

Samuel Merk, Colin Cramer, Naien Dai, Thorsten Bohl &
Marcus Syring

Faktorielle Validität der Einstellungen von Lehrkräften zu heterogenen Lerngruppen

Zusammenfassung

Der vorliegende Beitrag untersucht die Konstruktvalidität eines Instrumentariums, das kognitive, affektive und behaviourale Einstellungen sowie motivationale Orientierungen von $N = 879$ Lehramtsstudierenden bezüglich heterogener Lerngruppen erfasst. Dieses Instrumentarium erfasst Einstellungen und motivationale Orientierungen zu sozialer, leistungsbezogener und kultureller Heterogenität jeweils in den Dimensionen Kosten, Nutzen, negative Emotionen, eigener kompetenter Umgang und intrinsisch-motivationale Orientierung. Um die faktorielle Validität dieses Instrumentariums zu untersuchen wurden konfirmatorische Mehrebenenfaktorenanalysen durchgeführt, welche die theoretisch postulierte Struktur sowohl je Heterogenitätsform und Person (drei fünf-dimensionale Einstellungen, Within-person-Ebene) als auch je Person (fünf-dimensionale Einstellungen, über die Heterogenitätsformen gemittelt, Between-person-Ebene) simultan prüfen. Die theoretisch postulierte Struktur des Instruments konnte sowohl auf Within-person- als auch auf Between-person-Ebene (mit ebenenvarianten τ -kongenerischen Messmodellen) bestätigt werden. Eine Betrachtung von Intraklassenkorrelationskoeffizienten der Items und latenten Faktoren sowie der prädiktiven Effekte der Form von Heterogenität auf die Ausprägung der Dimensionen impliziert zudem, dass das Instrumentarium die Einstellungen der Lehramtsstudierenden sowohl in deren Dimensionen als auch in deren Formen zu differenzieren vermag, was als ein starker Beleg für die Konstruktvalidität des Instrumentes gewertet wird.

Jun.-Prof. Dr. Samuel Merk (corresponding author) · Prof. Dr. Colin Cramer · Naien Dai · Prof. Dr. Thorsten Bohl, Eberhard Karls Universität Tübingen, Institut für Erziehungswissenschaft, Münzgasse 22–30, 72070 Tübingen, Deutschland
E-Mail: samuel.merk@ife.uni-tuebingen.de
colin.cramer@uni-tuebingen.de
n.dai@student.uni-tuebingen.de
thorsten.bohl@uni-tuebingen.de

Dr. Marcus Syring, Ludwig-Maximilians-Universität München, Department für Pädagogik und Rehabilitation, Lehrstuhl für Schulpädagogik, Leopoldstraße 13, 80802 München, Deutschland
E-Mail: marcus.syring@edu.lmu.de

Schlagworte

Konstruktvalidität, Einstellungen, Lehramtsstudierende, Heterogenität

Factorial validity of attitudes towards heterogeneous student groups

Abstract

The current study investigates the construct validity of an instrument that assesses cognitive, affective and behavioral attitudes as well as intrinsic motivational orientation towards heterogeneous learning groups, held by $N = 879$ student teachers. The investigated instrument assesses attitudes and motivational orientations and contains five factors (utility, cost, negative emotions, competence and intrinsic motivation) which were assessed for three kinds (social, ethnic and performance) of heterogeneity. To investigate the factorial validity of this instrument we conducted multi-level confirmatory factor analyses to test the postulated structure at the within-person level (three five-dimensional measurements) and the between-person level (five-dimensional structure of person specific means over the three forms of heterogeneity) simultaneously. The results confirm the postulated structure on both levels, with level variant τ -congeneric measurement models. Closer inspection of intraclass correlation coefficients of the items and predictive effects of the form of heterogeneity on the five factors of attitudes towards heterogeneous learning groups imply, that the instrument can distinguish between the three forms of heterogeneity. We judge these findings as strong evidence for the construct validity of the instrument.

Keywords

Construct validity, Attitudes, Student teachers, Heterogeneity

1. Einleitung

Der Begriff *Heterogenität* wird seit einigen Jahren in Öffentlichkeit und Wissenschaft intensiv diskutiert. Er vermag durch seine Offenheit unterschiedliche Herausforderungen, die an Schule gestellt werden, zu bündeln. In dieser Offenheit besteht zugleich auch die Schwäche des Begriffs, der sich durch Unschärfe charakterisieren lässt und häufig überfrachtet wird (Budde, 2012), weil sich mit ihm hohe Anforderungen an das professionelle Handeln von Lehrkräften verbinden (Gebauer, McElvany & Klukas, 2013; HRK, 2015). In der Literatur werden motivational-affektive Orientierungen sowie Einstellungen von Lehrkräften als Voraussetzungen erfolgreichen unterrichtlichen Handelns von Lehrpersonen hohe Relevanz zugeschrieben (z.B. Baumert & Kunter, 2006; Shulman, 1986). Allerdings ist die Forschung bisher sehr undifferenziert und meist auf Leistungsheterogenität beschränkt (Rose & Gerkmann, 2015). Bislang existie-

ren einige Instrumente, die Einstellungen zur Heterogenität (und vor allem zur Inklusion) erfassen (vgl. hierzu z.B. die Schwerpunktthemen in der „Empirischen Sonderpädagogik“ 1/2016 oder der „Zeitschrift für Bildungsforschung“ 4/2014), jedoch fehlen – insbesondere bei deutschsprachigen Instrumenten – oft Untersuchungen zur deren Konstruktvalidität. Dies ist umso erstaunlicher, da die konzeptuellen Unterschiede in Instrumenten zur Erfassung von Einstellungen zu heterogenen Lerngruppen enorm sind. Beispielsweise umfasst der häufig verwendete EZI („Einstellungen zur Integration“; Kunz, Luder & Moretti, 2010) mit elf Items unter Einstellungen die Bereitschaft zur schulischen Förderung und Unterstützung sowie die soziale Integration. Dahingegen differenziert das hier untersuchte Instrument zur Erfassung heterogenitätsbezogener Einstellungen von (angehenden) Lehrkräften (Gebauer et al., 2013) diverse Formen von Heterogenität (soziale, kulturelle und leistungsbezogene Heterogenität) jeweils in mehreren Dimensionen (Nutzen, Kosten, negative Emotionen, Motivation, eigener kompetenter Umgang). Dieser Kontrast veranschaulicht die Notwendigkeit, die Konstruktvalidität von Einstellungen zu heterogenen Lerngruppen zu untersuchen, um Aussagen zur Angemessenheit der Skalenwertinterpretation treffen zu können (siehe Abschnitt Konstruktvalidität 2.1).

Dieser Beitrag überprüft die Konstruktvalidität des Instruments von Gebauer et al. (2013) auf der Grundlage konfirmatorischer Mehrebenen-Faktorenanalysen. Dazu wird zunächst auf das Konzept der Konstruktvalidität eingegangen, gefolgt von einer Beschreibung des untersuchten Instruments sowie dessen theoretischer und konzeptioneller Grundlagen.

2. Erfassung von Einstellungen zum Umgang mit Heterogenität

Im Folgenden wird zunächst auf das Konzept der Konstruktvalidität eingegangen, um im Anschluss daran das zu validierende Instrument von Gebauer et al. (2013) zu skizzieren. Das dem Instrument zugrundeliegende Konstrukt der Einstellungen bzw. motivationalen Orientierungen sowie die erfassten drei Formen von Heterogenität (soziale, kulturelle und leistungsbezogene) werden theoretisch und mit Blick auf den aktuellen Forschungsstand beschrieben.

