

Entwicklung und Validierung des Modularen Fragebogens zur Evaluation digitaler Lehr-Lern-Szenarien (MOFEDILLS)

Henrike Kärchner, Maren Gehle, Malte Schwinger

Zusammenfassung: Spätestens seit der Corona-Pandemie gehören digital gestützte Lehrveranstaltungen zum Alltag an deutschen Hochschulen. Im Zuge der Qualitätsentwicklung von Hochschullehre sollten Besonderheiten und Spezifika digitaler Lehre (z.B. räumliche Flexibilität) bei der Evaluation derartiger Lehrveranstaltungen angemessen berücksichtigt werden. Bisher verwendete Evaluationsinstrumente leisten dies nicht und bleiben zudem häufig einen Nachweis ihrer theoretischen Grundlage und psychometrischen Qualität schuldig. In der vorliegenden Studie werden daher die Entwicklung und psychometrischen Eigenschaften des eigens für digitale Lehr-Lernsettings konzipierten modularen Fragebogens MOFEDILLS vorgestellt. Evaluationsdaten aus vier hessischen Hochschulen (N = 1,604) stützen die angenommene faktorielle Struktur und Messinvarianz sowie Objektivität, Reliabilität und Validität des neuen Instruments.

Schlüsselwörter: Evaluation, digitale Lehre, Hochschule, psychometrische Eigenschaften, selbstreguliertes Lernen

Development and validation of the Modular Questionnaire for the Evaluation of Digital Teaching-Learning Scenarios (MOFEDILLS)

Summary: Ever since the Corona pandemic, digitally supported lectures have become part of everyday life at German universities. In the course of developing the quality of university teaching, the special features and specifics of digital teaching (e.g., room flexibility) should be adequately taken into account in the evaluation of such courses. Evaluation instruments used to date do not do this and also often fail to provide evidence of their theoretical basis and psychometric quality. This study therefore presents the development and psychometric properties of the modular questionnaire MOFEDILLS, which was designed specifically for digital teaching-learning settings. Evaluation data from four Hessian universities (N = 1,604) support the assumed factorial structure and measurement invariance as well as objectivity, reliability, and validity of the new instrument.

Keywords: evaluation, digital learning, university, psychometric properties, self-regulated learning

1 Einleitung

Zum Sommersemester 2020 erfolgte an den Universitäten eine rasante Umstellung von Präsenzlehrbetrieb auf digitale Distanz-Lehr-Lernformate (Dittler & Kreidl, 2021). Diese disruptive Entwicklung geschah zu einer Zeit, in der angesichts der bis dato ohnehin fortschreitenden Digitalisierung bereits seit längerem diskutiert wurde, inwiefern neue Anforderungen

an die Evaluation universitärer Lehre zu stellen sind (Gilch et al., 2019; Mitterauer, Pohlentz & Harris-Huemmert 2019). In Reaktion hierauf waren an verschiedenen Standorten Bemühungen zur Anpassung von Lehrevaluationen an die digitale Hochschullehre erkennbar (Harris-Hummert, Pohlentz & Mitterauer, 2018). Allerdings liegt unserer Kenntnis nach bislang kein universelles, theoretisch fundiertes und open-access basiertes Evaluationsinstrument zur Befragung von Studierenden vor, welches eigens für digitale Kontexte konstruiert wurde und situationsübergreifend digitale Lehr-Lernszenarien sowie deren relevante Outcomes (z.B. Lernzuwachs, selbstreguliertes Lernen) analysiert.

Zwar können mit den meisten vorhandenen Instrumenten diverse Fragestellungen im Rahmen der Qualitätssicherung der Lehre beantwortet werden, jedoch werden theoretisch und praktisch relevante Fragen im Bereich digitaler Lehre kaum berücksichtigt. Diese reichen von allgemeinen Wirkungsaspekten wie zeitliche und räumliche Flexibilität über spezifische Umsetzungsformate wie z.B. virtuelle Exkursionen bis zu der Frage, wie einzelne Bausteine digitaler Lehre sich sinnvoll zu einer neuen Lehrveranstaltung zusammensetzen lassen (z.B. Produktion eigener Lernvideos in Flipped-Classroom-Settings). Die Anforderung an ein modernes, an digitaler Lehre orientiertes Evaluationsinstrument besteht somit darin, zum einen durch spezifische Evaluationsmodule in Bezug auf einzelne Aspekte digitaler Lehre (z.B. Virtual Reality, Abstimmungssysteme, Videonutzung) neue Entwicklungen in diesem Bereich evaluativ begleiten zu können, zum anderen aber durch einen diese speziellen Formate verbindenden allgemeinen Evaluationsbereich die enorme Heterogenität an digitalen Lehrformaten abzubilden und vergleichbar machen zu können. Für diesen allgemeinen Teil bedarf es einer Fokussierung auf generelle Gelingensbedingungen digitaler Lehre, welche sich nur teilweise mit denen analoger Lehre überschneiden (vgl. Ulrich, 2020). Ein weiterer Schwachpunkt der bisherigen Forschung und evaluativen Praxis bezieht sich auf die häufig nur schwer beurteilbare Qualität bestehender Evaluationsinstrumente (Zumbach, Spinath, Schahn, Friedrich & Kögel, 2007). Gründe dafür sind u.a. ein stark limitierter Zugriff auf die Instrumente sowie ein gravierender Mangel an Veröffentlichungen zu ihrer Konstruktion und psychometrischen Qualität (vgl. Rindermann, 2003; Mittag, Mutz & Daniel, 2012). Vor dem Hintergrund der geschilderten Defizite und Herausforderungen besteht das Ziel der vorliegenden Arbeit darin, die Entwicklung und Konstruktion des Instruments *Modularer Fragebogen zur Evaluation von digitalen Lehr-Lern-Szenarien* (MOFEDILLS) zur Bewertung digitaler und teildigitalisierter Lehrveranstaltungen durch Studierende transparent und detailliert darzustellen. MOFEDILLS und alle verfügbaren Informationen über die Qualität des Instruments stehen allen potentiellen Anwender*innen unentgeltlich zur Verfügung (Kärchner, Gehle & Schwinger, 2020: <https://fragebogen-hochschullehre.de>). Im zweiten Teil des Artikels stellen wir die psychometrischen Eigenschaften des neuen Instruments vor, welche anhand einer großen Stichprobe von Studierenden aus vier hessischen Hochschulen bestimmt werden konnten. Ergänzend werden Zusammenhänge zwischen postulierten Gelingensbedingungen digital gestützter Lehre und wahrgenommenen Lernergebnissen berichtet.

2 Digitale Lehre: Konzeptualisierung und Befunde

Allgemein charakterisiert digitale Lehre, dass digitale Technologien in unterschiedlichen Ausprägungen und Dimensionen in Lehr-Lernangebote inkludiert werden¹. Besonders für Forschungs- und Evaluationsaktivitäten stellt jedoch die enorme Begriffsvielfalt im Bereich digitaler Lehre (u.a. M-Learning, E-Learning, E-Teaching, Blended Learning) und deren rasante Weiterentwicklung (Wannemacher, Jungermann, Scholz, Tercanli, & Villiez, 2016) eine grundlegende Herausforderung dar. In Anbetracht der zunehmend starken Verbreitung digitaler Lehr-Lernformen in der Lehrpraxis (vgl. Bauer et al., 2020; Getto, Hintze, & Kerres, 2018) und des großen Spektrums digitaler Lehr-Lernformate existieren vergleichsweise wenig zugängliche und aggregierte Erkenntnisse darüber, wie vollständig oder teilweise digitalisierte Lehrveranstaltungen (und deren Bestandteile) von Studierenden wahrgenommen werden (vgl. Dittler & Kreidl, 2020; Traus, Höffken, Thomas, Mangold & Schröer, 2020). Weitgehend unbekannt ist zudem, welche Merkmale qualitativ hochwertige digitale Lehrveranstaltungen aufweisen (sollen) und mit welchen lern- und leistungsrelevanten Outcomes diese assoziiert sind (vgl. Lehner & Sohm, 2021).

Durch die enorme Heterogenität digitaler Lernformate und den daraus entstehenden vielfältigen didaktischen und methodischen Kombinationsmöglichkeiten erscheint es schwierig, allgemeine Merkmale und Gelingensbedingungen erfolgreicher digital gestützter Lehre abzuleiten. Bisherige Studien zur Analyse digital gestützter Lehrveranstaltungen beziehen sich überwiegend auf einzelne Tools wie web- oder appbasierte Quiz-Anwendungen (z.B. Andrés, Sanchis & Poler, 2015) oder einzelne Lehrformate und -formen wie Inverted bzw. Flipped Classroom (z.B. Handke & Sperl, 2012). Viele solcher Best-Practice-Beispiele wurden bisher überwiegend anhand sehr variierender Stichprobengrößen in unterschiedlichen Hochschultypen, Fachdisziplinen und Ländern in meist speziellen hybriden oder multimodalen Lernsettings implementiert. Inwiefern dabei eine Generalisierung der Ergebnisse auf andere Kontexte und Zielgruppen möglich ist sowie insgesamt eine Vergleichbarkeit von Lehrevaluationsergebnissen gegeben ist, bleibt fraglich (vgl. Zumbach et al., 2007). Basierend auf dem aktuellen Stand zur analogen und digitalen Lehrforschung und theoretischen wie ökonomischen Überlegungen werden im Folgenden die für den hier präsentierten Fragebogen hergeleiteten verschiedenen Gelingensbedingungen betrachtet, welche in dieser Arbeit anhand von Evaluationsdaten erstmals validiert werden.

