem auffällig hohen oder niedrigen facility index, wenn also auffällig viele – oder aber auffällig wenige – Studierende eine Frage korrekt bzw. nicht korrekt beantwortet haben, konnten geläufige Formulierungsschwierigkeiten und Verbesserungsmöglichkeiten interaktiv identifiziert werden.

Trotz der doppelten Reflexion und möglicher Überarbeitungsschleifen blieben starke Qualitätsunterschiede im finalen Fragenpool erkennbar. In der mündlichen Modul-Abschlussprüfung wurden daher die von den Studierenden erstellten Fragen genutzt, um diese erneut (selbst-)kritisch und stoffbezogen zu reflektieren. In diesem Schritt ging es nicht darum, das Lernziel durch den eingeschränkten Fragetypen definieren zu lassen, sondern vielmehr darum, die Stärken und Schwächen der Fragen als Anlass für eine thematische Kontextualisierung zu nutzen. Durch diese erneute Thematisierung konnten die Fragen im Rahmen ihrer Prüfungsvorbereitung und -auswertung didaktisch nutzbar gemacht werden.<sup>1</sup>

In der Überprüfung und teilweisen Überarbeitung der von den Studierenden erstellten Fragen hatten sich zwei Herausforderungen in der Fragenerstellung besonders deutlich manifestiert. Die nachträgliche Auswertung von einem Sample von 194 Fragen aus den jeweils parallel laufenden literatur- und sprachdidaktischen Veranstaltungen des Moduls Fachdidaktik I lässt hier eine grobe Zuordnung zu. Einerseits wurden bei 28 Prozent der Fragen starke Mängel in der Funktionsweise oder Fragequalität festgestellt, welche eine Überarbeitung der Frage unabdingbar machten. Dies lässt sich damit erklären, dass ein Teil der Fragen

---

<sup>1</sup> Bei dieser Schilderung des didaktischen Settings handelt es sich um eine gekürzte Form, die andernorts ausführlicher beschrieben wurde. Vgl. Berg, 2020.

von den Studierenden als reine pro-forma-Fragen mit dem Zweck erstellt wurde, das Soll zu erfüllen.

Andererseits zeigte sich bei einem Großteil der verbleibenden Fragen, dass sich die Studierenden mehrheitlich an rein deklarativem Wissen orientiert hatten. Trotz der technischen Möglichkeiten, komplexe Sachverhalte als Ausgangslage für prozedurale oder metakognitive Fragen darzustellen, wurden diese nur bei weniger als 10 Prozent der Fragen genutzt. Neben den MC- und SC-Fragen, die sich auf Aufzählungen fokussierten, sind dabei insbesondere Lückentext-Fragen aufgefallen, die zu einem unkritischen Übernehmen bestehender Textfragmente mit mehr oder weniger arbiträren Lücken motiviert haben.

Vergleichbare Herausforderungen an die Fragenerstellung durch Lernende wurden bereits andernorts mit Hinblick auf das Problem, unterschiedliche Kompetenz- und Schwierigkeitsniveaus abbilden zu können, festgestellt (vgl. Papinczak et al., 2012). Diesem Umstand konnte in einer weiteren Durchführung des Konzeptes damit begegnet werden, dass anhand von Beispielfragen auf unterschiedlichen Kompetenzniveaus das Thema der Validität und Qualität von Quizfragen frühzeitig noch stärker thematisiert wurde. Zudem half den Studierenden die Ausrichtung an Fragen für Open-Book-Prüfungen im Zuge des digitalen Semesters dabei zu verstehen, was es bedeutet, Fragen nicht auf rein deklarative Wissensbestände hin zu konstruieren.

### **Entwicklung adaptiver Testsettings durch nachhaltige Fragenverwaltung**

In einer kurzfristigen Perspektive werden durch das Erstellen zusätzlicher Fragen durch Studierende viele Probleme der Fragenverwaltung verschärft. Während alte Fragenbestände regelmäßig auf ihre Aktualität überprüft und durch neue ergänzt werden müssen, kommen mit den studentisch generierten Fragen neue hinzu. Dies erschwert sowohl die Übersichtlichkeit als auch die Qualität der Fragen, wodurch sowohl auf technischer als auch auf didaktischer Seite eine Systematisierung der Fragenverwaltung mit der flexiblen Verwendung von Schlagworten und Taxonomien notwendig wird. Die direkte Einbindung in die Lehre stellt zugleich eine Gelegenheit dar, sich gezielt und offen mit der Qualitätsentwicklung von Fragenpools sowie mit dem Einsatz in neuen Lernsettings auseinanderzusetzen, und bietet mithin einen deutlichen Mehrwert gegenüber dem alleinigen Einsatz in kursbezogenen Selbstlerneinheiten und Prüfungen.

Obwohl ein strukturiertes Qualitätsmanagement durch die Lernplattform ILIAS bisher nur bedingt unterstützt wird, kann gerade eine frühzeitige und breite Bereitstellung von studentisch generierten Fragen in formativen Quizze zur steten Überarbeitung beitragen. Indem die Teststatistiken hinsichtlich des Anforderungsniveaus von Fragen oder häufigen Missverständnissen ausgewertet werden, können die Studierenden über die wiederholte Teilnahme an Testdurchläufen dazu beitragen, Qualität und Adäquanz einzelner Fragen genauer einzustufen, woraufhin diese ggf. aus dem Fragenpool entfernt oder aber entsprechend überarbeitet werden können.

Eine mögliche Weiterentwicklung, die eine längerfristige Nutzbarmachung der durch Dozierende und Studierende erstellten Fragen ermöglicht, sind adaptive Testsettings für den

Studien- oder Seiteneinstieg ins Lehramtsstudium Deutsch. In einem Lehrveranstaltungsunabhängigen Kurs können dabei themenspezifische Fragen auf unterschiedlichen Schwierigkeitsstufen zur Verfügung gestellt werden. Die Einteilung nach unterschiedlichen Schwierigkeits- resp. Kompetenzniveaus ermöglicht gerade durch die erhöhte Zahl studentisch erstellter Fragen die Möglichkeit, zufallsgenerierte und adaptive Lernsettings zu gestalten.

Ein erster Schritt dorthin ist die Aufschlüsselung unterschiedlicher Anforderungsniveaus der verwendeten Testaktivitäten in sogenannten „task classes“ (Merriënboer & Kirschner, 2018, S. 166f), die Test- oder Lernaktivitäten von vergleichbarer Schwierigkeit gruppieren. Die Anordnung unterschiedlicher Testaktivitäten entlang eines aufsteigenden Schwierigkeits- und Komplexitätsgrads bildet damit die Grundlage für ein adaptives Kurssetting. In der einfachsten Form sieht dies so aus, dass beispielsweise zuerst das Quiz auf Ebene 1 absolviert werden muss, bevor Ebene 2 freigeschaltet wird, deren erfolgreiches Absolvieren wiederum Ebene 3 freischalten würde. Durch entsprechende Hilfsmaterialien auf den jeweiligen Ebenen lässt sich dieser Prozess im Sinne eines digital unterstützten ‚scaffolding‘ (Sharma & Hannafin, 2007) zusätzlich anreichern und damit der Lernprozess stützen.

## **Fazit**

‚Peer-Quizzing‘ kann als eigenes Einsatzszenario von E-Assessment-Technologien flexibel und vielfältig eingesetzt werden. In diesem Beitrag wurde aufgezeigt, wie das studentische Erstellen von Test- und Übungsfragen nicht nur Potenziale für die Reflexionsfähigkeit der Lernenden birgt. Vielmehr lassen sich aus den daraus entstehenden erweiterten Fragesets komplexe, adaptive Übungsszenarien entwickeln. Damit rückt der Fokus beim Einsatz digitaler Hilfsmittel stärker auf den organisatorisch-didaktischen Einsatz, wohingegen die konkrete technische Umsetzung allenfalls der Erleichterung im Sinne der Usability und mithin der Steigerung der Akzeptanz dient.

Als zentrale Gelingensbedingungen für das Ausschöpfen dieser Potenziale konnten in der Analyse ausgewählter Fragesets insbesondere die Qualität der Fragen und eine gewisse Diversität hinsichtlich Kompetenz- und Schwierigkeitsstufen identifiziert werden. In der fortlaufenden Reflexion der Fragen in Lehrveranstaltungen und den mündlichen Prüfungen konnte zwar ein reflektiertes Qualitätsbewusstsein bei den Studierenden erzielt werden, allerdings bedurften die Fragen für eine spätere Nachnutzung der Überarbeitung. Wenngleich der integrierte Ansatz der Fragenerstellung im formativen Lernprozess also eine effiziente Generierung größerer Fragenpools erlaubt, so muss diese auch durch die gezielte Vermittlung von Gütekriterien anhand konkreter Beispiele auf unterschiedlichen Schwierigkeits- und Kompetenzstufen unterstützt werden. Auf diese Weise wird eine einfachere Wiederverwendbarkeit von digitalen Fragen sowohl in formativen als auch in diagnostisch-adaptiven und summativen Settings ermöglicht.

## Literatur

- Bell, T., Urhahne, D., Schanze, S. & Ploetzner, R. (2010). Collaborative Inquiry Learning: Models, tools, and challenges. *International Journal of Science Education*, 32(3), S. 349–377.
- Berg, G. (2020). Digitale Quiz-Didaktik in der Lehrer\*innenbildung. Konzept und Ergebnisse des Projekts [D3] Deutsch Didaktik Digital, In K. Kaspar, M. Becker-Mrotzek, S. Hofhues et al. (Hrsg.), *Bildung, Schule, Digitalisierung* (S. 152-157). Münster: Waxmann.
- Derr, K. (2021). Formatives E-Assessment und Diagnostik, In R. Küstermann, M. Kunkel, M. Mersch & A. Schreiber (Hrsg.). *Selbststudium im digitalen Wandel* (S. 127-140). Berlin: Springer Spektrum.
- Edelson, D. C. (2001). Learning-for-use: A framework for the design of technology-supported inquiry activities. *Journal of Research in Science Teaching*, 38(3), S. 355–385.
- Filip, A., Pudło, W. & Marchewka, D. (2018). Innovative Learning: Students in the Process of Exam Quizzes Building, In G. Ubachs & J.-A. Fenna (Hrsg.), *Blended and Online Learning "Changing the Educational Landscape". Overview of papers on Higher Education for the Future as presented during the Online, Open and Flexible Higher Education Conference in Aarhus* (S. 184–191). Aarhus.
- Hakulinen, L. & Korhonen, A. (2010). Making the most of using PeerWise in education. In Aalto University, Lifelong Learning Institute Dipoli (Hrsg.). *ReflekTori 2010 – Symposium of Engineering Education*, S. 57–67.
- King, A. (1992). Facilitating elaborative learning through guided student generated questioning. *Educational Psychologist*, 27(1), S. 111–126.
- King, A. (1989). Effects of self-questioning training on college students' comprehension of lectures. *Contemporary Educational Psychologist*, 14(4), S. 366–381.
- Kiron, N., Adaji, I., Long, J. & Vassileva, J. (2019): Tower of Questions (TOQ): A Serious Game for Peer Learning, In A. Liapis, G. Yannakakis, M. Gentile & M. Ninaus (Hrsg.), *Games and Learning Alliance. GALA 2019. Lecture Notes in Computer Science*, Springer International Publishing.
- Meister, D. M. & Oevel, G. (2017). E-Assessment in der Hochschulpraxis. Empfehlungen zur Verankerung von E-Assessments in NRW. Paderborn.
- Papinczak T., Peterson R., Babri A.S., Ward K., Kippers V. & Wilkinson D. (2012). Using student-generated questions for student-centred assessment. *Assessment & Evaluation in Higher Education*, 37(4), S. 439-452.
- Merriënboer, J. & Kirschner, P. (2018). Ten steps to complex learning. A systematic approach to four-component instructional design. Third Edition. New York, London: Routledge.
- Reinmann, G. (2009). *Fragen lernen ist (nicht) schwer* [Blog-Beitrag]. Abgerufen am 02.02.2019 von <https://gabi-reinmann.de/?p=1530>.