2.1 Konstruktvalidität

In den 1950er Jahren beauftragte die American Psychological Association (APA) Lee Cronbach, das bestehende Validitätskonzept weiterzuentwickeln. In einem diesbezüglich zentralen Werk (Cronbach & Meehl, 1955) ordnen die Autoren empirische Diagnostik als einen hypothetisch-deduktiven Ansatz ein und verknüpfen so erstmals explizit Theorie mit der Validität eines Testergebnisses. Dieser

Ansatz findet sich auch in aktuellen Definitionen von Konstruktvalidität wieder, etwa im Dorsch-Lexikon der Psychologie, wonach „Konstruktvalidität [...] in dem Maße gegeben [ist], in dem ein Instrument Daten erzeugt, die von einer empirisch bewährten oder überzeugenden Theorie vorhergesagt werden, in die das zu messende Konstrukt eingebunden ist.“ (Schmitt, 2014, S. 921). Im Anschluss an diese Definition haben sich drei empirische Ansätze zur Überprüfung der Konstruktvalidität durchgesetzt (Hartig, Frey & Jude, 2008): Die Untersuchung der (faktoriellen) Struktur der Daten, die das Instrument liefert, die Vorhersage von Itemschwierigkeiten sowie die Untersuchung von Zusammenhängen mit theoretisch proximalen/distalen Konstrukten (konvergente/diskriminante Validität). Die Grundidee der Untersuchung faktorieller Strukturen zur Überprüfung der Konstruktvalidität liegt darin, dass empirisch untersucht werden kann, wie gut die Daten eines Instruments zur theoretisch angenommenen Dimensionalität passen. Der Beitrag verfolgt diesen Ansatz mithilfe konfirmatorischer (Mehrebenen-) Faktorenanalysen, da die zum Konstrukt vorliegenden theoretischen Vorarbeiten für die beiden anderen Ansätze zu schwach ausgereift scheinen.

2.2 Das Konstrukt von Gebauer, McElvany und Klukas (2013)

Das Konstrukt von Gebauer et al. (2013) zur Erfassung von Überzeugungen zu heterogenen Lerngruppen orientiert sich zum einen inhaltlich an drei Formen von Heterogenität (kulturelle, soziale und leistungsbezogene) und zum anderen an vier Einstellungs-Dimensionen (Nutzen, Kosten, negative Emotionen, eigener kompetenter Umgang) und einer intrinsisch motivationalen Orientierung. Die Einstellungs-Dimensionen lassen sich wiederum bzgl. ihrer konstruktbezogenen Herkunft als kognitive (Nutzen, Kosten), affektive (negative Emotionen) und verhaltensbezogene (eigener kompetenter Umgang) Einstellungen klassifizieren (vgl. Eagly & Chaiken, 1993; siehe Tabelle 1 und Tabelle 2).

Tabelle 1: Struktur des Instruments zur Erfassung der Einstellungen bzgl. heterogener Schülerschaften

		Einstellungen				
		kognitiv		affektiv	behavioural	
		Nutzen	Kosten	negative Emotionen	eigener kompetenter Umgang	intrinsische motivationale Orientierung
Form der Heterogenität	sozial	5 Items	5 Items	4 Items	5 Items	5 Items
	kulturell	5 Items	5 Items	4 Items	5 Items	5 Items
	leistungsbezogen	5 Items	5 Items	4 Items	5 Items	5 Items

Tabelle 2: Beispielimtem aus dem Instrument zur Erfassung der Einstellungen bzgl. heterogener Schülerschaften

Form der Heterogenität	Fragebogenprompt	Beispielimtem	Dimension
kulturell	Haben kulturell-heterogene Lerngruppen einen Mehrwert für die Schüler/innen? Bitte lesen Sie die folgenden Aussagen durch und kreuzen Sie an, wie sehr diese Ihrer Meinung nach zutreffen. Schüler/innen profitieren durch Unterricht in kulturell heterogenen Lerngruppen hinsichtlich ihres schulischen Lernens.	Nutzen
Leistung	Aus der Perspektive der zukünftig Lehrenden: Inwieweit teilen Sie die folgenden Einschätzungen? Die leistungsbezogene Heterogenität in unseren Schulklassen erfordert von den Lehrkräften einen zusätzlichen Arbeitsaufwand.	Kosten
sozial	Wie zutreffend sind die folgenden Äußerungen? Im Hinblick auf das Unterrichten von sozial heterogenen Klassen fühle ich mich bezogen auf die Berücksichtigung der sozialen Differenzen der Schüler/innen unter Druck gesetzt.	negative Emotionen

2.3 Einstellungen, Überzeugungen und motivationale Orientierung

Einstellungen/Überzeugungen/Werthaltungen und motivationale Orientierungen sind im kompetenztheoretischen Ansatz von Professionalität im Lehrerinnen- und Lehrerberuf neben dem prominenten Professionswissen (Shulman, 1986) Dimensionen professioneller Handlungskompetenz von Lehrpersonen (Brunner, Kunter & Krauss, 2006; Blömeke et al., 2010). Insbesondere Einstellungen, Über-

zeugungen und Werthaltungen gelten als hoch funktional für die Wahrnehmung und individuelle Deutung von Anforderungssituationen (Fives & Buehl, 2012) und sind deshalb Prädiktoren guten Unterrichts und erfolgreichen Lernens (z. B. Helmke, 2004; Prenzel & Doll, 2002). Die Begriffe *Einstellungen*, *Überzeugungen* und *Werthaltungen* werden im deutschen Sprachgebrauch nur schwach differenziert verwendet und können als „implizite und explizite subjektive Konzepte, die in Bezug zum Gegenstand, sich selbst und dem Kontext als wahr angesehen werden“ (Gebauer et al., 2013) definiert werden (vgl. auch Eynde, Corte & Verschaffel, 2002). Derartige Einstellungen können sowohl kognitiver, als auch motivational-affektiver und verhaltensbezogener Art sein (vgl. Eagly & Chaiken, 1993) und beziehen sich ebenso wie die intrinsisch-motivationale Orientierung im vorliegenden Beitrag auf alle drei Formen von Heterogenität.

2.4 Heterogenität und deren Formen

Das Konstrukt von Gebauer et al. bezieht sich auf drei ausgewählte Formen von Heterogenität, welche in der öffentlichen und wissenschaftlichen Debatte derzeit besondere Beachtung finden (Budde, 2012; Sturm, 2016) und für Lehrkräfte eine große Herausforderung darstellen (Terhart, 2015).

Kulturelle Heterogenität. Unter kultureller Heterogenität kann die Vielfalt und Unterschiedlichkeit von Schülerinnen und Schülern mit Blick auf ihren kulturellen Hintergrund verstanden werden, worunter Aspekte wie Migrationsgeschichte, Erst-/Zweitsprache, kulturelle und religiöse Zugehörigkeit etc. zu fassen sind. Kulturelle Heterogenität wird oft mit einer familiären Migrationsgeschichte, unterschiedlicher ethnischer Herkunft, unterschiedlichen Interessen, Werten und Überzeugungen (Gebauer et al., 2013) in Verbindung gebracht, die für schulische Leistungen von Relevanz sein können (z. B. Baumert, Bos & Lehmann, 1997; Baumert et al., 2001).

