



Erfassung der Selbstwirksamkeitserwartungen angehender Geographielehrpersonen zum Experimentieren im Geographieunterricht – Vorstellung und Validierung eines Messinstruments

Capturing Prospective Geography Teachers' Self-Efficacy Beliefs towards Experimentation in Geography Classes – Presentation and Validation of a Measurement Instrument

Evaluación de las creencias de autoeficacia de los futuros profesores de Geografía en la formación para la experimentación en las enseñanzas de la Geografía – presentación y validación de un instrumento de medida

Hanna Velling  , Jan Christoph Schubert 

Zusammenfassung Den Selbstwirksamkeitserwartungen von Lehrpersonen wird eine zentrale Rolle für ihr unterrichtliches Handeln zugeschrieben. Daher werden Instrumente benötigt, die die Erfassung der Selbstwirksamkeitserwartungen ermöglichen. Vor diesem Hintergrund wird im Beitrag ein aus der Physikdidaktik adaptiertes Instrument zur Erhebung der Selbstwirksamkeitserwartungen zum Experimentieren im Geographieunterricht vorgestellt und anhand einer Stichprobe von $n=303$ Geographielehrstundentinnen und Geographielehrstundenten validiert. Die Ergebnisse konfirmatorischer Faktorenanalysen sowie von Korrelations- und Messinvarianzanalysen und Mittelwertvergleichen bestätigen die Konstruktvalidität und Reliabilität des Instruments. Mit Blick auf die Ausprägung der Selbstwirksamkeitserwartungen zeigen sich bei den angehenden Lehrpersonen mittlere bis positive Selbstwirksamkeitserwartungen zum Experimentieren im Geographieunterricht.

Schlüsselwörter Selbstwirksamkeitserwartungen, Experimentieren, angehende Geographielehrpersonen, Erhebungsinstrument, professionelle Handlungskompetenz

Abstract Teachers' self-efficacy beliefs play a central role in their classroom teaching, reason why there is need for instruments to capture them. This paper presents an instrument adapted from Physics Education for assessing self-efficacy beliefs towards experimenting in Geography. The paper also introduces the results of the validation process with initial Geography teacher education students ($n=303$). The results of confirmatory factor analyses, correlation and measurement invariance analyses, as well as mean value comparisons confirm the construct validity and reliability of the instrument. Concerning expression, the students show moderate to positive self-efficacy beliefs towards experimenting in Geography teaching.

Keywords self-efficacy beliefs, experiments, prospective geography teachers, questionnaire, teacher behavior

Resumen Las creencias de autoeficacia del profesorado juegan un papel central en sus actividades de enseñanza en el aula, por lo que se necesitan instrumentos que nos permite capturarlas. Este artículo presenta un instrumento adaptado desde la Didáctica de la Física para evaluar las creencias de autoeficacia acerca de experimentos en Geografía. El artículo también presenta los resultados del proceso de validación del dicho instrumento con estudiantes de formación inicial de profesorado de Geografía ($n=303$). Los resultados de los análisis factoriales confirmatorios, los análisis de correlación y de invarianza de medición, así como las comparaciones de valores medios confirman la validez de constructo y la confiabilidad del instrumento. En cuanto a la expresión, los estudiantes muestran creencias de autoeficacia de moderadas a positivas hacia experimentos en la enseñanza de la Geografía.

Palabras clave creencias de autoeficacia, experimentación, futuros profesores de Geografía, cuestionario, competencia profesional

1. Einleitung

Dem Experimentieren im Geographieunterricht werden insbesondere vor dem Hintergrund einer naturwissenschaftlichen Grundbildung (*Scientific Literacy*) zahlreiche Potenziale für Lernprozesse und -ergebnisse zugeschrieben. Experimente gelten als didaktischer Königsweg in den Naturwissenschaften und sind als Denk- und Arbeitsweise im (naturwissenschaftlichen) Geographieunterricht etabliert (u.a. LETHMATE, 2006; WILHELMI, 2012; OTTO & MÖNTER, 2015). So sehen auch geographische Bildungsstandards und Lehrpläne ihren Einsatz im Geographieunterricht vor (z.B. DGfG, 2020). Trotz normativer Vorgaben sowie eines Konsenses bezüglich dieser Potenziale gibt es Hinweise darauf, dass Experimente als Aspekt forschenden Lernens nur selten im Geographieunterricht berücksichtigt werden (HEMMER & HEMMER, 2010; HOF & HENNEMANN, 2013; MIENER & KÖHLER, 2013; HÖHNLE & SCHUBERT, 2016; MADDOX ET AL., 2018; XUAN ET AL., 2019). Gleichzeitig stellt ihr Einsatz bezüglich der Vorbereitung sowie unterrichtlichen Einbettung und Gestaltung hohe Anforderungen an Geographielehrpersonen (MIENER & KÖHLER, 2013; OTTO & MÖNTER, 2015; AKUMA & CALLAGHAN, 2019; FÖGELE ET AL., 2020). Neben externen Faktoren wie ungünstigen Rahmenbedingungen (beispielsweise eine mangelhafte Experimentierausstattung an Schulen) wirken so insbesondere auch interne Faktoren wie die didaktisch-methodische Qualifikation, das Fachwissen und die Selbstwirksamkeitserwartungen von (angehenden) Geographielehrpersonen mit Blick auf das Experimentieren als Hindernisse für deren Einsatz (CRAWFORD, 2007; DIBIASE & McDONALD, 2015; HÖHNLE & SCHUBERT, 2016; RAMNARAIN, 2016; AKUMA & CALLAGHAN, 2019).

Gerade in herausfordernden Situationen, die von der unterrichtlichen Routine abweichen – wie

2. Theoretische Grundlegung

Das Konstrukt der Selbstwirksamkeit ist auf die sozial-kognitive Theorie nach BANDURA (1997) zurückzuführen. Dieser versteht unter Selbstwirksamkeit „[...] die Überzeugung einer Person, über die Fähigkeiten und Mittel zu verfügen, um diejenigen Handlungen durchführen zu können, die notwendig sind, um ein definiertes Ziel zu erreichen – und zwar auch dann, wenn Barrieren zu überwinden sind“ (BAUMERT & KUNTER, 2006, S. 502; nach BANDURA, 1997). Die praktische Bedeutung von SWE liegt in ihrer regulierenden Wirkung auf die

dem Experimentieren im Geographieunterricht – kommt den Selbstwirksamkeitserwartungen (SWE) von Lehrpersonen als Teil ihrer professionellen Handlungskompetenz eine besondere Rolle zu. Sie gelten im Rahmen der sozial-kognitiven Theorie nach BANDURA (1997) als entscheidend für einerseits die Aufnahme von Handlungen, andererseits jedoch auch für die Ausdauer bzw. Anstrengung, mit der diese auch beim Auftreten von Hindernissen verfolgt werden (PAJARES, 1996; SCHWARZER & JERUSALEM, 2002), sodass dem Konstrukt eine zentrale Bedeutung für das professionelle Handeln von Lehrpersonen zugeschrieben wird (BAUMERT & KUNTER, 2006).

Der geographiedidaktischen Hochschullehre kommt vor diesem Hintergrund und im Rahmen der Professionalisierung zukünftiger Lehrpersonen für den Geographieunterricht die wichtige Aufgabe zu, den Aufbau angemessener SWE zum Experimentieren zu unterstützen. Um die Ausprägung der SWE sowie mögliche Veränderungen sichtbar machen zu können, bedarf es jedoch zunächst valider und gleichzeitig in Einsatz und Auswertung ökonomisch und universell für verschiedene Lernsituationen einsetzbarer Instrumente zur Messung dieser. Darüber hinaus kann ein solches Messinstrument auch gezielt in Seminarsituationen zum Einsatz kommen, um den Studentinnen und Studenten selbst ihre SWE bewusst und somit einer (Selbst-)Reflektion zugänglich zu machen (MARTSCHINKE ET AL., 2007; RABE ET AL., 2012). Im vorliegenden Beitrag wird ein solches Messinstrument zur Erhebung der SWE angehender Lehrpersonen zum Experimentieren im Geographieunterricht validiert und für den Einsatz in der geographiedidaktischen Hochschullehre und Forschung vorgestellt.

Handlungsintention und Zielsetzung einer Person. So beeinflusst sie, ob und mit welcher Ausdauer und Anstrengung komplexe und herausfordernde Handlungen, welche nicht durch einfache Routinen lösbar sind, überhaupt aufgenommen werden (PAJARES, 1996; SCHWARZER & JERUSALEM, 2002; RABE ET AL., 2012). Folglich wirken sich SWE gemeinsam mit dem erwarteten Ergebnis indirekt auf die konkrete Handlung einer Person sowie den Erfolg dieser Handlung aus. MEINHARDT ET AL. (2018, S. 35) beschreiben als Kern des Konstrukts von Bandura

zentrale Eigenschaften, welche sich teilweise bereits in der obigen Definition widerspiegeln und bei der Erhebung von SWE Berücksichtigung finden sollten:

- (a) Selbstreferentialität/Subjektivität (Selbsteinschätzung eigener Fähigkeiten),
- (b) Handlungsbezug (Bezug der Selbsteinschätzung auf konkrete Handlungen),
- (c) Schwierigkeitsbezug (Bezug der Selbsteinschätzung auf Abstufungen von Herausforderungen oder Hindernissen anstelle einfacher Routinen oder trivialer Handlungen, bspw. durch die Formulierung von Handlungsbarrieren) sowie
- (d) Situations- und Domänenspezifität (Abhängigkeit der Selbsteinschätzung von einer bestimmten Handlungssituation und Domäne).

2.1 Lehrkräfteselbstwirksamkeitserwartungen

Die Selbstwirksamkeit einer Person ist als motivationales und volitionales Konstrukt zu verstehen. Auch im Modell professioneller Handlungskompetenz von Lehrpersonen (COACTIV) von BAUMERT und KUNTER (2006) werden SWE neben dem Enthusiasmus innerhalb der motivationalen Orientierungen verortet. Gemeinsam mit der Fähigkeit zur Selbstregulation sind sie „für die psychische Dynamik des Handelns, die Aufrechterhaltung der Intention und die Überwachung und Regulation des beruflichen Handelns über einen langen Zeitraum verantwortlich“ (Baumert & Kunter, 2006, S. 501). Somit kommt SWE auch in Bezug auf das professionelle Handeln und die Entwicklung professioneller Kompetenzen von Lehrpersonen eine bedeutende Rolle zu. SWE gelten dabei als prinzipiell erlern- und veränderbar. Ein zentraler Gegenstand von Forschung liegt folglich im berufsbiographischen Verlauf der SWE von Lehrpersonen, u.a. mit Blick auf ihre Entwicklung und Veränderung im Laufe des Studiums sowie den Einfluss entsprechender Praxisphasen (WESS ET AL., 2020). Bezüglich letzterem berichten einige Studien von höheren Kompetenzeinschätzungen bei Studentinnen und Studenten im Zuge von begleiteten Praxiserfahrungen (u.a. SMOLLECK & MONGAN, 2011; SEUNG ET AL., 2019); andere Studien verzeichnen jedoch einen Rückgang der SWE der angehenden Lehrpersonen im Sinne eines sogenannten Praxischocks (u.a. TSCHANNEN-MORAN ET AL., 1998; HOY & SPERO, 2005; PENDERGAST ET AL., 2011; RABE ET AL., 2012), was unter anderem auf eine theoretische Selbstüberschätzung der Studentinnen und Studenten zu Beginn von Praxisphasen zurückzuführen sein könnte (PENDERGAST ET AL., 2011).