- Reitinger, J. (2014). *Forschendes Lernen. Theorie, Evaluation und Praxis in naturwissenschaftlichen Lernarrangements*. 2. unveränderte Auflage. Theorie und Praxis der Schulpädagogik, Band 12. Immenhausen: Prolog-Verlag.
- Ruedel, C., Schiefner, M., Noetzli, C. & Seiler Schiedt, E. (2007). Risikomanagement für E-Assessment, In M. Merkt, K. Mayrberger, R. Schulmeister, A. Sommer & van den Berk, I. (Hrsg.), *Studieren neu erfinden – Hochschulen neu denken* (S. 180-190). Münster u. a.: Waxmann.
- Schmees, M., Krüger, M. & Schaper, E. (2013). E-Assessments an Hochschulen: Ein vielschichtiges Thema. In M. Krüger & M. Schmees (Hrsg.), *E-Assessments in der Hochschullehre. Einführung, Positionen & Einsatzbeispiele*. Frankfurt am Main: Peter Lang.
- Sharma, P. & Hannafin, M. (2007). Scaffolding in Technology-Enhanced Learning Environments. *Interactive Learning Environments*, 15(1), S. 27–46.
- Song, D. (2016). Student-generated Questioning and Quality Questions: A Literature Review. *Research Journal of Educational Studies and Review* 5(2). S. 58–70.

**Benjamin Eugster**, M. A., studierte Sozialwissenschaften an der Universität Zürich und Bildungsmanagement an der Universität Duisburg-Essen. Seit 2020 leitet er das *Projekt [D-3] Deutsch Didaktik Digital* an der Martin-Luther-Universität Halle-Wittenberg. Zuvor war er an mehreren Schweizer Hochschulen in der digitalen Lehrentwicklung und im Projektmanagement tätig. Seine Arbeitsschwerpunkte liegen in den Bereichen Instructional Design, Kompetenzentwicklung und E-Assessment.

[benjamin.eugster@germanistik.uni-halle.de](mailto:benjamin.eugster@germanistik.uni-halle.de)

**René Barth**, M. A., studierte Germanistik und Soziologie und arbeitet seit 2016 an der Martin-Luther-Universität Halle-Wittenberg. Nach zwei Tätigkeiten als wissenschaftlicher Mitarbeiter am Lehrstuhl für Bildungssoziologie und im *Projekt [D-3] Deutsch Didaktik Digital* am germanistischen Institut hat er im April 2020 seine Arbeit im QL-B-Projekt *DikoLa – Digital kompetent im Lehramt* am Zentrum für Lehrer\*innenbildung aufgenommen, in dessen Rahmen er die Evaluation und das interne Monitoring betreut.

[rene.barth@zlb.uni-halle.de](mailto:rene.barth@zlb.uni-halle.de)

**Matthias Ballod** ist Professor für Sprach-, Literatur- und Informationsdidaktik an der Martin-Luther-Universität Halle-Wittenberg. Er hat in der Linguistik zu automatisierten Verfahren der Textanalyse promoviert und in der Didaktik zu „Wissenstransfer und Wissenstransformation“ habilitiert. Seine Forschungsinteressen umfassen die funktionale Integration digitaler Medien in Lehr-Lernkontexte von Schule und Universität ebenso wie organisationales und individuelles Wissensmanagement.

[matthias.ballod@germanistik.uni-halle.de](mailto:matthias.ballod@germanistik.uni-halle.de)

# Testbasiertes Lernen in der medizinisch-pharmakologischen Lehre

*Alp Aslan<sup>1,2</sup> und Joachim Neumann<sup>3</sup>*

<sup>1</sup> *Martin-Luther-Universität Halle-Wittenberg, Institut für Psychologie*

<sup>2</sup> *Technische Hochschule Rosenheim, Angewandte Psychologie*

<sup>3</sup> *Martin-Luther-Universität Halle-Wittenberg, Institut für Pharmakologie und Toxikologie*

Ziel des hier vorgestellten Projekts war es, den sogenannten *forward testing effect* (FTE) – den positiven Effekt des Testens zuvor gelernter Informationen auf das Lernen und Behalten nachfolgender Informationen – in einem anwendungsorientierten Setting zu untersuchen. Dazu lernten Medizinstudierende im Rahmen eines Seminars zur Pharmakologie zunächst zwei Listen mit pharmakologischen Aussagen (z. B. „Clozapin kann eine Agranulozytose bedingen“). Nach dem Lernen einer jeden Liste wurde die Liste entweder noch einmal zum Lernen präsentiert (Lernbedingung) oder die Liste wurde getestet (Testbedingung; z. B. „Clozapin kann eine \_\_\_\_\_ bedingen“). Anschließend wurde in beiden Bedingungen eine dritte (kritische) Liste gelernt und nach 1 min bzw. 7 Tagen getestet. Konsistent mit früheren Befunden zum FTE war die Behaltensleistung für die (kritische) Liste 3 in der Testbedingung signifikant höher als in der Lernbedingung. Dies galt unabhängig von der Länge des Behaltensintervalls. Die Ergebnisse deuten an, dass Tests den (sonst üblichen) Aufbau von proaktiver Interferenz reduzieren und dadurch das nachfolgende Lernen und Behalten von neuen Informationen begünstigen können. Der Befund, dass der FTE nicht auf artifizielle Laborsituationen beschränkt ist, sondern sich auch mit pädagogisch relevanten Lerninhalten und nach pädagogisch relevanten (längeren) Behaltensintervallen finden lässt, hat potentiell bedeutsame Implikationen für die (multimediale) Hochschullehre.

## Testbasiertes Lernen in der medizinisch-pharmakologischen Lehre

Die Martin-Luther-Universität (MLU) Halle-Wittenberg trägt der immer größer werdenden Bedeutung multimedialen Lernens und Lehrens durch ein eigens formuliertes Multimedia-Leitbild Rechnung. In der Präambel dieses Leitbilds heißt es unter anderem, dass die MLU „den Herausforderungen moderner Lehre mit innovativen Lehr-, Lern- und Prüfungsformen [begegnet]“ (Multimedia-Leitbild der Martin-Luther-Universität Halle-Wittenberg, 2015). Diese in der Präambel implizierte Trennbarkeit von Formen des Lehrens, Lernens und Prüfens erscheint auf den ersten Blick durchaus sinnvoll, da sie die Chronologie eines typischen Semesters an Universitäten widerspiegelt: Der zu vermittelnde Stoff wird zunächst gelehrt und gelernt, bevor er – in der Regel erst am Ende des Semesters – geprüft wird. Jüngere Befunde aus der kognitionspsychologischen Grundlagenforschung deuten jedoch an, dass die Trennlinie zwischen Lernen und Prüfen weit weniger scharf ist als oftmals angenommen, und dass Lernen nicht nur *vor*, sondern auch *während* und insbesondere *durch* Prüfungen stattfinden kann. Dieses Lernen anhand von Prüfungen, in der Literatur besser bekannt als *testbasiertes Lernen*, war der Ausgangspunkt des hier vorgestellten Kooperationsprojekts zwischen dem Institut für Psychologie und dem Institut für Pharmakologie und Toxikologie der MLU.<sup>1</sup>

Testbasiertes Lernen ist eine Sammelbezeichnung für eine Reihe von Befunden, von denen der sogenannte *backward testing effect* (BTE) sicherlich der bekannteste ist. Studien zum BTE zeigen, dass das (wiederholte) Testen von zuvor gelernten Inhalten das langfristige Behalten dieser Inhalte substanziell verbessern kann und dabei insbesondere effektiver ist als das wiederholte Lernen der Inhalte (z. B. Roediger & Karpicke 2006; Karpicke & Roediger, 2008). Der BTE wurde in der jüngeren Vergangenheit intensiv beforscht und hat sich dabei als ein sehr robuster und genereller Befund erwiesen, der unter Verwendung verschiedenster Lerninhalte und Lehr-Lern-Szenarien demonstriert werden konnte (z. B. Carpenter & Kelly, 2012; Karpicke & Roediger, 2008; McDaniel, Anderson, Derbish & Morrisette, 2007; für einen Überblick, siehe Roediger & Butler, 2011, oder Roediger, McDermott, & McDaniel, 2015).

Der Fokus des hier vorgestellten Projekts lag auf einem vergleichsweise weniger intensiv beforschten Testeffekt, dem sogenannten *forward testing effect* (FTE). In einem Standard-Laborexperiment zum FTE lernen drei Gruppen von Proband\*innen mehrere Listen von unzusammenhängenden Wörtern. Dieses massierte Lernen hat zum Ziel, ein hohes Ausmaß an proaktiver Interferenz zu erzeugen, womit der störende Einfluss von früher gelernten Inhalten auf das Lernen und Erinnern nachfolgender Inhalte gemeint ist (Anderson & Neely, 1996; Crowder, 1976; Underwood, 1957). Die drei Proband\*innengruppen unterscheiden sich in der Art der Aktivität, die auf die Präsentation einer jeden Liste folgt: Die erste Gruppe bearbeitet nach jeder Liste eine irrelevante Distraktoraufgabe; die zweite Gruppe bekommt jede Liste ein weiteres Mal zum Lernen präsentiert; die dritte Gruppe schließlich wird unmit-

---

<sup>1</sup> Das an der MLU Halle-Wittenberg initiierte Kooperationsprojekt wurde nach dem Wechsel des Erstautors an die TH Rosenheim fortgeführt.

telbar nach jeder Liste getestet. Anschließend lernen alle Proband\*innen eine weitere (kritische) Liste, die nach einem kurzen Behaltensintervall erinnert werden soll. Der typische Befund ist, dass die Gruppe, die nach den ersten Listen jeweils getestet wird, signifikant mehr Wörter der kritischen Liste erinnert und dabei signifikant weniger Intrusionsfehler aus den ersten Listen produziert als die beiden anderen Gruppen (z. B. Szpunar, McDermott, & Roediger, 2008). Dieses mittlerweile vielfach replizierte Ergebnismuster (z. B. Bufo & Aslan, 2018; Pastötter, Schicker, Niedernhuber, & Bäuml, 2011; Weinstein, McDermott, & Szpunar, 2011; Yang & Shanks, 2018; für einen Überblick, siehe Chan, Meissner, & Davis, 2018) deutet an, dass das Testen von zuvor gelernten Inhalten den (sonst üblichen) Aufbau von proaktiver Interferenz reduziert und dadurch das Lernen und Erinnern nachfolgender Inhalte verbessert. Die lern- und gedächtnisförderlichen Effekte von Tests scheinen somit nicht allein auf bereits gelernte (und getestete) Inhalte beschränkt zu sein, sondern sich auch auf noch zu lernende (und selber nicht getestete) Inhalte zu verallgemeinern.

Proaktive Interferenz spielt nicht nur in Laborexperimenten eine wichtige Rolle, sondern kann auch in vielen Alltagssituationen, insbesondere beim Lernen großer Mengen von Informationen, ein potenzielles Problem darstellen. Medizinstudierende beispielsweise müssen sich in Veranstaltungen zur Pharmakologie Wissen über eine große Anzahl an Wirkstoffen aneignen (Namen, Indikationen, Nebenwirkungen etc.). Auch in diesem Praxisbeispiel ist von einer signifikanten Zunahme proaktiver Interferenz während des Lernens auszugehen, derart, dass mit jedem bereits gelernten Wirkstoff das Lernen und Behalten weiterer Wirkstoffe für Studierende immer schwieriger werden sollte. Sowohl für Studierende als auch für Lehrende wäre es daher von hoher praktischer Relevanz zu wissen, ob durch den Einsatz von Tests auch in einem solch angewandten Setting proaktive Interferenz reduziert und dadurch das Lernen von Studierenden effektiver gestaltet werden könnte.