Soziale Heterogenität. Der Kapitaltheorie Bourdieus (2012) folgend entstehen aufgrund unterschiedlichen Ausmaßes an sozialem Kapital herkunftsbedingte Unterschiede, die wiederum von Relevanz für die Leistungen der Schülerinnen und Schüler sein können. Studien zur Auswirkung dieser sozialen Heterogenität auf Schule, Unterricht und das Lernen existieren bisher kaum (Gebauer et al., 2013). Einige Studien berichten von Zusammenhängen des sozialen Status mit Peer-Beziehungen, der sozialen Eingebundenheit und der Leistungsmotivation (z. B. Looser, 2011). Der oben erwähnten Theorie Bourdieus folgend, gibt es einen Zusammenhang von sozialem und ökonomischem Kapital, der in der Forschung oftmals als sozioökonomischer Status erfasst wird. Für den Zusammenhang des so operationalisierten sozialen Status der Herkunftsfamilie und der schulischen Leistung findet sich in zahlreichen Large-Scale-Studien wie z. B. IGLU, PISA oder TIMSS Evidenz (vgl. zusammenfassend z. B. Blossfeld et al., 2007).

Leistungsheterogenität. Eine dritte Form stellt die Leistungsheterogenität dar. Sie beschreibt die unterschiedliche Leistungsausprägung von Schülerinnen und

Schülern trotz gleichen Alters und gleicher Bildungsstufe (vgl. z. B. Blossfeld et al., 2007).

3. Ableitung der Fragestellung

Der Beitrag zielt darauf ab, die Konstruktvalidität des von Gebauer et al. (2013) entwickelten Instruments zur Erfassung der Einstellungen von Lehrkräften zu verschiedenen Dimensionen von Heterogenität zu untersuchen. Er konzentriert sich dabei auf die konfirmatorische Prüfung der faktoriellen Struktur des Instruments sowie der Generalisierbarkeit dieser über die Heterogenitätsformen (kulturelle, soziale und leistungsbezogene Heterogenität) hinweg. In bisherigen Publikationen (z. B. Hartwig, Schwabe, Gebauer & McElvany, 2017; Gebauer & McElvany, 2017) wurden keine konfirmatorischen Faktorenanalysen für die postulierte fünf-faktorielle Struktur in allen drei Heterogenitätsformen durchgeführt (weder konsekutiv je Heterogenitätsform noch simultan). Dies lässt die Konstruktvalidität des Instrumentes im Ungewissen, da beispielsweise unklar bleibt, ob die Skalen *Kosten* und *Nutzen*, tatsächlich als zwei verschiedene Merkmale interpretiert werden können, sich die fünf-faktorielle Struktur auch in den Mittelwerten der Skalen je Proband (über die Heterogenitätsformen hinweg) zeigt etc. Der vorliegende Beitrag ist darauf angelegt, Evidenz für diese Fragen und damit für die strukturelle Validität des Konstruktes von Gebauer et. al (2013) zu generieren, indem er den folgenden Forschungsfragen nachgeht:

1. Lässt sich die fünf-faktorielle Struktur des Instruments für jede Heterogenitätsform konfirmatorisch bestätigen? Aufgrund der Ergebnisse bei Gebauer et al. (2013) erwarten wir jeweils akzeptable Modellanpassungen bei substantieller Kovariation der Faktoren.
2. Lässt sich die postulierte fünf-faktorielle Struktur simultan für die Within- und between-person-Ebene generalisieren? In konfirmatorischen Mehrebenen-Faktorenanalysen erwarten wir erstens ebenfalls akzeptable Modellpassungen sowie zweitens substantielle Intraklassenkorrelationen der Items und latenten Faktoren als Evidenz für faktorielle Validität des gesamten untersuchten Instrumentes.
3. Unterscheiden sich Einstellungen von Lehrkräften zu heterogenen Schülerschaften zwischen den Formen der Heterogenität nach Kontrolle der personenspezifischen Mittelwerte? Wir erwarten, dass sich Unterschiede in den Einstellungen zu Heterogenität auch nach Kontrolle der personenspezifischen Mittelwerte absichern lassen. Insbesondere bei den negativ gepolten Dimensionen (Kosten, negative Emotionen) erwarten wir aufgrund früherer Befunde größere Mittelwerte für die Form der Leistungsheterogenität.

Diesen Forschungsfragen und Hypothesen zur strukturellen Validität von Einstellungen zu heterogenen Schülerschaften wird im Folgenden mit einer Serie von konfirmatorischen (Mehrebenen-)Faktorenanalysen nachgegangen.

4. Methode

4.1 Stichprobe

Zur Beantwortung der Forschungsfragen wurden Fragebogendaten von $N = 879$ Lehramtsstudierenden (28.7 % männlich) an sechs verschiedenen Hochschulstandorten in Südwestdeutschland herangezogen, wovon ca. die Hälfte in Studiengängen des gymnasialen Lehramts (47.6 %) eingeschrieben waren, während die andere Hälfte sich im Wesentlichen aus Studiengängen des Grund- und Hauptschullehramts (43.2 %) sowie des Realschullehramts (4.2 %) rekrutierte. Die Studierenden waren zumeist in höheren Studiensemestern ($M = 6.12$, $SD = 2.37$) und zwischen 20 und 46 Jahren alt ($M = 23.59$, $SD = 3.10$). Die Teilnehmerinnen und Teilnehmer nahmen freiwillig an der Studie teil und erhielten dafür keine Belohnungen. Die Rekrutierung erfolgte durch Dozierende in Präsenzsitzungen der jeweiligen Bildungseinrichtungen. Auch die Datenerhebung wurde in Präsenzsitzungen durchgeführt (Gruppenvorgaben).

4.2 Instrument

Einstellungen und motivationale Orientierungen bzgl. heterogener Lerngruppen werden mit dem zu validierenden Instrument von Gebauer et al. (2013) erhoben, dessen theoretische Struktur in Tabelle 1 beschrieben ist. Beispielhafte Frageprompts und Items sind Tabelle 2 zu entnehmen.

4.3 Statistische Analysen

Hierarchische Daten. Da das zu validierende Instrument dieselben Items für jede Form der Heterogenität einsetzt, entstehen hierarchisch genestete Daten: Jede Antwort ist einer Form von Heterogenität (Ebene-1) zugeordnet, wobei jede Person (Ebene-2) diese Items jeweils zu drei Formen der Heterogenität beantwortet. Diese hierarchische Nestung kann bei klassischen Verfahren mit Problemen wie z. B. verzerrten Parameterschätzungen einhergehen und muss deshalb in der statistischen Modellierung berücksichtigt werden.

Konfirmatorische Mehrebenen-Faktorenanalyse. Als eine adäquate multivariate Methode zur Beantwortung der Forschungsfragen 2 und 3 unter Berücksichtigung der hierarchischen Natur der Daten sehen wir konfirmatorische Mehrebenen-Faktorenanalysen (KMFA) an (Hox, 2010; Lüdtke, Trautwein, Schnyder & Niggli,

2007). Die Grundidee der KMFA besteht darin, den Wert der Antwort y_t einer Person bzgl. einer bestimmten Heterogenitätsform als eine Differenz dieser Antwort vom Mittelwert y_b aller Antworten dieser Person zu diesem Item (über die drei Formen von Heterogenität hinweg) und einem Abweichungswert y_w zu betrachten: $y_t = y_w + y_b$. Auch die Populations-Kovarianzmatrix aller Antworten Σ_T kann in eine heterogenitätsformspezifische Ebene-1 Kovarianzmatrix Σ_w und eine personenspezifische Ebene-2 Kovarianzmatrix Σ_b zerlegt werden. Dabei stellen Σ_b die Kovarianzen der Personenmittelwerte und Σ_w die Kovariationen der Abweichungsterme dar. Die konfirmatorische Prüfung faktorieller Strukturen erfolgt bei der KMFA nun auf beiden Ebenen simultan anhand der beiden Kovarianzmatrizen – wobei die Schätzung dieser komplexer ist als im Falle nur einer vorhandenen Ebene (Heck & Thomas, 2015). Evaluationskriterien für die Modellanpassung von KMFA (etwa anhand von Fit-Indices) gelten als noch nicht etabliert (Nagengast, Trautwein, Kelava & Lüdtke, 2013). Daher wird hier auf die Empfehlungen für herkömmliche konfirmatorische Faktorenanalysen zurückgegriffen (Hu & Bentler, 1999; Marsh, Hau & Wen, 2004).