Zahlreiche empirische Studien belegen den Einfluss der SWE von Lehrkräften auf praxisrelevante Aspekte, beispielsweise ihre Zielsetzungen im Unterricht (u.a. GOROZIDIS & PAPAIOANNOU, 2011; CHO & SHIM, 2013), ihren Einsatz von Unterrichtsstrategien sowie ihre Haltung gegenüber Innovation und Veränderung (u.a. HSIAO ET AL., 2011; GOROZIDIS & PAPAIOANNOU, 2011). Weniger selbstwirksame Lehrkräfte neigen in komplexen Situationen stärker zu einem Empfinden von Überforderung (u.a. SCHWARZER & WARNER, 2014) und bevorzugen folglich einfachere, aber sicherere Unterrichtsaktivitäten (u.a. TSCHANNEN-MORAN ET AL., 1998; APPLETON & KINDT, 2002). Lehrkräfte mit höherer Selbstwirksamkeit sind hingegen motivierter und empfinden eine höhere Verantwortung für einen insgesamt herausfordernden Unterricht (SCHWARZER & JERUSALEM, 2002; RABE ET AL., 2012). Auch für Lehrpersonen im naturwissenschaftlichen Bereich deutet sich an, dass höhere SWE mit der in die Unterrichtsplanung investierten Zeit korrelieren und die Qualität des Lehrens dahingehend positiv beeinflussen, dass innovative und herausfordernde Methoden eingesetzt werden (TSCHANNEN-MORAN ET AL., 1998; ROBERTS ET AL., 2001). Gerade in komplexen und herausfordernden Handlungsfeldern – wie dem Experimentieren im Geographieunterricht – könnte positiven SWE folglich eine zentrale Rolle zukommen.

2.2 Erhebung der Selbstwirksamkeitserwartungen (angehender) Lehrpersonen zum Experimentieren

Im Bereich von Lehrpersonen werden SWE auf allgemeiner Ebene, als allgemeine Lehrkräfte-SWE, jedoch auch fachbezogen sowie noch konkreter innerhalb eines Faches auf situationaler Ebene betrachtet. Bezüglich ihrer Handlungsnähe bzw. der Prädiktivität einer Messung von SWE ist es wichtig, den Grad an Spezifität, mit der diese erhoben werden, in den Blick zu nehmen. Allgemeine bzw. unangemessen definierte Messungen können umgekehrt zu einer Fehleinschätzung von Zusammenhängen beitragen, sodass der Grad an Spezifität einer Messung auch der konkreten Anforderung und dem konkreten Interessensbereich entsprechen sollte (PAJARES, 1996; RABE ET AL., 2012; WESS ET AL., 2020).

SWE von Lehrpersonen werden traditionellerweise in Form von Fragebögen zur Selbstauskunft erhoben, vereinzelt kommen auch qualitative Ansätze zum Einsatz (z.B. RENNER, 2021). Insbesondere im Bereich naturwissenschaftlichen Unterrichts existiert eine Vielzahl an Messinstrumenten zur Erhebung der SWE von Naturwissenschaftslehrpersonen, wie beispielsweise das weit verbreitete Instrument *Teaching Science as Inquiry* (TSI) (SMOLLECK ET

AL., 2006) zum forschenden Lernen sowie das *Science Teaching Beliefs Instrument* (STEBI) (ENOCHS & RIGGS, 1990). Letzteres wurde vielfach weiterentwickelt und adaptiert, beispielsweise für den deutschsprachigen Raum zur Erhebung von SWE zum Physik-Lehren (RIESE, 2009) oder zum Unterrichten von Geographie (BENT ET AL., 2017). Darüber hinaus sind geographiespezifische SWE mit wenigen Ausnahmen (z.B. SWE angehender Geographielehrpersonen zu geographischen Exkursionen, RENNER, 2021; medienbezogene Selbstwirksamkeitserwartungen angehender Geographielehrpersonen, SCHWEIZER & HORN, 2010) wenig untersucht. In keinem der genannten Instrumente werden dabei alle der zuvor beschriebenen zentralen Eigenschaften von SWE berücksichtigt, dies gilt insbesondere für den Bezug der Selbsteinschätzung auf Herausforderungen oder Hindernisse durch die Formulierung von Handlungsbarrieren (Schwierigkeitsbezug). Gleichzeitig werden die SWE in fast allen Arbeiten auf einer eher allgemeinen Ebene, also weder mit Bezug auf bestimmte Situationen noch auf konkrete Handlungen (Handlungsbezug und Situationspezifität) erfasst.

Für das vorliegende Vorhaben war jedoch gerade die Betrachtung der SWE auf dem Niveau der Domänen- und Situationspezifität des Experimentierens (Situation) im Geographieunterricht (Domäne) von Bedeutung. Für physikdidaktische Handlungsfelder liegt bereits ein Instrument vor, dass ein solches Spezifitätsniveau berücksichtigt (RABE ET AL., 2012; MEINHARDT ET AL., 2016, 2018). Dieses umfasst neben dem Experimentieren das Elementarisieren, den Umgang mit Aufgaben und den Umgang mit Schülervorstellungen sowie innerhalb jeden Handlungsfelds wiederum die Dimensionen *Planung* und *Durchführung*. Gleichzeitig sind die zuvor beschriebenen Eigenschaften von SWE bei der Entwicklung des Instruments einbezogen: Die Items besitzen alle den Itemstamm *Ich kann...* (Selbstreferentialität) und stellen komplexe Handlungen dar (Kompetenz- bzw. Handlungsbezug). Darüber hinaus werden konkrete Anforderungsbezüge realisiert und Schwierigkeiten in Form von Handlungsbarrieren (... , *auch wenn...*) verbalisiert (Schwierigkeitsbezug). Neben den acht Skalen stellen MEINHARDT ET AL. (2016, 2018) für jedes Handlungsfeld eine Kurzversion vor, welche – je nach Forschungsinteresse und Anwendungsfall sowie als Konsequenz von teilweise sehr hohen latenten Korrelationen zwischen der *Planungs-* und *Durchführungsdimension* – die globale Erfassung eines Handlungsfelds ohne Berücksichtigung der beiden Dimensionen ermöglicht.

Die jeweilige Kurzskaala wurde post-hoc anhand inhaltlicher Kriterien entwickelt, wobei u.a. bezüglich der Eindimensionalität (resp. Unterscheidung

zwischen den Dimensionen *Planung* und *Durchführung*) „[...] potentielle Testanwender in der Verantwortung [sind] ggf. weitere Validierungsmaßnahmen zu ergreifen“ (MEINHARDT ET AL., 2018, S. 145). Das physikdidaktische Instrument ist zwar situationsspezifisch, jedoch inhaltsunabhängig, also nicht auf einen physikalischen Inhaltsbereich (wie beispielsweise Optik oder Mechanik) bezogen, formuliert. Dadurch eignet es sich für eine Adaption in weitere Naturwissenschaftsdidaktiken, so auch in Bezug auf das Handlungsfeld Experimentieren für die Geographiedidaktik. Hier unterscheidet sich das Begriffsverständnis im Kern nicht von anderen Naturwissenschaftsdidaktiken, besitzt das Experimentieren im Geographieunterricht doch ebenfalls das Ziel des Erkennens von Ursache-Wirkungs-Beziehungen und allgemeinen Gesetzmäßigkeiten. Der Unterschied zu diesen Fächern ist vielmehr im Kontext zu verorten: Für die Fächer Chemie, Physik und Biologie beziehen sich Fragestellungen als Ausgangspunkt des Experimentierens vornehmlich auf natürliche Phänomene, während geographische Fragestellungen außerdem „Wechselwirkungen zwischen naturgeographischen Gegebenheiten und menschlichen Aktivitäten“ (DGFG, 2020, S. 5), also Beziehungen zwischen Mensch und Umwelt, einbeziehen können (OTTO ET AL., 2010; WILHELMI, 2012).

Auch ROSENDAHL ET AL. (2020) evaluieren ein neu entwickeltes GEO Lehr-Lern-Labor neben qualitativen Interviews anhand eines teilweise von MEINHARDT ET AL. (2016) sowie teilweise deduktiv u.a. aus BAUMERT und KUNTER (2006) abgeleiteten quantitativen Fragebogens zur Erhebung der SWE von Studentinnen und Studenten hinsichtlich des Einsatzes von Experimenten im Geographieunterricht. Die eindimensionale, bisher noch unveröffentlichte Gesamtskala bildet zwei konkrete, im Seminar fokussierte Bereiche (Beurteilen/Erstellen von Experimentieraufgaben und Erkennen/Beurteilen von Experimentierleistungen) ab, wobei der Schwerpunkt auf der Berücksichtigung heterogener Lernvoraussetzungen beim Experimentieren liegt. Im Rahmen des vorliegenden Beitrags soll jedoch gerade kein an ein spezifisches Seminarkonzept angepasstes, sondern ein universell für verschiedene Lern- und Experimentiersituationen einsetzbares Messinstrument zur Erhebung der SWE beim Experimentieren im Geographieunterricht bereitgestellt werden.

Mit Blick auf die Ausprägung schätzen die bei ROSENDAHL ET AL. (2020) befragten Studentinnen und Studenten ihre situationsspezifischen SWE zum Experimentieren im Geographieunterricht vor Besuch des Seminars im mittleren Bereich ein ($M=2.78$, $SD=0.51$ auf einer fünfstufigen Likertska-

la, $n=29$). Angehende Physiklerpersonen berichten von etwas höheren SWE zum Planen ($M=3.96$, $SD=0.77$ auf einer sechsstufigen Likertskala, $n=525$) sowie zum Durchführen von Experimenten ($M=4.12$, $SD=0.74$ auf einer sechsstufigen Likertskala, $n=525$), wobei ebenfalls befragte Referendarinnen und Referendare signifikant niedrigere Werte, Physiklerkräfte hingegen die höchsten

Werte aufweisen (MEINHARDT ET AL., 2016). Auch angehende Chemielehrpersonen besitzen etwas höhere SWE zum selbstgesteuerten Experimentieren ($M=3.07$ auf einer vierstufigen Likertskala, $n=29$) (EHLERT & TEPNER, 2023). Ähnliche Ergebnisse zeigen sich für Studentinnen und Studenten mit Blick auf SWE zum forschenden Lernen (u.a. SMOLLECK & MONGAN, 2011; SEUNG ET AL., 2019).

3. Zielstellung und Forschungsfragen

Das Ziel des vorliegenden Beitrags liegt in der Bereitstellung eines Messinstruments zur Erhebung der SWE von Lehramtsstudentinnen und Lehramtsstudenten zum Experimentieren im Geographieunterricht für die geographiedidaktische Hochschullehre und Forschung (Fragebogen vgl. [Anhang 1](#), [Anhang 2](#)). Konkret sollen die SWE durch einen validen und gleichzeitig im Einsatz und der Auswertung ökonomisch handhabbaren sowie universell in verschiedenen Lernsituationen (beispielsweise Seminarveranstaltungen oder (Schul-)Praktika) einsetzbaren Fragebogen, welcher zuverlässige Informationen zu den SWE von Studentinnen und Studenten und deren Veränderungen liefert, empirisch erfassbar gemacht werden. Hierfür wurde ein auf das Experimentieren bezogener Teil des Messinstruments von RABE ET AL. (2012) bzw. MEINHARDT ET AL. (2016, 2018) zur Erhebung der SWE in physikdidaktischen Handlungsfeldern geographiespezifisch adaptiert. Dieses Instrument schien insbesondere für eine Adaption geeignet, da es erstens die beschriebenen Eigenschaften zur Messung des Konstrukts, insbesondere die geforderte Situationspezifität bzw. Kontextabhängigkeit, erfüllt und zweitens inhaltsunabhängig entwickelt wurde.