Trotz seiner offensichtlichen Relevanz für die universitäre Lehr- und Lernpraxis wurde der FTE bislang meist nur unter eher artifiziellen Laborbedingungen und mit relativ alltagsfernen Lerninhalten (z. B. Listen von unzusammenhängenden Wörtern) untersucht (Aslan & Bäuml, 2016; Pastötter et al., 2011; Szpunar et al., 2008). Zudem, und dies ist ein Unterschied zum BTE, verwendeten die meisten bisherigen Studien zum FTE nur relativ kurze, aus pädagogisch-didaktischer Sicht eher uninteressante, Behaltensintervalle (30 sek bis max. 25 min; Aslan & Bäuml, 2016; Chan, Manley, Davis, & Szpunar, 2018). Ziel des hier vorgestellten Kooperationsprojekts war es daher, einen empirisch-experimentellen Beitrag zur Anwendungsrelevanz des FTE für die Hochschullehre zu liefern. Dazu wurde im Rahmen von Pharmakologie-Veranstaltungen für Medizinstudenten konkret untersucht, inwieweit sich der FTE auch unter Verwendung von pharmakologischen Lerninhalten und für die Lehr- und Lernpraxis relevanteren Behaltensintervallen finden lässt.

## **Methode**

### **Studienteilnehmende**

Es nahmen insgesamt 151 Medizinstudierende (109 weiblich, 42 männlich; Durchschnittsalter: 23.6 Jahre) an der Studie teil. Die Teilnahme war freiwillig und fand im Rahmen von mehreren Parallelseminaren zur Allgemeinen Pharmakologie (6. Semester) statt. Entsprechend der natürlichen Zusammensetzung der Seminare wurden die Teilnehmenden in Gruppen von 15-25 Personen getestet.

### **Material**

Das Lernmaterial bestand aus drei Listen (Liste A, Liste B, Liste C) mit je 10 zufällig gewählten, pharmakologisch sinnvollen Aussagen (z. B. „Clozapin kann eine Agranulozytose bedingen“ oder „Cephalosporine erhöhen die Wirksamkeit von Phenprocoumon“).

### **Versuchsdesign**

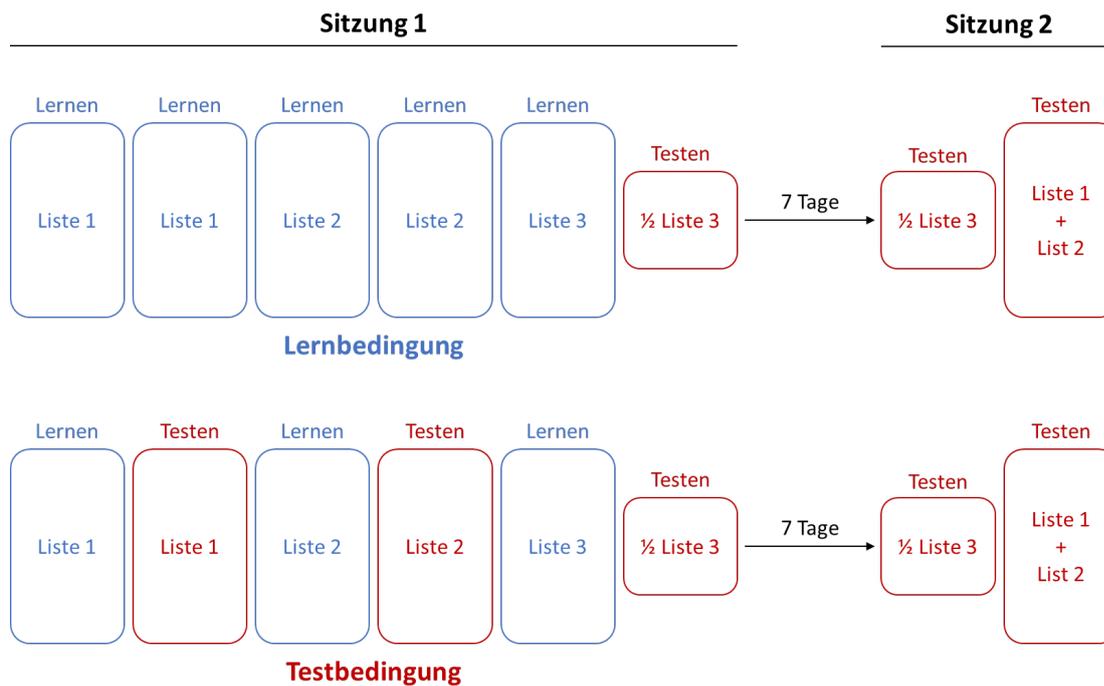
Das Experiment hatte ein 2x2x3 gemischt-faktorielles Design mit den Zwischensubjektfaktoren Bedingung (Lernbedingung, Testbedingung) und Behaltensintervall (1 min, 7 Tage) und dem Innersubjektfaktor Liste (Liste 1, Liste 2, Liste 3). Die Zuordnung der drei Lernlisten (A, B, C) zu den drei Listenpositionen wurde über die Teilnehmendengruppen hinweg ausbalanciert. Als abhängige Variable wurde die Behaltensleistung als Prozentsatz korrekt vervollständigter Aussagen in den drei Listen erhoben. Es wurden auch intra-experimentelle und extra-experimentelle Intrusionen erfasst; diese werden hier aber aus Gründen der Prägnanz nicht berichtet.<sup>2</sup>

### **Versuchsablauf**

Die Teilnehmenden der Studie wurden vorab informiert, dass es sich um eine Untersuchung zur Effektivität von Lernstrategien handelt, in der sie mehrere Listen von pharmakologischen Aussagen lernen und sich für einen späteren Erinnerungstest merken sollen. Die Teilnehmenden wurden zudem informiert, dass nach dem Lernen einer jeden Liste die Liste nach dem Zufallsprinzip entweder ein weiteres Mal zum Lernen präsentiert oder aber getestet werden könnte. Tatsächlich jedoch war die Zuordnung von Teilnehmenden zu den beiden Bedingungen vorab festgelegt.

---

<sup>2</sup> Eine intra-experimentelle Intrusion liegt vor, wenn eine Aussage (fälschlicherweise) mit einem Wirkstoff vervollständigt wird, der Teil einer anderen, im Experiment verwendeten Aussage ist. Eine extra-experimentelle Intrusion liegt vor, wenn eine Aussage (fälschlicherweise) mit einem Wirkstoff vervollständigt wird, der nicht Teil einer der im Experiment verwendeten Aussagen ist.



**Abbildung 1.** Schematische Skizze des Versuchsablaufs für die Lernbedingung (oben) und die Testbedingung (unten). Die beiden Bedingungen unterschieden sich lediglich darin, ob die Listen 1 und 2 nach ihrer ersten Präsentation noch einmal zum Lernen präsentiert oder getestet wurden

Das Experiment wurde in zwei getrennten Sitzungen im Abstand von 7 Tagen durchgeführt (siehe **Abbildung 1**). In der ersten Sitzung wurden zunächst die Aussagen der ersten beiden Listen sequentiell und in zufälliger Reihenfolge in einem Takt von 10 sec pro Aussage präsentiert. Die Präsentation erfolgte computergestützt mittels PowerPoint-Folien, welche über einen Beamer auf eine Leinwand projiziert wurden. Die Proband\*innen wurden gebeten, sich jede einzelne Aussage genau durchzulesen und sich für einen späteren Test zu merken. Nach der Präsentation der jeweils letzten Aussage einer Liste erhielten die Proband\*innen einfache Additions- und Subtraktionsaufgaben, die sie für eine Dauer von 1 min bearbeiten sollten. Im Anschluss an diese irrelevante Distraktoraufgabe wurde die Liste – je nach Bedingung – entweder ein zweites Mal präsentiert (Lernbedingung), oder die Liste wurde getestet (Testbedingung). In der Lernbedingung wurden die Aussagen der Liste auf dieselbe Art und Weise wie bei ihrer ersten Präsentation dargeboten, und die Teilnehmenden wurden aufgefordert, die Aussagen ein weiteres Mal zu lernen. In der Testbedingung hingegen wurden die Aussagen unvollständig dargeboten (z. B. „Clozapin kann eine \_\_\_\_\_ bedingen“ oder „Cephalosporine erhöhen die Wirksamkeit von \_\_\_\_\_“), und die Proband\*innen wurden aufgefordert, die Aussagen jeweils innerhalb von 10 sek korrekt zu vervollständigen. Es wurde betont, dass es möglicherweise mehrere inhaltlich korrekte Antworten geben könne, dass die Aufgabe der Teilnehmenden jedoch darin bestehe, die Lücke ausschließlich mit dem tatsächlich zuvor präsentierten Wort zu ergänzen. Im Anschluss daran wurde allen Teilnehmenden die dritte (kritische) Liste präsentiert und, nach einem Behaltensintervall von 1 min, sowohl in der Test- als auch in der Lernbedingung(!) direkt getestet. Dieser Test bezog sich jedoch in beiden Bedingungen nur auf eine zufällig gewählte Hälfte der Aussagen (d. h. 5 von

10 Aussagen), die verbleibenden 5 Aussagen von Liste 3 wurden zu Beginn der zweiten Sitzung 7 Tage später getestet. Abschließend wurden in der zweiten Sitzung in beiden Bedingungen auch die insgesamt 20 Aussagen von Liste 1 und 2 (noch einmal) in zufälliger Reihenfolge getestet.

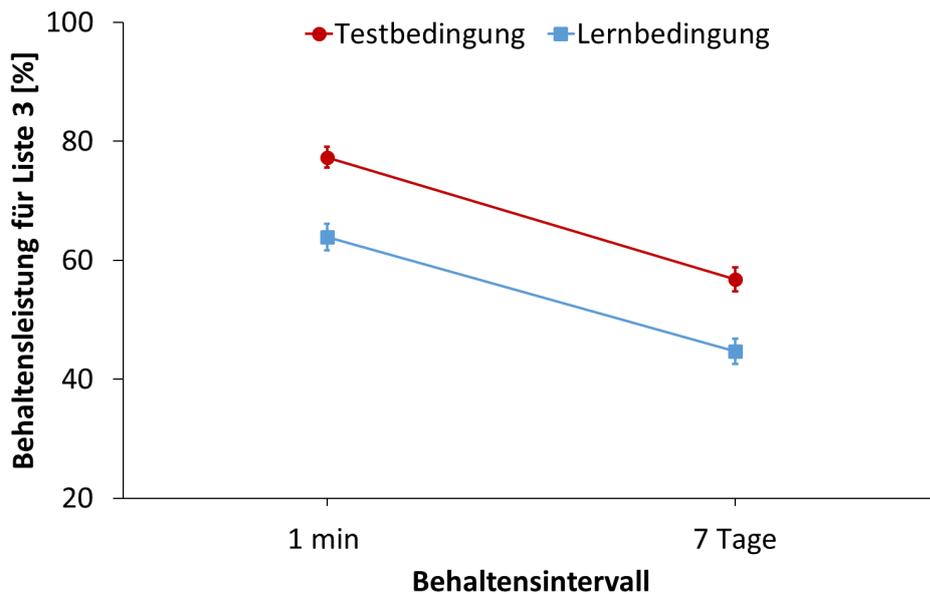
## Ergebnisse

In einem ersten Schritt wurde zunächst die mittlere (unmittelbare) Behaltensleistung der drei Listen (Liste 1, Liste 2, Liste 3) in der Testbedingung miteinander verglichen (siehe **Tabelle 1**). Eine einfaktorielle Varianzanalyse mit Messwiederholung ergab keinen signifikanten Haupteffekt des Innersubjektfaktors Liste,  $F(1, 146) < 1$ , was darauf hindeutet, dass die Behaltensleistung über die drei Listen hinweg unverändert blieb und es in der Testbedingung somit zu keinem Aufbau von proaktiver Interferenz kam.