Während die im Folgenden dargestellten herkömmlichen konfirmatorischen Faktorenanalysen zur Beantwortung von Forschungsfrage 1 mit dem Paket lavaan (Rosseel, 2012) der freien und quelloffenen Statistiksoftware R (R Core Team 2014) durchgeführt wurden, kam zur Schätzung der Parameter der KMFA-Modelle die Software MPlus 7 (Muthén & Muthén, 2012) zum Einsatz. Dieses Softwarepaket bietet unter anderem die Möglichkeit, mit Hilfe des Full-Information-Maximum-Likelihood-Schätzverfahrens (bzw. robuster Varianten) fehlenden Werten mit modellimmanenten Methoden zu begegnen, was bei einem vorliegenden Prozentsatz fehlender Werte bis 2,3 % als ein angemessenes Verfahren gilt (Acock, 2012) und deshalb für die hier berichteten Analysen eingesetzt wurde.

5. Ergebnisse

Die Ergebnisse der konfirmatorischen (Mehrebenen-)Faktorenanalysen werden nach Forschungsfragen gegliedert dargestellt. Deskriptive Ergebnisse der Skalenergebnisse sowie Maße der internen Konsistenz werden im Anschluss an die Ergebnisse zur ersten Forschungsfrage dargestellt, da Skalenbildung und Reliabilitätsschätzung eine Dimensionalitätsprüfung voraussetzen (Dunn, Baguley & Brunson, 2014).

5.1 Dimensionalität je Heterogenitätsform (Forschungsfrage 1)

In einem ersten Analyseschritt wurde je Heterogenitätsform ein fünfdimensionales Modell mit τ -kongenerischen Messmodellen spezifiziert, in dem keine Nebenladungen und Residualkovarianzen zugelassen wurden. Diese Modelle ergaben akzeptable Fit-Indices (Hu & Bentler, 1999; Marsh, Hau & Wen, 2004) für zwei Formen der Heterogenität (kulturelle Heterogenität: $\chi^2 = 756.712$, $df = 265$,

CFI = .918, TLI = .907, RMSEA = .048, SRMR = .049, Leistungsheterogenität: $\chi^2 = 1056.735$, $df = 265$, CFI = .902, TLI = .889, RMSEA = .061, SRMR = .061), für die soziale Heterogenität verfehlen die Fit-Indices klassische Cut-Off-Werte teils knapp ($\chi^2 = 1191.426$, $df = 265$, CFI = .891, TLI = .877, RMSEA = .065, SRMR = .057). Nach Freisetzen einer Residualkovarianz zweier Items eines Faktors, deren Auswahl anhand von Modifikationsindizes erfolgte, verbesserte sich die Modelpassung auf ein befriedigendes Niveau ($\chi^2 = 1013.202$, $df = 264$, CFI = .912, TLI = .900, RMSEA = .059, SRMR = .054).

5.2 Deskriptive Ergebnisse und interne Konsistenz

In Tabelle 3 sind klassische Verteilungskennwerte der eingesetzten Items, aufgeschlüsselt nach Heterogenitätsform, dargestellt. Die Ergebnisse scheinen deskriptiv mit denen von Gebauer et al. (2013) vergleichbar zu sein: Studierende schöpfen die Skala meist aus, ohne dass Decken- oder Bodeneffekte erkennbar wären. Die Standardabweichung liegt i. d. R. bei der Hälfte einer Likertstufe, wobei die arithmetischen Mittel der Skalenwerte (nach Rekodierung der negativ gepolten Dimensionen Leistung und Kosten) mit einer Ausnahme im Bereich der Zustimmung liegen. Durchschnittlich weisen die untersuchten Studierenden also positiv konnotierte Einstellungen auf: Die arithmetischen Mittel für Nutzen, Kompetenz und Motivation liegen im Bereich der Zustimmung der Likert-Skalen, die arithmetischen Mittel der Dimensionen Kosten und negative Emotionen (mit einer Ausnahme) im Bereich der Ablehnung dieser Skala.

Tabelle 3: Parameter zur Beschreibung der Verteilung der Skalenwerte

Dimension	Form	<i>N</i>	<i>M</i>	<i>SD</i>	Min.	Max.
Wert	kulturell	879	3.2	0.42	1	4
	leistungsbezogen	877	3.0	0.50	1	4
	sozial	869	3.1	0.50	1	4
Kosten	kulturell	879	2.4	0.54	1	4
	leistungsbezogen	876	2.7	0.60	1	4
	sozial	869	2.3	0.59	1	4
negative Emotionen	kulturell	878	2.0	0.55	1	4
	leistungsbezogen	876	2.3	0.62	1	4
	sozial	867	2.0	0.61	1	4
eigener kompetenter Umgang	kulturell	874	2.9	0.39	1.6	4
	leistungsbezogen	868	3.0	0.44	1.4	4
	sozial	867	3.0	0.45	1.8	4
intrinsische motivationale Orientierung	kulturell	878	3.1	0.50	1	4
	leistungsbezogen	872	2.9	0.56	1	4
	sozial	871	2.8	0.57	1	4

Anmerkungen: *Min.* und *Max.* stellen empirische Werte dar (theoretischer Wertebereich = [1; 4]).

Die Schätzung der Reliabilität erfolgte aufgrund der bestätigten τ -kongenerischen Messmodelle mit Hilfe des Koeffizienten McDonalds ω (Dunn et al., 2014), da der vielfach gebrauchte Koeffizient Cronbachs α mindestens essentiell τ -äquivalente Messmodelle voraussetzt, auch wenn diese Voraussetzung in der Forschungspraxis kaum geprüft wird (Green & Yang, 2009). Tabelle 4 gibt eine Übersicht über die Reliabilität der fünf Skalen je Heterogenitätsform. Da die Ausprägungen des Koeffizienten McDonalds ω analog zu jenen Cronbachs α zu interpretieren sind, kann mit zwei akzeptablen Ausnahmen von insgesamt guter interner Konsistenz gesprochen werden.

