

JERO
Journal for Educational Research Online
Journal für Bildungsforschung Online
www.j-e-r-o.com

Executive Editors:

Nele McElvany, Institut für Schulentwicklungsforschung, Technische Universität Dortmund

Cornelia Gräsel, Institut für Bildungsforschung in der School of Education, Bergische Universität Wuppertal

Editorial Office:

Andreas Sander, Institut für Schulentwicklungsforschung, Technische Universität Dortmund, Phone: 0049 (0)2 31/75 58 292, E-mail: editorialoffice@j-e-r-o.com

Editors:

Rolf Becker (Bern), Oliver Böhm-Kasper (Bielefeld), Wolfgang Böttcher (Münster), Martin Bonsen (Münster), Wilfried Bos (Dortmund), Hartmut Ditton (München), Cornelia Gräsel (Wuppertal), Stephan Huber (Zug), Olaf Köller (Kiel), Sabine Krolak-Schwerdt (Luxemburg), Harm Kuper (Berlin), Katharina Maag Merki (Zürich), Romain Martin (Luxemburg), Nele McElvany (Dortmund), Florian H. Müller (Klagenfurt), R. Daniel Muijs (Southampton), Jost Reinecke (Bielefeld), Kerstin Schneider (Wuppertal), Claudia Schuchart (Wuppertal), Wolfram Schulz (Camberwell), Birgit Spinath (Heidelberg), Petra Stanat (Berlin), Jan Van Damme (Leuven), Stefanie van Ophuysen (Münster), Andreas Voss (Hamburg), Horst Weishaupt (Frankfurt), Ludger Wößmann (München), Olga Zlatkin-Troitschanskaia (Mainz)

Please register as a reader under www.j-e-r-o.com if you want to be informed about the publication of each new issue. There is a spring edition and an autumn edition each year. The subscription of the journal is free of charge.

Please register as an author on the website www.j-e-r-o.com if you want to submit your article considering the submission guidelines.

ISSN 1866-6671

© Waxmann Verlag GmbH 2018
Steinfurter Straße 555, 48159 Münster, Germany
Phone: 0049 (0)2 51/2 65 04 0, Fax: 0049 (0)2 51/2 65 04 26,
Internet: www.waxmann.com, E-mail: info@waxmann.com

Advertising: Waxmann Verlag GmbH, Maike Rösmann: roesmann@waxmann.com
Setting: Stoddart Satz- und Layoutservice, Münster
Web Layout: Matthias Grunert, Münster
All rights reserved

Contents

Article General

<i>Sarantis Tachtsoglou & Johannes König</i> The Influence of Learning Opportunities during Initial Teacher Education on the Pedagogical Knowledge of Pre-Service EFL Teachers	3
<i>Samuel Merk, Colin Cramer, Naien Dai, Thorsten Bohl & Marcus Syring</i> Factorial Validity of Attitudes Towards Heterogeneous Student Groups.....	34
<i>Anita Sandmeier, Martin Gubler & Silvio Herzog</i> Teachers' career mobility – a structural overview.....	54
<i>Eveline Wittmann</i> Indicator development in federal states – An analysis from the perspective of governance in the context of Canadian educational reporting	74
<i>Marianne Schübach, Wim Nieuwenboom, Lukas Frei & Benjamin von Allmen</i> Open all-day schools as a means to promote mathematical skills and to reduce social or cultural disparities?	93

Inhalt

Artikel Allgemein

- Sarantis Tachtsoglou & Johannes König*
Der Einfluss von Lerngelegenheiten in der Lehrerausbildung
auf das pädagogische Wissen angehender Englischlehrkräfte..... 3
- Samuel Merk, Colin Cramer, Naien Dai, Thorsten Bohl &
Marcus Syring*
Faktorielle Validität der Einstellungen von Lehrkräften
zu heterogenen Lerngruppen 34
- Anita Sandmeier, Martin Gubler & Silvio Herzog*
Berufliche Mobilität von Lehrpersonen –
Ein strukturierter Überblick über das Forschungsfeld 54
- Eveline Wittmann*
Indikatorenentwicklung im Föderalstaat –
Governancetheoretisch fundierte Analysen im Rahmen
der kanadischen Bildungsberichterstattung..... 74
- Marianne Schübach, Wim Nieuwenboom, Lukas Frei &
Benjamin von Allmen*
Offene Tagesschulen als Mittel um Mathematikleistungen
zu fördern und sozial oder kulturell bedingte Disparitäten
zu verringern?..... 93

Sarantis Tachtsoglou & Johannes König

Der Einfluss von Lerngelegenheiten in der Lehrerausbildung auf das pädagogische Wissen angehender Englischlehrkräfte

Zusammenfassung

Die Studie untersucht die Wirkung der formellen Lerngelegenheiten in der Lehramtsausbildung auf das pädagogische Wissen von angehenden Lehrkräften des Faches Englisch. Zu diesem Zweck wurden im Rahmen der PKE-Studie (Professionelle Kompetenz von Englischlehrkräften: Fachdidaktisches Wissen angehender Englischlehrkräfte - Konzeption, Messung, Validierung) Daten von 216 Masterstudierenden und 228 Referendaren in NRW erhoben. Die Ergebnisse einer Varianzanalyse zeigen zunächst, dass in Bezug auf die Ausbildungsphase Referendare über ein höheres Niveau pädagogischen Wissens verfügen als Masterstudierende. Dieser Unterschied ist statistisch signifikant und beträgt ungefähr eine Standardabweichung. Ergebnisse aus multivariaten Analysen im Strukturgleichungsansatz zeigen, dass die Ausbildungsphase ebenfalls das pädagogische Wissen positiv beeinflusst. Dieser Effekt wird jedoch über proximale, differenzierte Maße zu Merkmalen der pädagogischen Lerngelegenheiten durch die angehenden Lehrkräfte vermittelt. Insgesamt sprechen die Ergebnisse für die Wirksamkeit der Lehrerausbildung im Fach Englisch.