*Tabelle 1. Mittlere Behaltensleistung (in Prozent) für die drei Listen in der Testbedingung (Standardfehler in Klammern)*

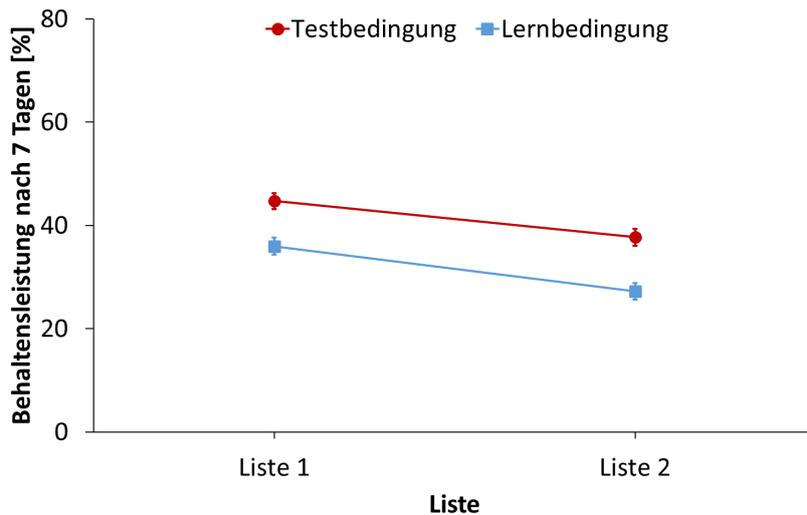
Liste 1	Liste 2	Liste 3
76.0 (1.4)	75.0 (1.6)	77.3 (1.6)

In einem zweiten Schritt wurde die mittlere Behaltensleistung für Liste 3 in Abhängigkeit der Bedingung (Lernbedingung, Testbedingung) und des Behaltensintervalls (1 min, 7 Tage) untersucht (siehe **Abbildung 2**). Eine zweifaktorielle 2×2-Varianzanalyse ergab signifikante Haupteffekte des Zwischensubjektfaktors Bedingung ( $F(1, 149) = 32.3, p < .001$ ) und des Innersubjektfaktors Behaltensintervall ( $F(1, 149) = 116.5, p < .001$ ). Der Effekt des Behaltensintervalls spiegelt das „normale“ Vergessen über die Zeit wider; der Effekt der Bedingung reflektiert das Vorliegen eines FTEs, d. h. eine bessere Behaltensleistung in der Testbedingung im Vergleich zur Lernbedingung. Der Interaktionseffekt zwischen den beiden Faktoren war nicht signifikant,  $F(1, 149) < 1$ , was bedeutet, dass die Größe des FTEs nach den beiden Behaltensintervallen vergleichbar war. Einzelvergleiche bestätigten, dass sowohl nach 1 min ( $t(149) = 4.7, p < .001$ ) als auch nach 7 Tagen ( $t(149) = 4.1, p < .001$ ) ein signifikanter FTE vorlag.



**Abbildung 2.** Mittlere Behaltensleistung (in Prozent) als Funktion der Bedingung und des Behaltensintervalls (Standardfehler in Klammern)

In einem dritten und letzten Schritt schließlich wurde die mittlere Behaltensleistung für die Listen 1 und 2 in den beiden Bedingungen (Lernbedingung, Testbedingung) nach 7 Tagen untersucht (siehe **Abbildung 3**). Eine zweifaktorielle 2×2-Varianzanalyse ergab signifikante Haupteffekte des Zwischensubjektfaktors Bedingung ( $F(1, 149) = 28.9, p < .001$ ), des Innersubjektfaktors Liste ( $F(1, 149) = 32.9, p < .001$ ), aber keine Interaktion zwischen den beiden Faktoren,  $F(1, 149) < 1$ . Der Effekt der Liste spiegelt einen sogenannten *Primacy*-Effekt auf Listenebene wider, d. h. einen Behaltensvorteil für die zuerst gelernte Liste 1 im Vergleich zur später gelernten Liste 2. Der Effekt der Bedingung reflektiert einen typischen BTE, d. h. die Listen 1 und 2 wurden insgesamt besser erinnert, wenn sie nach dem erstmaligen Lernen getestet – statt wiederholt gelernt – wurden. Die fehlende Interaktion zwischen den beiden Faktoren schließlich spiegelt die Tatsache wider, dass die Größe des BTEs für beide Listen vergleichbar war. Einzelvergleiche bestätigten, dass sowohl für Liste 1 ( $t(149) = 3.9, p < .001$ ) als auch für Liste 2 ( $t(149) = 4.6, p < .001$ ) ein signifikanter BTE vorlag.



**Abbildung 3.** Mittlere Behaltensleistung (in Prozent) für Liste 1 und 2 als Funktion der Bedingung (Standardfehler in Klammern).

## Diskussion

Ziel des hier vorgestellten Kooperationsprojekts zwischen dem Institut für Psychologie und dem Institut für Pharmakologie und Toxikologie der MLU war es, die Relevanz des FTE im Kontext der universitären Lehr- und Lernpraxis zu untersuchen. Die zentralen Ergebnisse des Projekts und deren Implikationen lassen sich wie folgt zusammenfassen:

1. Die (unmittelbare) Behaltensleistung für eine Liste von pharmakologischen Aussagen (Liste 3) war signifikant höher, wenn zwei zuvor präsentierte Listen von Aussagen (Listen 1 und 2) nach deren Präsentation jeweils getestet – statt jeweils wiederholt gelernt – wurden. Dieser Befund repliziert und erweitert die Ergebnisse früherer Studien zum FTE, indem er zeigt, dass Tests den (sonst üblichen) Aufbau von proaktiver Interferenz nicht nur beim Lernen von simplen Wortlisten, sondern auch beim Lernen pädagogisch relevanter Informationen reduzieren können.
2. Innerhalb der Testbedingung blieb die (unmittelbare) Behaltensleistung über die drei Listen hinweg weitgehend konstant, d. h. die pharmakologischen Aussagen der späteren Listen 2 und 3 wurden genauso gut gelernt und erinnert wie die Aussagen der Liste 1. Dieser Befund deutet an, dass Tests – zumindest unter den in dieser Studie gegebenen Bedingungen – den Aufbau von proaktiver Interferenz nicht nur reduzieren, sondern gänzlich eliminieren können.
3. Der Behaltensvorteil der Testbedingung gegenüber der Lernbedingung zeigte sich nicht nur nach einem kurzen Behaltensintervall von 1 min, sondern insbesondere auch noch nach einem längeren Behaltensintervall von 7 Tagen. Die Ergebnisse der vorliegenden Studie replizieren somit den in der Literatur bekannten Befund

eines unmittelbaren FTEs, gehen aber darüber hinaus, indem sie (erstmalig) andeuten, dass der Effekt auch über einen längeren Zeitraum hinweg gefunden werden kann.

4. Die Behaltensleistung für Listen von pharmakologischen Aussagen (Listen 1 und 2) war signifikant höher, wenn die Listen nach ihrer ursprünglichen Präsentation 7 Tage zuvor jeweils getestet – statt jeweils wiederholt gelernt – wurden. Dieser Befund spiegelt einen vom FTE zu unterscheidenden, ebenfalls langanhaltenden Testeffekt, den sogenannten BTE, wider.

Die Ergebnisse des hier vorgestellten Kooperationsprojekts sind vielversprechend und haben potentiell bedeutsame Implikationen für die (multimediale) Hochschullehre. Konsistent mit Befunden aus früheren Laborexperimenten zum BTE (z. B. Hogan & Kintsch, 1971; Thompson, Wenger, & Bartling, 1978; Wheeler, Ewers, & Buonanno, 2003) und FTE (z. B. Aslan & Bäuml, 2016; Pastötter et al., 2011; Szpunar et al., 2008) zeigen sie, dass das Testen (bzw. Prüfen) von zuvor gelernten Inhalten gleich in zweierlei Hinsicht förderlich für das Lernen von Studierenden sein kann: Zum einen verbessert es das langfristige Behalten der getesteten Informationen (BTE), gleichzeitig aber erleichtert es auch das Lernen und Behalten neuer (nicht getesteter) Informationen (FTE). Der Befund, dass diese beiden Effekte auch mit pädagogisch relevanten Lerninhalten und nach pädagogisch bedeutsamen (längeren) Behaltensintervallen beobachtet werden können, ist von großer Bedeutung für die universitäre Lehr- und Lernpraxis, da er eine häufigere Nutzung von Tests in der (multimedialen) Hochschullehre nahelegt. Tests, seien es mündliche oder schriftliche Prüfungen, können und sollten den vorliegenden Ergebnissen zufolge nicht nur am Ende eines Semesters eingesetzt werden, um Wissen zu diagnostizieren, sondern bereits während des Semesters, um Studierende dabei zu unterstützen, Wissen aufzubauen und langfristig zu behalten.

## Literatur

- Aslan, A., & Bäuml, K.-H. T. (2016). Testing enhances subsequent learning in older but not in younger elementary school children. *Developmental Science, 19*, 992-998.
- Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory: Handbook of perception and cognition* (2nd ed., pp. 237-313). San Diego, CA: Academic Press.
- Bufe, J., & Aslan, A. (2018). Desirable difficulties in spatial learning: Testing enhances subsequent learning of spatial information. *Frontiers in Psychology: Cognition, 9*, 1701.
- Carpenter, S. K., & Kelly, J. W. (2012). Tests enhance retention and transfer of spatial learning. *Psychonomic Bulletin & Review, 19*, 443-448.
- Chan, J. C. K., Manley, K. D., Davis, S. D., & Szpunar, K. K. (2018). Testing potentiates new learning across a retention interval and a lag: A strategy change perspective. *Journal of Memory and Language, 102*, 83-96.

- Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin*, *144*, 1111-1146.
- Crowder, R. G. (1976). Principles of learning and memory. Oxford, England: Lawrence Erlbaum.
- Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 562-567.
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science* *319*, 966-968.
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*, 494-513.
- Multimedia-Leitbild der Martin-Luther-Universität Halle-Wittenberg (2015). <https://www.llz.uni-halle.de/multimedia-leitbild/>
- Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology Learning, Memory, & Cognition*, *37*, 287-297.
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20-27.
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science*, *15*, 249-255.
- Roediger, H. L., McDermott, K. B., & McDaniel, M. A. (2015). Using testing to improve learning and memory. In M. A. Gernsbacher, R. Pew, L. Hough & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 69-78). New York, NY: Worth Publishing Co.
- Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal Experimental Psychology: Learning, Memory & Cognition*, *34*, 1392-1399.
- Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 210-221.
- Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, *64*, 49-60.
- Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face-name learning. *Psychonomic Bulletin & Review*, *18*, 518-523.
- Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*, 571-580.
- Yang, C., & Shanks, D. R. (2018). The forward testing effect: Interim testing enhances inductive learning. *Journal Experimental Psychology: Learning, Memory & Cognition*, *44*, 485-492.

**Alp Aslan** leitete bis August 2020 die Abteilung für Entwicklungspsychologie an der Martin-Luther-Universität Halle-Wittenberg und ist seit September 2020 Professor für die Grundlagen der Angewandten Psychologie an der Technischen Hochschule Rosenheim. Seine Forschungsschwerpunkte liegen an den Schnittstellen zwischen Allgemeiner Psychologie, Entwicklungspsychologie und Pädagogischer Psychologie und umfassen sowohl grundlagenorientierte als auch anwendungsorientierte Fragestellungen zu den Themengebieten Lernen/Lehren, Gedächtnis und kognitive Entwicklung über die Lebensspanne.

[alp.aslan@th-rosenheim.de](mailto:alp.aslan@th-rosenheim.de)

**Joachim Neumann** ist Professor für Pharmakologie und Toxikologie, Facharzt für Pharmakologie und Klinische Pharmakologie sowie Erasmus Koordinator für den Bereich Medizin an der Martin-Luther-Universität Halle-Wittenberg. Er arbeitet hauptsächlich im Bereich der Grundlagenforschung zur Herz-Kreislauf-Pharmakologie mit verschiedenen Forschungsansätzen. Neben dieser Grundlagenforschung beschäftigt er sich mit dem Lernverhalten von Studierenden mit dem Ziel, Erkenntnisse zur Verbesserung der Lehre in der Pharmakologie und Toxikologie zu gewinnen.