Im Rahmen der Adaption muss – mit Blick auf die neue Zielgruppe der Geographielehramtsstudentinnen und Geographielehramtsstudenten und die Auswahl der post-hoc entwickelten und

weiterführend zu validierenden Kurzversion – die Validität und Reliabilität des geographiespezifischen Instruments überprüft werden. Die zentrale Forschungsfrage des Beitrags lautet folglich:

Können die Selbstwirksamkeitserwartungen angehender Lehrpersonen zum Experimentieren im Geographieunterricht anhand des geographiespezifisch adaptierten Messinstruments von MEINHARDT ET AL. (2016) valide und reliabel erhoben werden?

Zusätzlich kann die im Rahmen der Validierung stattfindende Datenerhebung erste Hinweise auf die Ausprägung der SWE angehender Geographielehrpersonen liefern. Der Beitrag geht folglich zusätzlich der folgenden Fragestellung nach: *Wie sind die Selbstwirksamkeitserwartungen zum Experimentieren im Geographieunterricht bei den untersuchten angehenden Lehrpersonen ausgeprägt?* Da das Experimentieren im Geographieunterricht von (angehenden) Lehrpersonen als herausfordernd wahrgenommen wird (MIENER & KÖHLER, 2013; HÖHNLE & SCHUBERT, 2016; FÖGELE ET AL., 2020) sowie angesichts der vorgestellten empirischen Befunde zu Ausprägungen der SWE angehender Lehrpersonen zum Experimentieren (u.a. MEINHARDT ET AL., 2016; ROSENDAHL ET AL., 2020; EHLERT & TEPNER, 2023) gehen wir von geringen bis mittleren SWE der Studentinnen und Studenten zum Experimentieren im Geographieunterricht aus.

4. Methodik

Da das Begriffsverständnis von Experimenten in der Geographiedidaktik im Kern mit dem in der Physik- sowie in den weiteren Naturwissenschaftsdidaktiken vergleichbar ist (vgl. Kap. 2.2), wird bezüglich der Inhaltsvalidität des adaptierten Messinstruments auf die im Rahmen der Instrumententwicklung durch MEINHARDT ET AL. (2018) durchgeführten Expertinnen- und Expertenbefragungen und Interviewstudien mit den drei Befragungsgruppen (Physiklehramtsstudentinnen und Physiklehramtsstudenten, Physiklehramtsanwärterinnen

und Physiklehramtsanwärter und Physiklerpersonen) verwiesen.

Im Rahmen der vorliegenden Studie wurde folglich auf weiterführende qualitative Pretests verzichtet. So wird mit Blick auf die neue Stichprobe der Geographielehramtsstudentinnen und Geographielehramtsstudenten sowie die post-hoc anhand inhaltlicher Kriterien entwickelte, beide Dimensionen umfassende Kurzversion des Instruments der Fokus im vorliegenden Beitrag auf die Überprüfung der Konstruktvalidität gelegt. Hierbei

muss u.a. die Dimensionalität der Kurzversion in Bezug auf die *Planungs-* und *Durchführungsdimension* (neun Items umfassende eindimensionale Gesamtskala vs. zweidimensionale Skala mit den Dimensionen *Planung* [vier Items] und *Durchführung* [fünf Items], Fig. 1) überprüft werden.

Weiterführend werden, um die Frage nach der Reliabilität und Validität des adaptierten Messinstruments zu beantworten, analog zu RABE ET AL. (2012) bzw. MEINHARDT ET AL. (2016, 2018) folgende zentrale Schritte unternommen: Überprüfung der Reliabilität, Evaluation der Eignung für verschiedene Subgruppen (Geschlecht, Schulform, Praxiserfahrung), Überprüfung der Abgrenzbarkeit zu weiteren Konstrukten sowie Mittelwertvergleiche verschiedener Gruppe von Studentinnen und Studenten. Dazu wird zunächst die Stichprobe und das Messinstrument beschrieben, bevor das Vorgehen bei den Validierungsschritten genauer dargestellt wird.

4.1 Stichprobe

An der Befragung nahmen $n=303$ Geographielehramtsstudentinnen und Geographielehramtsstudenten teil. Zur Online-Datenerhebung zwischen Juni und November 2021 wurde die Software SoSci Survey (LEINER, 2021) genutzt, indem den Studentinnen und Studenten ein Link zur freiwilligen Teilnahme an der Studie zugänglich gemacht wurde. Die Stichprobe teilte sich in 70% Studentinnen ($n=211$), 24% Studenten ($n=72$) sowie 1% mit der Angabe divers ($n=2$) (6% mit fehlender Angabe, $n=18$), was der bundesweiten Geschlechterverteilung im Lehramt im Wintersemester 2021/22 entspricht (STATISTISCHES BUNDESAMT, 2022). Geschlechtsspezifische Unterschiede bezüglich der SWE zum Experimentieren im Geographieunterricht bestanden nicht ($t(281)=-1.29$; $p>.05$). Das Durchschnittsalter betrug zum Zeitpunkt der

Erhebung 22,26 Jahre ($SD=3.70$) und die Studentinnen und Studenten besuchten das erste bis 13. Fachsemester ($M=4.62$, $SD=3.04$; $n=99$ im ersten und zweiten Fachsemester [35%], $n=51$ im dritten und vierten Fachsemester [18%], $n=56$ im fünften und sechsten Fachsemester [18,5%], $n=48$ im siebten und achten Fachsemester [15,9%] und $n=29$ im neunten Fachsemester oder höher [9,6%], 6,6% mit fehlender Angabe). Jeweils 37% studierten Grundschullehramt ($n=106$) und Gymnasiallehramt ($n=105$) und jeweils 13% Mittelschullehramt ($n=36$) und Realschullehramt ($n=38$) (6% mit fehlender Angabe, $n=18$). Etwas über die Hälfte der Studentinnen und Studenten ($n=157$) hatte zum Zeitpunkt der Befragung noch keine unterrichtspraktische Erfahrung im Rahmen des studienbegleitenden fachdidaktischen Praktikums im Studium gesammelt. Insgesamt wurden Studentinnen und Studenten an sieben Universitäten in Bayern befragt.

4.2 Messinstrumente

Zur Erhebung der SWE angehender Lehrpersonen zum Experimentieren im Geographieunterricht wurde das bereits beschriebene physikbezogene Messinstrument von MEINHARDT ET AL. (2016) geographiespezifisch adaptiert. Dieses Instrument wurde theoriegeleitet entworfen und mehrschrittig für die drei Befragungsgruppen weiterentwickelt und validiert. Für eine ausführliche Beschreibung der Entwicklung und Validierung des Messinstruments inklusive Konstruktionsvorschriften und Itemkonstruktionsregeln wird auf RABE ET AL. (2012), MEINHARDT ET AL. (2018) sowie die Skalendokumentation in MEINHARDT ET AL. (2016) verwiesen.

Für unser Vorhaben wurde aus den vier Handlungsfeldern das Feld Experimentieren sowie die aus neun Items bestehende Kurzversion des Instruments ausgewählt. Diese deckt die folgenden Themenfelder innerhalb der *Planung* und *Durchführung*

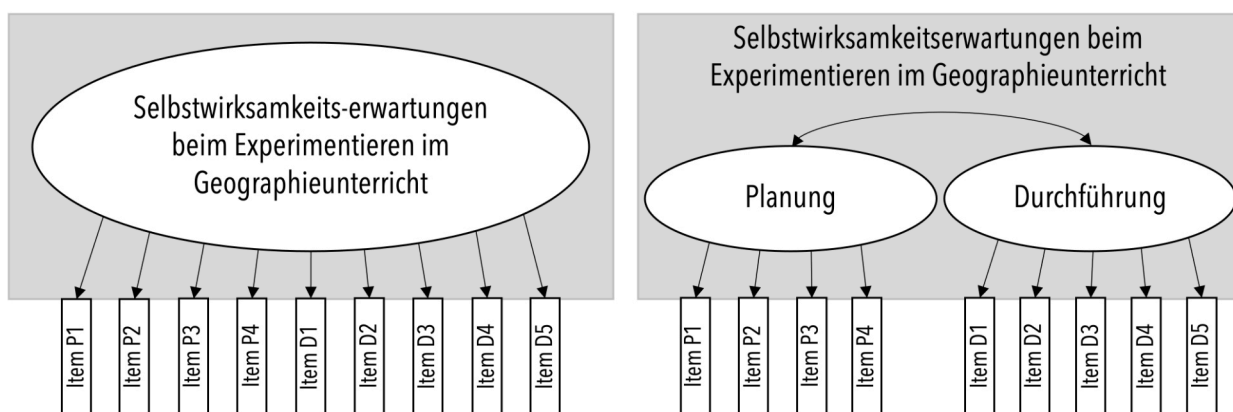


Fig. 1. Eindimensionales Modell mit Globalfaktor (Modell 1) (links) vs. zweidimensionales Modell mit Unterscheidung zwischen der *Planungs-* und *Durchführungsdimension* (Modell 2) (rechts) (Ovale: latente Skalen; Rechtecke: Einzelitems innerhalb der Skalen SWE beim Experimentieren im Geographieunterricht - *Planung* [Item P1-P4] bzw. - *Durchführung* [Item D1-D5]) (Quelle: Autorin und Autor)

zung von Experimenten ab: Auswahl von Experimenten (*exp1, exp7*), Variation von Experimenten (*exp2*), Zusammenstellung von Experimenten (*exp3*), Durchführung/Variation von Experimenten (*exd1, exd2, exd5*), Ermöglichung von Reflexionsprozessen (Theorie-Experiment-Beziehung) (*exd3*), Inszenierung von Experimenten (*exd8*). Zur geographiespezifischen Adaption des Messinstruments wurden zunächst die physikbezogenen Formulierungen jeweils auf die Geographie übertragen. Hierfür wurden Begriffe wie *Physik(unterricht)* durch *Geographie(unterricht)* ersetzt (bspw. *Ich kann in meiner Unterrichtsplanung zu den Lernzielen passende Experimente aufbauen, auch wenn die Geographie [Physik] dafür schlecht ausgestattet ist. [exp1]*). Andere Items ohne physikspezifische Formulierung (bspw. *exd2*) konnten ohne Anpassung beibehalten werden. Die genauen Itemformulierungen sind [Fig. 3](#) zu entnehmen, der gesamte Fragebogen zu den SWE zum Experimentieren im Geographieunterricht ist im Anhang beigefügt ([Anhang 1, Anhang 2](#)).

Das so entstandene geographiespezifische Instrument wurde weiterführend im Rahmen eines geographiedidaktischen Seminars zum Experimentieren an einer kleineren studentischen Stichprobe ($n=18$) getestet, um die angepassten Items hinsichtlich ihrer grundsätzlichen Eignung anhand von Itemkennwerten (Itemtrennschärfe und Itemschwierigkeiten) zu prüfen. Da die Ergebnisse des quantitativen Pretests auf grundsätzliche Eignung der adaptierten Items hinwiesen, wurde das Instrument unverändert eingesetzt.