3 Allgemeine Gelingensbedingungen und Merkmale qualitativ hochwertiger digitaler Lehre

In der Literatur existieren verschiedene Typologien (z.B. Staufenbiel, 1999) und darauf aufbauende Modelle (z.B. Rindermann, 2001, 2003; Ullrich, 2020), die versuchen, alle potenziellen Faktoren und deren Kriterien, die qualitativ hochwertige (analoge) Hochschullehre charakterisieren und beeinflussen könnten, zu identifizieren und abzubilden. Das Vier-Kompo-

1 Dieser breiten Definition folgt auch die Konstruktion von MOFEDILLS, was u.a. die Aufteilung in einen allgemeinen Teil und einen Teil mit spezifischen Modulen für verschiedene digitale Lernformate und -tools begründet (s. Abschnitt 6).

nentenmodell nach Rindermann (2001, 2003) beschreibt multidimensionale Beziehungen verschiedener Bestandteile der Lehre als drei verschiedene Gelingens- bzw. Bedingungsfaktoren (Dozierende, Studierende und Kontextvariablen) und den Lehrerfolg (vgl. auch Ulrich, 2020). Allerdings bleibt im Detail die Evidenz der einzelnen Bestandteile der Bedingungsfaktoren sowie ihrer häufig nicht weiter spezifizierten Zusammenhänge untereinander schwer zu beurteilen, und es ist unklar, inwiefern eine umfassende empirische und systematische Testung des gesamten Modells bereits erfolgt ist. Ebenso liegen keine Erkenntnisse über die Gewichtung der aufgelisteten einzelnen Bedingungsfaktoren nach Rindermann (2001, 2003) vor. Dennoch ermöglicht dieses Modell eine heuristische Abbildung der Multidimensionalität digitaler Hochschullehre, welche durch spezifische Erkenntnisse zu digitalen Lehr-Lernsettings modifiziert und erweitert werden kann, weshalb wir es als theoretische Grundlage für MOFEDILLS verwenden (vgl. Abbildung 1).

In unserem adaptierten Rahmenmodell entsteht eine erfolgreiche digital gestützte Lehrveranstaltung aus dem wechselseitigen Zusammenspiel der vier Determinanten „Studierende“ (u.a. Mitarbeit, Fehlzeiten), „Lehrende“ (u.a. Gesamturteil mit Lehrperson und Kommunikation), „Rahmenbedingungen der Lehrveranstaltung“ (u.a. Art, Tempo und Stoffumfang der Veranstaltung) und „Gelingensbedingungen digital gestützter Lehre“, welche den resultierenden Lernerfolg der Studierenden (Lernzuwachs, Gesamturteil der Veranstaltungsqualität, Partizipation und selbstreguliertes Lernen einschließlich Motivation) beeinflussen sollen.

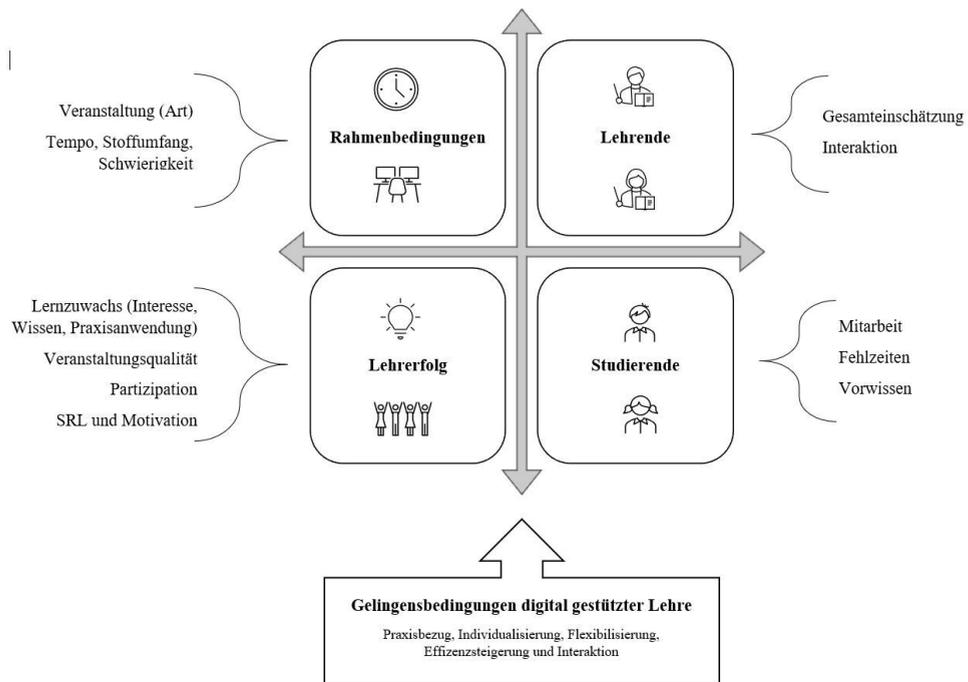


Abbildung 1: Adaptiertes multidimensionales Faktorenmodell des Lehrerfolgs im digitalen Lehr- und Lernkontext (angelehnt an Rindermann, 2003)

Diese Aspekte stehen jedoch nicht für sich allein, sondern hängen miteinander zusammen. Beispielsweise hängt das Ausmaß an studentischer Mitarbeit nicht nur von individuellen Interessen der Studierenden ab, sondern wird auch maßgeblich durch die Lehrqualität, also passende Instruktionen, Feedback etc. beeinflusst (Hattie, 2009; Ulrich, 2020). In besonderem Maße zeigen Studien, wie bedeutsam die Herstellung von sozialer Interaktion zwischen Studierenden untereinander, aber auch mit dem Dozierenden, für den hochschulischen Lernerfolg ist (Schneider & Preckel, 2017). Ergänzend wird häufig verschiedenen Kontextvariablen von Veranstaltungen (z.B. Thema, Fächer, Veranstaltungstyp, Gruppengröße) eine beeinflussende Rolle auf die Veranstaltungsbeurteilung durch Studierende zugeschrieben. Zusätzlich werden externe Faktoren und Rahmenbedingungen wie beispielsweise Ausbildungsinhalte und Prüfungsordnungen der jeweiligen Studiengänge sowie die Ausstattung von Technik und der Bibliotheken beschrieben, die insgesamt ebenfalls mit einer erfolgreichen Hochschullehre assoziiert zu sein scheinen (Rindermann, 2003; Staufenbiel, 2001). Wenn gleich ein Konsens darüber besteht, dass solche externen Faktoren (z.B. Art der Veranstaltung, Tempo der Veranstaltung) wichtig sind und einen Einfluss auf die Qualität der Lehrveranstaltung haben können, existieren wenig empirisch aussagekräftige Befunde dazu, in welcher Form und in welchem Ausmaß sie im Vergleich zu anderen Faktoren zum Erfolg der Lehrveranstaltung beitragen. Insgesamt versuchen Autor*innen solcher Modelle und Ansätze, sich der komplexen Realität der Hochschullehre möglichst weit anzunähern. Inwiefern sich bisherige Perspektiven und Ansätze zur Charakterisierung „guter“ analoger Hochschullehre (Schneider & Preckel, 2017; Ulrich, 2020) auf digitale Lernkontexte an Hochschulen übertragen lassen, wurde unserer Kenntnis nach bislang nicht umfassend untersucht.

Zahlreiche Arbeiten fokussieren auf einzelne spezifische Besonderheiten und Vorteile digitaler Lehre (z.B. Individualisierung von Lernprozessen und Vielfalt der Lernangebote, Henderson, Selwyn & Aston, 2017). Davon ausgehend und basierend auf dem adaptierten Komponentenmodell in Abbildung 1 sowie den identifizierten Dimensionen „Interaktion“ und „Individualisierung“ digitalisierter Lernelemente und -formate nach Wannemacher et al. (2016) lassen sich insgesamt häufig postulierte und zentrale Gelingensbedingungen (GB) digitaler Lehrangebote ableiten. Zu diesen zählen die räumliche und zeitliche Flexibilität beim Lernen (GB Flexibilität), die Möglichkeit, selbstbestimmt zu lernen (GB Individualisierung), Kontakt- und Austauschmöglichkeiten zwischen verschiedenen Beteiligten (GB Interaktion) sowie die Verfügbarkeit von Materialien und Technik und die Abwesenheit von potenziell damit verbundenen Störungen (GB Effizienzsteigerung). Ausgehend von der Erwartungswert-Theorie (Eccles & Wigfield, 2020) kann zudem die durch Studierende wahrgenommene persönliche Nützlichkeit der Veranstaltung als weitere Gelingensbedingung betrachtet werden genauso wie die für Studierende besonders relevanten Praxisbezüge (Bargel, Müßig-Trapp & Willige, 2007) in den Veranstaltungen (GB Praxisbezug). Demnach sollte eine gelungene und durch die Studierenden positiv eingeschätzte digital gestützte Veranstaltung so gestaltet sein, dass Studierende flexibel (örtlich und zeitlich), individualisiert (Tempo, Art und Weise, Anknüpfung an Vorwissen) und effizient (Bereitstellung von Materialien, keine Störungen) lernen können. Zudem sollen die vermittelten Inhalte einen Praxis- und Forschungsbezug aufweisen und von den Studierenden als nützlich wahrgenommen werden. Ebenso sollen die eingesetzten (digitalen) Lerntools, -werkzeuge usw. die Zusammenarbeit unter den Studierenden, deren Zusammenarbeit sowie den Austausch mit Dozierenden fördern und erleichtern. Je stärker die Gelingensbedingungen in einer Lehrveranstaltung berücksichtigt werden und von den Studierenden als gegeben wahrgenommen werden, desto moti-

vierter und selbstregulierter sollten sich die Studierenden wahrnehmen und ebenso eine Zunahme des Wissens und des Interesses bezüglich der Themen der Veranstaltung sowie der Fähigkeit, die Veranstaltungsinhalte anzuwenden, feststellen. Ob sich die genannten Gelingensbedingungen erfolgreicher digitaler Lehre empirisch abbilden lassen und mit der Gesamtzufriedenheit einer digitalen Lehrveranstaltung assoziiert sind, wurde unseres Wissens nach bislang nicht empirisch geprüft.

Die beschriebenen Aspekte spielen zwar möglicherweise auch in traditioneller Lehre eine bedeutsame Rolle, allerdings wären hier eventuell andere Gelingensbedingungen noch zentraler, wie z.B. die didaktische Instruktion (Schneider & Preckel, 2017; Ulrich, 2020). Zudem ist an dieser Stelle relevant, dass sich zwar häufig die Ziele einer Veranstaltung in Präsenz vs. digital ähneln dürften, aber der Weg zur Erreichung dieser Ziele anders bzw. schwerer oder einfacher sein sollte. So dürfte es beispielsweise in einer Lehrveranstaltung zu den Themen Beratung und Mediation in Präsenz einfacher sein, anhand von Rollenspielen etc. den Praxisbezug des Stoffes zu verdeutlichen, während das gleiche Ziel über digitale Kanäle sicher schwerer zu realisieren ist (vor allem die nicht automatisch gegebene soziale Interaktionsmöglichkeit bei digitaler Lehre macht Gelingensbedingungen wie Praxisbezug und Interaktion daher auch so zentral). Im Gegenzug lässt sich durch digitale Optionen wie Vorlesungsvideos, die man jederzeit pausieren kann, wiederholen kann und zu jeder Tages- und Nachtzeit anschauen kann, der Grad an Autonomie und Individualisierung beim Lernen deutlich leichter steuern als in traditionellen Präsenzvorlesungen – in beiden Settings ist das Autonomieerleben im Lernprozess gleichermaßen zentral für die Herstellung und Aufrechterhaltung intrinsischer Motivation (Ryan & Deci, 2017).