Schlagwörter

Lerngelegenheiten, Pädagogisches Wissen, Wirksamkeit der Lehrerbildung

Dr. Sarantis Tachtsoglou (Korrespondenzautor), Bundeskriminalamt, Äppelallee 45, 65203 Wiesbaden, Deutschland
E-Mail: sarantis.tachtsoglou@bka.bund.de

Prof. Dr. Johannes König, Department Erziehungs- und Sozialwissenschaften, Universität zu Köln, Gronewaldstr. 2, 50931 Köln, Deutschland
E-Mail: johannes.koenig@uni-koeln.de

The Influence of Learning Opportunities during Initial Teacher Education on the Pedagogical Knowledge of Pre-Service EFL Teachers

Abstract

The aim of this study is to investigate the effect of formal learning opportunities during initial teacher education on the pedagogical knowledge of pre-service EFL (English as a Foreign Language) teachers. To accomplish this aim, we sampled within the PKE-study (Professional Competence of EFL Teachers) pre-service teachers in the first phase during their Master studies at university, and in the last year of their induction phase in North Rhine-Westphalia. The results of the analysis of variance show initially that regarding the studying phase pre-service EFL teachers in the induction phase have a higher level of pedagogical knowledge than pre-service EFL teachers in the first phase. This difference is statistically significant and is approximately one standard deviation. At the multivariate level, the results of structural equation modeling show that the studying phase also affects the pedagogical knowledge positively. This effect, however, is mediated by more proximal and differentiated pedagogical-specific learning opportunities realized by the pre-service teachers. Overall, our analysis results demonstrate the effectiveness of teacher training for EFL.

Keywords

Opportunities to learn, Pedagogical knowledge, Effectiveness of teacher education

1. Einleitung

Pädagogisches Wissen wird übereinstimmend neben fachdidaktischem Wissen und Fachwissen als wesentlicher Bestandteil des professionellen Wissens von Lehrkräften (Grossman & Richert, 1988; Terhart, 1993; Bromme, 1997) sowie als Kernbereich ihrer professionellen Kompetenz (Baumert & Kunter, 2006) angesehen. Laut Shulman (1987) bezieht sich das pädagogische Wissen (*General Pedagogical Knowledge*) auf „those broad principles and strategies of classroom management and organization that appear to transcend subject matter“ (S. 8). An diese Definition schließen die Standards für die Lehrerbildung der Kultusministerkonferenz (KMK) von 2004 an:

„Lehrerinnen und Lehrer sind Fachleute für das Lehren und Lernen. Ihre Kernaufgabe ist die gezielte und nach wissenschaftlichen Erkenntnissen gestaltete Planung, Organisation und Reflexion von Lehr- und Lernprozessen sowie ihre individuelle Bewertung und systemische Evaluation. Die berufliche Qualität von Lehrkräften entscheidet sich an der Qualität ihres Unterrichts.“ (KMK, 2004, S. 3)

In der Folge wird der universitären Lehrerbildung eine zentrale Rolle zugeschrieben, wenn die Vermittlung professionsbezogenen Wissens und dadurch pädagogischen Wissens zur Debatte stehen (Terhart, 2012). Die in der KMK (2004) formulierten Standards für die Lehrerbildung markieren zugleich eine Ergänzung der Input- um die Output-Orientierung der Lehrerbildung in Deutschland. Dadurch werden nicht nur Zielsetzungen festgelegt, die während der Lehrerbildung erreicht werden sollen, sondern es geht, zumindest konzeptionell, auch um die Frage, ob und inwiefern die gesetzten Ziele nach einem bestimmten Ausbildungsabschnitt tatsächlich erreicht werden (König & Blömeke, 2009), d. h., ob die Lehrerbildung wirksam ist (Terhart, 2012). Bezogen auf die Forschung zum pädagogischen Wissen sind erst in jüngster Zeit nennenswerte Projekte durchgeführt worden, wie ein systematisches Review der OECD belegt (König, 2014a; vgl. auch Voss, Kunina-Habenicht, Hoehne & Kunter, 2015). Zu erwähnen sind vor allem Studien wie MT21 (Blömeke et al., 2008), TEDS-M (Blömeke et al., 2010), TEDS-FU (König, Blömeke & Kaiser, 2015), TEDS-LT (Blömeke et al., 2011), TEDS-Telekom (Buchholtz et al., 2011), SPEE (Seifert & Schaper, 2010), COACTIV-R (Voss, Kunter & Baumert, 2011), LEK (König & Seifert, 2012), LEK-R (König, Buchholtz & Dohmen, 2015), DIDAKTUM (König & Rothland, 2012), BilWiss (Kunina-Habenicht et al., 2013), CME (König & Lebens, 2012) und EMW (König & Rothland, 2015). Ungefähr die Hälfte von diesen Studien (COACTIV-R, MT21, TEDS-M, TEDS-LT, TEDS-FU, TEDS-Telekom) fokussieren auf das pädagogische Wissen von Lehramtsstudierenden des Faches Mathematik, während fünf (MT21, TEDS-Telekom, LEK, EMW, CME) auf die zeitliche Entwicklung des pädagogischen Wissens abzielen, wobei ein Vergleich zwischen Lehramtsstudierenden und Referendaren nur von den Studien MT21 und CME geleistet wird. Dieser Vergleich ist jedoch wichtig, wenn der Erfolg der universitären Lehrerbildung detailliert beurteilt werden soll.