Da davon auszugehen ist, dass die Vorstellungen der befragten Studentinnen und Studenten von Experimenten im wissenschaftlichen Sinne abweichen könnten (LETHMATE, 2006; MIENER & KÖHLER, 2013; FÖGELE ET AL., 2020), wurde zudem ein knapper Informationsblock mit Definition und Beispielen zu naturwissenschaftlichen Arbeitsweisen im Geographieunterricht (Untersuchungen, Modelle, Experimente) formuliert und dem Fragebogen gemeinsam mit dem Hinweis, sich bei der Beantwortung der Items erstens auf die darin angeführte Definition von Experimenten und zweitens auf das Experimentieren speziell im Geographieunterricht zu beziehen, vorangestellt ([Anhang 1](#)). Ziel dessen war es, ein gemeinsames Verständnis der Begrifflichkeiten zu schaffen sowie die Domänen- und Situationsspezifität der Items hervorzuheben.

Neben der Kurzskala zu SWE zum Experimentieren wurden im Fragebogen demographische Informationen sowie das Fähigkeitsselbstkonzept in Bezug auf Physische Geographie (FSK-PG, 4 Items, z.B. *Es fällt mir leicht, neue physisch-geographische Inhalte zu verstehen.*) und SWE im Hinblick auf das Unterrichten physisch-geographischer

Themen (SWE-PGU, 3 Items, z.B. *Ich fühle mich kompetent genug, physisch-geographische Themen im Geographieunterricht zu behandeln.*) erfasst. Dazu wurden sprachlich auf physisch-geographische Anteile des Geographieunterrichts angepasste Skalen aus KAUERTZ ET AL. (2013) verwendet. Die Skalen bezogen sich ausschließlich auf den physisch-geographischen Anteil der Geographie, da Experimente im Geographieunterricht üblicherweise, wenn auch nicht ausschließlich, mit physiogeographischem Bezug durchgeführt werden. Ziel des Einsatzes war es, diese analog zu MEINHARDT ET AL. (2018) als externale Validierungsfacetten zu nutzen, wobei niedrige bis mittlere Korrelationen als Hinweis auf die konvergente Validität, gleichzeitig jedoch auch auf die Abgrenzbarkeit der SWE zum Experimentieren zu weiteren ähnlichen Konstrukten gewertet wurden. Alle Items waren auf einer fünfstufigen Likert-Skala (1=*stimme gar nicht zu* bis 5=*stimme vollkommen zu*) mit grafischer Unterstützung zur Darstellung der Äquidistanz der Items zu beantworten.

4.3 Statistische Analysen

Für die klassische Item- und Skalenanalyse sowie inferenzstatistische Verfahren wurde die Software SPSS Statistics 29 genutzt, die konfirmatorischen Faktorenanalysen, Messinvarianzevaluation und latenten Korrelationen wurden in R (version 4.1.0) anhand der Pakete *lavaan* (ROSSEEL, 2012) und *semTools* (JORGENSEN ET AL., 2018) berechnet. Die Analyse fehlender Werte brachte keine Datenlücken hervor.

Im Rahmen der Konstruktvalidierung des adaptierten Instruments wurden zunächst konfirmatorische Faktorenanalysen (KFA) durchgeführt. Zur Beurteilung der Anpassungsgüte der Modelle wurden der Satorra-Bentler χ^2 -Test als absoluter Fit-Index, der *Comparative Fit Index* (CFI), der *Tucker-Lewis-Index* (TLI), der *Root Mean Square Error Of Approximation* (RMSEA) mit 90%-Konfidenzintervall sowie der Standardised Root Mean Square Residual (SRMR) herangezogen (BROWN, 2015). Als Cutoff-Werte für die jeweiligen Fit-Indizes wurden die Empfehlungen von HU und BENTLER (1999; siehe auch BROWN, 2015) verwendet: $CFI/TLI \geq .95$, $RMSEA \leq .06$ und $SRMR \leq .08$ und $\chi^2/df < 2$. Außerdem wurden folgende Werte als akzeptabler Fit beurteilt: $CFI/TLI \geq .90$, $RMSEA \leq .08$ und $SRMR \leq .11$ (BENTLER, 1995; HU & BENTLER, 1999). Da für alle Items anhand des Kolmogorov-Smirnov- und des Shapiro-Wilk-Tests eine signifikante Abweichung der Daten von der Normalverteilung angezeigt wurde (linksschiefe Verteilung), wurden alle Berechnungen mit dem robusten MLR-Schätzer (ro-

bust maximum likelihood estimation) durchgeführt (BROWN, 2015). Der erste Indikator jeden Faktors wurde für die Faktorenidentifikation auf 1 festgelegt, alle anderen Ladungen der Faktoren wurden frei geschätzt.

Zur Überprüfung der Dimensionalität des Konstrukts (Fig. 1) wurden zwei separate KFA-Modelle berechnet: erstens ein eindimensionales Modell (Globalfaktor ohne Subskalen, Modell 1) und zweitens ein alternatives zweidimensionales Modell zur Abbildung des Konstrukts in den Dimensionen *Planung* und *Durchführung*, in welchem beide Faktoren frei miteinander korrelieren konnten (Modell 2). Zur statistischen Ermittlung des vorzuziehenden Modells wurden beide konkurrierenden Modelle anhand eines Satorra-Bentler skalierten χ^2 -Differenzentest verglichen. Darüber hinaus wurden mögliche Modellanpassungen in Form von Residualkorrelationen anhand von Modifikationsindizes betrachtet, anschließend zur statistischen Absicherung mittels χ^2 -Differenzentest miteinander verglichen und bei inhaltlicher Interpretierbarkeit in die jeweiligen Modelle integriert. Für die Beurteilung der internen Konsistenz wurden Cronbach's Alpha und die Faktorreliabilität (kongenerische Reliabilität, gemessen durch $\rho_{\epsilon} \geq .60$; HAIR ET AL., 2014) herangezogen.

Im Zuge der Konstruktvalidierung wurde weiterführend analog zu MEINHARDT ET AL. (2018) anhand von Korrelationsanalysen geprüft, ob die geographiebezogenen SWE gegenüber anderen, benachbarten Konstrukten (vgl. Kap. 4.2) abgrenzbar sind. Darüber hinaus wurden - mit dem Ziel ein valides Instrument für den Einsatz in Seminarsituationen mit heterogenen Lerngruppen (bspw. in Bezug auf die studierte Schulform und Praxiserfahrung) zu erhalten - Messinvarianzevaluationen (In-

varianz der Indikatorintercepts) für verschiedene personenbezogene Variablen durchgeführt. Hierfür wurden im Rahmen der KFA Mehrgruppen-Faktorenanalysen durchgeführt, in denen zur Bestimmung des Grads an Messinvarianz schrittweise nach dem step-up-Ansatz zunehmende Restriktionen unternommen wurden: (1) äquivalente Ladungsmuster und Anzahl der Faktoren (konfigurale Invarianz), (2) äquivalente Faktorladungen (metrische Invarianz), (3) äquivalente Intercepts/Regressionskonstanten (skalare/starke Invarianz) (BROWN, 2015; SCHWAB & HELM, 2015). Das Vorliegen von skalarer Messinvarianz gilt als Voraussetzung für die Interpretation latenter Mittelwerte und Korrelationen, da somit itemspezifische Schwierigkeitsunterschiede zwischen den Gruppen ausgeschlossen werden können (SCHWAB & HELM, 2015). Die Modellfit-Vergleiche zwischen den schrittweise restriktiveren Modellen wurden anhand des Satorra-Bentler-skalierten χ^2 -Differenzentest sowie basierend auf Chen's Faustregel (CHEN, 2007) evaluiert, nach der Messinvarianz vorliegt, sofern der RMSEA nicht um .015 Einheiten steigt und der CFI nicht um .01 Einheiten sinkt.

Abschließend wurden zur Konstruktvalidierung anhand von t-Tests Mittelwertvergleiche bezüglich der Praxiserfahrungen der Studentinnen und Studenten durchgeführt. Hierbei wurden, da „[...] in der Literatur von einem Rückgang der SWE bei ersten Unterrichtserfahrungen berichtet wird“ (RABE ET AL., 2012, S. 309), niedrigere SWE bei Studentinnen und Studenten, welche im Studium bereits Praxiserfahrungen im Rahmen des studienbegleitenden fachdidaktischen Praktikums gesammelt hatten, erwartet und als Hinweis auf die Validität des Messinstruments gewertet (RABE ET AL., 2012).

5. Ergebnisse

5.1 Konfirmatorische Faktorenanalysen, Dimensionsvergleich und Reliabilität

Die Fit-Indizes der genesteten Modelle sowie die Modellvergleiche sind in Fig. 2 dargestellt. Die Indikatoren luden sowohl für das ein- als auch für das zweidimensionale Modell jeweils signifikant und substantiell auf die jeweiligen Faktoren (Fig. 3). Das eindimensionale Modell 1 wies zunächst einen unzureichenden Modellfit auf, für das zweidimensionale Modell 2 wurde zunächst ein akzeptabler Modellfit angezeigt. Auch ein Vergleich anhand des χ^2 -Differenzentests wies auf eine statistisch bessere Passung der Daten zu dem zweidimensionalen Modell 2 hin. Dabei fiel die latente Korrelation zwischen der *Planungs-* und *Durchführungsdimension*

mit $r = .86$ sehr hoch, wenn auch unter der von MEINHARDT ET AL. (2016) berichteten latenten Korrelation, aus. Folglich bildet das zweidimensionale Modell die Daten besser ab als eine Gesamtskala zum Experimentieren.

Um signifikante Verbesserungen der zunächst lediglich akzeptablen bzw. unzureichenden Modelle zu bestimmen, wurden die Modifikationsindizes beider Modelle betrachtet. Besonders auffällig war das Item *exp3* der Dimension *Planung*, da es die höchste Summe aller Modifikationsindizes in Form von Residualkorrelationen mit mehreren Items aufwies. Als Konsequenz kam entweder ein Entfernen des Items aus der Skala oder ein Zulassen inhaltlich interpretierbarer Residualkorrelationen in Betracht. Ein ledigliches Entfernen des

Modell	$\chi^2_{SB}(p)$	df	χ^2/df	CFI	TLI	RMSEA [90% CI]	SRMR	Modell- vergleich	$\Delta\chi^2_{SB}(p)$	Δdf
Modell 1: eindimensionales Modell	77.530 ($<.001$)	27	2.87	.91	.88	.09 [.06, .11]	.05	1 vs. 1b	14.854 ($<.001$)	2
Modell 1b: revidiertes eindimensionales Modell	62.888 ($<.001$)	25	2.52	.93	.90	.08 [.05, .10]	.05	1b vs. 2b	14.399 ($<.001$)	1
Modell 2: zweidimensionales Modell	68.930 ($<.001$)	26	2.65	.93	.90	.08 [.06, .10]	.05	1 vs. 2	7.168 ($<.01$)	1
Modell 2b: revidiertes zweidimensionales Modell	44.888 ($<.01$)	24	1.87	.96	.95	.06 [.03, .08]	.05	2 vs. 2b	24.526 ($<.001$)	2

χ^2_{SB} : Satorra-Bentler skaliertes χ^2 test; df: Freiheitsgrade; CFI: Comparative Fit Index; TLI: Tucker-Lewis-Index; RMSEA: Root Mean Square Error of Approximation; SRMR: Standardized Root Mean Residual; $\Delta\chi^2_{SB}$ bzw. Δdf = Änderung im χ^2 bzw. den Freiheitsgraden zwischen den beiden verglichenen Modellen.