Tabelle 4: Punktschätzung der Reliabilitätskoeffizienten McDonalds Omega

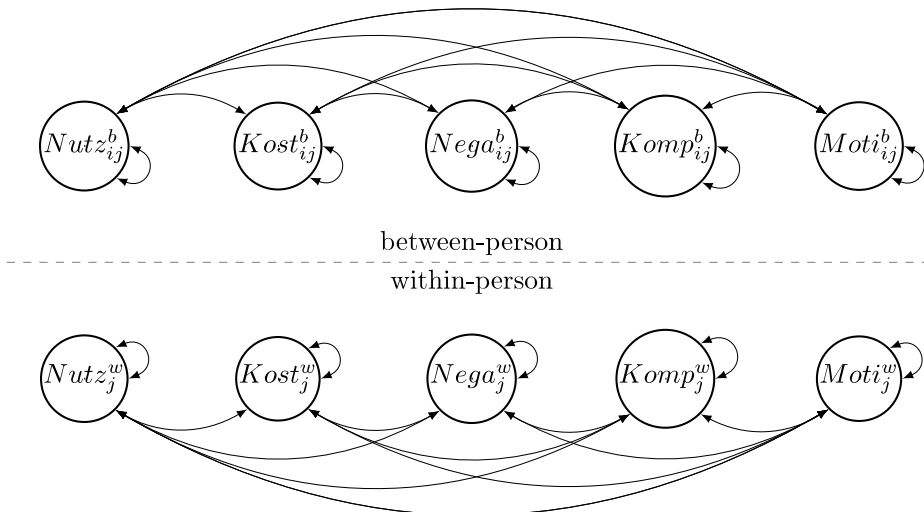
Form	McDonalds ω			
	eigener kompetenter Umgang	Kosten	negative Emotionen	intrinsische motivationale Orientierung
kulturell	.70	.72	.81	.86
leistungsbezogen	.78	.79	.86	.87
sozial	.78	.79	.88	.87

5.3 Dimensionalität bei simultaner Modellierung der Ebenen (Forschungsfrage 2)

Die konfirmatorische Bestätigung der fünf-faktoriellen Struktur, die zur Bearbeitung von Forschungsfrage 1 konsekutiv für alle drei Formen der Heterogenität durchgeführt wurde, lässt offen, ob sich diese Struktur auch für die Personenebene bestätigen lässt. Dieser Frage wurde jeweils nachgegangen, indem zunächst ein Modell spezifiziert wurde, das simultan auf beiden Ebenen eine fünf-faktorielle Struktur mit τ -kongenerischen Messmodellen annimmt (siehe Abbildung 1). Lässt man bei einem solchen Modell auf Ebene-1 drei und auf Ebene-2 sieben Residualkovarianzen (ausgewählt anhand von Modifikationsindizes) innerhalb der Faktoren zu, erhält man Fit-Indices, die für eine sehr gute Modellanpassung auf der Ebene der Heterogenitätsformen und eine akzeptable Passung der Daten zum Modell auf der Personenebene sprechen ($\chi^2 = 1772.8$, $df = 522$, CFI = .936, TLI = .927, RMSEA = .030, $SRMR_{within} = .050$, $SRMR_{between} = .088$). Die erhobenen Daten sprechen also nicht nur für die fünf-faktorielle Struktur der Einstellungen zu den drei einzelnen Heterogenitätsformen, es lässt sich auch nachweisen, dass die Annahme einer fünf-faktoriellen Struktur über die Heterogenitätsformen hinweg (Between-person-Ebene) gerechtfertigt ist, was die Forschungsfrage 2 diesbezüglich positiv bestätigt.

Die im Rahmen dieser Analyse ebenfalls erhaltenen Intraklassenkorrelationskoeffizienten (ICC1) der einzelnen Items sind weiterhin ein quantitatives Maß der Heterogenitätsformspezifität der Einstellungen, da sie als der Anteil der Gesamtvarianz des Items interpretierbar sind, welcher auf Ebene 1 zu lokalisieren ist (Lüdtke, Robitzsch, Trautwein & Kunter, 2009). Weil diese ICC1-Koeffizienten einen Wertebereich zwischen .142 und .549 aufweisen ($M = 0.33$, $SD = 0.08$) kann konstatiert werden, dass die Items zwar mehrheitlich heterogenitätsformspezifische Varianzanteile aufweisen, jedoch gleichzeitig ein substantieller Anteil auf Personenebene zu lokalisieren ist. Dies ist ebenfalls konsistent zu den in Forschungsfrage 2 formulierten Hypothesen.

Abbildung 1: Konfirmatorische Mehrebenen-Faktorenanalyse zur Beantwortung der Forschungsfrage 2. Die Messmodelle sind aus Gründen der Übersichtlichkeit nicht dargestellt.



Anmerkungen. Nutz = Nutzen, Kost = Kosten, Nega = negative Emotionen, Komp = eigener kompetenter Umgang, Moti = intrinsisch motivationale Orientierung.

Um eine Intraklassenkorrelation der latenten Faktoren zu berechnen, ist es notwendig, die Messmodelle ebeneninvariant zu spezifizieren. Erst dann ist auch gewährleistet, dass die Faktoren auf beiden Ebenen wirklich in gleicher Art und Weise interpretierbar sind. Ein solches Modell ergab eine gute Modellpassung für die globalen Fit-Indices, der Wert für $SRMR_{\text{between}}$ schließt eine Missspezifikation des Modells auf Personenebene jedoch nicht aus ($\chi^2 = 1977.901$, $df = 541$, $CFI = .927$, $TLI = .919$, $RMSEA = .032$, $SRMR_{\text{within}} = .055$, $SRMR_{\text{between}} = .108$). Mit entsprechender Vorsicht sind die ICC der latenten Faktoren zu interpretieren, welche sich zu .79 (Wert), .03 (Kosten), .26 (Negative Emotionen), .60 (Kompetenz) und .48 (Motivation) ergeben. Mit einem χ^2 -Differenztest wurde schließlich inferenzstatistisch entschieden, welches der beiden Modelle zu bevorzugen ist. Das

Ergebnis spricht mit $\Delta\chi^2(19) = 175.793$, $p < .000$ für das weniger restringierte Modell (welches unterschiedliche Messmodelle auf beiden Ebenen zulässt). Damit kann die diesbezügliche Hypothese zu Forschungsfrage 2 nicht eindeutig beantwortet werden.

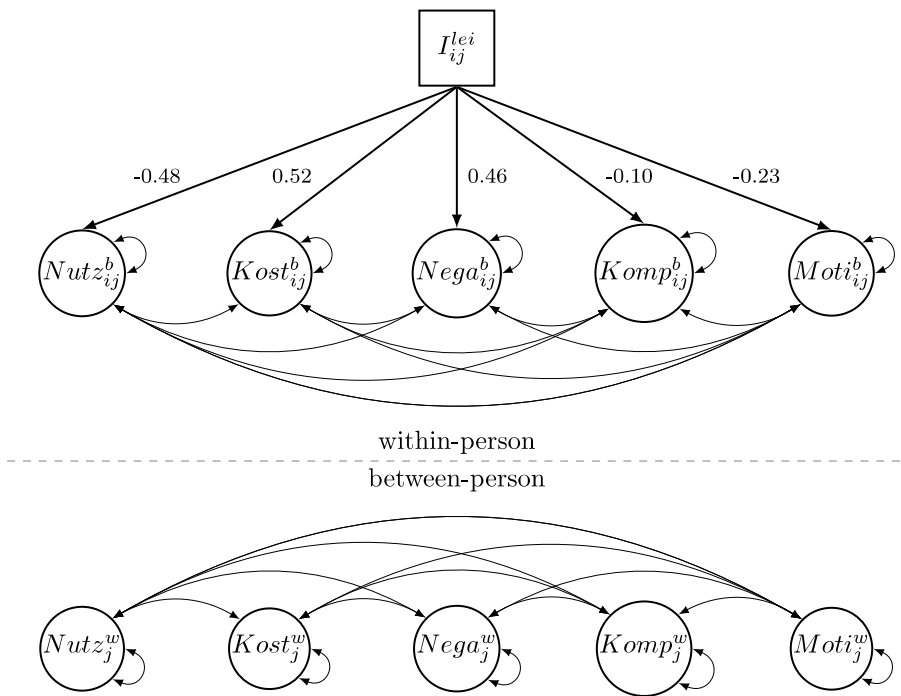
5.4 Differenzierung zwischen Heterogenitätsformen (Forschungsfrage 3)