An dieser Problemstellung setzt die Studie *Professionelle Kompetenz von angehenden Englischlehrkräften* (PKE) an, die von der *Deutschen Forschungsgemeinschaft* (DFG, Geschäftszeichen: KO3947/6-1) gefördert wurde. Mit dieser Studie wurden im Sommer 2015 angehende Englischlehrkräfte für die Sekundarstufe (Lehrämter Gymnasium/ Gesamtschule und Haupt-/Real-/Gesamtschule) aus erster (Master) und zweiter Ausbildungsphase (Referendariat) in Nordrhein-Westfalen (NRW) zu ihrem fachlichen, fachdidaktischen und pädagogischen Wissen mithilfe von standardisierten Papier-Bleistift-Tests getestet. Die Testung des pädagogischen Wissens bezieht sich dabei auf das im Kontext von TEDS-M entwickelte Messinstrument. In diesem Beitrag werden auf der Basis des Datensatzes der PKE-Studie zwei Fragestellungen untersucht. Erstens wird gefragt, ob es signifikante Unterschiede im pädagogischen Wissen zwischen Master-Lehramtsstudierenden und Referendaren gibt. Da die beiden Ausbildungsphasen nur distale Indikatoren für die Variation in der organisatorisch-curricularen Gestaltung des Lehramtsstudiums darstellen, wird zweitens mithilfe von regressionsanalytischen Strukturgleichungsmodellen der Einfluss der Lerngelegenheiten – bezogen einerseits auf die in Vorlesungen und Seminaren behandelten

pädagogischen Ausbildungsinhalte sowie andererseits auf die durchgeführten lernprozessbezogenen schulpraktischen Tätigkeiten (vgl. Abschnitt 2.3 und 3.2.2) – untersucht. Diese Forschungsfrage baut auf der Annahme auf, dass der Erwerb professioneller Kompetenz von angehenden Lehrkräften durch spezifische Merkmale von Lerngelegenheiten des jeweiligen Ausbildungsprogramms beeinflusst wird (u.a. Helmke, 2007; Tatto et al., 2008; Blömeke, Kaiser & Lehmann, 2010; König & Seifert, 2012), die weitaus differenzierter zu analysieren sind als es die formale Unterscheidung in theoretische und praktische Ausbildungsphase ermöglicht (vgl. Floden, 2015). Sie zielt unter anderem darauf ab, ein zusätzliches Argument für die Validität des von uns eingesetzten Messinstruments zur Erhebung des pädagogischen Wissens zu liefern (vgl. Voss et al., 2015; König & Klemenz, 2015).

2. Forschungsgrundlage und theoretischer Hintergrund

2.1 Pädagogisches Wissen und Lehrerexpertise

Die Annahmen und Befunde der Forschung zu Lehrerexpertise (Baumert & Kunter, 2006; Blömeke, Kaiser & Lehmann, 2008; Tatto et al., 2008) stehen im Einklang mit den Ausführungen von Shulman (1987). Bezogen auf die Lehrerausbildungsforschung ist somit das Wissen und Können der Lehrenden in den Vordergrund gerückt, welche für die Bewältigung zentraler Aufgaben – wie das Unterrichten – benötigt werden (Helmke, 2007; Krauss & Bruckmaier, 2014). Lehrpersonen gelten als Experten, wenn Sie über einen längeren Zeitraum hinweg in ihrem Gebiet fundiert auf ihrem Wissen herausragende Leistungen erbringen (Gruber, 2001). Dabei gilt Lehrerexpertise nicht als eine angeborene Eigenschaft des Individuums, sondern sie besteht aus erworbenen Fähigkeiten. Speziell wird angenommen, „dass die (erfolgreiche) Tätigkeit von Lehrkräften auf Wissen und Können beruht, das in der Ausbildung in theoretischen und praktischen Phasen gewonnen und dann durch die Berufserfahrung weiterentwickelt wurde“ (Bromme, 2008, S. 159). Der Expertenansatz fokussiert somit auf das Wissen als Grundlage des Handelns (Bromme, 2008).

Sehr prominent in diesem Zusammenhang ist ein Modell zur Entwicklung der Lehrerexpertise, das von Berliner (2001, 2004) basierend auf der Arbeit von Dreyfus und Dreyfus (1986) aufgestellt worden ist (vgl. z.B. detailliert König, 2010). Demnach weist Lehrerexpertise folgende Stadien auf: (a) Novize (*novice*) (b) Fortgeschrittener Anfänger (*advanced beginner*) (c) Kompetent Handelnder (*competent performer*), (d) Profiliertes (*proficiency*) und (e) Experte (*expertise*). Studienanfänger können laut diesem Modell vor allem aufgrund ihrer fehlenden Unterrichtserfahrungen in Bezug auf ihr pädagogisches Wissen als Novizen bezeichnet werden, die sich ein Grundverständnis von Unterricht (z.B. wesentliche Begriffe und Konzepte) aneignen müssen. Sie sollen aufgrund ihrer Kenntnisse