4 Lern- und Leistungsergebnisse digitaler Lehrveranstaltungen

Generell stellt der Lehrerfolg operationalisiert als Lernerfolg oder Lerngewinn von Lernenden einen wichtigen Indikator der zu ermittelnden Lehrqualität dar (vgl. Metz-Göckel, Kamphans & Scholkmann, 2012). Lernerfolg wird in Fragebögen unterschiedlich erfasst, teils allgemein über die absolute Menge gelernter Inhalte und ihren Zugewinn, teils detaillierter in Bezug auf Thema und Lernziele der Veranstaltung (z.B. Zumbach et al., 2007, FEVOR: Staufenbiel, 2001). Zumeist wird bei der Messung des Lernerfolgs der Studierenden auch die Wissenszunahme (z.B. MoGLi-K: Treppesch, Hense & Raser, 2015) nach der Veranstaltung erfasst. Lernerfolg wird in dem hier vorgestellten Fragebogen im Sinne einer Entwicklungsperspektive über die Zunahme des eigenen Wissens und Interesse über die behandelten Themen und die Fähigkeit, die Inhalte der Veranstaltung in der Praxis anzuwenden, konzeptualisiert (Lernzuwachs).

Neben dem Erzielen von Lernzuwachs ist ein weiteres erstrebenswertes Ziel in Lehrveranstaltungen die Einübung und Förderung selbstregulierten Lernens (vgl. Peverly, Brobst, Graham & Shaw, 2003; Schober et al., 2016). Studierende unterscheiden sich dahingehend, inwieweit sie den komplexen Prozess des Wissenserwerbs und dessen Abruf selbstständig gestalten können. Dabei steuern sie aktiv und selbstständig ihre Kognitionen, Gefühle und motivationalen Einstellungen und richten ihr Verhalten anhand ihrer persönlichen Ziele und Umweltbedingungen aus (Zimmermann, 2000). Diese Fähigkeit stellt eine Grundvoraussetzung für das Erreichen von eigenen Zielen im Hochschulkontext dar und ist eindeutig mit

Erfolg im Studium assoziiert (z.B. Boekaerts, 1997; Cassidy, 2011). Daher wird in unserem Fragebogen selbstreguliertes Lernen (u.a. Motivation, Planung, Self-monitoring als Komponenten der Selbstregulation nach Schmitz, 2001; Schmitz & Schmidt, 2007) als weiterer wichtiger Leistungsoutcome digitaler Lehrveranstaltungen berücksichtigt. Als zentralen Bestandteil des selbstregulierten Lernens betonen zahlreiche Studienergebnisse die Bedeutsamkeit der akademischen Motivation für erfolgreiches Lernen (z.B. Boekaerts, 1997; Hattie, 2009; Pintrich, 2000). Generell hängen Motivation und Lernerfolg überwiegend indirekt (z.B. vermittelt über Kurswahlen) zusammen (z.B. Steinmayr, Weidinger, Schwinger & Spinath, 2019). So dürfte es motivierten Studierenden im Gegensatz zu weniger motivierten, aber gleich intelligenten Kommiliton*innen leichter fallen, sich über den gesamten Veranstaltungszeitraum mit den Lehrinhalten zu beschäftigen. Zusätzlich werden sie dabei mutmaßlich mehr Freude und Sinnhaftigkeit empfinden, was insgesamt mit einer positiven Lernerfahrung einhergeht. Dadurch dürfte wiederum die Bereitschaft der motivierten Studierenden steigen, sich anzustrengen, sich neues Wissen anzueignen und die erlernten Inhalte in einer weiteren freiwillig gewählten Lehrveranstaltung zu vertiefen. Durch die Berücksichtigung verschiedener Komponenten selbstregulierten Lernens (z.B. Motivation) als Lehrveranstaltungoutcome erhalten Lehrende differenziertere Rückmeldungen und können davon ausgehend Optimierungsbedarf für Ihre Lehrveranstaltung ableiten.

Während zahlreiche Studien die Relevanz selbstregulierten Lernens unterstreichen, liegen kaum Arbeiten vor, welche die Bedeutung von Partizipation im Kontext der Hochschullehre adressieren. Anwesenheit und aktive Mitarbeit in Veranstaltungen gilt traditionell als uneingeschränkt positiv und wird in vielen Lehrevaluationsinstrumenten standardmäßig miterfasst (z.B. Staufienbiel, 2000; MoGLi-K: Treppesch et al., 2015). Demnach ist erfolgreiches Lernen wahrscheinlicher, wenn die Studierenden – insbesondere in teildigitalisierten Lernsettings – einen Großteil der Sitzungen anwesend waren (Partizipation) und die Gestaltung der Veranstaltung eine aktive Auseinandersetzung mit den Lerninhalten förderte. Wir nehmen somit an, dass sich Studierende bei einer gelungenen und durch sie positiv eingeschätzten Veranstaltung aktiv mit den Inhalten auseinandersetzen und den Sitzungen beiwohnen.

5 Evaluationsinstrumente für digitale Lehrveranstaltungen

An deutschen Hochschulen werden im Rahmen von Lehrevaluationen zahlreiche Feedbackinstrumente zur Befragung von Studierenden (z.B. TRIL; vgl. Gollwitzer, Kranz & Vogel, 2006) eingesetzt. Diese ursprünglich für Präsenzlehre entwickelten Instrumente unterscheiden sich u.a. hinsichtlich ihres Konstruktionshintergrundes, ihres Umfangs und Untersuchungsgegenstands sowie – sofern überhaupt bekannt – ihrer psychometrischen Qualität (Zumbach et al., 2007). Einen vollständigen Überblick über existierende Instrumente zu erhalten ist jedoch kaum möglich, da nicht alle Verfahren an den jeweiligen Standorten zugänglich und einsehbar sind. Seit einiger Zeit werden zunehmend Evaluationsinstrumente in modularisierter Form angeboten, um flexibler die Bedarfe unterschiedlicher Anwender*innen und Kontexte (z.B. Vorlesung vs. Seminar) berücksichtigen zu können (z.B. MoGLi-K, Treppesch et al., 2015; LVE Fragebögen der Goethe-Universität Frankfurt, o. D.).

Mit Hilfe modular aufgebauter Instrumente können vor allem Besonderheiten (z.B. spezielle Veranstaltungsdesigns) oder neuartige Entwicklungen (z.B. E-Learning) einfacher als

bisher in vorhandene Instrumente integriert werden. Viele der ursprünglich für den traditionellen Lehrkontext konstruierten Fragebögen (z.B. FEVOR, Staufenbiel 2000, 2001) wurden spätestens im Zuge der Corona-Pandemie um einzelne Items oder Module zur digitalen Lehre erweitert (z.B. Referat Lehrevaluation an der Philipps-Universität Marburg, 2021; MoGli – Modul "E-Learning": Treppesch et al., 2015). Bislang mangelt es jedoch an psychometrisch überprüften Fragebögen, die zur Evaluation von digitalen Lehr-Lernangeboten geeignet sind und die versuchen, Weiterentwicklungen der Lehre in den letzten Jahren adäquat und flexibel abzubilden. Aus diesem Grund entwickelten wir den Modularen Fragebogen zur Evaluation von digitalen Lehr-Lern-Szenarien (MOFEDILLS) zur Analyse und Ermittlung der studentischen Wahrnehmung digitaler und teildigitalisierter Lehr- Lernveranstaltungen und stellen ihn in dieser Arbeit detailliert vor.

6 Modularer Fragebogen zur Evaluation digitaler Lehr-Lern-Szenarien (MOFEDILLS)

Das multidimensionale und modularisierte Instrument MOFEDILLS² ist für die Evaluation von teilweise oder vollständig digitalisierten Lehr-Lernszenarien im Hochschulkontext (z.B. Vorlesungen, Seminare, Praktika, Lehrprojekte) entwickelt worden. Zugrunde liegt dem Instrument eine breite und allgemein gültige Begriffsdefinition digitaler Hochschullehre, welche die Integration digitaler Technologien jeglicher Ausprägung und Dimension in Lehr-Lernangebote umfasst. Dieser breiten Definition folgend würden somit z.B. online bereitgestellte, asynchron abrufbare Vorlesungsformate, welche spätestens seit der Corona-Pandemie flächendeckend eingeführt wurden, genauso unter diesen Begriff fallen wie hoch spezialisierte Seminarformate im Bereich Virtual Reality. Die Idee von MOFEDILLS besteht darin, zum einen durch spezifische Evaluationsmodule in Bezug auf einzelne Aspekte digitaler Lehre (z.B. Virtual Reality, Abstimmungssysteme, Videonutzung) neue Entwicklungen in diesem Bereich evaluativ begleiten zu können, zum anderen aber durch einen diese speziellen Formate verbindenden allgemeinen Evaluationsbereich die enorme Heterogenität an digitalen Lehrformaten abzubilden und vergleichbar machen zu können. Diese doppelte Funktion von MOFEDILLS ermöglicht es Lehrenden sowohl kriteriales Feedback zur eigenen, spezifisch ausgearbeiteten Lehrveranstaltung zu erhalten (Beispielitem Virtual Reality: „Der Einsatz der VR hat mir geholfen, komplexe Zusammenhänge zu erfahren.“) als auch eine Einordnung im sozialen Vergleich mit anderen digitalen Lehrformaten vornehmen zu können.