Um der Frage nachzugehen, inwiefern das vorliegende Instrument die theoretische Differenzierung zwischen den verschiedenen Formen der Heterogenität (konstrukt-)valide zu unterscheiden in der Lage ist, kann auf das Ergebnis der Intraklassenkorrelationen der latenten Faktoren des vorherigen Abschnitts verwiesen werden. Deren stark variierende Ausprägung gibt bereits Hinweise darauf, dass in einigen Faktoren große within-person Varianz vorliegt. Es bleibt jedoch die Frage, ob sich diese within-person Varianz auch in inferenzstatistisch absicherbaren Mittelwertunterschieden niederschlägt. Dem wurde nachgegangen, indem das KMFA-Modell mit ebenenvarianten Messmodellen um effektkodierte Indikatorkovariaten für die soziale Heterogenität und die Leistungsheterogenität auf der within-person Ebene erweitert wurde, so dass sich auf dieser Ebene ein klassisches Multiple Indication Multiple Causation (MIMIC) Modell (Jöreskog & Goldberger, 1975) ergab. Dabei ist die Wahl der Referenzkategorie und der daraus resultierenden Indikatorvariablen arbiträr und ohne inhaltliche Bedeutung, da die jeweiligen Ergebnisse bei anderer Wahl informationsgleich wären (Eid, Gollwitzer & Schmitt, 2013). Jedoch wies dieses Modell Konvergenzprobleme auf, weshalb die Koeffizienten zweier MIMIC-Modelle mit jeweils einer Indikatorvariablen geschätzt wurden (siehe Abbildung 2 und 3). Diese wiesen jeweils einen, in der KMFA gut bekannten Heywood-Case auf (kleine, nicht signifikant von Null verschiedene Residualvarianz), welchem nach den Empfehlungen von Chen, Bollen, Paxton, Curran und Kirby (2001) durch eine Fixierung auf Null entgegnet wurde. Die so erhaltenen Modelle zeigten jeweils gute Modellanpassungen (MIMIC-Modell 1: $\chi^2 = 2007.047$, $df = 542$, CFI = .928, TLI = .917, RMSEA = .032, $SRMR_{within} = .052$, $SRMR_{between} = .078$, MIMIC-Modell 2: $\chi^2 = 2107.47$, $df = 542$, CFI = .923, TLI = .911, RMSEA = .033, $SRMR_{within} = .052$, $SRMR_{between} = .086$).

Da die Indikatorvariablen effektkodiert wurden (um das auf den Random-Intercepts basierende Ebene 2 Modell stringent zu vorherigen Modellen interpretieren zu können), sind die Effekte als Unterschiede in der Heterogenitätsform der entsprechenden Indikatorvariable und dem Mittelwert über alle Heterogenitätsformen hinweg zu interpretieren. Nach Abbildung 2 und 3 resultiert daraus eine Bestätigung der Hypothese größerer negativer Emotionen und Kosten bzgl. der Leistungsheterogenität mit mittleren Effektstärken nach Cohen (1988). Die Einstellungen gegenüber leistungsheterogenen Schülergruppen gehen bei den untersuchten Lehramtsstudierenden außerdem mit etwas geringerer Kompetenzselbsteinschätzungen, aber deutlich geringerer Motivation und

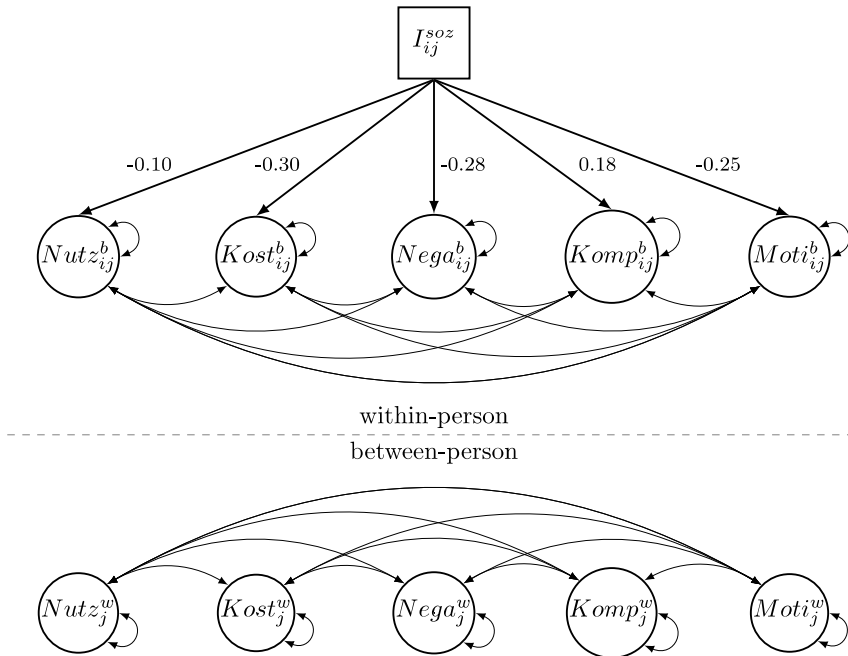
Wertschätzung einher. Die Indikatorvariable für sozial heterogene Schülergruppen weist ausnahmslos signifikante, kleine Effekte auf: Während die untersuchten Lehramtsstudierenden gegenüber ihrer durchschnittlichen Einstellung (über alle Heterogenitätsformen hinweg) geringeren Nutzen, geringere Kosten, weniger negative Emotionen und Motivation entgegenbringen, schätzen sie Kompetenz im Umgang mit sozialer Heterogenität größer ein.

Abbildung 2: MIMIC-Modell 1 zur Beantwortung von Forschungsfrage 3. Die y-standardisierten Pfadkoeffizienten entsprechen den standardisierten Mittelwertsunterschieden zwischen den Einstellungen bezüglich der Leistungsheterogenität und dem Gesamtmittelwert der Einstellungen über alle drei Heterogenitätsformen hinweg.



Anmerkungen. Nutz = Nutzen, Kost = Kosten, Nega = negative Emotionen, Komp = eigener kompetenter Umgang, Moti = intrinsisch motivationale Orientierung. Alle dargestellten Pfadkoeffizienten sind hochsignifikant ($p < .001$).

Abbildung 3: MIMIC-Modell 2 zur Beantwortung von Forschungsfrage 3. Die y-standardisierten Pfadkoeffizienten entsprechen den standardisierten Mittelwertsunterschieden zwischen den Einstellungen bezüglich der sozialen Heterogenität und dem Gesamtmittelwert der Einstellungen über alle drei Heterogenitätsformen hinweg.



Anmerkungen. Nutz = Nutzen, Kost = Kosten, Nega = negative Emotionen, Komp = eigener kompetenter Umgang, Moti = intrinsisch motivationale Orientierung. Alle dargestellten Pfadkoeffizienten sind hoch signifikant ($p < .001$).

6. Diskussion und Implikationen

Im Folgenden werden die Ergebnisse der konfirmatorischen (Mehrebenen-) Faktorenanalysen zur Bewertung der Konstruktvalidität des Instruments von Gebauer et al. (2013) zusammenfassend diskutiert. Der Beitrag schließt mit einem kurzen Ausblick.