[joachim.neumann@medizin.uni-halle.de](mailto:joachim.neumann@medizin.uni-halle.de)

# 4

## Innovationen beim digitalen Prüfen

# Detection of Cheating on E-Exams – The Performance of Transferred Detection Rules

*Jochen Ranger<sup>1</sup>, Nico Schmidt<sup>1</sup> und Anett Wolgast<sup>2</sup>*

*<sup>1</sup> Martin-Luther-University Halle-Wittenberg, Institute of Psychology*

*<sup>2</sup> University of Applied Sciences FHM Hannover, Department of Psychology*

Data forensics is a field in psychometrics that deals with the detection of cheating on exams. A recent trend in data forensics is the application of techniques from machine learning. This, however, is impeded by the fact that machine learning requires large data sets. Such data sets are often not available in educational assessment. A solution might consist in combining knowledge from different sources. In this paper, we investigate whether this approach is promising. We analyze whether a detector of cheating can be transferred from one exam to another. We train a flexible discriminant model to separate cheaters from regular responders with a first training data set. In the flexible discriminant model, we use standard indicators of cheating, but also new ones based on process data as predictors of cheating. We then apply the flexible discriminant model to new data (different test material, different forms of cheating). In the training data set, the flexible discriminant model is capable to separate cheaters from regular responders moderately well (sensitivity: 0.68 / specificity: 0.98). The transfer of the flexible discriminant model to the new data set is less successful. The specificity remains high, but the sensitivity drops significantly. This is due to concept drift, a change of the relation between the predictors and the tendency to cheat.

## Detection of Cheating on E-Exams – The Performance of Transferred Detection Rules

Cheating is a serious threat to the validity of academic assessment. Examinees that cheat by answer copying, collusion with other examinees or the acquisition of test content before the exam may get exam scores that overrate their true level of proficiency. Although cheating is seen as ethically wrong by most students, it is widespread in higher education. More than 40 % of all high and college school students report having cheated at least once during their academic career (Bernardi et al., 2008; Cizek & Wollack, 2017a). The prevalence of cheating might even increase in the future as recent inventions like micro cameras or smart watches make it easier to cheat. This requires measures to prevent, or at least, detect cheating.

In psychometrics, the term data forensics is used to denote all activities that aim at the detection of test fraud; for an overview of data forensics see the monographs by Cizek and Wollack (2017b), Kingston and Clark (2014), and Wollack and Fremer (2013) or the manuscript of Sinharay (2018b). In this manuscript, we focus on one specific problem in data forensics, the detection of cheaters. The detection of cheaters requires a classification rule that states which examinees should be regarded as regular responders and which examinees as cheaters. The classification rule consists in an algorithm that derives the response mode – cheating or regular responding – from aspects of an examinee’s test results (e.g., response pattern) and test taking behavior (e.g., response times). Mathematically, the detection of cheating consists in mapping points of the space spanned by predictors of cheating to one of the two response modes. Although such a mapping can be generated by conventional statistical models (e.g., logistic regression model), there has been increased interest in the application of algorithms from machine learning recently (Burlak et al., 2006; Kim et al., 2017; Man et al., 2019).

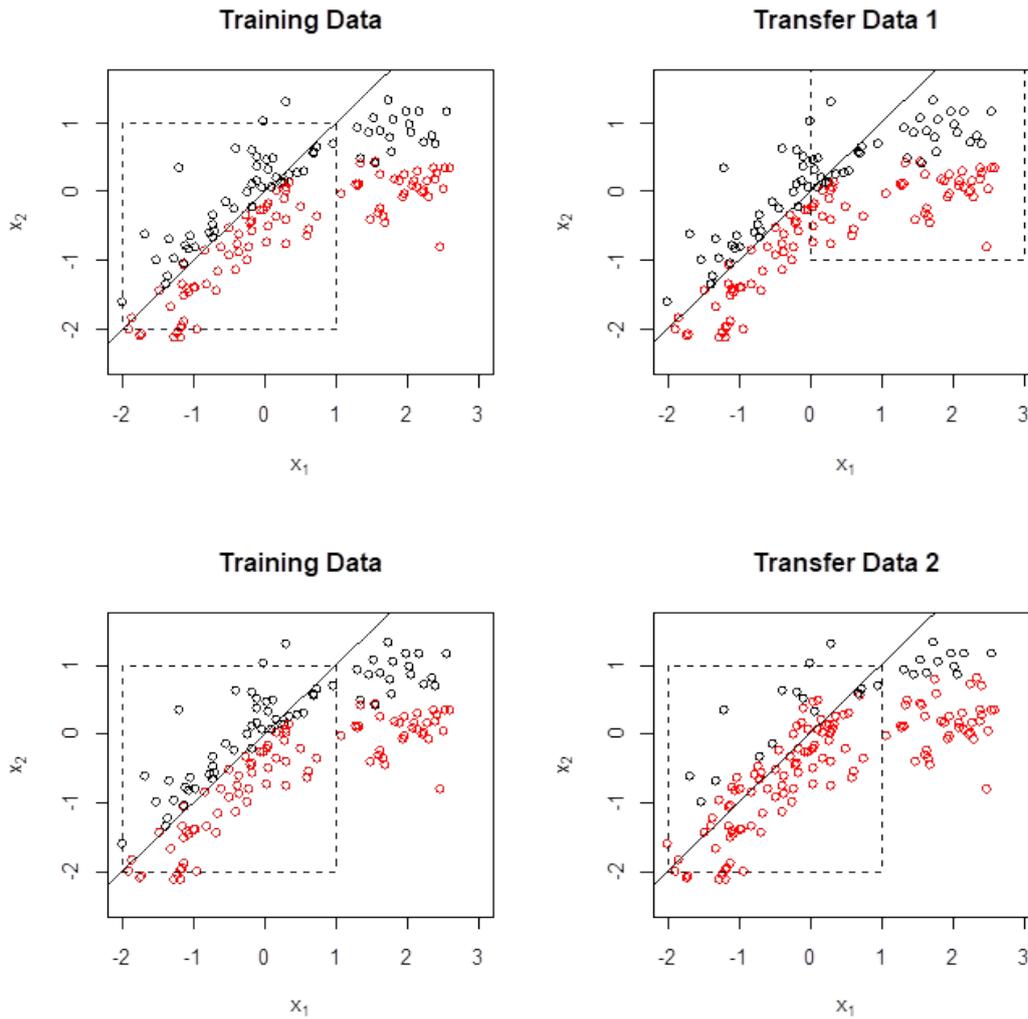
In data forensics, two different forms of machine learning are employed (e.g., Man et al., 2019). *Supervised* machine learning aims at discovering a rule that best separates a known group of regular responders from a known group of cheaters. Once discovered, the rule can be employed to new examinees with unknown response mode. In *unsupervised* machine learning, the sample of all examinees is divided into regular responders and cheaters. Unsupervised machine learning is exploratory in the sense that no initial sample with labeled examinees is required. Although both forms of machine learning have their merits, they soon reach their limit in educational assessment. Supervised machine learning requires labeled cases, that is, an initial sample with known cheaters. A sample with labeled cases is usually not available in practice. Unsupervised machine learning requires samples with large sample size and does not perform well when cheaters and regular responders are hard to separate. Unsupervised machine learning is also unspecific in the sense that it aims at identifying groups of examinees, irrespective whether these are cheaters, rapid guessers or simply examinees that behave differently than the remaining ones.

In this paper, we investigate the applicability of a third form of machine learning. We investigate whether it is possible to transfer a classification rule from an initial training data

set to a new transfer data set. We proceed as follows. We first train a supervised learning machine to separate cheaters from regular responders with an initial training data set. We then use the trained machine in order to classify the examinees of a new transfer data set. This proceeding is an improvement over unsupervised machine learning as the classification of the cheaters in the transfer data set is not exploratory. It also does not require labeled examinees in the transfer data set, as supervised machine learning would do.

Knowledge transfer from one area of application to another area has received increased interest in machine learning (e.g., Pan & Yang, 2010; Weiss et al., 2016). Being able to transfer knowledge requires several prerequisites. First of all, there has to be knowledge that can be transferred. In the present context, this requires the existence of indicators of cheating that – provided that they are combined in the right way – are capable to separate cheaters from regular responders in the training data set well. Then, there has to be a way to transfer the classification rule to the transfer data set. This requires that the indicators used in the training data set are also defined in the transfer data set. Finally, the classification rule should be valid in the transfer data set. This requires that the interpretation of the indicators is context independent in the sense that the values that characterize cheating do not depend on characteristics of the exam (e.g., the number and format of the questions). In summary, the success of transfer learning depends on the similarity of the training data set and the transfer data set.

In practice, not all of these prerequisites are met. Often, there is no absolute similarity between the training and the transfer data set. Two cases have to be distinguished, sample selection bias and concept shift. *Sample selection bias* (aka covariate shift) refers to the case that the distribution of the predictors (in the present case, the indicators of cheating) differs between the training and the transfer data set. This case is illustrated in **Figure 1** in the first row. There, the values of examinees on two predictors are visualized. The examinees can be classified into one of two groups. The group is indicated by the color (red/black). Let us assume that in the training data set (**Figure 1**, Row 1, left), only the examinees in the dotted box are observed. The straight line would be able to separate the groups well. If we observe different examinees, for example, by moving the dotted box as in the transfer data set 1 (**Figure 1**, Row 1, right), the straight line does not separate the groups well anymore. Although sample selection appears to be problematic on first sight, it can be shown that it does not have serious consequences for knowledge transfer under certain conditions (Jiang, 2008; Zadrozny, 2004). *Concept shift* refers to the case that the relation between the criterion and the predictors changes. Concept shift is illustrated in **Figure 1** in the second row. Here, the distribution of the predictors does not change from the training data set to the transfer data set, but the conditional probability of belonging to one of the two groups changes. As a consequence, the straight line is not capable to separate the groups well in the transfer data set 2 (**Figure 1**, Row 2, right). Concept shift undermines the success of knowledge transfer (Gama et al., 2014).



**Figure 1.** Illustration of the Effects of Sample Selection (Transfer Data I) and Concept Drift (Transfer Data II) for the Separation of Two Groups on Basis of Two Predictors  
**Note.** The figure visualizes the values of examinees from two groups (red dots, black dots) on two predictors ( $x_1$ ,  $x_2$ ). The examinees in the dotted area are observed whereas all values outside the dotted area are assumed to be missing. The straight line represents a separating boundary that is used in order to classify the examinees into one of the two groups.

In this paper, we make several contributions. We first propose new indicators of cheating on basis of the response times on exams. The proposed indicators parallel indices of person fit developed in nonparametric item response theory and should meet the requirement of context independent interpretation. We then investigate whether the indicators can be combined to a classification rule that separates regular responders from cheaters. For this purpose, we use a flexible discriminant model. We finally analyze whether the classification rule can be used for the detection of other forms of cheating or in different examination conditions. We investigate whether there is evidence for sample selection or concept shift. The outline of the paper is as follows. In the theoretical part of the paper, we describe the new indicators of cheating and flexible discriminant analysis. In the empirical part of the paper,

we investigate the performance of the indicators, the predictive accuracy of the flexible discriminant model in the training data set and the applicability of the trained discriminant model to new forms of cheating and new data.

## Indicators of Cheating on Basis of Response Times

Numerous indicators of cheating have been proposed up to now. The most popular indicators assess the plausibility of a response pattern, the amount of item revisions or the similarity of response patterns from different examinees; see Cizek and Wollack (2017b) for an overview. Although these indices have their merits and are capable to detect some forms of cheating, they are insensitive to others. Preknowledge in the form that an examinee has access to all questions before taking the exam does manifest itself neither in an unusual response pattern, a high similarity with other response patterns, nor a large amount of item revisions. The detection of such forms of cheating requires additional information like, for example, process data. The most popular form of process data are the times needed to respond to the questions in an exam.