Der Fragebogen für Studierende gliedert sich in zwei Teile: die Basismodule und die spezifischen Module für verschiedene digitale Lernformate und -tools (s. Abbildung 2). Die Basismodule erfassen mit insgesamt 38 Items (a) allgemeine Beurteilungen von Lehr-Lernszenarien (Gesamtbeurteilung der Veranstaltung), (b) Gelingensbedingungen digitaler Lehrangebote (Individualisierung, Flexibilisierung, Praxisbezug, Effizienzsteigerung und Interaktion) und (c) resultierende Lern- und Leistungsergebnisse (Lernzuwachs, Partizipation, Selbstreguliertes Lernen). Im zweiten Teil des Fragebogens stehen einzelne Module zur de-

2 Electronic Supplementary Materials (ESM) wie alle Items des eingesetzten Fragebogens, ergänzende psychometrische Kennwerte, spezifische Zusammensetzungen der Stichprobe etc. können unter https://osf.io/4b3vf/?view_only=1ca25289396345faa23e693f57db36ab oder unter <http://is.gd/mofedills> abgerufen werden.

taillierten Analyse von verbreiteten Bestandteilen digitaler Lern-Lehrsznarien zur Verfügung (z.B. Selbstlernmodule, Videos, Abstimmungssysteme, Game-Based-Learning). Zusätzlich unterschieden wird bei einigen Bausteinen des zweiten Teils, ob beispielsweise Videos zum Lernen in der Veranstaltung lediglich eingesetzt (Konsum) oder auch von den Studierenden selbst erstellt (Produktion) wurden. Die einzelnen Bausteine enthalten jeweils zwischen 6 und 15 Items. Die Items werden auf einer 5-stufigen Likertskala (*trifft gar nicht zu* bis *trifft voll zu*) oder entsprechenden 5-stufigen semantischen Differentialen (z.B. „Die Schwierigkeit dieser Veranstaltung war ... viel zu gering [1] ... viel zu hoch [5]) beantwortet.

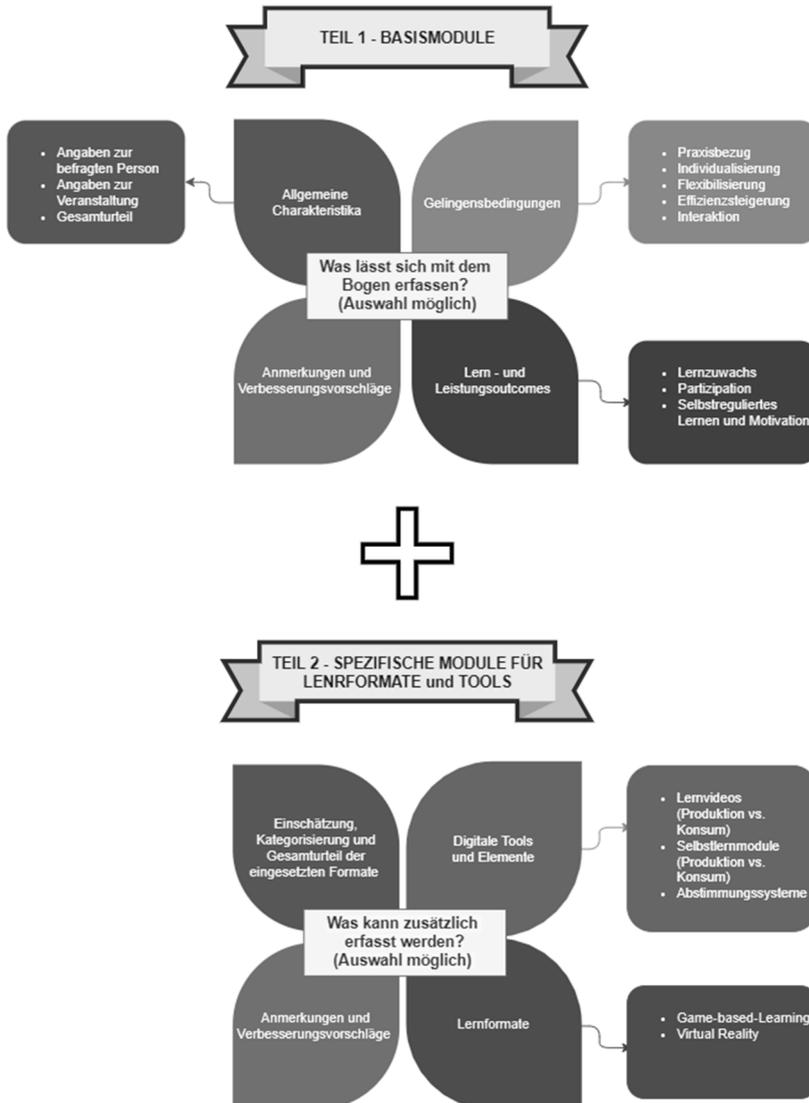


Abbildung 2: Module und Inhalte des Fragebogens zur Evaluation von digitalen Lehr-Lern-Szenarien

Somit können Lehrende den ersten Teil des Fragebogens vollständig nutzen und je nach Bedarf kein weiteres, ein oder mehrere Module des zweiten Teils verwenden. Die Auswahl der Module können sie passend zu den in ihrer Veranstaltung eingesetzten spezifischen digitalen Lernformaten, -tools oder -elementen vornehmen. So lassen sich auch spezifische, neuartige digitale Lehrformen und -formate oder zukünftige Entwicklungen in diesem Bereich analysieren. Zusätzlich lassen sich bei Bedarf weitere Spezifikationen wie beispielweise die Art der digitalen Veranstaltung (u.a. Vorlesung, Seminar, Kurs) oder die Klassifikation digitaler Anteile (u.a. Art, Kategorisierung, Merkmale, Setting, Zweck) im zweiten Teil vornehmen. Damit besteht im Nachhinein die Möglichkeit einer Bottom-up-Aggregation aller evaluierten Veranstaltungen für einen allgemeineren Erkenntnisgewinn über die angebotene Lehre. Bedingt durch die absichtliche Konstruktion insbesondere des ersten allgemeineren Teils des Fragebogens ist der Einsatz des Instruments ebenso in Lehrveranstaltungen mit analogen Anteilen (z.B. teildigitalisierte Lehr- und Lernangebote) möglich, da die allgemeinen Charakteristika wie auch die Lern- und Leistungsergebnisse für jedes Lehr-Lernformat von Bedeutung sind. Insgesamt ermöglicht die Modularität des Fragebogens einen flexiblen, ökonomischen und anwendungsbezogenen Einsatz, bei dem sogar individuelle und praxisnahe Fragestellungen der Lehrenden sowie unterschiedliche Lernoutcomes berücksichtigt und betrachtet werden können.

7 Methode

7.1 Entwicklung, Konstruktion und Einsatz des Fragebogens

Auf Basis einschlägiger Literatur, bereits bestehender Evaluationsinstrumente (z.B. Staufenbiel, 2000), des Bedarfs der Lehrenden und Lehrverantwortlichen sowie der Erfahrungen aller Kooperationspartner wurde im Rahmen des Verbundprojekts „Digital gestütztes Lehren und Lernen in Hessen“ der Modulare Fragebogen zur Evaluation von digitalen Lehr-Lern-Szenarien aus der Studierendenperspektive (MOFEDILLS; Kärchner et al., 2020) entwickelt (s. ESM). Grundlage für die Konstruktion der Module des zweiten Teils und deren Auswahl waren jeweils einschlägige Literaturarbeiten zu den jeweiligen digitalen Bestandteilen (z.B. Videos: Brame, 2016, Laurillard, 2002; für eine vollständige Übersicht s. ESM) und die erhobenen Bedarfe von Lehrenden während der Pilotierungsphase (Wintersemester 2019/2020), in welcher MOFEDILLS in digital gestützten Lehrveranstaltungsevaluationen an der Philipps-Universität Marburg erprobt und anschließend auf Basis der Rückmeldungen weiterentwickelt wurde. Eine besondere Herausforderung bestand hierbei darin, die immense Heterogenität der digitalen Lehrveranstaltungen im Allgemeinen und ihrer (fach-)spezifischen Umsetzungen im Besonderen abbilden zu können ohne die angestrebte ökonomische Durchführbarkeit (maximale Bearbeitungsdauer 15 Minuten) zu gefährden.

7.2 Datenerhebung

Die Online-Versionen des Fragebogens (s. ESM) wurden durch die jeweiligen Evaluationsverantwortlichen an den beteiligten Hochschulen im Rahmen der regulären Lehrveranstaltungsevaluationen über die standardmäßig eingesetzte Befragungssoftware EvaSys oder in Einzelfällen SoSci Survey im Sommersemester 2020 (Juli-August) unter Einhaltung der gültigen Datenschutzvorgaben eingesetzt. Das Evaluationsvorgehen orientierte sich eng an den Standards für Evaluation der DeGEval (2016) und erfolgte in enger Abstimmung mit den Evaluationsverantwortlichen an den Standorten. Alle Lehrende, die sich im Vorfeld für eine Lehrveranstaltungsevaluation mit diesem Instrument angemeldet hatten, erhielten eine E-Mail ihrer zentralen Evaluationseinrichtung. Diese Mail beinhaltete alle Informationen zum Evaluationsvorhaben und einen Link zur Evaluationsbefragung der jeweiligen Lehrveranstaltung. Diesen Link zur Umfrage leiteten die Lehrenden an ihre Studierenden weiter. Die Studierenden wurden zu Beginn über das Ziel der Befragung und sämtliche Rahmenbedingungen der Evaluation (Anonymisierung und Verarbeitung der Daten usw.) aufgeklärt. Sie füllten die Fragebögen nach Abfrage über die Einwilligung zur Teilnahme entweder im Rahmen der jeweiligen Lehrveranstaltung oder im Setting ihrer Wahl aus. Die Auswahl und Kombination der Module des Fragebogens konnten die Lehrenden entweder selbstständig für ihre jeweilige Lehrveranstaltung vornehmen oder aus Kombinationsvorschlägen von den jeweiligen Einrichtungen auswählen. Für alle Lehrenden und Studierenden war die Teilnahme freiwillig, und es wurden keinerlei externe Anreize zur Evaluationsteilnahme vergeben. Alle Lehrenden erhielten von den Evaluationseinrichtungen einige Wochen später einen automatisiert erstellten Rückmeldebericht mit den anonymisierten und aggregierten Evaluationsergebnissen für ihre jeweilige Veranstaltung.

7.3 Stichprobe

Nach Dateninspektion und -aufbereitung umfasste die Stichprobe insgesamt 1,604 Studierende (922 weiblich, 668 männlich, 14 divers, Alter: $M = 23.07$, $SD = 4.52$) verschiedener Fachrichtungen von vier hessischen Hochschulen (Philipps-Universität Marburg, Technische Hochschule Mittelhessen, Goethe-Universität Frankfurt und Justus-Liebig-Universität Gießen) aus insgesamt 157 verschiedenen Lehrveranstaltungen. Die überwiegende Mehrheit (81.3% der Studierenden) füllte die Befragung in Bezug auf eine einzelne Lehrveranstaltung aus. Unter den weiteren Personen war ein Studierender, der zehn verschiedene Veranstaltungen bewertete, sowie drei Studierende, die vier verschiedene Veranstaltungen evaluierten. Der übrige Teil der Studierenden füllte die Befragung mehrmals für zwei oder drei verschiedene Lehrveranstaltungen aus. Eine detaillierte Aufstellung der vertretenen Fächergruppen und Lehrveranstaltungen kann dem elektronischen Supplement entnommen werden.