Fig. 2. Modellfits der konkurrierenden und revidierten Messmodelle: Ergebnisse der konfirmatorischen Faktorenanalysen und Modellvergleiche (Quelle: Autorin und Autor)

Items zur Verbesserung des Modellfits wurde auf Grundlage inhaltlicher Überlegungen jedoch verworfen: das Messinstrument sollte erstens lediglich angepasst und nicht in seiner Struktur verändert werden, zweitens deckt das Item als einziges das Themenfeld Zusammenstellen von Experimenten ab. Beim Überprüfen der Residualkorrelationen ergaben sich jedoch nur zwei inhaltlich sinnvoll interpretierbare Anpassungen des Modells: zunächst wurde eine Residualkorrelation zwischen Item *exp3* und *exd3* zugelassen, da beide Items die gleiche Handlungsbarriere besitzen ([...] *auch wenn ich [bei der Planung] unter Zeitdruck stehe*). Zudem wurde ein methodischer Effekt vermutet, da beide Items – möglicherweise aufgrund dieser als *schwieriger* beurteilten Handlungsbarriere – die höchste Itemschwierigkeit aufwiesen. Die Residualkorrelation zwischen beiden Indikatoren betrug im eindimensionalen Modell $r=.25$ bzw. $r=.31$ im zweidimensionalen Modell. Darüber hinaus erschien eine weitere Residualkorrelation zwischen den Items *exp3* und *exd1* inhaltlich interpretierbar, da letzteres trotz Operationalisierung innerhalb der Dimension *Durchführung* in seiner Handlungsbarriere den Aspekt der *Planung* enthält ([...] *auch wenn ich es vorher nicht geplant hatte*). Beiden Items ist folglich ein limitierender Faktor im Bereich der *Planung* gemeinsam. Hier betrug die Residualkorrelation im eindimensionalen Modell $r=.14$ bzw. $r=.20$ im zweidimensionalen Modell. Nach Zulassen beider Residualkorrelationen wies das eindimensionale Modell 1b einen akzeptablen, das zweidimensionale Modell 2b einen guten Modell-Fit auf. Auch ein Vergleich der revidierten Modelle an-

hand des χ^2 -Differenzentests zeigt die statistisch beste Passung der Daten zu dem revidierten zweidimensionalen Modell 2b an (Fig. 2). Die Korrelation zwischen beiden Dimensionen war mit $r=.80$ weiterhin hoch, dennoch lag sie unter der Korrelation in Modell 2. Angesichts des akzeptablen Modell-Fits des eindimensionalen Modells 1b und der hohen Korrelation zwischen beiden latenten Dimensionen *Planung* und *Durchführung* ist auch der Einsatz des Messinstruments als Globalskala denkbar.

Die Reliabilitätsmaße (Fig. 4) waren sowohl für Cronbach's Alpha als auch für die Faktorreliabilität für das Gesamtmodell 2 gut und in einem ähnlichen Bereich wie bei MEINHARDT ET AL. (2016). Für die beiden Skalen des zweidimensionalen Modells 2b lagen beide Werte im akzeptablen Bereich, wobei die niedrigeren Werte der beiden Skalen auf die niedrigere Itemanzahl zurückgeführt werden. Innerhalb der Skala SWE zum Experimentieren – *Planung* würde die Reliabilität durch Entfernen des Items *exp3* steigen, dennoch wurde sich auf Grundlage inhaltlicher Überlegungen erneut für eine Beibehaltung des Items entschieden. Auf Einzelitemebene lagen die Itemschwierigkeiten zwischen $P=.54$ und $P=.71$, die Item-Skala-Korrelationen lagen für alle Items zwischen $r_{it}=.39$ und $r_{it}=.58$.

5.2 Messinvarianzevaluation

Die Messinvarianzevaluationen (skalare Messinvarianz bzw. Invarianz der Intercepts) wurden mit dem revidierten zweidimensionalen Modell 2b, welches die beste Passung zu den Daten aufwies, in Bezug auf das Geschlecht, die studierte Schul-

Abk.	Itemformulierung	M (SD)	λ Modell 1b (SE)	λ Modell 2b (SE)
Items der Skala Experimentieren – Planung				
exp1	Ich kann in meiner Unterrichtsplanung zu den Lernzielen passende Experimente aufbauen, auch wenn die Geographie dafür schlecht ausgestattet ist.	3.14 (.92)	.51 (.06)	.57 (.06)
exp2	Ich kann bei meiner Unterrichtsplanung ein Experiment ggf. so variieren, dass ich es in einer Übungsphase einsetzen kann, auch wenn ich es bisher nur als Einstiegsexperiment genutzt habe.	3.45 (.97)	.62 (.06)	.73 (.06)
exp3	Ich kann Schülerexperimente so zusammenstellen, dass die Fähigkeiten meiner Schülerinnen und Schüler auf verschiedenen Niveaus gefördert werden, auch wenn ich bei der Planung unter Zeitdruck stehe.	3.00 (1.11)	.46 (.07)	.52 (.07)
exp7	Ich kann für ein geographisches Experiment begründet entscheiden, ob es didaktisch sinnvoller ist, es als Demonstrations- oder Schülerexperiment einzuplanen, auch wenn ich das Experiment noch nicht eingesetzt habe.	3.51 (1.01)	.45 (.07)	.42 (.08)
Items der Skala Experimentieren – Durchführung				
exd1	Ich kann geographische Experimente an interessante Impulse meiner Schülerinnen und Schüler anpassen, auch wenn ich das vorher nicht geplant hatte.	3.35 (1.01)	.60 (.06)	.61 (.06)
exd2	Ich kann ein Demonstrationsexperiment für meine Schülerinnen und Schüler nachvollziehbar durchführen, auch wenn es sich um eine komplexe Versuchsanordnung handelt.	3.58 (1.02)	.63 (.07)	.65 (.06)
exd3	Ich kann beim Experimentieren spontan mit den Schülerinnen und Schüler über das Wechselspiel von Theorie und Experiment reflektieren, auch wenn ich den Anlass nicht vorgesehen hatte.	3.54 (1.01)	.69 (.06)	.72 (.06)
exd5	Ich kann ein Experiment, das im Geographieunterricht nicht auf Anhieb funktioniert, zum Laufen bringen, auch wenn ich unter Zeitdruck stehe.	3.01 (.96)	.59 (.06)	.60 (.06)
exd8	Ich kann ein Experiment so inszenieren, dass meine Schülerinnen und Schüler motiviert sind mitzuarbeiten, auch wenn es sich um eine unbeliebte Randstunde handelt.	3.82 (.95)	.61 (.07)	.60 (.07)

Items adaptiert nach MEINHARD ET AL. (2016) bei fünfstufiger Likert-Skala mit Ausprägungen *stimme gar nicht zu* (1) bis *stimme vollkommen zu* (5); λ = Faktorladung; SE = Standardfehler; alle Faktorladungen sind auf dem $p < .001$ -Niveau statistisch signifikant.

Fig. 3. Items, Mittelwerte und Standardabweichungen (in Klammern) sowie Faktorladungen mit Standardfehler (in Klammern) für das eindimensionale revidierte Modell 1b und das zweidimensionale revidierte Modell 2b (Quelle: Autorin und Autor)

form und die Praxiserfahrung im Studium durchgeführt. Um Messinvarianz in Bezug auf das Geschlecht zu testen, wurden die Daten der beiden Personen, welche *divers* angaben, als *Missing* kodiert. In Bezug auf die studierte Schulform wurden wie auch bei MEINHARDT ET AL. (2016) die beiden Gruppen Realschule und Mittelschule (Sekundarstufe I) zusammengefasst. Die Mehrgruppen-Analysen konnten skalare Messinvarianz in Bezug auf das Geschlecht und die studierte Schulform zeigen, da eine Betrachtung der Modellfit-Indizes keine substantiellen Unterschiede zwischen den Gruppen anzeigte. In Bezug auf die Praxiserfahrung während des Studiums war der χ^2 -Test, welcher das skalare

Modell und das weniger restriktive metrische Modell verglichen, signifikant ($\Delta\chi^2=21.840$, $\Delta df=7$, $p<.01$). Auch die weiteren Indizes zeigten substantielle Unterschiede zwischen den Modellen an ($\Delta CFI=.026$, $\Delta RMSEA=.009$). Deshalb wurde untersucht, ob die skalare Invarianz zumindest partiell erreicht werden konnte. Bei freier Schätzung des Intercepts des Items *exp3* zeigte sich partielle Messinvarianz in Bezug auf die Praxiserfahrung im Studium ($\Delta\chi^2=12.037$, $\Delta df=6$, $p>.05$). Diese Ergebnisse lassen den Schluss zu, dass das Messinstrument für alle untersuchten Gruppen gültig und für Vergleiche zwischen diesen geeignet ist (BROWN, 2015).

Konstrukt	M (SD)	α	ρ_c	SWE-PGU	FSK-PG
SWE zum Experimentieren – Gesamtskala	3.38 (.64)	.82	.80	.48*** (.04)	.36*** (.04)
SWE zum Experimentieren – Planung	3.27 (.69)	.63	.65	.24* (.04)	.21*(.04)
SWE zum Experimentieren – Durchführung	3.46 (.71)	.77	.78	.55*** (.05)	.40*** (.05)
SWE zum Unterrichten von physisch-geographischen Themen (SWE-PGU)	3.87 (.82)	.81	.83	-	.74*** (.05)
Fähigkeitsselbstkonzept bzgl. Physischer Geographie (FSK-PG)	3.61 (.81)	.84	.83	-	-

*** $p < .001$; ** $p < .01$; * $p < .05$; $N = 279$; fünfstufige Likert-Skala mit Ausprägungen *stimme gar nicht zu* (1) bis *stimme vollkommen zu* (5); α = Cronbach's Alpha; ρ_c = Faktorreliabilität (kongenerische Reliabilität).

Fig. 4. Mittelwerte, Standardabweichungen (in Klammern) und Reliabilitäten (Cronbach's Alpha und Faktorreliabilität) der revidierten Skalen sowie latente Korrelationen mit externalen Facetten (Quelle: Autorin und Autor)

5.3 Korrelative Validierung

Zur Überprüfung der Abgrenzbarkeit der SWE zum Experimentieren gegen weitere, benachbarte Konstrukte wurden im Rahmen der KFA Korrelationsanalysen durchgeführt. Es ergaben sich erwartungsgemäß für die Gesamtskala sowie für beide Dimensionen moderate Korrelationen mit dem Fähigkeitsselbstkonzept bezüglich Physischer Geographie sowie etwas höhere Korrelationen mit dem enger vergleichbaren Konstrukt SWE zum Unterrichten von physisch-geographischen Themen (Fig. 4). Dies deutet im Sinne einer konvergenten Validität darauf hin, dass die erhobenen Konstrukte zwar Gemeinsamkeiten besitzen, jedoch hinrei-

chend gut gegen benachbarte Konstrukte abgrenzbar sind. Folglich messen sie unterschiedliche Konstrukte, was außerdem erste Hinweise auf die Diskriminanzvalidität liefert. Innerhalb des zweidimensionalen Modells lagen die Korrelationen beider externaler Facetten mit der Dimension *Durchführung* höher als mit der Dimension *Planung*. Dies erschien plausibel, da die Skala SWE-PGU vorwiegend auf die Durchführung physisch-geographischen Unterrichts abzielt.