und ersten Erfahrungen mit dem Unterrichten, die sie auch zunehmend reflektieren, bis zum Ende ihrer Ausbildung das zweite Stadium des Fortgeschrittenen Anfängers erreicht haben. Während des Berufseinstiegs kann die dritte Stufe des Kompetent Handelnden, der auch volle Verantwortung für Unterrichtshandeln übernimmt, erreicht werden, auf die dann weitere Stufen folgen (Berliner 2001, 2004). Da die Basis der Expertise theoretisch-formales Wissen ist, das häufig anhand der Unterteilung von Shulman (1986) operationalisiert wird (vgl. Bromme, 2008), und da pädagogisches Wissen ein wesentlicher Bestandteil dieses theoretisch-formalen Wissens darstellt (Baumert & Kunter, 2006), kann man anhand des Lehrerexpertise-Ansatzes erwarten, dass im Laufe der Lehrerausbildung das pädagogische Wissen der Studierenden zunimmt. Dies gilt unter der Voraussetzung, dass entsprechende Lerngelegenheiten im Laufe der Lehrerausbildung bereitgestellt werden. Bezogen auf unsere Fragestellung können Masterstudierende aufgrund ihrer fehlenden Praxiserfahrung als Novizen, Referendare im letzten Jahr ihrer zweiten Ausbildungsphase, die umfangreiche Praxiserfahrung ermöglicht, als Fortgeschrittene Anfänger bezeichnet werden.

2.2 Messung und Skalierung von pädagogischem Wissen

Wie oben erwähnt (vgl. Abschnitt 1) basiert die Testung des pädagogischen Wissens in diesem Beitrag auf dem in TEDS-M entwickelten Testinstrument. Als standardisierter Papier-Bleistift-Test wurde der Test sehr detailliert in mehreren Publikationen beschrieben (vgl. König & Blömeke, 2009; König & Blömeke, 2010a; König, 2012; König & Klemenz, 2015; König et al., 2016), sodass wir an dieser Stelle nur die Grundzüge des Testinstruments darstellen.

Determinierend für die Konzeption und Strukturierung des Tests ist die auf der Allgemeinen Didaktik und Unterrichtsforschung basierende Erkenntnis, dass das Unterrichten die Kernaufgabe von Lehrpersonen ist (KMK, 2004; König, 2010). Obwohl der Test ursprünglich für die Erfassung von Lernergebnissen am Ende der Lehrerausbildung konzipiert war, konnte in früheren Studien belegt werden, dass sich das Testinstrument auch für die Zielgruppe von Lehramtsstudierenden während der universitären Ausbildung eignet (Kaiser et al., 2010; König, 2012). Die in dieser Studie verwendete Testversion besteht aus insgesamt 40 Items, welche indikativ für die Erfassung von vier beruflichen Anforderungen sind, mit denen sich Lehrpersonen aus fachübergreifender, pädagogischer Perspektive beim Unterrichten konfrontiert sehen: (a) Umgang mit Heterogenität (b) Strukturierung von Unterricht und (c) Klassenführung/Motivierung und Leistungsbeurteilung. Neben diesen Inhaltsbereichen beinhaltet das Testinstrument eine Item-Klassifikation, die auf die Erfassung von drei Dimensionen kognitiver Prozesse rekurrierend auf die von Anderson und Krathwohl (2001) revidierte und erweiterte Bloomsche Taxonomie Bezug nimmt: (a) Erinnern (b) Verstehen/Analysieren und (c) Kreieren:

„(1) *Erinnern*. Zur Bearbeitung von Testaufgaben, die diesem kognitiven Prozess zugeordnet werden, wird Wissen benötigt, das in Erinnerung gerufen werden muss (z.B. Lehrbuchwissen und bzw. oder Erfahrungswissen); die Bearbeitung benötigt in dieser Hinsicht kognitive Prozesse wie das Nennen bzw. Erkennen. Testaufgaben dieses Typs fordern bei der Befragung angehende Lehrer auf,

- eine definitorische Nennung zu geben,
- eine Aufzählung von Elementen eines Phänomens, Begriffs oder Konzepts vorzunehmen oder
- einen Begriff oder ein Konzept zu erkennen oder zu identifizieren.

(2) *Verstehen/analysieren*. Zur Bearbeitung dieser Testaufgaben wird Wissen benötigt, das in Erinnerung gerufen werden muss (z.B. Lehrbuchwissen und bzw. oder Erfahrungswissen), das darüber hinaus aber auch mit einer Problemstellung in Verbindung gebracht werden muss; die Bearbeitung benötigt in dieser Hinsicht kognitive Prozesse wie das Beurteilen, Erklären bzw. Vergleichen. Bei der Bearbeitung der Testaufgaben dieser Dimension wurde den angehenden Lehrpersonen abverlangt,

- einen Sachverhalt, ein Konzept, einen Aspekt zu erklären oder zu charakterisieren,
- einen Begriff oder ein Konzept auszuwählen und damit eine Erklärung abzugeben oder
- einen Sachverhalt, eine Situation, eine Falldarstellung, einen oder mehrere Begriffe zu vergleichen, zu kategorisieren, zu ordnen oder zu interpretieren.