7.4 Statistische Analysen

Die psychometrischen Analysen wurden getrennt für den Basisteil des Instruments und die jeweiligen speziellen Module des zweiten Teils durchgeführt. Innerhalb des zweiten Teils wurde zudem jeder Modulbaustein einzeln analysiert. Mittels konfirmatorischer Faktorenanalysen wurden die theoretisch postulierten Faktorenstrukturen von MOFEDILLS mit dem

Programm Mplus 8.6 (Muthén & Muthén, 1998-2021) überprüft. Hierbei wurde in den verschiedenen Faktormodellen zur Korrektur der oben beschriebenen Verletzung der Unabhängigkeitsannahme der Probanden der robuste Maximum-Likelihood-Schätzer (MLR) in Kombination mit dem Befehl TYPE = COMPLEX (Asparouhov & Muthén, 2010) verwendet. Zunächst wurde untersucht, inwiefern sich das für den Basisteil angenommene Modell mit neun Faktoren (*Gesamturteil: Anforderungen, Gesamturteil: Eindruck, Praxisbezug, Individualisierung, Flexibilisierung, Effizienzsteigerung, Interaktion, selbstreguliertes Lernen und Lernzuwachs*) in den Daten abbilden ließ. Das Gesamturteil für die Veranstaltung in Form einer Schulnotenskala wurde nicht in die konfirmatorischen Faktorenanalysen eingeschlossen, sondern später als Validitätskriterium verwendet. Von den Faktorenanalysen ausgenommen wurden zudem zwei Items zum Partizipationsverhalten, wovon eines ein Freitext-Item ist. Da Freitext-Items aber nicht in eine konfirmatorische Faktorenanalyse einfließen können und die Berechnung eines Faktors auf Grundlage nur einer beobachteten Variablen statistisch nicht sinnvoll ist, wurde die gesamte Skala ausgeschlossen. Anschließend wurde die angenommene 1-Faktorenstruktur für die spezifischen Module *Abstimmungssystem, Selbstlernmodul Nutzung, Selbstlernmodul Produktion* und *Videonutzung* konfirmatorisch überprüft. Alle Faktorladungen wurden frei geschätzt und die Varianzen der latenten Variablen auf 1 fixiert. Zur Bewertung der Modellanpassung wurden der Confirmatory Fit Index (CFI) und der Standardized Root Mean Square Residual (SRMR) sowie Root Mean Square Error of Approximation (RMSEA) herangezogen. Ein $CFI \geq .90$, ein SRMR oder RMSEA $\leq .08$ sprechen für einen akzeptablen Modellfit, währenddessen ein $CFI \geq .95$, ein SRMR oder RMSEA $\leq .05$ einem guten Modellfit entsprechen (Jackson, Gillespy & Purc-Stephenson, 2009; Schermelleh-Engel, Moosbrugger & Müller, 2003).

Analyse und Vergleich von Evaluationsdaten aus verschiedenen Hochschulen setzen voraus, dass die Items messinvariant für verschiedene Subgruppen von Studierenden sind, d.h. dasselbe zugrundeliegende Konstrukt messen. Für den Basisteil des Fragebogens wurde daher nach dem Step-Up-Ansatz (Brown, 2006) die Messinvarianz für verschiedene Hochschulen, Fächergruppen und Geschlechter bestimmt. Da der Großteil der Daten von Studierenden der Philipps-Universität Marburg und der Technischen Hochschule Mittelhessen stammte, konnte nur ein Vergleich zwischen diesen beiden Hochschulen erfolgen. Die im Gesamtdatensatz erfassten Veranstaltungen wurden jeweils einer von vier übergeordneten Fächergruppen zugewiesen („Rechts-, Wirtschafts- und Sozialwissenschaften“, „Sport- und Erziehungswissenschaften“, „Sprach- und Kulturwissenschaften, Kunst, Kunstwissenschaften“, „Mathematik, Naturwissenschaften, Humanmedizin/Gesundheitswissenschaften, Ingenieurwissenschaften“, s. ESM). Zunächst wurde die *konfigurale Messinvarianz* überprüft, welche dann gegeben ist, wenn ein Konstrukt in allen Gruppen durch die gleichen Items repräsentiert ist. Im nächsten Schritt wurde die *metrische Messinvarianz* durch die Gleichsetzung aller Faktorladungen über die Gruppen hinweg getestet. Wird diese Annahme bestätigt, kann davon ausgegangen werden, dass in den untersuchten Subpopulationen die latenten Konstrukte die gleiche inhaltliche Bedeutung haben. Zum Schluss wurde das Ausmaß an *skalarer Messinvarianz* durch die zusätzliche Fixierung der entsprechenden Intercepts ermittelt. Bei Nachweis dieser Messinvarianz kann davon ausgegangen werden, dass keine itemspezifischen Schwierigkeitsunterschiede zwischen den Gruppen bestehen. Die Ausprägung in der latenten Variablen kann dann zwischen den Gruppen verglichen werden (Schwab & Helm, 2015). Es erfolgte eine schrittweise Betrachtung der Unterschiede in den Modellfits mit Hilfe des

Satorra-Bentler skalierten χ^2 -Differenztests (bei Verwendung des MLR-Schätzers) und der zugehörigen Interpretationsheuristik von Chen (2007).

8 Psychometrische Eigenschaften von MOFEDILLS

8.1 Konformatorische Faktorenanalysen

Das angenommene 9-Faktoren-Modell für den Basisteil des Fragebogens wies eine akzeptable Anpassung an die Daten auf (s. Tabelle 1 und Tabelle A1 im Anhang). Auch die jeweils vermuteten 1-Faktorenstrukturen der spezifischen Module *Abstimmungssystem*, *Selbstlernmodul Nutzung*, *Selbstlernmodul Produktion* und *Videonutzung* zeigten einen befriedigenden bis guten Modell-Fit mit Ausnahme des Moduls „Abstimmungssysteme“ (s. Tabelle 1).

Tabelle 1: Konfirmatorische Faktorenanalysen

Module	N	χ^2 [df]; p	CFI	SRMR	RMSEA
Teil 1 Basismodul	1,603	1372.905 [263]; < .001	.930	.050	.051
Teil 2 AS	91	171.485 [65]; < .001	.866	.050	.134
Teil 2 SN	146	74.612 [44]; < .001	.961	.044	.072
Teil 2 VN	680	254.816 [54]; < .001	.939	.052	.074
Teil 2 SP	110	13.631 [5]; .017	.968	.024	.125

Anmerkung: Teil 1 = alle Basismodule des 1. Teils des modularen Fragebogens (9-Faktoren-Lösung). Teil 2 (jeweils 1-Faktor-Lösung): AS = Baustein Abstimmungssystem; SN = Baustein Selbstlernmodul Nutzung; VN = Baustein Videonutzung; SP = Baustein Selbstlernmodul Produktion.

Mit wenigen Ausnahmen lagen die Faktorladungen über alle Items hinweg in einem hohen bis sehr hohen Bereich (s. ESM). Für die Bausteine Virtual Reality, Game-based Learning und Videoproduktion lagen auf Grund geringer Nachfrage keine für die Analysen ausreichenden Daten vor.

8.2 Reliabilitäten, Trennschärfen und Itemschwierigkeiten

Zur Beurteilung der Skalenreliabilitäten für den Basisteil wurde der Reliabilitäts-Koeffizient McDonald's Omega (ω ; McDonald, 1999) herangezogen. Die Reliabilitäten der Skalen lagen mit Ausnahme der Skalen GB Effizienzsteigerung und GB Flexibilisierung in einem zufriedenstellenden bis sehr guten Bereich ($.73 \leq \omega \leq .87$; vgl. Tabelle A2 im Anhang). Als weiterer wichtiger Itemkennwert wurden die Trennschärfen der Items berechnet. Unter der Trennschärfe eines Items versteht man die Korrelation eines Items mit dem Gesamtergebnis eines Testinstruments, was eine Einschätzung ermöglicht, wie gut ein Item zwischen Personen mit niedriger und hoher Merkmalsausprägung trennt. In dieser Studie wurde bei der Berechnung die übliche „part-whole-Korrektur“ vorgenommen, d.h. der Gesamtestwert wurde für die Trennschärfenanalyse ohne das betreffende Item berechnet. Alle Trennschärfen befanden sich im anzustrebenden Bereich von $KT > .40$, wobei die Trennschärfen im zweiten Teil im Vergleich zum ersten Teil zumeist etwas höher lagen (s. Tabelle A1). Die Itemschwierigkeiten variierten über die Skalen hinweg zwischen $\rho = .40$ bis $\rho = .81$. Die internen Konsistenzen der spezifischen Module des zweiten Teils des Fragebogens reichten von $\omega = .92$ bei der

Skala „Videonutzung“ bis $\omega = .98$ bei der Skala „Abstimmungssystem“ und können somit als sehr gut bezeichnet werden (vgl. ESM). Die Itemschwierigkeiten variierten über die Skalen hinweg zwischen $\rho = .38$ und $\rho = .86$.

8.3 Kriteriumsvalidität und Inhaltsvalidität

Die Skala „Gesamturteil der Veranstaltung“ in Schulnoten wurde in dieser Studie als Hinweis auf die Kriteriumsvalidität herangezogen und weist mit Ausnahme der Skala „Gesamturteil Anforderung“ erwartungskonforme, moderat signifikante Zusammenhangsmuster mit allen Skalen auf (vgl. Tabelle A2). Interessant ist, dass je höher Studierende die Anforderungen der evaluierten Lehrveranstaltung einschätzen, desto schlechter fällt ihr Gesamturteil über die Veranstaltung aus. Allerdings ist dieser Zusammenhang als klein einzuschätzen (Cohen, 1988).

Um einen ersten Eindruck zu gewinnen, wieviel Varianz im Kriterium durch die verschiedenen Dimensionen des Fragebogens erklärt werden kann, wurde eine multiple Regression zur Vorhersage des Gesamturteils (GU) der Veranstaltung (als Schulnotenskala) berechnet. Die in Tabelle 2 dargestellten Befunde zeigen, dass insgesamt 36% der Kriteriumsvarianz erklärt werden konnte. Als signifikante Prädiktoren erwiesen sich hierbei die Skalen GU Anforderung, GU Eindruck, GB Interaktion, GB Effizienzsteigerung, SRL und Lernzuwachs.