5.4 Validierung durch erste Mittelwertvergleiche

Abschließend wurden zur Konstruktvalidierung Mittelwertvergleiche für die *Planungs-* und *Durch-*

Konstrukt	Gruppe	M (SD)	p	Effektstärke d
SWE zum Experimentieren -Planung	mit Praxis	3.22 (.69)	.021*	.24
	ohne Praxis	3.38 (.66)		
SWE zum Experimentieren - Durchführung	mit Praxis	3.43 (.70)	.056	.19
	ohne Praxis	3.56 (.68)		

* $p < .05$ (einseitig); fünfstufige Likert-Skala mit Ausprägungen *stimme gar nicht zu* (1) bis *stimme vollkommen zu* (5); $n=285$.

Fig. 5. Mittelwerte und Standardabweichungen (in Klammern) für Studentinnen und Studenten mit und ohne Praxiserfahrung im Studium sowie Ergebnisse des Vergleichs beider Gruppen (Quelle: Autorin und Autor)

führungsdimension für Studentinnen und Studenten mit und ohne Praxiserfahrungen im Rahmen des Lehramtsstudiums durchgeführt. Diese zeigten für beide Dimensionen erwartungsgemäß leicht höhere SWE bei Studentinnen und Studenten ohne Praxiserfahrung an, wobei nur für die Dimension *Planung* ein signifikanter geringer Effekt vorlag (Fig. 5). Die niedrigeren SWE bei Studentinnen und Studenten mit Praxiserfahrungen im Studium können als erster Hinweis auf die Validität der Skalen gedeutet werden (RABE ET AL., 2012).

6. Diskussion

6.1 Zusammenfassung und Implikationen

Ziel des vorliegenden Beitrags war die Validierung eines Messinstruments zur Erhebung der SWE angehender Lehrpersonen zum Experimentieren im Geographieunterricht. Mit dem aus der Physikdidaktik geographiespezifisch adaptierten Fragebogen liegt ein valides und reliables Messinstrument für den Einsatz in der geographiedidaktischen Hochschullehre und Forschung vor (vgl. Anhang 2). Anhand des Messinstruments können die SWE angehender Geographielehrpersonen hinsichtlich der beiden Dimensionen *Planung* und *Durchführung* von Experimenten im Geographieunterricht erfasst werden. Die Ergebnisse der KFA sowie die Modellvergleiche sprechen für eine solche zweidimensionale Erhebung der SWE zum Experimentieren, wobei angesichts der hohen Korrelation der beiden Dimensionen sowie abhängig vom Anwendungsfall und Erkenntnisinteresse aus statistischer Perspektive ebenfalls eine eindimensionale Erhebung der SWE anhand einer Globalskala möglich ist. Das adaptierte Messinstrument erfüllt die von MEINHARDT ET AL. (2018) beschriebenen Eigenschaften des Konstrukts der SWE nach BANDURA (1997). Insbesondere werden die SWE auf einem domänen- und situationsspezifischen Niveau erfasst, was für eine größere Handlungsnähe spricht (PAJARES, 1996; WESS ET AL., 2020).

Die SWE von Lehrpersonen besitzen gerade in herausfordernden Situationen – wie dem Experimentieren im Geographieunterricht – einen großen Einfluss auf ihr unterrichtliches Handeln. Vor diesem Hintergrund und hinsichtlich des in der Literatur berichteten Absinkens von SWE im Rahmen von Praxiserfahrungen (TSCHANNEN-MORAN ET AL., 1998; PENDERGAST ET AL., 2011; WESS ET AL., 2020) kommt fachdidaktischen Hochschulangeboten zum Aufbau angemessener SWE zum Experimen-

5.5 Ausprägungen der Selbstwirksamkeitserwartungen der angehenden Lehrpersonen zum Experimentieren im Geographieunterricht

Die Studentinnen und Studenten schätzen ihre Selbstwirksamkeit zum Experimentieren im Geographieunterricht insgesamt eher positiv ein ($M=3.38$; $SD=0.64$) (Fig. 4). Für die Dimension *Durchführung* ($M=3.46$; $SD=.071$) lag der Wert signifikant höher als für die Dimension *Planung* ($M=3.27$; $SD=0.69$; $t(302)=-4.67$, $p<.001$, $d=.27$), wobei alle drei Werte signifikant von der theoretischen Skalenmitte von 3 abwichen ($d=.39-.65$).

tieren im Geographieunterricht eine besondere Bedeutung zu.

Der vorgestellte Fragebogen kann dazu dienen, die Wirksamkeit solcher geographiedidaktischen Hochschulangebote empirisch abzusichern und Veränderungen der SWE in deren Verlauf zu erfassen. Dabei ermöglicht er mit neun Items eine ökonomische Handhabbarkeit für den Einsatz und die Auswertung in der geographiedidaktischen Hochschullehre und Forschung. Der Fragebogen ist erstens unabhängig vom Inhalt formuliert, also nicht auf einen geographischen Inhaltsbereich (wie z.B. Experimente im Bereich Hydrologie oder zum Klimawandel) fokussiert, und zweitens an keine spezifische Lern- und Experimentiersituation gekoppelt. Somit ist er universell in verschiedenen universitären Lehrformaten wie beispielsweise in geographiedidaktischen Seminaren oder praktischen Lehrveranstaltungen mit dem Schwerpunkt Experimentieren einsetzbar. Zudem ist er für den Einsatz in heterogenen studentischen Lerngruppen geeignet, da angesichts (partieller) skalarer Messinvarianz für beide Skalen hinsichtlich der Schulform, des Geschlechts und der Praxiserfahrung von einer äquivalenten Messung in den verschiedenen Gruppen von Studentinnen und Studenten ausgegangen werden kann. Weitere Argumente für die Reliabilität und Validität des Messinstruments liefern die Ergebnisse der Korrelationsanalysen mit externalen Facetten und der Reliabilitätsanalyse. Auch die gefundenen theoriekonformen, wenn auch eher geringen, Mittelwertunterschiede der SWE hinsichtlich der Praxiserfahrung der Studentinnen und Studenten können als erster Hinweis auf die Validität des Messinstruments gedeutet werden. Weiterführend muss der Einfluss von Praxiserfahrungen auf die SWE im Bereich des Experimentierens in Folgestudien wie Längsschnittuntersuchungen zum Verlauf der SWE in

den verschiedenen Phasen der Lehrkräftebildung abgesichert werden. Die Unterschiede deuten zudem auf die Eignung des Instruments zur Messung der Veränderungen der SWE, beispielsweise im Rahmen von Praxiserfahrungen, hin.

Die untersuchten Studentinnen und Studenten weisen insgesamt positive SWE zum Experimentieren im Geographieunterricht auf, was vor dem Hintergrund der Literatur zum Experimentieren im Geographieunterricht zunächst verwunderlich ist und auf eine bereits häufig berichtete theoretische Selbstüberschätzung eigener SWE durch Studentinnen und Studenten hindeuten könnte (TSCHANNEN-MORAN ET AL., 1998; PENDERGAST ET AL., 2011; RABE ET AL., 2012). ROTHLAND (2018) begreift solche häufig hohen Kompetenzeinschätzungen von Studentinnen und Studenten in diesem Sinne vielmehr als ideales Selbstkonzept der eigenen angehenden Lehrperson denn als realistische Selbsteinschätzung. Wiederum realistischer könnten so die Einschätzungen derjenigen befragten Studentinnen und Studenten ausfallen, welche bereits ein Schulpraktikum absolviert und somit Praxiserfahrungen gesammelt haben und geringere SWE zum Experimentieren im Geographieunterricht angeben.

Eine solche Selbstüberschätzung eigener Fähigkeiten von Studentinnen und Studenten kann einen Praxisschock im Rahmen erster Praxiserfahrungen begünstigen (WESS ET AL., 2020), sodass – neben der Vermittlung fachdidaktischen Wissens zum Experimentieren im Geographieunterricht – dem Aufbau angemessener (im Sinne von *realistischeren*) SWE im Rahmen des Studiums und deren schrittweiser Stärkung vor dem Eintritt ins Referendariat Bedeutung zukommt. In diesem Zusammenhang wird auch der Bewusstmachung und Auseinandersetzung mit den eigenen SWE eine positive Wirkung zugeschrieben (MARTSCHINKE ET AL., 2007; RABE ET AL., 2012). Anhand des vorgestellten Instruments können im Rahmen universitärer Lehrveranstaltungen individuelle sowie gruppenspezifische Auswertungen der SWE der Studentinnen und Studenten zum Experimentieren vorgenommen werden. So können den Studentinnen und Studenten einerseits ihre eigenen SWE bewusst und somit einer (Selbst-)Reflexion zugänglich gemacht werden. Andererseits können die Ergebnisse durch die Verortung der eigenen SWE innerhalb der Gruppe als Diskussionsanlass dienen. Ein solcher selbstreflexiver Zugang kann möglicherweise zur Entwicklung von realistischeren und somit angemesseneren SWE durch die Studentinnen und Studenten beitragen (MARTSCHINKE ET AL., 2007). Auch ROTHLAND (2018, S. 491) versteht Selbstreflexion in diesem Sinne „[...] als Vergleich zwischen realem und idealem Selbstkonzept“ mit dem Ziel, dass

„[...] die Vorstellungen von den eigenen berufsbezogenen Fähigkeiten und Fertigkeiten [im Rahmen der Praxiserfahrung] irritiert, in Zweifel gezogen und die Kompetenzeinschätzungen hin zu einem realen Selbstkonzept im Ergebnis abnehmen würden“ (ROTHLAND, 2018, S. 483).

Als wichtigste Quelle von SWE gelten eigene Handlungsergebnisse bzw. Erfolgserfahrungen (BANDURA, 1997), sodass darüber hinaus zur Stärkung der SWE gerade solche Praxisphasen im Studium in den Vordergrund rücken. Eine unzureichende Begleitung der Studentinnen und Studenten während dieser kann jedoch ein Absinken von SWE im Sinne eines Praxisschocks begünstigen. Positive Wirkungen auf die SWE von Studentinnen und Studenten können Praxiselemente im Studium jedoch insbesondere dann haben, wenn sie in Form von Mentoring oder konstruktivem Feedback theoriebasiert begleitet werden (WESS ET AL., 2020). Dabei können zur Komplexitätsreduktion zunächst Schülerinnen und Schüler in sog. Lehr-Lern-Laboren in Lehrveranstaltungen an der Universität eingebunden werden. Mit Blick auf die Stärkung der SWE angehender Lehrpersonen hinsichtlich des Experimentierens in heterogenen Lerngruppen liefern ROSENDAHL ET AL. (2020) erste Hinweise auf wirksame komplexitätsreduzierende Maßnahmen. Die befragten Studentinnen und Studenten nehmen beim Besuch des GEO-Lehr-Lern-Labors insbesondere Unterstützungsmaßnahmen durch die Dozentin sowie durch Kommilitoninnen und Kommilitonen als förderlich für ihre Kompetenzentwicklung wahr.