Response times are closely related to the response process. Irregular ways of responding might be indicated by unusual response time patterns. Response times have been used to assess person fit by van der Linden and Guo (2008), Marianti et al. 2014, Fox and Marianti (2017) and Sinharay (2018a). Response times have also been used for the detection of cheating directly (Man et al., 2018; Ranger et al., 2020; Sinharay, 2020; van der Linden, 2009; Wang et al., 2018). Although these approaches have their merits, they also have limitations. Most of them are model based, that is, require strong assumptions about the response time distribution. They are also not robust against outliers. This limits their usefulness in practice. In this paper, we suggest two new indices – the UT1 index and the  $HT^T$  index – that do not require strong model assumptions and are robust against outliers. The indices parallel two popular indices of person fit developed in nonparametric item response theory.

**UT1 Index.** The UT1 index assesses the Guttman homogeneity of an examinee's response time pattern. The concept of Guttman homogeneity was developed in nonparametric item response theory. A response pattern is denoted as Guttman homogeneous in case solving a difficult question implies that all easier questions have been solved as well. The concept of Guttman homogeneity can be extended to the response times directly. Response patterns are Guttman homogeneous when questions with lower time demand are solved faster than questions with higher time demand. The amount of Guttman homogeneity of a response time pattern can be quantified as follows. First, the questions of the exam are sorted with respect to their time demand represented by the median response time. In case an examinee's response time pattern is perfectly Guttman homogeneous, his/her response times should be strictly increasing now. A Guttman error occurs when a response time violates the monotony relation. The UT1 index is the ratio of the number of Guttman errors in a response

time pattern to the maximal number of Guttman errors that could have occurred. The maximal number of Guttman errors is  $G \cdot (G - 1) / 2$ , where  $G$  denotes the number of the questions in the exam.

We exemplify the calculation of the UT1 index on the basis of the data contained in **Table 1**. In **Table 1**, the response times of an examinee in a test of five questions (second row) are reported. The questions have been sorted according to the median response times in the sample of all examinees. The median response times are reported in the first row.

*Table 1. Response Times (RT) of an Examinee in an Exam Consisting of Five Questions as well as the Median Response Times of the Questions*

Question	1	2	3	4	5
Median RT	10.1	12.3	14.8	20.7	28.3
Examinee RT	12.9	14.8	14.6	25.7	30.5

**Note.** Questions are sorted with respect to their time demand (Median RT).

The examinee has a low work pace as his/her response times are larger than average in all items. The response time profile is not unusual. The examinee's response times are sorted almost perfectly. The examinee is faster in the items with low time demand and slower in item with high time demand. There is one Guttman error (Question 2 – Question 3) where the response times are not sorted correctly. The maximal number of Guttman errors would be  $G \cdot (G - 1) / 2 = 5 \cdot (5 - 1) / 2 = 10$  when all pairs would be sorted incorrectly. The UT1 index is  $UT1 = 1 / 10 = 0.1$ . The UT1 index is related to the Kendall rank correlation  $\tau$  between the response times of an examinee and the median response times via the relation  $UT1 = (1 - \tau) / 2$ .

Cheating usually does not require full problem processing when responding. An examinee with preknowledge, for example, simply recalls the solution instead of generating it actively. This implies that the response times in the compromised questions should be short. In case only a subset of the questions is affected, the response time pattern becomes irregular and the UT1 index high; note that in its essence the UT1 index is very similar to the U1 index suggested by van der Flier (1977) for the responses.

**HT<sup>T</sup> Index.** The HT<sup>T</sup> index is an extension of the H<sup>T</sup> index (Sijtsma, 1986) to response times. The HT<sup>T</sup> index assesses how similar the response time pattern of an examinee is to the remaining response time patterns in the sample. The index is defined as the average of all  $\tau$  correlations between an examinee's response time pattern and the response time patterns of all other examinees. In theory, as it is done by Sijtsma (1986), one could relate the average  $\tau$  correlation to the maximal value that could be achieved given the total testing time, but this will not change the results much.

We exemplify the calculation of the HT<sup>T</sup> index on basis of the data given in **Table 2**. **Table 2** contains the response time patterns of three examinees in a test of five questions.

The response time profiles of Examinee 2 and Examinee 3 are very similar apart from a difference in overall speed. Examinee 1 has a different response time profile as he/she responds fast to questions that were answered slowly by Examinee 2 and Examinee 3. The response times of Examinee 1 and Examinee 2 correlate with  $\tau = -0.60$ ; the response times of Examinee 1 and Examinee 3 with  $\tau = -0.40$ . The  $HT^T$  index is the average of all  $\tau$  correlations  $HT^T = \bar{\tau} = -0.50$ .

**Table 2.** Response Times (RT) of Three Examinees in an Exam Consisting of Five Questions

Question	1	2	3	4	5
Examinee 1	15.1	12.3	9.8	9.7	10.3
Examinee 2	10.9	11.8	13.6	17.7	30.5
Examinee 3	14.8	17.2	20.5	19.8	33.2

In case an examinee behaves differently and the different behavior manifests itself in a different response time pattern, the  $HT^T$  index is low. As cheating might alter the time an examinee has to spend on answering the questions, the  $HT^T$  index might be low in cheaters.

### Detection of Cheaters by Means of Flexible Discriminant Analysis

Indicators of cheating are usually imperfect in the sense that there is no cut point that separates regular responders from cheaters perfectly. The detection of cheating can be improved by combining several weak indicators. A high test score is not suspicious on its own, but might become a cause of concern in combination with very short response times. In order to take several indicators of cheating into account at the same time, one needs an algorithm that specifies how the indicators have to be combined. Such an algorithm can be derived with machine learning. The problem consists in finding a map of the input space into the output space that assigns specific values of the cheating indicators to one of the two response modes. This assignment partitions the input space into a subspace corresponding to the cheaters and a subspace corresponding to the regular responders and constitutes the classification rule. In this paper, we determine the classification rule by means of flexible discriminant analysis (Hastie et al., 2014, pp. 440–449).

*Flexible* discriminant analysis is an extension of *linear* discriminant analysis. In *linear* discriminant analysis, the linear combination of the predictors is sought that best separates the groups. In the present case with just two groups (cheaters/regular responders), this problem can be interpreted as the least squares problem

$$\min_{\beta} \left( \sum_{i=1}^n (\theta(y_i) - \mathbf{x}_i' \beta) \right)^2 \quad (1)$$

In Equation (1),  $\theta(y_i)$  denotes a numerical code that indicates the response mode  $y_i$  of an examinee,  $\mathbf{x}_i$  the vector containing the indicators of cheating and  $\boldsymbol{\beta}$  the vector of the unknown regression coefficients. The sum runs over the  $i = 1, \dots, n$  examinees in the sample. The unknown regression coefficients are determined by minimizing the sum of squared differences between the criterion  $\theta(y_i)$  and the prediction  $\mathbf{x}_i' \boldsymbol{\beta}$  (Hastie et al., 2014, p. 440).

In linear discriminant analysis, the linear predictor  $\eta_i = \mathbf{x}_i' \boldsymbol{\beta}$  is used in order to classify examinees as cheaters or regular responders. The classification rule assigns examinees to the group with the closest centroid. This classification rule partitions the input space spanned by the indicators linearly. With two indicators  $x_1$  and  $x_2$ , for example, the input plane formed by  $x_1$  and  $x_2$  is simply divided by a straight line into two regions that correspond to cheaters and regular responders.

*Flexible* discriminant analysis extends linear discriminant analysis in two ways. First, the linear predictor  $\eta_i = \mathbf{x}_i' \boldsymbol{\beta}$  is replaced by a nonlinear (e.g., polynomial) function of the indicators  $\eta_i^* = f(\mathbf{x}_i)' \boldsymbol{\beta}$ . Second, a penalty is introduced that smooths the nonlinear function. In contrast to linear discriminant analysis, flexible discriminant analysis is capable to learn nonlinear classification rules. Although flexible discriminant analysis still assumes a linear separation of the groups with respect to the composite  $\eta_i^* = f(\mathbf{x}_i)' \boldsymbol{\beta}$  (extended input space), the separation is nonlinear in the original input space spanned by the predictors  $\mathbf{x}_i$ .

## Method

The aim of the empirical study was threefold. Firstly, we wanted to investigate the performance of process data in comparison to the classical indices of person fit in the detection of cheaters. Secondly, we wanted to investigate whether the flexible discriminant model provides a classification rule that separates regular responders and cheaters well. Thirdly, we wanted to investigate whether a classification rule determined with one data set can be generalized to other forms of cheating or other exams.

## Participants

Three data sets were considered. With the first data set (training data set), we trained a flexible discriminant model to separate cheaters from regular responders. The trained discriminant model was then used to detect cheaters in two new data sets (transfer data set 1 / transfer data set 2). The training data set and the transfer data set 1 were collected for the study by the authors. The transfer data set 2 was downloaded from the internet. In all data sets, the cheaters were known. Due to lack of space, we only give a rudimentary description of the data; for more details, see Ranger et al. (2020).

**Training Data Set.** The training data set consisted of the responses and the response times of 297 education undergraduate students in a mock exam on psychology. The mock exam was scheduled one week before the final exam the students had to take in order to pass a course on psychology. The exam consisted of 20 questions with multiple choice response format and was computer assisted. Two hundred seventy-five students responded

regularly, 22 students had preknowledge. Preknowledge was induced intentionally as follows. One day before the mock exam, the students had the chance to examine 8 of the 20 questions and discuss the correct solution. The questions were presented during an ostensibly unrelated study the students were recruited for. Although the students received the questions, they were not aware that the questions would be part of the mock exam the next day. When selecting the questions to be leaked, we tried to avoid any systematic differences between the leaked and uncompromised questions with respect to their time demand (as assessed by the word count) and difficulty (as assessed by expert ratings).

**Transfer Data Set 1.** The first transfer data set consisted of the responses and the response times of 414 students in a mock exam. The mock exam was identical to the one used in the training data set. The sample, however, consisted of a different cohort of students. This time, 397 students responded regularly. The remaining 17 students had preknowledge. Preknowledge was induced intentionally as follows. One day before the mock exam, the students had the chance to examine 8 of the 20 questions of the exam and discuss the correct solution. The questions were exposed during a supposedly unrelated study the students were recruited for. The students were informed that the questions were taken from the item pool of their lecturer and could possibly be used in the final exam. The leaked questions in the transfer data set 1 were different to the leaked questions in the training data set. The leaked questions had a lower time demand and a lower item difficulty in the transfer data set 1 than in the training data set.

**Transfer Data Set 2.** The second transfer data set was a subset of a credentialing data set (Cizek & Wollack, 2017a). The data set consisted of the responses and response times in a credentialing test provided in two forms, each consisting of 170 scored items and 10 pretest items. Here, the analysis is restricted to the 87 items that were contained in both forms. We analyzed the data of all examinees from India. The subset contained 318 examinees. Thirty-eight examinees are known to have obtained live test content prior to the examination.

## Analysis

We determined the exam score, the total testing time and the two indicators of cheating ( $UT1$ ,  $HT^T$ ) proposed in the previous section. We additionally calculated the  $H^T$  index and the  $U1$  index that are frequently used in order to assess the regularity of the responses; see Meijer and Sijtsma (2001) for more details. These indices were included as a benchmark. For sake of clarity, we will denote these indices as  $UX1$  and  $HX^T$  in the following.

In a descriptive analysis, we explored whether the distribution of the responses, the response times and the indicators of cheating were different in regular responders and cheaters. We then trained a flexible discriminant model to separate cheaters from regular responders with the training data set. For sake of simplicity, we used only two indicators, namely the  $UT1$  and the  $UX1$  index. Including further indicators did not change the performance much, but complicates the interpretation of the findings. For the flexible discriminant analysis, the two indicators were expanded to multivariate adaptive regression splines

(mars). The flexible discriminant analysis was fitted with the package *mda* in the statistical software environment R (Hastie et al., 2013). We used the default options with respect to the penalty term, degree of the splines and the interactions. The trained flexible discriminant model was used to predict the response mode in the training data set, the first transfer data set and the second transfer data set.

## Results

### Descriptive Statistics

The averages and standard deviations of the sum score, the total testing time and the four indicators of cheating are reported in **Table 3** for the three data sets.