6.1 Bewertung der Konstruktvalidität

Die durchgeführten konfirmatorischen (Mehrebenen-)Faktorenanalysen ergaben eine Bestätigung der fünfdimensionalen Struktur (Nutzen, Kosten, negative Emotionen, Kompetenz, Motivation) der Einstellungen zu heterogenen Schülerschaften. Diese Faktorenstruktur konnte sowohl für jede der drei Heterogenitäts-

formen (soziale, kulturelle, leistungsbezogene) getrennt nachgewiesen werden (Forschungsfrage 1) als auch simultan auf der Ebene der Heterogenitätsformen und der Personen (Forschungsfrage 2). Schließlich zeigte sich, dass die vorhandene substantielle Varianz innerhalb der Personen (über die Heterogenitätsformen hinweg) systematisch anhand der Heterogenitätsform prädiert werden konnte (Forschungsfrage 3).

Was bedeuten diese Ergebnisse für die Konstruktvalidität und die theoretische Grundlegung des untersuchten Instrumentes? Wie in der Einleitung dargestellt, unterliegt das Konzept der Konstruktvalidität einem deduktiven Ansatz mit der Grundidee, dass eine empirische Bestätigung (a priori getroffener) theoretischer Annahmen eines Konstrukts dessen (Konstrukt-)Validität erhöhen können. Dies ist im vorliegenden Fall für Annahmen über die Struktur des Instruments gelungen und zwar im Rahmen eines Mehrebenenansatzes. Dies stellt u. E. einerseits aufgrund folgender drei Beobachtungen ein starkes Argument für die Konstruktvalidität des untersuchten Instruments dar: Erstens lässt sich die anhand explorativer Faktorenanalysen gefundene dimensionale Struktur selbst hoch etablierter Instrumente oft nicht konfirmatorisch bestätigen – es gilt als kaum möglich, dass ein fünfdimensionales Instrument mit je fünf Items pro Faktor gute Modellanpassung zeigt (Marsh, 2007). Zweitens weisen die Ergebnisse der MIMIC-Modelle darauf hin, dass die simultane Bestätigung der faktoriellen Struktur nicht in der Weise trivial ist, als dass sie lediglich eine Messwiederholung abbildet. Vielmehr bestätigt sie die theoretische Annahme, dass zwischen den Einstellungen zu verschiedenen Formen der Heterogenität unterschieden werden muss – wie vom Konstrukt theoretisch postuliert. Ein drittes, kontrastives Argument: Die Bestätigung der faktoriellen Struktur fällt der Literatur zufolge insbesondere für Überzeugungskonstrukte schwer (Fives & Buehl, 2008).

Andererseits zeigen die im folgenden Abschnitt angeführten Limitationen, dass diese Evidenz für die Konstruktvalidität des Instrumentes nur ein Baustein für die theoretische Grundlegung des Instrumentes darstellt. Sie beantwortet beispielsweise nicht die Frage, ob es legitim ist, exakt die vorliegenden Formen von Heterogenität (kulturelle, leistungsbezogene und soziale) zu untersuchen.

6.2 Limitationen

Wenngleich die hier generierten Ergebnisse im vorherigen Absatz als überzeugende Evidenz für die faktorielle Validität des untersuchten Instrumentes gewertet wurden, liegen dieser Bewertung einige limitierende Annahmen zu Grunde, welche im Folgenden ebenso aufgeführt werden sollen, wie problematische Annahmen der Einstellungsforschung generell. Zunächst sei einschränkend angemerkt, dass eine Bestätigung faktorieller Annahmen eine notwendige Bedingung für Konstruktvalidität darstellt: Sie lässt letztlich offen, was das untersuchte Instrumentarium inhaltlich erfasst. Hier bieten sich in weiteren Untersuchungen Analysen der konvergenten/divergenten Validität an. Ebenfalls li-

mitierend berücksichtigt werden muss, dass die vorliegende Evidenz für die faktorielle Validität nichts über die Exhaustivität der erfassten Dimensionen (und der gewählten Formen) der Überzeugungen zu heterogenen Lerngruppen (siehe für einen Überblick zur Inklusion Ruberg & Porsch, 2017) sowie über deren tatsächliche Existenz (Reifikationsproblem; siehe Döring & Bortz, 2015, S. 232) aussagt. Bezüglich dieser Fragestellungen erscheinen etwa kognitive Interview-Methoden als geeignete methodische Ansätze für Studien, welche anstreben diese weiteren Aspekte der Validität dieses Instruments zu untersuchen. Zudem soll an dieser Stelle noch erwähnt werden, dass die hier untersuchte faktorielle Validität (natürlich) keinerlei Aussagen über die Genese und Handlungsrelevanz der Überzeugungen macht, wenngleich die grundlegende Bedeutsamkeit von Überzeugungen von Lehrerinnen und Lehrern für deren Handeln (für einen Überblick siehe Fives & Buehl, 2012; für eine kritische Auseinandersetzung siehe Gasterstädt & Urban, 2016) unter anderem als Motivation für diesen Beitrag aufgeführt wurde. Schließlich gilt es zu berücksichtigen, dass es sich bei den Probandinnen und Probanden der vorliegenden Studie um Lehramtsstudierende handelt. Eine Replikation der Befunde bei Lehrerinnen und Lehrern ist daher wünschenswert.

6.3 Ausblick

Die Ergebnisse (vor allem der dritten Forschungsfrage) zeigen, dass Lehramtsstudierende nicht per se eine Einstellung zum Umgang mit Heterogenität haben, sondern dass heterogenitätsbezogene Einstellungen stark nach der jeweiligen Form von Heterogenität variieren.

Während Trautmann und Wischer (2011) Heterogenität generell als Berufserschwernis beschreiben, kann somit festgehalten werden, dass Leistungsheterogenität am stärksten mit negativen Einstellungen einhergeht. Weiterführende Studien zu differenzierten Einstellungen Lehramtsstudierender zu heterogenen Lerngruppen haben damit potenziell Bedeutung für Fragen der Gestaltung institutionalisierter Lehrerbildung. Solche Forschung setzt allerdings ein valides Instrumentarium voraus – eine Grundlage für eine solche Forschung wurde durch die vorliegend nachgewiesene Konstruktvalidität des Instruments von Gebauer et al. (2013) geschaffen.

Literatur

- Acock, A. C. (2012). What to do about missing values. In H. Cooper et al. (Hrsg.), *A handbook of research methods in psychology, Vol 3: Data analysis and research publication* (S. 27–50). Washington, DC: APA.
- Baumert, J., Bos, W. & Lehmann, R. (2000). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn: Bd. 2. Mathematische und phy-*