Tabelle 4: Regression zur Vorhersage des Gesamturteils der Veranstaltung

Skalen	β (SE)	p	95 % CI
GU Anforderung	.287 (.031)	< .001	.207, .021
GU Eindruck	-.199 (.041)	< .001	-.280, -.119
GB Praxisbezug	-.040 (.031)	.202	-.100, .021
GB Individualisierung	-.019 (.032)	.558	-.081, .044
GB Flexibilisierung	.064 (.034)	.063	-.003, .131
GB Interaktion	-.077 (.029)	.007	-.133, -.021
GB Effizienzsteigerung	-.132 (.036)	< .001	-.203, -.062
SRL	-.079 (.039)	.041	-.156, -.003
Lernzuwachs	-.239 (.035)	< .001	-.308, -.170

Anmerkungen: AV: Gesamturteil der Veranstaltung in Schulnoten, d.h. niedrige Werte implizieren eine positivere Bewertung. Aufgeklärte Varianz: $R^2 = .359$. GU = Gesamturteil, GB = Gelingensbedingung, SRL = Selbst-reguliertes Lernen. * $p < .05$, ** $p < .01$.

Zur Sicherung der Inhaltsvalidität wurden über ein Jahr lang diverse Feedbackschleifen mit den Kooperationspartnern des Zentrums für kooperatives Lehren und Lernen (ZekoLL) an der THM und der zentralen E-Learning Einrichtung studiumdigitale an der Goethe-Universität Frankfurt durchgeführt. So wurde im Laufe der Gespräche jedes einzelne Item, dessen Formulierung und inhaltliche Passung überprüft.

8.4 Messinvarianz

Die Überprüfung von Messinvarianz hinsichtlich des Geschlechts der Studierenden zeigte, dass trotz der signifikanten Ergebnisse in den χ^2 -Tests (Differenz skalar vs. metrisch:

$\Delta\chi^2 [17] = 34.987, p = .001$) aufgrund der hohen Sensitivität des X^2 Tests für große Stichproben skalare Invarianz nach Chen (2007) angenommen werden kann ($\Delta RMSEA = .000$, $\Delta SRMR = .000$, $\Delta CFI = .001$). Zudem konnte die These gestützt werden, dass der Basisteil des Fragebogens messinvariant hinsichtlich der Hochschulen THM und PUM ist (Differenz skalar vs. metrisch: $\Delta\chi^2 [17] = 121.555, p = .001$, $\Delta RMSEA = .001$, $\Delta SRMR = .002$, $\Delta CFI = -.006$). Das bedeutet, dass keine itemspezifischen Schwierigkeitsunterschiede zwischen den Geschlechtern und den beiden genannten Hochschulen bestehen (Schwab & Helm, 2015). Ferner ist mindestens metrische Invarianz zwischen allen Studiengängen anzunehmen (Differenz skalar vs. metrisch: $\Delta\chi^2 [51] = 269.294, p = .001$, $\Delta RMSEA = .003$, $\Delta SRMR = .002$, $\Delta CFI = -.012$). Es kann somit davon ausgegangen werden, dass über alle Studiengänge hinweg die latenten Konstrukte die gleiche inhaltliche Bedeutung haben (Schwab & Helm, 2015). Details der Messinvarianzprüfungen können dem ESM entnommen werden.

8.5 Interkorrelationen

Die Skalen „Gesamturteil Anforderung“ und „Gesamturteil Eindruck“ korrelieren negativ miteinander. Die „Gelingensbedingungen“ hängen untereinander positiv zusammen und sind zudem alle mit den Lernoutcomes „SRL“ sowie „Lernzuwachs“ positiv assoziiert. Die Skala „Gesamteindruck Anforderungen“ korreliert signifikant negativ mit allen anderen Skalen des vollständigen Fragebogens, hingegen korreliert die Skala „Gesamturteil Eindruck“ mit allen anderen Skalen signifikant positiv. Alle untersuchten Interkorrelationen sind signifikant und können Tabelle 3 entnommen werden, in der sich zudem die Mittelwerte und Standardabweichungen der Skalen finden.

9 Diskussion

Diese Studie beschreibt die Entwicklung und psychometrischen Eigenschaften von MOFEDILLS, einem theoretisch fundierten, open-access basierten, modularen multidimensionalen Instrument zur Evaluation digitaler und teildigitalisierter Lehr-Lernszenarien im Hochschulkontext. Ziel war es, ein anwendungsfreundliches Instrument zu konstruieren, welches die Bedarfe aller Akteure von Lehrevaluationen berücksichtigt und dabei flexibel einsetzbar bleibt, sodass möglichst alle (teil-)digitalen Lehrveranstaltungen und deren relevante Outcomes wie Lernzuwachs und selbstreguliertes Lernen umfassend und trotzdem ökonomisch erfasst werden können.

Die hier untersuchten Evaluationsdaten verschiedener hessischer Hochschulen stützen die Annahme einer hohen psychometrischen Qualität und Eignung des modularen Fragebogens für seinen Einsatzzweck. So konnte gezeigt werden, dass der Fragebogen insgesamt sowohl faktoriell valide als auch messinvariant über verschiedene Geschlechter, Hochschulen und Fächergruppen hinweg ist. Eine kleinere Unstimmigkeit zeigte sich bei der konfirmatorischen Faktorenanalyse zum Modul Abstimmungssystem. Trotz suboptimalen Modellfits waren die Faktorladungen jedoch durchgängig hoch, so dass die Annahme der Eindimensionalität hier zunächst beibehalten werden kann, jedoch in weiteren Untersuchungen mit größeren Substichproben detaillierter analysiert werden sollte. Insgesamt weisen die teil-

weise sehr hohen Reliabilitäten der Skalen des zweiten Teils darauf hin, dass manche Skalen ohne größere psychometrische Einbußen um einige Items gekürzt werden könnten. Vor entsprechenden Entscheidungen sollten allerdings zunächst höhere Fallzahlen für diese spezifischen Skalen erhoben werden, und es sollte geprüft werden, inwiefern die hier berichteten Ergebnisse zu den Skaleneigenschaften repliziert werden können.

Unsere Studienergebnisse stützen die vermutete Bedeutsamkeit der bislang selten empirisch untersuchten Gelingensbedingungen digitaler Lehrveranstaltungen (Helmke, 1996; Rindermann 2001, 2003; Ullrich, 2020). So zeigten sich unseren Erwartungen entsprechend signifikante Zusammenhänge zwischen den jeweiligen Gelingensbedingungen Praxisbezug, Individualisierung, Flexibilisierung, Interaktion, Effizienzsteigerung und der Gesamtzufriedenheit mit der Veranstaltung. Je eher die Gelingensbedingung als gegeben bewertet wurde (z.B. problemlose Interaktion aller Beteiligten), desto positiver wurde die evaluierte Veranstaltung bewertet. In Einklang mit der einschlägigen Literatur (z.B. Steinmayr et al., 2019; Zimmerman, 2000) war der subjektive Lernzuwachs am stärksten mit der Skala „selbstreguliertes Lernen“ assoziiert. Ebenfalls hohe Korrelationen zeigten sich zwischen Lernzuwachs und der Skala „Praxisbezug“, was die theoretisch angenommene Bedeutsamkeit dieses Faktors für Lehr-Lernsettings im Allgemeinen unterstreicht (vgl. Bargel et al., 2007). Die Ergebnisse der multiplen Regression zur Vorhersage der Schulnote für die Veranstaltung zeigten, dass durch die MOFEDILLS-Skalen 36% der Kriteriumsvarianz erklärt werden konnte. Da derartige Kriteriumsvariablen von zahlreichen weiteren, nicht in einem Evaluationsinstrument abbildbaren Faktoren beeinflusst werden (z.B. Vorwissen, überdauernde Interessen, Bewertungstendenzen etc.), kann der hier durch die Fragebogenskalen aufgeklärte Varianzanteil als durchaus substantiell und als Beleg für die prognostische Validität von MOFEDILLS gewertet werden.

In Bezug auf verschiedene Gütekriterien psychometrischer Tests fällt die Betrachtung insgesamt sehr positiv aus. Die Modularität des Fragebogens stellt einen flexiblen, ökonomischen und individuellen Einsatz sicher. Die Objektivität ist in Bezug auf die Durchführung und Auswertung aufgrund empfohlener Instruktionen, vorgegebener Antwortformate und Codierungen sowie Auswertungshinweise gegeben (vgl. ESM). Die ersten Befunde zur Konstruktvalidität von MOFEDILLS erscheinen vielversprechend, wenngleich weitere Überprüfungen notwendig sind. Zum einen deuten die theoretisch hergeleiteten und in wiederholten Feedbackprozessen mit den beteiligten E-Learning Einrichtungen adaptierten Items auf eine hohe Inhaltsvalidität hin. Zum anderen unterstützt das erwartungskonforme Korrelationsmuster der Skalen mit dem Gesamturteil der Veranstaltung die konvergente Validität des Fragebogens. Trotz einer großen hochschulübergreifenden Stichprobe lagen für die Module Virtual Reality, Game-based Learning und Videoproduktion keine für die Analysen ausreichenden Daten vor, so dass die Nützlichkeit dieser Module im Evaluationskontext digitaler Lehre noch in zukünftigen Studien geprüft werden muss. In Bezug auf inhaltliche Zusammenhänge wäre es wünschenswert, in kommenden Arbeiten genauer zu eruieren, welche digitalen Gelingensbedingungen am stärksten Lernerfolg und selbstreguliertes Lernen vorherzusagen.