**Table 3.** Mean (Standard Deviation) of the Total Testing Time (*T*), the Test Score (*X*) and Four Indicators of Cheating (*HTT*, *HXT*, *UT1*, *UX1*) for Two Groups of Examinees in Three Data Sets

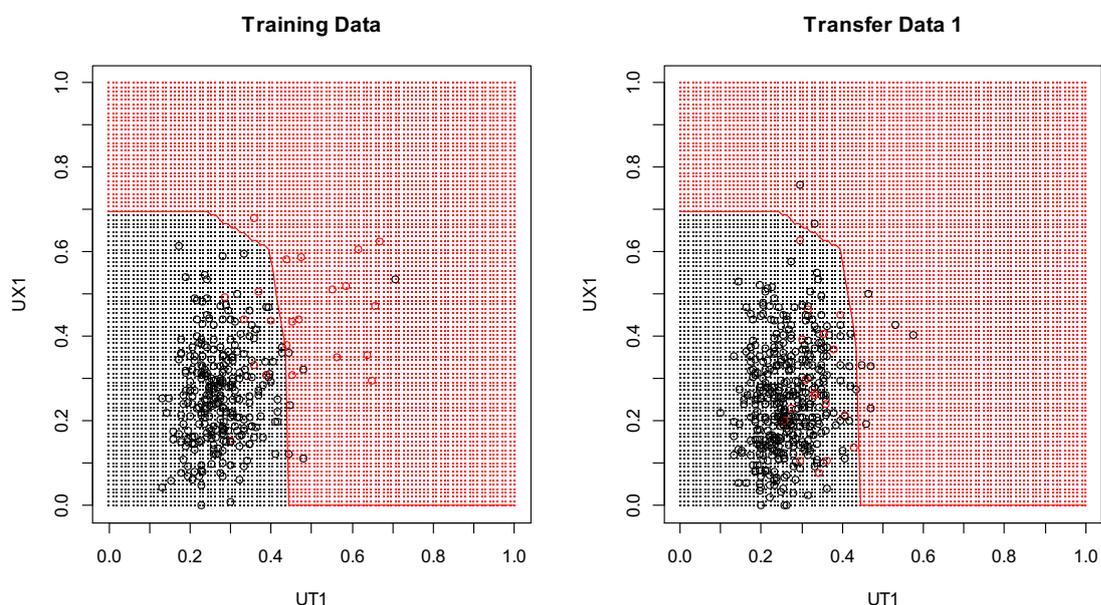
	Training Data		Transfer Data 1		Transfer Data 2	
	RR	CHT	RR	CHT	RR	CHT
<i>T</i>	653.85 (192.86)	521.12 (166.94)	642.13 (222.59)	517.51 (211.60)	5928.05 (480.86)	5626.26 (517.76)
<i>X</i>	11.07 (2.43)	14.36 (1.40)	11.31 (2.48)	15.00 (1.77)	56.07 (9.51)	63.76 (12.29)
<i>HTT</i>	0.25 (0.07)	0.02 (0.14)	0.29 (0.08)	0.21 (0.06)	0.22 (0.04)	0.23 (0.04)
<i>HXT</i>	0.20 (0.10)	0.08 (0.10)	0.23 (0.11)	0.19 (0.15)	0.19 (0.06)	0.23 (0.06)
<i>UT1</i>	0.27 (0.07)	0.47 (0.12)	0.26 (0.07)	0.34 (0.05)	0.30 (0.04)	0.29 (0.04)
<i>UX1</i>	0.26 (0.11)	0.44 (0.12)	0.25 (0.12)	0.29 (0.15)	0.26 (0.07)	0.23 (0.07)

**Note.** RR: Regular responders; CHT: Cheaters. *N* = 297 in the Training Data; *N* = 414 in the Transfer Data Set 1; *N* = 318 in the Transfer Data Set 2.

The results in **Table 3** corroborate that preknowledge inflates the test score and reduces the total testing time. The indicators of cheating do not separate the cheaters and the regular responders in all data sets equally well. In the training data set, the group specific averages of the indicators differ by more than one standard deviation. In the first transfer data set, the differences are much smaller. In the second transfer data set, there are hardly any differences. The indicators based on the responses and the indicators based on the response times have almost the same discriminatory power.

The flexible discriminant analysis performed well in the training data set. Of the 22 cheaters, 15 were identified and 7 were missed (sensitivity: 0.68). Of the 275 regular responders, 269 were correctly classified and 6 were misclassified (Specificity: 0.98). The separating boundary implied by the trained flexible discriminant model in the original predictor space (UT1-UX1 plane) is visualized in **Figure 2** on the left side. The red-dotted area represents all constellations of UT1 and UX1 that are classified as cheaters. The black-dotted area represents all constellations of UT1 and UX1 that are classified as regular responders. The separating boundary visualizes the classification rule that classifies all examinees with an UT1 index higher than 0.43 or an UX1 index higher than 0.70 as a cheater. The examinees in the learning data set are represented by dots. Black dots represent regular responders; red dots the cheaters. All black dots in the red area and all red dots in the black area are misclassified.

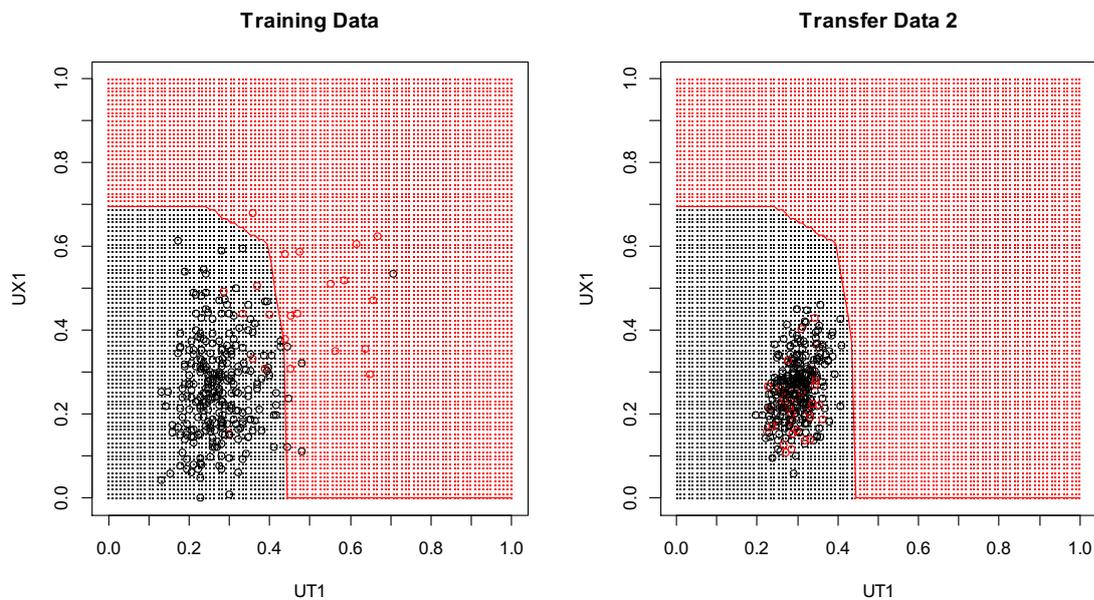
The classification rule discovered in the training data set was applied to the examinees of the first transfer data set. The results are visualized in **Figure 2** on the right side. The examinees are represented by dots. The color indicates the response mode. In the regular responders, the distribution of the UX1 and the UT1 index does not change much from the training data set to the first transfer data set. This is not surprising as the both data sets are based on the same exam. As a consequence, the classification rule maintains its high specificity after the transfer. Nine of the 397 regular responders are misclassified as cheaters (specificity: 0.98). In the cheaters, however, the distribution of the UX1 and the UT1 index changes. The values of the cheaters on the UX1 and UT1 index are generally lower in the first transfer data set than in the training data set. This reduces the separability of the groups. None of the cheaters is detected when the classification rule is transferred (sensitivity: 0.00).



**Figure 2.** Values of Examinees on two Indicators of Cheating in Two Data Sets as well as a Separating Boundary Provided by a Flexible Discriminant Analysis in Order to Separate Cheaters from Regular Responders

**Note.** Regular responders are represented as black dots, cheaters as red dots. The red line is the separating boundary derived with a flexible discriminant analysis and the training data. All examinees in the black area would be classified as regular responders and all examinees in the red area as cheaters.

The classification rule discovered in the training data set was also applied to the examinees of the second transfer data set. The results are visualized in **Figure 3** on the right. We also visualize the results for the training data set on the left for sake of comparison. In both data sets, the levels of the indicators are similar for the regular responders. The standard deviation, however, is smaller in the second transfer data set than in the training data set. The transfer of the classification rule from the training data set to the second transfer data set is successful in the sense that none of the regular responders is misclassified as a cheater (specificity: 1.00). The transferred cut-off value of the UT1 index seems to be adequate. The transferred cut-off value of the UX1 index could be lower. On average, the cheaters have lower values on the indices in the second transfer data set than in the training data set. This undermines the separation of regular responders and cheaters. The transferred classification rule is not capable to detect any cheater (sensitivity: 0.00).



**Figure 3.** Values of Examinees on two Indicators of Cheating in Two Data Sets as well as a Separating Boundary Provided by a Flexible Discriminant Analysis in Order to Separate Cheaters from Regular Responders

**Note.** Regular responders are represented as black dots, cheaters as red dots. The red line is the separating boundary derived with a flexible discriminant analysis and the training data. All examinees in the black area would be classified as regular responders and all examinees in the red area as cheaters.

## Discussion

Fraud is a serious problem that occurs in educational testing, but also in credit card transactions, insurance claims and usage of telecommunication services, for example. Recently, there has been intensive research into data mining-based fraud detection (e.g., Bolton &

Hand, 2001; Phua et al., 2010). Some of the methods developed in other fields might be useful in educational assessment, too. In educational assessment, however, the amount of data is small, at least when examinations are considered separately. This is suboptimal as data mining methods require large data sets in order to detect general patterns. A solution might consist in pooling the data from several examinations in a large data base. In this paper, we did some preliminary research on the prospects of the success of this endeavor. We investigated whether a classification rule can be generalized to different forms of cheating and different exams. The capability to transfer knowledge between exams would be the first step in implementing a general system for cheating detection. Results however were not entirely satisfactory.

We demonstrated that it is possible to separate cheaters and regular responders in some cases. In the training data set, the indicators of cheating based on the responses and the new indicators of cheating based on the response times had substantial discriminatory power. In all indicators, the mean differences between cheaters and regular responders were between one and two standard deviations. This corresponds to an effect size between  $d = 1$  and  $d = 2$  (Cohen, 1988). A flexible discriminant analysis was capable to separate cheaters and regular responders with a sensitivity of 0.68 and a specificity of 0.98.

In the first and second transfer data set, however, the indicators had low or no discriminatory power. The flexible discriminant model that worked well in the training data set was incapable to separate cheaters from regular responders in the transfer data sets. The low success in transferring the classification rule could be caused by sample selection or by concept shift. In the present case, there seems to be a concept shift from the training data set to the transfer data sets. **Figure 2** and **Figure 3** strongly suggest that there is no alternative separating boundary in the transfer data sets that separates regular responders from cheaters well, at least for the indicators of cheating that we considered. Hence, any endeavor to adapt the separating boundary from the training data set to the two transfer data sets will have little success. This questions the success of more sophisticated techniques of transfer learning (e.g., ensemble learning) proposed in the literature.