6.2 Limitationen

Eine zentrale Limitierung des Messinstruments liegt in der Erfassung der SWE in dem Format des Fragebogens mit geschlossenem Antwortformat. Durch die Wahl dieses Erhebungsformates und die Auswahl der Kurzversion des Instruments von MEINHARDT ET AL. (2016) für die Adaption muss in Kauf genommen werden, dass der Fragebogen nur SWE zu einzelnen Bereichen der Vorbereitung und des Einsatzes von Experimenten im Geographieunterricht abbilden kann. Beispielsweise beinhaltet die Kurzversion keine Items zur Entwicklung von Experimenten oder der Unterstützung beim Experimentieren. Gleichzeitig lag das zentrale Ziel in einem für universitäre Lehrveranstaltungen im Einsatz und der Auswertung zeitökonomisch handhabbaren Messinstrument zur Erfassung der SWE zum Experimentieren im Geographieunterricht, weshalb wir die Entscheidung für das geschlossene Fragebogenformat und die Auswahl der Kurzversion unter Inkaufnahme der Limitationen getroffen haben.

Das zweidimensionale Modell mit Unterscheidung zwischen den Dimensionen *Durchführung* und *Planung* von Experimenten erreicht einen guten Modell-Fit und erweist sich auch mit Bezug auf die weiteren betrachteten Validitätsargumente als robustes Instrument. Dennoch liegen die Reliabilitäten der beiden einzelnen Skalen lediglich im akzeptablen Bereich, was jedoch erstens auf die relativ niedrige Anzahl an Items innerhalb der beiden Skalen (4 bzw. 5 Items/Skala) (GÄDE ET AL., 2020) und zweitens auf das Item *exp3* zurückgeht. Dieses Item erwies sich auch mit Blick auf die Modifikationsindizes sowie die Messinvarianz der Skalen bezüglich der Praxiserfahrung als problematisch. Da Entscheidungen zum Entfernen von Items nicht auf rein statistischer, sondern ebenfalls inhaltlicher Grundlage getroffen werden sollten (BROWN, 2015), wurde das Item mit Blick auf die Inhaltsvalidität des Messinstrumentes beibehalten.

Obwohl in der Einleitung dem Fragebogen ein Informationsblock mit Definitionen und Beispielen von naturwissenschaftlichen Arbeitsweisen vorangestellt war, kann darüber hinaus nicht ausgeschlossen werden, dass die Vorstellungen der Studentinnen und Studenten von Experimenten im Detail dennoch abweichen (LETHMATE, 2006; FÖGELE ET AL., 2020). Deshalb sollte in einer weiterführenden qualitativen Testung das Verständnis der befragten Studentinnen und Studenten bezüglich des Experimentierens im wissenschaftlichen Sinne erhoben werden sowie in Folgestudien die Vorstellungen

der Studentinnen und Studenten von Experimenten gemeinsam mit ihren SWE zum Experimentieren erfasst werden.

6.3 Ausblick

Im Rahmen weiterführender Studien ist es notwendig, das vorgestellte Messinstrument für im Referendariat und sich im Beruf befindliche Lehrpersonen zu validieren. Dies ermöglicht die vergleichende Betrachtung der SWE von Lehrpersonen in verschiedenen Phasen der Lehrkräftebildung bzw. Längsschnittuntersuchungen zum Verlauf der SWE sowie eine Absicherung der zuvor angestellten Überlegungen (u.a. zur Selbstreflexion sowie der Rolle von SWE beim Experimentieren im Geographieunterricht). Darüber hinaus kann das vorgestellte Instrument insbesondere mit Blick auf die Diskrepanz zwischen Potenzialen und geringen Einsatzhäufigkeiten von Experimenten für die geographiedidaktische Forschung dazu dienen, mögliche Zusammenhänge mit tatsächlichem unterrichtlichen Handeln, fachdidaktischem Wissen und professionellen Überzeugungen von Lehrpersonen bezüglich des Experimentierens (VELLING ET AL., 2022) zu untersuchen. In diesem Zusammenhang kommt insbesondere der situationsspezifischen Erfassung der SWE durch das vorgestellte Messinstrument eine wichtige Rolle zu (PAJARES, 1996; SCHWARZER & JERUSALEM, 2002; Rabe et al., 2012).

Literatur

- AKUMA, F. V., & CALLAGHAN, R. (2019). [Characterising Extrinsic Challenges Linked to the Design and Implementation of Inquiry-Based Practical Work](#). *Research in Science Education*, 49(6), 1677–1706.
- APPLETON, K., & KINDT, I. (2002). [Beginning Elementary Teachers' Development as Teachers of Science](#). *Journal of Science Teacher Education*, 13(1), 43–61.
- BANDURA, A. (1997). *Self-Efficacy: The Exercise of Control*. W. H. Freeman and Company.
- BAUMERT, J., & KUNTER, M. (2006). [Stichwort: Professionelle Kompetenz von Lehrkräften](#). *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- BENT, G. J., BAKX, A., & BROK, P. D. (2017). [Primary Education Teachers' Self-Efficacy Beliefs for Teaching Geography Lessons](#). *International Research in Geographical and Environmental Education*, 26(2), 150–165.
- BENTLER, P. M. (1995). *EQS Structural Equations Program Manual*. Multivariate Software.
- BROWN, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. *Methodology in the Social Sciences*. The Guilford Press.
- CHEN, F. F. (2007). [Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance](#). *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504.
- CHO, Y., & SHIM, S. S. (2013). [Predicting Teachers' Achievement Goals for Teaching: The Role of Perceived School Goal Structure and Teachers' Sense of Efficacy](#). *Teaching and Teacher Education*, 32, 12–21.
- CRAWFORD, B. A. (2007). [Learning to Teach Science as Inquiry in the Rough and Tumble of Practice](#). *Journal of Research in Science Teaching*, 44(4), 613–642.
- DGfG (DEUTSCHE GESELLSCHAFT FÜR GEOGRAPHIE) (2020). *Bildungsstandards im Fach Geographie für den Mittleren Schulabschluss*. DGfG.

- DI BIASE, W., & McDONALD, J. R. (2015). [Science Teacher Attitudes Toward Inquiry-Based Teaching and Learning](#). *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 88(2), 29-38.
- EHLERT, L., & TEPNER, O. (2023). *Erfassung der Selbstwirksamkeitserwartung beim selbstgesteuerten Experimentieren*. Jahrestagung GDGP. Aachen.
- ENOCHS, L. G., & RIGGS, I. M. (1990). [Further Development of an Elementary Science Teaching Efficacy Belief Instrument: A Preservice Elementary Scale](#). *School Science and Mathematics*, 90(8), 694-706.
- FÖGELE, J., LUBER, L., & MEHREN, R. (2020). [Types of Student Teachers between Theory and Practice in the Subject of Geography: Reconstructing Orientations Using the Example of a Seminar on Experiments](#). *Journal of Geography in Higher Education*, 44(2), 310-333.
- GÄDE, J. C., SCHERMELLEH-ENGEL, K., & WERNER, C. S. (2020). Klassische Methoden der Reliabilitäts-schätzung. In H. MOOSBRUGGER & A. KELAVA (Hg.), *Testtheorie und Fragebogenkonstruktion* (S. 305-334). Springer.
- GOROZIDIS, G., & PAPAIOANNOU, A. (2011). [Teachers' Self-efficacy, Achievement Goals, Attitudes and Intentions to Implement the New Greek Physical Education Curriculum](#). *European Physical Education Review*, 17(2), 231-253.
- HAIR, J. F., BLACK, W. C., BABIN, B. J., & ANDERSON, R. E. (2014). *Multivariate Data Analysis*. Pearson.
- HEMMER, I., & HEMMER, M. (Hg.) (2010). *Schüler-interesse an Themen, Regionen und Arbeitsweisen des Geographieunterrichts: Ergebnisse der empirischen Forschung und deren Konsequenzen für die Unterrichtspraxis*. Geographiedidaktische Forschungen, Band 46. HGD.
- HOF, S., & HENNEMANN, S. (2013). [Geographie-lehrerinnen und -lehrer im Spannungsfeld zwischen erworbenen und geforderten Kompetenzen: Eine empirische Studie zur zweiphasigen Lehramtsausbildung](#). *Geographie und ihre Didaktik*, 41(2), 57-80.
- 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*, 2-3(142/143), 153-161.
- HOY, A. W., & SPERO, R. B. (2005). [Changes in Teacher Efficacy during the Early Years of Teaching: A Comparison of Four Measures](#). *Teaching and Teacher Education*, 21(4), 343-356.
- HSIAO, H.-C., CHANG, J.-C., TU, Y.-L., & CHEN, S.-C. (2011). The Impact of Self-Efficacy on Innovative Work Behavior for Teachers. *International Journal of Social Science and Humanity*, 1(1), 31-36.
- HU, L., & BENTLER, P. M. (1999). [Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives](#). *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- JORGENSEN, T. D., PORNPRASERTMANIT, S., SCHOEMANN, A. M., & ROSSEEL, Y. (2018). *semTools* (Version 0.5-1) [Computer software].
- KAUERTZ, A., KLEICKMANN, T., EWERHARDY, A., FRICKE, K., LANGE, K., OHLE, A., POLLMEIER, K., TRÖBST, S., WALPER, L., FISCHER, H. E., & MÖLLER, K. (2013). *Dokumentation der Erhebungsinstrumente im Projekt PLUS*. nwu-essen.
- LEINER, D. J. (2021). *SoSci Survey* (Version 3.2.23) [Computer software].
- LETHMATE, J. (2006). Experimentelle Lehrformen und Scientific Literacy. *Praxis Geographie*, 36(11), 4-11.
- MADDOX, L. E., HOWELL, J. B., & SAYE, J. W. (2018). [Designing Geographic Inquiry: Preparing Secondary Students For Citizenship](#). *Journal of Geography*, 117(6), 254-268.
- MARTSCHINKE, S., KOPP, B., & HALLITZKY, M. (2007). Die Bedeutung von selbstreflexivem und forschendem Lernen in der Ausbildung für Lehramtsstudierende. *PAradigma*, 1, 6-16.
- MEINHARDT, C., RABE, T., & KREY, O. (2016). *Selbstwirksamkeitserwartungen in physikdidaktischen Handlungsfeldern*. Skalendokumentation. MLU.
- MEINHARDT, C., RABE, T., & KREY, O. (2018). [Formulierung eines evidenzbasierten Validitätsarguments am Beispiel der Erfassung physikdidaktischer Selbstwirksamkeitserwartungen mit einem neu entwickelten Instrument](#). *Zeitschrift für Didaktik der Naturwissenschaften (ZfDN)*, 24(1), 131-150.
- MIENER, J. P., & KÖHLER, K. A. (2013). *Experimentelle Arbeitsweisen im Geographieunterricht. Vorstellungen von Geographielehrern zu Chancen und Barrieren*. AV Akademiker Verlag.
- OTTO, K.-H., & MÖNTER, L. O. (2015). Scientific Literacy im Geographieunterricht fördern: Experimentelle Lehr-/Lernformen und Modellexperimente. *geographie heute*, 36(322), 2-7.
- OTTO, K.-H., MÖNTER, L. O., HOF, S., & WIRTH, J. (2010). [Das geographische Experiment im Kontext empirischer Lehr-/Lernforschung](#). *Geographie und ihre Didaktik | Journal of Geography Education*, 38(3), 133-145.