(3) *Kreieren*. Zur Bearbeitung der Aufgaben, die dieser Dimension zugeordnet werden, wird wiederum Wissen benötigt, das in Erinnerung gerufen werden muss (z.B. Lehrbuchwissen und bzw. oder Erfahrungswissen), das darüber hinaus aber auch mit einer Problemstellung und einer typischen Situation in Verbindung gebracht werden muss; die Bearbeitung benötigt in dieser Hinsicht kognitive Prozesse wie das Anwenden. Im Gegensatz zur vorherigen Dimension des Verstehens und Analysierens wird den Befragten für die Bearbeitung jedoch abverlangt,

- Handlungsoptionen zu entwickeln und zu formulieren, die ihnen für die Bewältigung einer typischen Situation zur Verfügung stehen,
- praktisches Wissen und Können, das als propositionale mentale Repräsentationen beschrieben werden kann, zu explizieren“ (König, 2012).

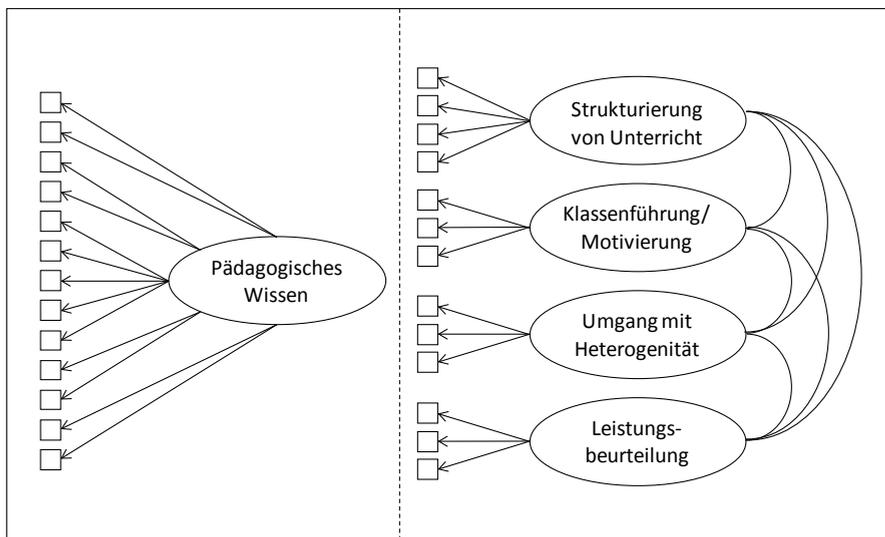
Berufsbezogene Inhaltsbereiche und Dimensionen kognitiver Prozesse bilden somit eine Matrix (vgl. Abbildung 1), welche die Basis zur Messung von pädagogischem Wissen darstellen.

Abbildung 1: Inhaltsdimensionen und Kognitive Prozesse von pädagogischem Wissen

	Erinnern	Verstehen/ Analysieren	Kreieren
Umgang mit Heterogenität			
Strukturierung von Unterricht			
Klassenführung/Motivierung			
Leistungsbeurteilung			

Die Skalierung des pädagogischen Wissens im Rahmen unserer Studie ist in König et al. (2016) sehr detailliert dargestellt. Es wurden Methoden der IRT-Skalierung mithilfe der Software *ConQuest* (Version 3.0) eingesetzt, um die Struktur des pädagogischen Wissens (Abbildung 2) zu testen. Zuerst wurde ein eindimensionales Rasch-Modell (*Partial Credit*-Modell) geschätzt. Danach wurde ein mehrdimensionales Modell (*Random Coefficient Multinomial Logit*-Modell, Rost, 1996; Wu & Adams, 2006) mit vier latenten Variablen (vgl. Abbildung 2) geschätzt. Die Schätzung der Parameter in beiden Modellen erfolgte anhand des *marginalen Maximum-Likelihood-Schätzverfahrens*. Ein Vergleich zwischen den beiden Modellen zeigte einen besseren Modell-Fit für das vierdimensionale Modell. Basierend auf dem vierdimensionalen Modell stellen die *EAP*-Werte (*Expected A Posteriori* Schätzung, vgl. De Ayala, Schafer & Sava-Bolesta, 1995) die Personenparameter zur Operationalisierung des pädagogischen Wissens in Bezug auf die folgenden bivariaten Analysen dar (vgl. Abschnitt 3.2.1 und 4.1).

Abbildung 2: Schematische Darstellung des eindimensionalen Modells (links) und des vierdimensionalen Modells (rechts) pädagogischen Wissens



Belege für die Konstruktvalidität des in der vorliegenden Studie getesteten pädagogischen Wissens wurden ebenfalls von König et al. (2016) erbracht. Die Mehrdimensionalität des Lehrerwissens von angehenden Englischlehrkräften wurde anhand konfirmatorischer Faktorenanalysen (erster Ordnung) bestätigt: Ein dreidimensionales Faktorenmodell bestehend aus den latenten Variablen Fachwissen, fachdidaktisches Wissen und pädagogisches Wissen von angehenden Englischlehrkräften zeigt einen signifikant besseren Model-Fit als ein eindimensionales Modell, das nur aus einer generischen latenten Variablen zum Lehrerwissen besteht. Bei dem dreidimensionalen Modell sind die jeweiligen latenten Variablen anhand von zwei bezüglich des Fachwissens, drei bezüglich des fachdidaktischen Wissens und vier bezüglich des pädagogischen Wissens manifesten Variablen erfasst, welche die Anzahl der Subdimensionen pro latente Variable widerspiegeln. Die manifesten Variablen beziehen sich auf die tatsächlichen Testleistungen der Studierenden zu der jeweiligen Subdimension (detailliert König et al., 2016). Bei dem dreidimensionalen Modell beträgt die latente Korrelation zwischen pädagogischem Wissen und fachdidaktischem Wissen $r = 0.77$ und ist statistisch signifikant höher als die Korrelation zwischen pädagogischem Wissen und Fachwissen ($r = 0.40$). Diese Korrelationen stimmen mit den theoretischen Überlegungen zur Struktur des Lehrerwissens überein.