## References

- Bernardi, R., Baca, A., Landers, K., & Witek, M. (2008). Methods of cheating and deterrents to classroom cheating: An international study. *Ethics & Behavior, 18*, 373–391. <https://doi.org/10.1080/10508420701713030>
- Bolton, R., & Hand, D. (2001). Unsupervised profiling methods for fraud detection. *Proc. Credit Scoring and Credit Control, VII*, 5–7. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.24.5743>
- Burlak, G., Hernández, J., Ochoa, A., & Muñoz, J. (2006). *The use of data mining to determine cheating in online student assessment*. Paper presented at Electronics, Robotics and Automotive Mechanics Conference: CERMA'06, Cuernavaca, Mexico. <https://doi.org/10.1109/CERMA.2006.91>

- Cizek, G., & Wollack, J. (2017a). Exploring cheating on tests: The context, the concern, and the challenges. In G. Cizek & J. Wollack (Eds.), *Handbook of Quantitative Methods for Detecting Cheating on Tests* (pp. 3–19). Routledge. <https://doi.org/10.4324/9781315743097>
- Cizek, G., & Wollack, J. (Eds.) (2017b). *Handbook of quantitative methods for detecting cheating on tests*. Routledge. <https://doi.org/10.4324/9781315743097>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum.
- Fox, J.-P., & Marianti, S. (2017). Person-fit statistics for joint models for accuracy and speed. *Journal of Educational Measurement*, 54, 243–262. <https://doi.org/10.1111/jedm.12143>
- Gama, J., Zliobaite, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46, 4. <https://doi.org/10.1145/2523813>
- Hastie, T., Tibishrani, R., & Friedman, J. (2014). *The elements of statistical learning: Data mining, inference, and prediction*. Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Hastie, T., Tibishrani, R., Leisch, F., Hornik, K., & Ripley, B. (2013). *mda: Mixture and flexible discriminant analysis*. R package version 0.4-4. <http://CRAN.R-project.org/package=mda>
- Jiang, J. (2008). A literature survey on domain adaptation of statistical classifiers (Corpus ID: 1145283). University of Illinois at Urbana-Champaign. [http://www.mysmu.edu/faculty/jingjiang/papers/da\\_survey.pdf](http://www.mysmu.edu/faculty/jingjiang/papers/da_survey.pdf)
- Kim, D., Woo, A., & Dickison, P. (2017). Identifying and investigating aberrant responses using psychometrics-based and machine learning-based approaches. In G. Cizek & J. Wollack (Eds.), *Handbook of Quantitative Methods for Detecting Cheating on Tests* (pp. 71–97). Routledge. <https://doi.org/10.4324/9781315743097>
- Kingston, N., & Clark, A. (Eds.) (2014). *Test Fraud: Statistical detection and methodology*. Routledge. <https://doi.org/10.4324/9781315884677>
- Man, K., Harring, J., & Sinharay, S. (2019). Use of data mining methods to detect test fraud. *Journal of Educational Measurement*, 56, 251–279. <https://doi.org/10.1111/jedm.12208>
- Man, K., Harring, J., Ouyang, Y., & Thomas, S. (2018). Response time based nonparametric Kullback-Leibler divergence measure for detecting aberrant test-taking behavior. *International Journal of Testing*, 18, 155–177. <https://doi.org/10.1080/15305058.2018.1429446>
- Mariani, S., Fox, J., Avetisyan, M., & Veldkamp, B. (2014). Testing for aberrant behavior in response time modeling. *Journal of Educational and Behavioral Statistics*, 39, 426–451. <https://doi.org/10.3102/1076998614559412>
- Meijer, R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement*, 25, 107–135. <https://doi.org/10.1177/01466210122031957>
- Pan, S., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22, 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>

- Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). *A comprehensive survey of data mining-based fraud detection research*. <http://arxiv.org/ftp/arxiv/papers/1009/1009.6119.pdf>
- Ranger, J., Schmidt, N., & Wolgast, A. (2020). The detection of cheating on e-exams in higher education – The performance of several old and some new indicators. *Frontiers in Psychology, In Press*.
- Sijtsma, K. (1986). A coefficient of deviance of response patterns. *Kwantitatieve Methoden, 7*, 131–145.
- Sinharay, S. (2020). Detection of item preknowledge using response times. *Applied Psychological Measurement, 44*, 376–392. <https://doi.org/10.1177/0146621620909893>
- Sinharay, S. (2018a). A new person-fit statistic for the lognormal model for response times. *Journal of Educational Measurement, 55*, 457–476. <https://doi.org/10.1111/jedm.12188>
- Sinharay, S. (2018b). Application of Bayesian methods for detecting fraudulent behavior on tests. *Measurement: Interdisciplinary Research and Perspective, 16*, 100–113. <https://doi.org/10.1080/15366367.2018.1437308>
- van der Flier, H. (1977). Environmental factors and deviant response patterns. In Y. Poortinga (Ed). *Basic problems in cross-cultural psychology* (pp. 30-35). Swets and Zeitlinger.
- van der Linden, W. (2009). A bivariate lognormal response-time model for the detection of collusion between test takers. *Journal of Educational and Behavioral Statistics, 34*, 378–394. <https://doi.org/10.3102/1076998609332107>
- van der Linden, W., & Guo, F. (2008). Bayesian procedures for identifying aberrant response-time patterns in adaptive testing. *Psychometrika, 73*, 365–384. <https://doi.org/10.1007/s11336-007-9046-8>
- Wang, C., Xu, G., & Shang, Z. (2018). A two-stage approach to differentiating normal and aberrant behavior in computer based testing. *Psychometrika, 83*, 223–254. <https://doi.org/10.1007/s11336-016-9525-x>
- Weiss, K., Khoshgoftaar, T., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data, 3*, 9. <https://doi.org/10.1186/s40537-016-0043-6>
- Wollack, J., & Fremer, J. (Eds.) (2013). *Handbook of test security*. Routledge. <https://doi.org/10.4324/9780203664803>
- Zadrozny, B. (2004). *Learning and evaluating classifiers under sample selection bias*. Paper presented at the ICML '04: Proceedings of the twenty-first international conference on Machine learning, Banff, Canada. <https://doi.org/10.1145/1015330.1015425>

## Zusammenfassung

Der Begriff Datenforensik bezeichnet in der Psychometrie das Arbeitsgebiet, das die Detektion von Betrugsversuchen bei Tests oder Instrumenten der Leistungsdiagnostik umfasst. Die Detektion beruht dabei auf der Identifikation von auffälligen Antwortmustern und außergewöhnlichen Leistungssprüngen sowie der Nutzung von während der Klausur aufgezeichneten Verhaltensdaten (Antwortlatenzen, Antwortrevisionen, Blickbewegung). Letzteres setzt eine elektronische Testdurchführung voraus. Ein neuer Ansatz in der Datenforensik ist die Anwendung von Techniken aus dem Bereich des maschinellen Lernens. Dabei werden Algorithmen trainiert, anhand von Test- und Verhaltensdaten ehrliche Klausurteilnehmer von unehrlichen zu unterscheiden. Die erfolgreiche Anwendung von Techniken des maschinellen Lernens erfordert jedoch weitaus größere Datensätzen, als sie im akademischen Kontext üblicherweise verfügbar sind. Eine Lösung dieses Problems könnte darin bestehen, bereits vorhandenes Wissen über die Detektion von Betrugsversuchen auf neue Detektionsaufgaben anzuwenden. In diesem Artikel wird untersucht, ob ein Algorithmus, der mit Hilfe eines ersten Datensatzes trainiert wurde, auf einen neuen Datensatz übertragen werden kann. Hierzu wurde eine Untersuchung durchgeführt, welche aus zwei Phasen bestand. In der Trainingsphase wurde ein flexibles Diskriminanzmodell mit Hilfe eines ersten Datensatzes darauf trainiert, unehrliche Klausurteilnehmenden zu identifizieren. Die zum Training verwendeten Daten (Trainingsdatensatz) bestanden aus den Antworten und Antwortlatenzen von Teilnehmenden einer E-Klausur. Zudem waren im Trainingsdatensatz einzelne Teilnehmende markiert, welche vor oder während der Klausur unzulässigen Zugang zu Klausuraufgaben oder Klausurlösungen hatten. Diese Teilnehmenden bildeten die Klasse der unehrlichen Teilnehmenden, welche vom Diskriminanzmodell erkannt werden sollten. Die Antworten und Antwortlatenzen der Teilnehmende wurden zur weiteren Analyse zu Indikatoren kombiniert. Auf Grundlage der Indikatoren wurde dann durch das Diskriminanzmodell eine Regel gebildet, wie ehrliche von unehrlichen Teilnehmenden zu unterscheiden seien (Überwachtes Lernen). In der anschließenden Transferphase wurde die anhand des Trainingsdatensatzes erlernte Regel auf einen neuen Datensatz (Transferdatensatz) angewandt. Der neue Datensatz bestand ebenfalls aus den Antworten und Antwortlatenzen von Teilnehmenden einer E-Klausur. Form und Inhalt der zweiten E-Klausur (Transferdatensatz) unterschied sich jedoch deutlich von Form und Inhalt der ersten E-Klausur (Trainingsdatensatz). Jeder Teilnehmende der zweiten Klausur wurde durch das angepasste Diskriminanzmodell und die dadurch repräsentierte Regel als ehrlichen oder unehrlichen Teilnehmenden klassifiziert. Auch im zweiten Datensatz war bekannt, ob ein Teilnehmender unzulässigen Zugang zu Klausuraufgaben oder Klausurlösungen hatte. Diese Information wurde jedoch nur zur Validierung der Klassifikation genutzt. Im Trainingsdatensatz konnte das Diskriminanzmodell die ehrlichen Teilnehmenden von den unehrlichen Teilnehmenden relativ gut unterscheiden (Sensitivität: 0.68 / Spezifität: 0.98). Der Transfer des Diskriminanzmodells und die dadurch repräsentierte Regel auf den zweiten Datensatz war jedoch weniger erfolgreich. Zwar war die Spezifität auch bei den Transferdaten hoch. Die Sensitivität jedoch fiel extrem gering aus. Dieser Befund legt nahe, dass Algorithmen zur Detektion von Betrugsversuchen nicht einfach von einer Klausur auf eine andere Klausur übertragen werden können. Die Ursache hierfür scheint in einer Bedeutungsveränderung der Indikatoren zu liegen. Was bei einer Klausur auffällig ist, muss bei ei-

ner anderen Klausur nicht zwangsläufig indikativ für unehrliches Verhalten sein. Mathematisch bedeutet dies, dass die Beziehung zwischen den Indikatoren und der Auffälligkeit des Antwortverhaltens nicht Kontext invariant ist. Dieses Phänomen wird im Data Mining als Konzept Drift bezeichnet. Das Vorliegen eines Konzept Drifts erschwert den Transfer von Wissen und Regeln über verschiedene Kontexte erheblich.

**Jochen Ranger** promovierte in Psychologie und ist wissenschaftlicher Mitarbeiter in der Abteilung Methodenlehre am Institut für Psychologie der Martin-Luther-Universität Halle-Wittenberg. Sein Forschungsschwerpunkt liegt im Bereich der Psychometrie. Von ihm stammen Arbeiten zur Bewertung des Fits latenter Traitmodelle und zur Modellierung von Reaktionszeiten bei psychologischen Tests.

[jochen.ranger@psych.uni-halle.de](mailto:jochen.ranger@psych.uni-halle.de)

**Anett Wolgast** ist Professorin für Psychologie, mit dem Schwerpunkt Pädagogische Psychologie, an der Fachhochschule des Mittelstands in Hannover. In ihrer Forschung befasst sie sich mit Zusammenhängen zwischen Leistungsanforderungen und sozialem Verhalten in realen und digitalen Bildungskontexten. Ein Fokus ihrer Forschung liegt auf der Bereitschaft zur sozialen Perspektivübernahme und ihrer Bedeutung für das akademische und soziale Lernen in Schulen und Hochschulen.

[anett.wolgast@gmail.com](mailto:anett.wolgast@gmail.com)

**Nico Schmidt** ist Masterstudent an der Martin-Luther-Universität Halle-Wittenberg. Seine Forschungsinteressen liegen im Bereich der Führungseffizienz. Er ist Experte für die Softwareumsetzung experimentalpsychologischer Experimente.

[nico.schmidt2@student.uni-halle.de](mailto:nico.schmidt2@student.uni-halle.de)

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung



ZENTRUM FÜR MULTIMEDIALES  
LEHREN UND LERNEN

Gemeinsames Bund-Länder-Programm für bessere Studienbedingungen und mehr Qualität in der Lehre. Dieses Vorhaben wird aus Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01PL17065 gefördert.