- PAJARES, F. (1996). [Self-Efficacy Beliefs in Academic Settings](#). *Review of Educational Research*, 66(4), 543-578.
- PENDERGAST, D., GARVIS, S., & KEOGH, J. (2011). [Pre-Service Student-Teacher Self-efficacy Beliefs: An Insight Into the Making of Teachers](#). *Australian Journal of Teacher Education*, 36(12), 4.
- RABE, T., MEINHARDT, C., & KREY, O. (2012). Entwicklung eines Instruments zur Erhebung von Selbstwirksamkeitserwartungen in physikdidaktischen Handlungsfeldern. *Zeitschrift für Didaktik der Naturwissenschaften (ZfDN)*, 18, 293-315.
- RAMNARAIN, U. (2016). [Understanding the Influence of Intrinsic and Extrinsic Factors on Inquiry-based Science Education at Township Schools in South Africa](#). *Journal of Research in Science Teaching*, 53(4), 598-619.
- RENNER, T. (2021). *Selbstwirksamkeitserwartungen von Lehramtsstudierenden zu geographischen Exkursionen - eine qualitative, längsschnittliche Interviewstudie* (Dissertation).
- RIESE, J. (2009). *Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften*. Logos Verlag.
- ROBERTS, J. K., HENSON, R. K., THARP, B. Z., & MORENO, N. P. (2001). [An Examination of Change in Teacher Self-Efficacy Beliefs in Science Education Based on the Duration of Inservice Activities](#). *Journal of Science Teacher Education*, 12(3), 199-213.
- ROSENDAHL, N., HEMMER, M., & SCHRÜFER, G. (2020). Mit Vielfalt experimentieren: Professionalisierung angehender Lehrkräfte im GEO-Lehr-Lern-Labor. In R. KÜRTE, G. GREEFRATH & M. HAMMANN (Hg.), *Komplexitätsreduktion in Lehr-Lern-Laboren: Innovative Lehrformate in der Lehrerbildung zum Umgang mit Heterogenität und Inklusion* (S. 137-161). Waxmann.
- ROSSEEL, Y. (2012). [lavaan: An R Package for Structural Equation Modeling](#). *Journal of Statistical Software*, 48(2), 1-36.
- ROTHLAND, M. (2018). [Yes, we can! Anmerkungen zur trügerischen „Kompetenzentwicklung“ von Lehramtsstudierenden im Praxissemester](#). *BzL-Beiträge zur Lehrerinnen und Lehrerbildung*, 36(3), 482-495.
- SCHWAB, S., & HELM, C. (2015). Überprüfung von Messinvarianz mittels CFA und DIF-Analysen. *Empirische Sonderpädagogik*, 3, 175-193.
- SCHWARZER, R., & JERUSALEM, M. (2002). Das Konzept der Selbstwirksamkeit. In M. JERUSALEM & D. HOPF (Hg.), *Zeitschrift für Pädagogik Beiheft: Bd. 44. Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen* (S. 28-53). Beltz.
- SCHWARZER, R., & WARNER, L. M. (2014). Forschung zur Selbstwirksamkeit bei Lehrerinnen und Lehrern. In E. TERHART, H. BENNEWITZ & M. ROTHLAND (Hg.), *Handbuch der Forschung zum Lehrerberuf* (S. 618-634). Waxmann.
- SCHWEIZER, K., & HORN, M. (2010). Subjektive Überzeugungen, Selbstwirksamkeit und Medien: Welchen Einfluss hat die medienbezogene Selbstwirksamkeitserwartung Lehramtsstudierender der Geographie? In B. SCHWARZ, P. NENNIGER & R. S. JÄGER (Hg.), *Erziehungswissenschaftliche Forschung - nachhaltige Bildung. Beiträge zur 5. DGfE-Sektionstagung "Empirische Bildungsforschung"/AEPF-KBBB im Frühjahr 2009* (S. 197-203). Verlag Empirische Pädagogik e.V.
- SEUNG, E., PARK, S., & LEE, M.-A. (2019). [The Impact of a Summer Camp-Based Science Methods Course on Preservice Teachers' Self-Efficacy in Teaching Science as Inquiry](#). *Journal of Science Teacher Education*, 30(8), 872-889.
- SMOLLECK, L., ZEMBAL-SAUL, C., & YODER, E. P. (2006). [The Development and Validation of an Instrument to Measure Preservice Teachers' Self-Efficacy in Regard to The Teaching of Science as Inquiry](#). *Journal of Science Teacher Education*, 17(2), 137-163.
- SMOLLECK, L. A., & MONGAN, A. M. (2011). [Changes in Preservice Teachers' Self-Efficacy: From Science Methods to Student Teaching](#). *Journal of Educational and Developmental Psychology*, 1(1), 133-145.
- STATISTISCHES BUNDESAMT (2022). *Bildung und Kultur - Studierende an Hochschulen: Wintersemester 2021/2022. Fachserie 11: Reihe 4.1*. Statistisches Bundesamt Wiesbaden.
- TSCHANNEN-MORAN, M., HOY, A. W., & HOY, W. K. (1998). [Teacher Efficacy: Its Meaning and Measure](#). *Review of Educational Research*, 68(2), 202-248.
- VELLING, H., SCHUBERT, J. C., & NAUJOKS-SCHÖBER, N. (2022). [Beliefs angehender Lehrpersonen zum Experimentieren im Geographieunterricht: Entwicklung und Pilotierung eines Fragebogens](#). *Zeitschrift für Didaktik der Naturwissenschaften (ZfDN)*, 28(1), 1-16.
- WESS, R., PRIEMER, B., WEUSMANN, B., LUDWIG, T., SORGE, S., & NEUMANN, I. (2020). [Der Verlauf von lehrbezogenen Selbstwirksamkeitserwartungen angehender MINT-Lehrkräfte im Studium](#). *Zeitschrift für Pädagogische Psychologie*, 34(3-4), 221-238.
- WILHELMI, V. (2012). Die Experimentelle Lernform: Herausforderung des kompetenzorientierten

Geographieunterrichts. *Praxis Geographie*,
42(7/8), 4-8.

XUAN, X., JIN, Q., JO, I., DUAN, Y., & KIM, M. (2019).
[The Potential Contribution of Geography](#)

[Curriculum to Scientific Literacy](#). *Journal of
Geography*, 118(5), 185-196.

Autorin und Autor

✉ **Hanna Velling**

Friedrich-Alexander-Universität Erlangen-Nürnberg
Lehrstuhl für Didaktik der Geographie
Regensburger Str. 160
90478 Nürnberg
hanna.velling@fau.de

Prof. Dr. Jan Christoph Schubert

Friedrich-Alexander-Universität Erlangen-Nürnberg
Lehrstuhl für Didaktik der Geographie
Regensburger Str. 160
90478 Nürnberg
jan.christoph.schubert@fau.de

Anhang 1. Erste Seite des vorgestellten Fragebogens zu Selbstwirksamkeitserwartungen zum Experimentieren im Geographieunterricht, realisiert in *SoSci Survey* (LEINER, 2021) (Quelle: Autorin und Autor; Instruktionstext nach LETHMATE, 2006)

Experimente im Geographieunterricht

Im Folgenden werden Ihnen Aussagen präsentiert, die sich speziell auf das **Experimentieren im Geographieunterricht** beziehen. Denken Sie bitte an die unten angeführte Definition von Experimenten und beziehen Sie sich bei der Beantwortung der Aussagen **nur auf das Experimentieren im Geographieunterricht**.

i In der fachdidaktischen Diskussion unterscheidet man innerhalb der naturwissenschaftlichen Arbeitsweisen zwischen Untersuchungen, Experimenten und Modellen.

- Unter einer **Untersuchung** versteht man dabei Messungen mit Hilfsmitteln wie Niederschlags-, Temperatur- und pH-Wert-Messungen sowie typische Nachweisreaktionen von beispielsweise Kalk.
- **Modelle** dienen der Veranschaulichung durch eine vereinfachte Reduktion realer Phänomene (z.B. Aufbau eines Vulkans, Höhenstufenmodell, Planetarium).
- **Experimente** hingegen verfolgen den Zweck durch systematische Variation künstlich hergestellter und veränderbarer Bedingungen Ursache-Wirkungs-Zusammenhänge nachzuweisen (z.B. Einfluss der Bodenart auf das Wasserhaltevermögen, Einfluss der Temperatur auf das Luftvolumen, Einfluss der Bodenfeuchtigkeit auf die Erwärmung des Bodens).

Weiter

Anhang 2. Zweite Seite des vorgestellten Fragebogens zu Selbstwirksamkeitserwartungen zum Experimentieren im Geographieunterricht, realisiert in *SoSci Survey* (LEINER, 2021) (Quelle: Autorin und Autor; Items und Instruktion adaptiert nach MEINHARDT ET AL., 2016)

Wir bitten Sie, sich im Folgenden **selbst einzuschätzen**. Es handelt sich um Situationen, die sich auf das Handeln von Geographielehrpersonen beziehen. Es kann sein, dass Sie diese Situationen noch nicht selbst als angehende Geographielehrperson erlebt haben. Versuchen Sie bitte dennoch sich **in die Situationen hineinzusetzen** und entscheiden Sie, ob Sie mit Ihrem **heutigen** (!) Wissen und Können diese Situation bewältigen können oder nicht. Wir bitten Sie, sich möglichst **realistisch** einzuschätzen.

Selbsteinschätzung beim Experimentieren im Geographieunterricht

Wie schätzen Sie sich ein?

stimme gar nicht zu stimme vollkommen zu

Ich kann ein Experiment, das im Geographieunterricht nicht auf Anhieb funktioniert, zum Laufen bringen, auch wenn ich unter Zeitdruck stehe.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Ich kann beim Experimentieren spontan mit den Schüler*innen über das Wechselspiel von Theorie und Experiment reflektieren, auch wenn ich den Anlass nicht vorgesehen hatte.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Ich kann bei meiner Unterrichtsplanung ein Experiment gegebenenfalls so variieren, dass ich es in einer Übungsphase einsetzen kann, auch wenn ich es bisher nur als Einstiegsexperiment genutzt habe.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Ich kann geographische Experimente an interessante Impulse meiner Schüler*innen anpassen, auch wenn ich das vorher nicht geplant hatte.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Ich kann ein Experiment so inszenieren, dass meine Schüler*innen motiviert sind mitzuarbeiten, auch wenn es sich um eine unbeliebte Randstunde handelt.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Ich kann in meiner Unterrichtsplanung zu den Lernzielen passende Experimente aufbauen, auch wenn die Geographie dafür schlecht ausgestattet ist.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Ich kann ein Demonstrationsexperiment für meine Schüler*innen nachvollziehbar durchführen, auch wenn es sich um eine komplexe Versuchsanordnung handelt.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Ich kann Schülerexperimente so zusammenstellen, dass die Fähigkeiten meiner Schüler*innen auf verschiedenen Niveaus gefördert werden, auch wenn ich bei der Planung unter Zeitdruck stehe.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Ich kann für ein geographisches Experiment begründet entscheiden, ob es didaktisch sinnvoller ist, es als Demonstrations- oder Schülerexperiment einzuplanen, auch wenn ich das Experiment noch nicht eingesetzt habe.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

Zurück

Weiter