2.3 Lerngelegenheiten und der Erwerb von pädagogischem Wissen

In der vorliegenden Studie folgen wir dem für die Lehrerbildung spezifizierten Ansatz, den Erwerb professioneller Kompetenz von angehenden Lehrkräften durch das Angebot und die Nutzung von Lerngelegenheiten des jeweiligen Ausbildungsprogramms zu beschreiben (u. a. Helmke, 2007; Tatto et al., 2008; Blömeke et al., 2010; König & Seifert, 2012; Voss et al., 2015). Der Grundgedanke dabei ist, dass die universitäre Lehrerausbildung wie auch das Referendariat bzw. der Vorbereitungsdienst ein Angebot darstellen, das nicht direkt zu der Entwicklung der professionellen Kompetenz der angehenden Lehrkräfte führt, sondern dass deren Entwicklung durch die individuelle Nutzung dieses Angebots durch die Studierenden bzw. Referendare determiniert wird. Dies dient uns nachfolgend als Grundlage, Merkmale der Lerngelegenheiten über die Berichte der angehenden Lehrkräfte zu erfassen und in ihrer individuellen Ausprägung als Prädiktor für das zu erwerbende Wissen zu verwenden.

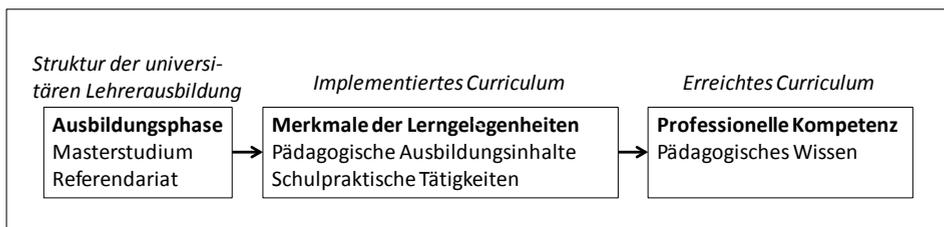
Lerngelegenheiten lassen sich im Allgemeinen in formelle und informelle Lerngelegenheiten¹ unterteilen (Eraut, 2004; Bauer et al., 2010; Kunina-Habenicht et al., 2013). Die formellen Lerngelegenheiten, auf die nachfolgend fokussiert

¹ Ferner sind die non-formalen Lerngelegenheiten zu erwähnen. Bei diesen handelt es sich um Lernprozesse, die nicht explizit institutionalisiert sind wie z.B. unabhängig von Lehrveranstaltungen stattfindender Arbeitsgruppen von Studierenden (vgl. Kunter et al. 2011).

wird, beziehen sich auf institutionalisierte Bildungseinrichtungen, die qualifizierte Abschlüsse ermöglichen (Blömeke et al., 2010). Die informellen Lerngelegenheiten dagegen finden in alltäglichen Situationen statt, wobei das Lernen aus der Sicht der Lernenden nicht notwendigerweise als Lernprozess wahrgenommen wird (Werquin, 2010; Hofmann, 2015). Die formellen Lerngelegenheiten lassen sich wiederum in das *intendierte*, *implementierte* und *erreichte Curriculum* unterteilen (McDonnell, 1995; Hascher, 2014). Das intendierte Curriculum wird üblicherweise in Studien- und Prüfungsordnungen konkretisiert und als verbindlich definiert. Es kann als Angebot betrachtet werden, auf das angehende Lehrkräfte in ihrer Ausbildung treffen (König, Blömeke & Kaiser, 2010). Das implementierte Curriculum beschreibt hingegen, welche Lehrangebote von den Studierenden tatsächlich genutzt werden (Darge, Schreiber, König & Seifert., 2012; König et al., 2010), und fokussiert somit auf die individuelle Nutzung des Lehrangebots durch die Studierenden (König, 2014b). Das erreichte Curriculum wiederum resultiert aus dem intendierten und implementierten Curriculum. In der vorliegenden Studie erfassen wir es durch das getestete pädagogische Wissen.

Theoretisch lässt sich der Einfluss der Lerngelegenheiten auf das pädagogische Wissen mithilfe von Angebots-Nutzungs-Modellen sehr gut begründen. Diese Modelle gehen vor allem auf die Ausführungen von Fend (1980) und Helmke (2007) zurück, wobei Helmke den Fokus auf der Erklärung der Wirksamkeit des Unterrichts in Bezug auf die Schülerschaft legt. Basierend auf den Vorarbeiten von Helmke haben z. B. Kunter et al. (2011) sowie König und Seifert (2012) erweiterte Angebots-Nutzungs-Modelle zur Erklärung der Entwicklung professioneller Kompetenz von (angehenden) Lehrkräften ausgearbeitet. All diesen Modellen liegt die Annahme zugrunde, dass die universitäre Lehrerausbildung ein Angebot darstellt, das nicht direkt zu der Entwicklung der professionellen Kompetenz der angehenden Lehrkräfte führt, sondern dass die eigene Entwicklung durch die individuelle Nutzung dieses Angebots durch die Studierenden determiniert wird (vgl. Abbildung 3).