Obwohl die vorliegende Studie insgesamt viele Belege für die psychometrische Qualität und Nützlichkeit von MOFEDILLS für die Evaluation digital gestützter Lehrveranstaltungen liefert, unterliegen die vorgestellten Befunde gewissen Einschränkungen. So erfolgte die Datenerhebung während der coronabedingt besonderen Lehrsituation im SoSe 2020, was verschiedene Bewertungsverzerrungen nach sich gezogen haben könnte (z.B. positivere Bewer-

tung als normalerweise, wenn die Lehre zumindest online stattfinden konnte). Zudem fehlte aufgrund der bis dato kaum existenten digitalen Lehre bei der Beurteilung vermutlich ein Vergleichsmaßstab (Gollwitzer et al., 2006). Eine weitere Limitation unserer Studie bezieht sich auf die Schwierigkeit, die Repräsentativität der erhobenen Daten zu beurteilen. Dem entgegen stehen allerdings die im Vergleich zu regulären Semestern höheren Rücklaufquoten. Die Validität von Selbsteinschätzungen zur Erfassung der Qualität der Lehre wird schon lange diskutiert und häufig bezweifelt (z.B. Kromrey, 2001; Marsh & Roche, 1997). Künftige Studien könnten mit der parallel entwickelten, aber bisher nicht systematisch überprüften Lehrendenversion von MOFEDILLS (Kärchner et al., 2020) die Einschätzungen der Lehrenden abfragen und im Sinne einer empirischen Beobachterübereinstimmung miteinbeziehen. Inwieweit der Fragebogen (zumindest die Module des ersten Teils) auch für den Einsatz in vollständig analogen Veranstaltungen – sofern diese durch den digitalen Wandel der Bildungskultur an Hochschulen (z.B. Blömer, Voigt & Hoppe, 2020) noch existieren – geeignet ist, können zukünftige Studien adressieren. Darüber hinaus wären gezielte und systematische inhaltliche Überprüfungen der angenommenen Bedingungsfaktoren und dessen Facetten erfolgreicher Lehre sowie Testungen der theoretischen Modelle zukünftig wünschenswert.

Zusammenfassend befürworten die vorliegenden Befunde den regulären Einsatz von MOFEDILLS im Rahmen von Evaluationen digital gestützter Lehre. Dies ist von großer Bedeutung, da hierfür im deutschen Sprachraum bislang kaum fundierte Instrumente verfügbar sind. Durch den Einsatz des hier vorgestellten Instruments MOFEDILLS können Lehrende wertvolle individualisierte Rückmeldungen der Studierenden zu ihren digitalen Lehrveranstaltungen erhalten.

Supplementmaterial

Im Beitrag wird verschiedentlich auf elektronisch bereitgestellte Zusatzmaterialien (ESM) hingewiesen. Diese können unter https://osf.io/4b3vf/?view_only=1ca25289396345faa23e693f57db36ab oder unter <http://is.gd/mofedills> abgerufen werden.

Literatur

- Andrés, B., Sanchis, R. & Poler, R. (2015, 2. – 4. März). *Quiz game applications to review the concepts learnt in class: An application at the University context*. Proceedings of INTED 2015 Conference, Madrid, Spanien, 5654–5662. Verfügbar unter <http://hdl.handle.net/10251/66300>
- Asparouhov, T. & Muthén, B. (2010). *Resampling methods in Mplus for complex survey data*. Verfügbar unter https://www.statmodel.com/download/Resampling_Methods5.pdf
- Bargel, T., Müßig-Trapp, P. & Willige, J. (2007). *Studienqualitätsmonitor 2007 – Studienqualität und Studiengebühren*. Hannover: HIS. Verfügbar unter <http://www.hisbus.de/faq/pdf/SQM2007.pdf>
- Bauer, R., Hafer, J., Hofhues, S., Schiefner-Rohs, M., Thillosen, A., Volk, B. & Wannemacher, K. (Hrsg.) (2020). *Vom E-Learning zur Digitalisierung – Mythen, Realitäten, Perspektiven*. Reihe Medien in der Wissenschaft (Band 76). Münster: Waxmann.
- Blömer, L., Voigt, C. & Hoppe, U. (2020). Corona-Pandemie als Treiber digitaler Hochschullehre. In R. Zender, D. Ifenthaler, T. Leonhardt & C. Schumacher (Hrsg.), *DELFI 2020 – Die 18. Facha-*

- gung Bildungstechnologien der Gesellschaft für Informatik e.V. – Komplettband. (S. 343–348). Bonn: Gesellschaft für Informatik.
- Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning & Instruction*, 7, 161–186. [https://doi.org/10.1016/S0959-4752\(96\)00015-1](https://doi.org/10.1016/S0959-4752(96)00015-1)
- Brame, C. J. (2016). Effective educational videos: Principles and guidelines for maximizing student learning from video content. *CBE-Life Sciences Education*, 15(4), 1–6. <https://doi.org/10.1187/cbe.16-03-0125>
- Brown, T. A. (2006). *Confirmatory factor analysis*. New York, NY: Guilford.
- Cassidy, S. (2011). Self-regulated learning in higher education: Identifying key component processes. *Studies in Higher Education*, 36, 989–1000. <https://doi.org/10.1080/03075079.2010.503269>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Auflage). Hillsdale, NJ: Lawrence Erlbaum.
- DeGEval – Gesellschaft für Evaluation (2016). *Standards für Evaluation* (1. Revision). Verfügbar unter https://www.degeval.org/fileadmin/Publikationen/DeGEval-Standards_fuer_Evaluation.pdf
- Dittler, U. & Kreidl, C. (2020). Ergebnisse einer trinationalen Umfrage unter Studierenden aus Deutschland, Österreich und der Schweiz zu ihren Erfahrungen mit der Online-Lehre des Sommersemesters 2020 – Ergebnisse der empirischen Studie zum Thema „Auswirkungen der Corona-Krise auf die Hochschullehre“. Verfügbar unter <https://opus.hs-furtwangen.de/frontdoor/deliver/index/docId/6737/file/Gesamtergebnisse+Corona+Hochschullehre.pdf>
- Dittler U. & Kreidl C. (2021). Eine kurze Chronologie der Covid-19-Pandemie im Frühjahr 2020. In U. Dittler & C. Kreidl (Hrsg.), *Wie Corona die Hochschullehre verändert* (S. 1–13). Wiesbaden: Springer Gabler. <https://doi.org/10.1007/978-3-658-32609-8>
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive and sociocultural perspective on motivation. *Contemporary Educational Psychology*, 61. <https://doi.org/10.1016/j.cedpsych.2020.101859>
- Getto, B., Hintze, P. & Kerres, M. (Hrsg.) (2018). *Digitalisierung und Hochschulentwicklung: Proceedings zur 26. Tagung der Gesellschaft für Medien in der Wissenschaft e.V.* Münster: Waxmann.
- Gilch, H., Beise, A. S., Krempkow, R., Müller, M., Stratmann, F. & Wannemacher, K. (2019). *Digitalisierung der Hochschulen: Ergebnisse einer Schwerpunktstudie für die Expertenkommission Forschung und Innovation* (Studien zum deutschen Innovationssystem, 14). Verfügbar unter <http://hdl.handle.net/10419/194284>
- Gollwitzer, M., Kranz, D. & Vogel, E. (2006). Die Validität studentischer Lehrveranstaltungsevaluierungen und ihre Nützlichkeit für die Verbesserung der Hochschullehre: Neuere Befunde zu den Gütekriterien des „Trierer Inventars zur Lehrevaluation“ (TRIL). In G. Krampen & H. Zayer (Hrsg.), *Psychologiedidaktik und Evaluation* (S. 90–104). Göttingen: Hogrefe Verlag.
- Handke, J. & Sperl, A. (Hrsg.) (2012). *Das Inverted Classroom Model: Begleitband zur ersten deutschen ICM-Konferenz*. München: Oldenbourg Wissenschaftsverlag.
- Harris-Huemmert, S., Pohlenz, P. & Mitterauer, L. (Hrsg.) (2018). *Digitalisierung der Hochschullehre: Neue Anforderungen an die Evaluation?* Münster: Waxmann.
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Oxford (GB): Routledge.
- Helmke, A. (1996). Studentische Evaluation der Lehre – Sackgassen und Perspektiven. *Zeitschrift für Pädagogische Psychologie*, 10, 181–186.
- Henderson, M., Selwyn, N. & Aston, R. (2017). What works and why? Student perceptions of ‘useful’ digital technology in university teaching and learning. *Studies in Higher Education*, 42(8), 1567–1579. <https://doi.org/10.1080/03075079.2015.1007946>
- Jackson, D. L., Gillespy, J. A. & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14, 6–23.

- Kärchner, H., Gehle, M. & Schwinger, M. (2020, 10. November). *eduValuation+ – Evaluationsinstrumente für die Hochschule – Modularer Fragebogen zur Evaluation digitaler Lehr-Lern-Szenarien (MOFEDILLS)*. Verfügbar unter <https://fragebogen-hochschullehre.de>
- Kromrey, H. (2001). Evaluation von Lehre und Studium. Anforderungen an Methodik und Design. In C. Spiel. (Hrsg.), *Evaluation universitärer Lehre – Zwischen Qualitätsmanagement und Selbstzweck* (S. 21–60). Münster: Waxmann.
- Laurillard, D. (2002). *Rethinking university teaching: A conversational framework for the effective use of learning technologies*. New York: Taylor & Francis.
- Lehner, M. & Sohm, K. (2021). Qualität, didaktische Methodik und Digitalität. In U. Dittler & C. Kreidl (Hrsg.), *Wie Corona die Hochschullehre verändert* (S.339–350). Wiesbaden: Springer Gabler. https://doi.org/10.1007/978-3-658-32609-8_21
- LVE Fragebögen der Goethe-Universität Frankfurt (o. D.). *Lehrveranstaltungsevaluation – Basisfragebögen, Ergänzungen zum Feedback für die Lehrperson, Ergänzungen zu organisatorischen Bedingungen/Kontextfaktoren*. studiumdigitale – Zentrale eLearning-Einrichtung. Verfügbar unter <https://www.studiumdigitale.uni-frankfurt.de/58533111/Evaluation>
- Marsh, H. W. & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist*, 52, 1187–1197.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.
- Metz-Göckel, S., Kamphans, D. S. M. & Scholkmann, A. (2012). Hochschuldidaktische Forschung zur Lehrqualität und Lernwirksamkeit. *Zeitschrift für Erziehungswissenschaft*, 15(2), 213–232. <https://doi.org/10.1007/s11618-012-0274-z>
- Mittag, S., Mutz, R. & Daniel, H.-D. (2012). Anforderungen an Qualitätssicherungsinstrumente für Lehre und Studium an Hochschulen. Ergebnisse einer Meta-Evaluation an der ETH Zürich. In Bayerisches Staatsinstitut für Hochschulforschung (Hrsg.), *Beiträge zur Hochschulforschung. Thema: Qualitätssicherung in Lehre und Forschung* (S. 8–31). Verfügbar unter <https://d-nb.info/1029232334/34#page=12>
- Mitterauer, L., Pohlenz, P. & Harris-Huermann, S. (Hrsg.) (2019). *Systeme im Wandel. Hochschulen auf neuen Wegen*. Münster: Waxmann.
- Muthén, L. K. & Muthén, B. O. (1998–2021). *Mplus user's guide* (8. Auflage). Los Angeles, CA: Muthén & Muthén.
- Peverly, S. T., Brobst, K. E., Graham, M. & Shaw, R. (2003). College adults are not good at self-regulation: A study on the relationship of self-regulation, note taking, and test taking. *Journal of Educational Psychology*, 95(2), 335–346. <https://doi.org/10.1037/0022-0663.95.2.335>
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Hrsg.), *Handbook of self-regulation* (S. 451–502). San Diego, CA: Academic Press. <https://doi.org/10.1016/B978-012109890-2/50043-3>
- Referat Lehrevaluation an der Philipps-Universität Marburg (2021). *Fragenblock – optionale Fragen zur digitalen Lehre an der Phillips-Universität Marburg*. <https://www.uni-marburg.de/de/universitaet/administration/verwaltung/dezernat3/dez3b/lehrevaluation/lehrevaluation/download-ordner/fragendigitalelehre.pdf>
- Rindermann, H. (2001). *Lehrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen mit einem Beitrag zur Evaluation computerbasierter Unterrichts*. Landau: Verlag Empirische Pädagogik.
- Rindermann, H. (2003). Lehrevaluation an Hochschulen: Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und seine Evaluation. *Zeitschrift für Evaluation*, 2, 233–256.
- Ryan, R. M. & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, 8(2), 23–74.