- sikalische Kompetenzen am Ende der gymnasialen Oberstufe. Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.) (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469–520.
- Blömeke, S., Suhl, U., Kaiser, G., Felbrich, A., Schmotz, C. & Lehmann, R. (2010). Lerngelegenheiten und Kompetenzerwerb angehender Mathematiklehrkräfte im internationalen Vergleich. *Unterrichtswissenschaft*, 38 (1), 29–50.
- Blossfeld, H.-P., Bos, W., Lenzen, D., Müller-Böling, D., Oelkers, J., Prenzel, M. et al. (Hrsg.). (2007). *Bildungsgerechtigkeit. Jahresgutachten 2007*. Wiesbaden: VS Verlag.
- Bourdieu, P. (2012). Ökonomisches Kapital kulturelles Kapital, soziales Kapital. In U. Bauer, U. H. Bittlingmayer & A. Scherr (Hrsg.), *Handbuch Bildungs- und Erziehungssoziologie* (S. 229–242). Berlin: Springer.
- Brunner, M., Kunter, M. & Krauss, S. (2006). Die professionelle Kompetenz von Mathematiklehrkräften: Konzeptualisierung, Erfassen und Bedeutung für den Unterricht. Eine Zwischenbilanz des COACTIV-Projekts. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms*. (S. 54–82). Münster: Waxmann.
- Budde, J. (2012). Problematisierende Perspektiven auf Heterogenität als ambivalentes Thema der Schul- und Unterrichtsforschung. *Zeitschrift für Pädagogik*, 58 (4), 522–540.
- Chen, F., Bollen, K. A., Paxton, P., Curran, P. J. & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological methods & research*, 29 (4), 468–508.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52 (4), 281–302.
- Döring, N. & Bortz, J. (2015). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin: Springer.
- Dunn, T. J., Baguley, T. & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105 (3), 309–412.
- Eagly, A. & Chaiken, S. (1993). *The psychology of attitudes*. Orlando, FL: Harcourt Brace Jovanovich College Publishers.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2013). *Statistik und Forschungsmethoden: Lehrbuch. Mit Online-Materialien* (3. Auflage). Weinheim: Beltz.
- Eynde, P. O., Corte, E. D. & Verschaffel, L. (2002). Framing students' mathematics-related beliefs. In G. C. Leder, E. Pehkonen & G. Törner (Hrsg.), *Beliefs: A hidden variable in mathematics education?* (S. 13–37). Berlin: Springer.
- Fives, H. & Buehl, M. M. (2008). What do teachers believe? Developing a framework for examining beliefs about teachers' knowledge and ability. *Contemporary Educational Psychology*, 33 (2), 134–176.
- Fives, H. & Buehl, M. M. (2012). Spring cleaning for the “messy” construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham & T. Urda (Hrsg.), *APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors* (S. 471–499). Washington, DC: APA.

- Gasterstädt, J., & Urban, M. (2016). Einstellung zu Inklusion? Implikationen aus Sicht qualitativer Forschung im Kontext der Entwicklung inklusiver Schulen. *Empirische Sonderpädagogik*, 1 (1), 54–66.
- Gebauer, M. M., & McElvany, N. (2017). Zur Bedeutsamkeit unterrichtsbezogener heterogenitätsspezifischer Einstellungen angehender Lehrkräfte für intendiertes Unterrichtsverhalten. *Psychologie in Erziehung und Unterricht*, 64 (3), 163–180.
- Gebauer, M. M., McElvany, N. & Klukas, S. (2013). Einstellungen von Lehramtsanwärterinnen und Lehramtsanwärttern zum Umgang mit heterogenen Schülergruppen in Schule und Unterricht. *Jahrbuch der Schulentwicklung*, 17, 191–216.
- Green, S. B. & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, 74, 121–135.
- Hartig, J., Frey, A. & Jude, N. (2008). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 135–163). Berlin: Springer.
- Hartwig, S. J., Schwabe, F., Gebauer, M. M. & McElvany, N. (2017). Wie beurteilen Lehrkräfte und Lehramtsstudierende Leistungsheterogenität? Ausprägungen, Zusammenhänge und Prädiktoren von Einstellungen und Motivation. *Psychologie in Erziehung und Unterricht*, 64 (2), 94–108.
- Heck, R. H. & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques: Mlm and sem approaches using mplus*. London: Routledge.
- Helmke, A. (2004). *Unterrichtsqualität erfassen, bewerten, verbessern*. Seelze: Kallmeyer.
- Hox, J.J. (2010). *Multilevel analysis: Techniques and applications*. London: Routledge.
- HRK – Hochschulrektorenkonferenz. (2015). *Lehrerbildung für eine Schule der Vielfalt. Gemeinsame Empfehlung von Hochschulrektorenkonferenz und Kultusministerkonferenz*. Zugriff am 16.05.2017 unter <https://www.hrk.de/positionen/gesamtliste-beschluesse/position/convention/lehrerbildung-fuer-eine-schule-der-vielfalt/>
- 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.
- Jöreskog, K. G. & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70, 631–639.
- Kunz, A., Luder, R. & Moretti, M. (2010). Die Messung von Einstellungen zur Integration (EZI). *Empirische Sonderpädagogik*, 2 (3), 83–94.
- Looser, D. (2011). *Soziale Beziehungen und Leistungsmotivation. Die Bedeutung von Bezugspersonen für die längerfristige Aufrechterhaltung der Lern- und Leistungsmotivation*. Opladen: Budrich UniPress.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology*, 34 (2), 120–131.
- Lüdtke, O., Trautwein, U., Schnyder, I. & Niggli, A. (2007). Simultane Analysen auf Schüler- und Klassenebene: Eine Demonstration der konfirmatorischen Mehrebenen-Faktorenanalyse zur Analyse von Schülerwahrnehmungen am Beispiel der Hausaufgabenvergabe. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 39, 1–11.
- Marsh, H. W. (2007). Application of confirmatory factor analysis and structural equation modeling in sport and exercise psychology. In G. Tenenbaum & R. C. Eklund (Hrsg.), *Handbook of sport psychology* (S. 774–798). Hoboken, NJ: Wiley.
- Marsh, H. W., Hau, K.-T. & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, 11 (3), 320–341.

- Muthén, L. K. & Muthén, B. O. (2012). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Nagengast, B., Trautwein, U., Kelava, A. & Lüdtke, O. (2013). Synergistic effects of expectancy and value on homework engagement: The case for a within-person perspective. *Multivariate Behavioral Research*, 48 (3), 428–460.
- Prenzel, M. & Doll, J. (Hrsg.). (2002). *Bildungsqualität von Schule: Schulische und außer-schulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen*. Weinheim: Beltz.
- R Core Team. (2014). *R: A language and environment for statistical computing*. Zugriff am 16.05.2017 unter <http://www.r-project.org/>
- Rose, N. & Gerkmann, A. (2015). Differenzierung unter Schüler_innen im reformierten Sekundarunterricht – oder: warum wir vorwiegend ‚Leistung‘ beobachten, wenn wir nach ‚Differenz‘ fragen. *Zeitschrift für qualitative Forschung*, 16 (2), 191–210.
- Rosseel, Y. (2012). lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, 48 (2), 1–36.
- Ruberg, C. & Porsch, R. (2017). Einstellungen von Lehramtsstudierenden und Lehrkräften zur schulischen Inklusion. Ein systematisches Review deutschsprachiger Forschungsarbeiten. *Zeitschrift für Pädagogik*, 63 (4), 394–415.
- Schmitt, M. (2014). Konstruktvalidität. In M. A. Wirtz (Hrsg.), *Dorsch – Lexikon der Psychologie* (S. 921). Bern: Hans Huber.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15 (2), 4–14.
- Sturm, T. (2016). *Lehrbuch Heterogenität in der Schule*. Stuttgart: UTB.
- Terhart, E. (2015). Umgang mit Heterogenität: Anforderung an Professionalisierungsprozesse. In C. Fischer (Hrsg.), *(Keine) Angst vor Inklusion. Herausforderungen und Chancen gemeinsamen Lernens in der Schule* (S. 69–85). Münster: Waxmann.
- Trautmann, M. & Wischer, B. (2011). *Heterogenität in der Schule. Eine kritische Einführung*. Wiesbaden: VS Verlag.