Abbildung 3: Untersuchungsmodell zum Einfluss der Ausbildungsphase und der Merkmale der Lerngelegenheiten auf das pädagogische Wissen



Im Fokus unserer Studie steht somit die individuelle Wahrnehmung der angehenden Lehrkräfte bezüglich der dargebotenen Lerngelegenheiten. Als wesentlichen Begründungszusammenhang betrachten wir hierbei, dass im Lehramtsstudium

je nach Ausbildungsphase, Ausbildungsgang und Standort den Studierenden eine große Wahlfreiheit in Bezug auf die zu besuchenden Veranstaltungen eingeräumt wird, die zum Erwerb von professionellen Kompetenzen führen soll (Terhart, 2012; König, Tachtsoglou, Darge & Lünemann, 2014). Bezogen auf den Erwerb von pädagogischem Wissen dürften diejenigen inhaltlichen Komponenten der Lehrausbildung von Bedeutung sein, die sich auf fachübergreifende und schulpraktische Lerngelegenheiten beziehen. Dies sind einerseits die pädagogischen Ausbildungsinhalte und andererseits die lernprozessbezogenen schulpraktischen Tätigkeiten (vgl. Abbildung 3).

Während die pädagogischen Ausbildungsinhalte zentrale Aspekte der pädagogischen Lehre abbilden, erfassen die lernprozessbezogenen schulpraktischen Tätigkeiten stärker die Erprobung des Unterrichts. Die schulpraktischen Lerngelegenheiten werden seit mehr als zehn Jahren in alle auf die Lehrerausbildung bezogenen Studienprogramme und -standorte durch die verbindlichen Vorgaben der KMK (2004, 2015) zur Gestaltung von Curricula eingeführt bzw. erweitert oder neu strukturiert (Arnold, 2014). Sie umfassen in Bezug auf die erste Phase der Lehrerausbildung (Bachelor/Master) mindestens zwei konsekutiv angelegte Praktika, wobei das erste sich auf erziehungs- und bildungswissenschaftliche Bereiche bezieht und das zweite den Fokus auf die Fachdidaktik legt. Wie die Analyse von Curriculumsdokumenten von Arnold (2014) zeigt, beinhaltet das erste Praktikum an den deutschen Hochschulen einen beträchtlichen Anteil an Konzepten der Allgemeinen Didaktik. Da sowohl die pädagogischen Ausbildungsinhalte als auch die schulpraktischen Tätigkeiten die curriculare Struktur zentraler Komponenten der Lehrerausbildung in Zusammenhang mit dem Erwerb von pädagogischem Wissen widerspiegeln (Cochran-Smith & Zeichner, 2005; Schmidt, Cogan & Houang, 2011), kann erwartet werden, dass die individuelle Nutzung dieser beiden Lerngelegenheitsbereiche positiv mit dem Erwerb pädagogischen Wissens zusammenhängt. Des Weiteren postulieren die Angebots-Nutzungs-Modelle, dass die individuellen Eingangsvoraussetzungen die professionelle Kompetenz der (angehenden) Lehrkräfte sowohl direkt als auch indirekt über die individuelle Nutzung der Lerngelegenheiten beeinflussen (Kunter et al., 2011; Terhart, 2012; König & Seifert, 2012; Kunina-Habenicht et al., 2013; König, 2014b). Diese umfassen sowohl soziodemografische Merkmale (wie z. B. das Geschlecht, Alter, Migrationsstatus) als auch kognitive Leistungsvoraussetzungen (wie z. B. die Abiturnote). Der Grundgedanke dabei ist, dass bestimmte Persönlichkeitsmerkmale, wie z. B. kognitive Fähigkeiten oder Personenmerkmale, die auf die jeweilige soziale Lage der Personen zurückzuführen sind, ebenfalls interindividuelle Unterschiede in der professionellen Kompetenz von Lehrkräften erklären können (Kunter et al., 2011). Da der Fokus dieses Artikels auf dem Einfluss von Merkmalen der Lerngelegenheiten auf das pädagogische Wissen liegt, werden wir in die Analysen (vgl. Abschnitt 3.2.3) die individuellen Eingangsvoraussetzungen betreffenden Variablen als Kontrollvariablen einbeziehen.

Während die Forschungslage zum pädagogischen Wissen im Allgemeinen während der letzten zehn Jahre erheblich zugenommen hat, lässt sich der