- Schmitz, B. (2001). Self-Monitoring zur Unterstützung des Transfers einer Schulung in Selbstregulation für Studierende: Eine prozessanalytische Untersuchung. *Zeitschrift für Pädagogische Psychologie*, 15(3–4), 181–197. <https://doi.org/10.1024/1010-0652.15.34.181>
- Schmitz, B. & Schmidt, M. (2007). Einführung in die Selbstregulation. In M. Landmann & B. Schmitz (Hrsg.), *Selbstregulation erfolgreich fördern. Praxisnahe Trainingsprogramme für effektives Lernen* (S. 9–18). Stuttgart: Kohlhammer.
- Schneider, M. & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, 143, 565–600.
- Schober, B., Jöstl, G., Klug, J., Wimmer, B., Spiel, C., Steuer, G., Schmitz, B., Ziegler, A. & Dresel, M. (2016). Kompetenzen zum selbstregulierten Lernen an Hochschulen – Das Projekt PRO-SRL. In BMBF (Hrsg.), *Bildungsforschung 2020. Zwischen wissenschaftlicher Exzellenz und gesellschaftlicher Verantwortung* (Bildungsforschung, Bd. 42, S. 184–192). Berlin.
- Schwab, S. & Helm, C. (2015). Überprüfung von Messinvarianz mittels CFA- und DIF- Analysen. *Empirische Sonderpädagogik*, 7(3), 175–193.
- Staufenbiel, T. (1999). Personalentwicklung. In C. Graf Hoyos & D. Frey (Hrsg.), *Lehrbuch Arbeits- und Organisationspsychologie* (S. 510–525). Weinheim: Psychologie Verlags Union.
- Staufenbiel, T. (2000). Fragebogen zur Evaluation von universitären Lehrveranstaltungen durch Studierende und Lehrende. *Diagnostica*, 46, 169–181. <https://doi.org/10.1026//0012-1924.46.4.169>
- Staufenbiel, T. (2001). Universitätsweite Evaluation von Lehrveranstaltungen in Marburg: Vorgehen, Instrumente, Ergebnisse. In E. Keiner (Hrsg.), *Evaluation (in) der Erziehungswissenschaft* (S. 43–61). Weinheim: Beltz.
- Steinmayr, R., Weidinger, A. F., Schwinger, M. & Spinath, B. (2019). The importance of students' motivation for their academic achievement – replicating and extending previous findings. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.01730>
- Traus, A., Höffken, K., Thomas, S., Mangold, K. & Schröer, W. (2020). *Stu.di.Co. – Studieren digital in Zeiten von Corona*. Hildesheim: Universitätsverlag Hildesheim. <https://dx.doi.org/10.18442/150>
- Treppesch, C., Hense, J. & Raser, M. (2015). *Modulares Gießener verhaltensbasiertes Lehrveranstaltungs-rückmeldungsinstrument (MoGLi)*. Servicestelle Lehrevaluation der Justus-Liebig-Universität Gießen. Verfügbar unter: <https://www.uni-giessen.de/org/admin/stab/stl/servicestelle/infomogli>
- Ullrich, I. (2020). *Gute Lehre in der Hochschule. Praxistipps zur Planung und Gestaltung von Lehrveranstaltungen*. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-31070-7>
- Wannemacher, K., Jungermann, I., Scholz, J., Tercanli, H. & Villiez, A. von (2016). *Digitale Lernszenarien im Hochschulbereich*. Arbeitspapier Nr. 15. Berlin: Hochschulforum Digitalisierung. Verfügbar unter https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD%20AP%20Nr%2015_Digitale%20Lernszenarien.pdf
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Hrsg.), *Handbook of self-regulation: Theory, research and applications* (S. 13–29). San Diego, CA: Academic Press.
- Zumbach, J., Spinath, B., Schahn, J., Friedrich, M. & Kögel, M. (2007). Entwicklung einer Kurzskaala zur Lehrevaluation. In M. Krämer, S. Preiser & K. Brusdeylins (Hrsg.), *Psychologiedidaktik und Evaluation VI* (S. 317–325). Göttingen: Vandenhoeck & Ruprecht.

Kontakt

Henrike Kärchner · Maren Gehle · Prof. Dr. Malte Schwinger
Institut für Psychologie, AE Pädagogische Psychologie
Philipps-Universität Marburg
Gutenbergstr. 18
35032 Marburg
E-Mail: henrike.kaerchner@uni-marburg.de
E-Mail: maren.gehle@uni-marburg.de
E-Mail: malte.schwinger@uni-marburg.de

Anhang

Tabelle A1: Itemkennwerte und Faktorladungen für den Basisteil von MOFEDILLS

Skala	Item- kennung	Itemname	<i>M</i>	<i>SD</i>	<i>T</i>	<i>p</i> (%)	Faktor- ladungen
Gesamturteil	GU_01	Tempo	3.01	0.75	.61	50.25	.74
Veranstaltung	GU_02	Stoffumfang	3.19	0.81	.60	54.75	.72
Anforderungen	GU_03	Schwierigkeit	3.04	0.76	.60	51.00	.71
Gesamturteil	GU_04	Zufriedenheit	3.88	1.29	.74	72.00	.84
Veranstaltung		Dozierende					
Eindruck	GU_05	Empfehlung LV	3.57	1.31	.74	64.50	.88
GB	GBPB_01	Nützlichkeit	3.73	1.25	.59	68.25	.84
Praxisbezug	GBPB_02	Berufsnähe	2.97	1.33	.59	49.25	.62
	GBPB_03	Forschungsnähe	3.26	1.29	.52	56.50	.59
GB	GBID_01	Lernart	3.46	1.35	.71	61.50	.84
Individualisierung	GBID_02	Lerntempo	3.33	1.35	.66	58.25	.82
	GBID_03	Anknüpfung Vorwissen	3.42	1.24	.45	60.50	.55
GB	GBFL_01	Räumlich	3.68	1.64	.49	67.00	.58
Flexibilisierung	GBFL_02	Zeitlich	4.07	1.38	.49	76.75	.82
GB	GBE_01	Verfügbarkeit	3.82	1.35	.35	70.50	.87
Effizienzsteigerung	GBE_02	<i>Störungen</i>	4.25	1.09	.35	81.25	.41
GB	GBIN_01	Kontakt	3.28	1.15	.78	57.00	.86
Interaktion		Teilnehmer/innen					
	GBIN_02	Zusammenarbeit Teilnehmer/innen	3.00	1.15	.77	50.00	.83
	GBIN_03	Kontakt Dozierende	3.35	1.21	.72	58.75	.80
Selbst- reguliertes Lernen	SR1_01	Motivation	3.48	1.25	.73	62.00	.79
	SR1_02	Planung	3.41	1.19	.78	60.25	.83
	SR1_03	Monitoring	3.52	1.22	.71	63.00	.71
	SR1_04	Kontrolle	3.33	1.20	.70	58.25	.76
	SR1_05	Reflexion	3.37	1.27	.71	59.25	.74
Lernzuwachs	LZ1_01	Wissen	3.78	1.22	.70	69.50	.82
	LZ1_02	Interesse	3.38	1.24	.71	59.50	.79
	LZ1_03	Praxis	3.17	1.22	.64	54.25	.70

Anmerkungen: *T* = Trennschärfe, *p* = Itemschwierigkeit in Prozent (%). Negativ gepolte Items sind kursiv geschrieben.

Tabelle A2: Reliabilitäten, Mittelwerte, Standardabweichungen und Interkorrelationen aller Skalen

Skalen	ω	M (SD)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) GU Veranstaltung ^a	–	2.14 (1.11)	1								
(2) GU Anforderung	.76	3.08 (0.66)	.16**	1							
(3) GU Eindruck	.85	2.11 (1.24)	-.43**	-.44**	1						
(4) GB Praxisbezug	.73	3.34 (1.08)	-.34**	-.31**	.54**	1					
(5) GB Individualisierung	.78	3.40 (1.10)	-.30**	-.41**	.52**	.48**	1				
(6) GB Flexibilisierung	.67	3.91 (1.32)	-.16**	-.37**	.50**	.35**	.51**	1			
(7) GB Interaktion	.87	3.14 (1.10)	-.35**	-.30**	.46**	.43**	.39**	.31**	1		
(8) GB Effizienz- steigerung	.60	4.00 (1.07)	-.31**	-.36**	.54**	.42**	.44**	.49**	.40**	1	
(9) SRL	.87	3.42 (1.01)	-.42**	-.40**	.64**	.56**	.59**	.48**	.53**	.53**	1
(10) Lern- zuwachs	.82	3.46 (1.04)	-.44**	-.34**	.65**	.58**	.50**	.41**	.44**	.45**	.71**

Anmerkungen: GU = Gesamturteil, GB = Gelingensbedingung, SRL = Selbstreguliertes Lernen. * $p < .05$, ** $p < .01$. ^a Skala Gesamturteil der Veranstaltung in Schulnoten, d.h. niedrige Werte implizieren eine positivere Bewertung. ω = McDonalds Omega.