Forschungsstand zum Einfluss der Lerngelegenheiten auf das pädagogische Wissen (vorausgesetzt man lässt Selbsteinschätzungsstudien außer Acht, vgl. König & Tachtsoglou, 2012; Hascher, 2014) als defizitär charakterisieren (vgl. Kunter et al., 2011; Schmidt et al., 2011; Terhart, 2012; Floden, 2015; Voss et al. 2015). Beispielsweise liegen Studien vor, die hauptsächlich auf den Einfluss von strukturellen Unterschieden in der Lehrerausbildung oder auf den Einfluss der Studiendauer auf das pädagogische Wissen fokussieren. In diesem Zusammenhang zeigen die Befunde aus MT21 (Blömeke et al., 2008, $n = 878$) basierend auf einem Vergleich von drei Kohorten (Grundstudium, Hauptstudium, Referendariat), dass das pädagogische Wissen im Laufe der Lehrerausbildung zunimmt. Empirische Befunde aus TEDS-M verdeutlichen, dass angehende Primarstufenlehrkräfte im letzten Jahr der universitären Lehrerausbildung mit einer Mathematik-Lehrberechtigung für die Klassen 1 bis 4 signifikant höhere Durchschnittswerte aufweisen als angehende Lehrkräfte, die sich für ein stufenübergreifendes Lehramt qualifizieren (König & Blömeke, 2010b). Befunde aus der Studie COACTIV-R ($n = 746$) demonstrieren (Voss et al., 2011), dass das pädagogische Wissen von Referendaren der Mathematik im zweiten Jahr des Vorbereitungsdienstes signifikant höher ist als das pädagogische Wissen von Referendaren im ersten Jahr des Vorbereitungsdienstes, wobei Mittelwertunterschiede mit praktischer Relevanz ($d \geq 2$) nur für die Subdimension des pädagogischen Wissens „Klassenführung“ zu konstatieren waren. Darüber hinaus weisen Referendare mit einem Lehramtsstudium höhere Werte im pädagogischen Wissen auf als Quereinsteiger (d. h. Referendare ohne Lehramtsstudium). Gleiches ist in der BilWiss-Studie festzustellen: Kunina-Habenicht et al. (2013) fanden basierend auf den Daten von 3.273 Referendaren in NRW, dass Referendare mit einem Lehramtsstudium über höhere Werte des pädagogischen Wissens verfügen als Quereinsteiger. Dagegen gibt es keinen statistisch signifikanten Zusammenhang zwischen der belegten Veranstaltungsanzahl und dem pädagogischen Wissen. Ferner fanden Großschedl, Harms, Kleickmann und Glowinski (2015) im Rahmen der KiL-Studie basierend auf einer Stichprobe von 274 angehenden Biologielehrkräften keinen signifikanten Unterschied im pädagogischen Wissen zwischen angehenden Gymnasiallehrkräften und angehenden Lehrkräften anderer Ausbildungsgänge (ebd.). Außerdem gibt es keinen signifikanten Unterschied zwischen Lehramtsstudierenden mit und ohne Lehrerfahrung (ebd.). Die Studiendauer hat ebenfalls keinen Effekt auf die Entwicklung des pädagogischen Wissens (ebd.).

Des Weiteren gibt es einige Studien, die den Fokus auf den Einfluss der individuellen Nutzung der Lerngelegenheiten auf das pädagogische Wissen legen. Zu erwähnen ist vor allem die Studie LEK (König, Tachtsoglou & Seifert, 2012), deren Analyseergebnisse einen signifikanten Zuwachs des mittleren pädagogischen Wissens im Laufe des Lehramtsstudiums (erhoben zweimal, im ersten und vierten Semester) offenbaren, der standortunabhängig ist. Allerdings ist dieser Zuwachs nicht auf die das pädagogische Wissen bezogene Anzahl der besuchten Lehrveranstaltungen zurückzuführen, sondern es zeigt sich vielmehr, dass an einigen Standorten die das pädagogische Wissen betreffenden Ausbildungsinhalte

den Zuwachs des pädagogischen Wissens zum Teil erklären können. Ebenfalls für einen signifikanten Wissenszuwachs von deutschen und österreichischen Lehramtsstudierenden über zwei Messzeitpunkte (1. und 5. Semester) sprechen die Befunde aus EMW (König & Klemenz, 2015), welche eine große praktische Bedeutsamkeit besitzen. Zudem weisen Studierende des Ausbildungsgangs *Lehrämter der Grundschule bzw. Primarstufe* signifikant höhere Werte des pädagogischen Wissens auf als Studierende des Ausbildungsgangs *Gymnasiallehramt bzw. Lehramt für höhere Schulen*. In dieser Studie wird auch der Effekt der schulpraktischen Lerngelegenheiten und der Ausbildungsinhalte berücksichtigt. Die Analyseergebnisse zeigen, dass Studierende, die über umfangreichere schulpraktische Erfahrungen in den Bereichen *Pädagogische Handlungssituationen durchführen, Theorien auf Situationen beziehen* und *Mit Situationen analytisch-reflexiv umgehen* berichteten, stärkere Wissenszuwächse in der Subdimension des pädagogischen Wissens *Kreieren/Generieren von Handlungssituationen* aufweisen. Pädagogische Ausbildungsinhalte haben ebenfalls spezifische Einflüsse auf den Zuwachs des pädagogischen Wissens, und zwar auf der individuellen Ebene der Studierenden wie auch auf der Ebene von Studiengängen innerhalb der Hochschulen (König, Ligtvoet, Klemenz & Rothland, 2017a). Dagegen zeigen die Ergebnisse der BilWiss-Studie (Kunina-Habenicht et al., 2013), dass das implementierte Curriculum gemessen an der belegten Veranstaltungsanzahl keinen Effekt auf das pädagogische Wissen hat. Watson, Seifert und Schaper (2017) haben zuletzt die Entwicklung des pädagogischen Wissens im Bachelor-Studium von 289 Lehramtsstudierenden in Zusammenhang mit der individuellen Nutzung der Lerngelegenheiten untersucht. Es zeigt sich, dass der Umfang an besuchten Lehrveranstaltungen mit einem hohen Wissenszuwachs einhergeht. Des Weiteren weisen Studierende, die angaben, sich intensiver in ihrem bildungswissenschaftlichen Studium mit Aspekten des Beurteilens und Diagnostizierens beschäftigt zu haben, einen höheren Wissenszuwachs zum zweiten Messzeitpunkt auf (Watson, 2015).

Wie aus den oben zitierten Studien ersichtlich ist, fokussieren die meisten auf den Einfluss von strukturellen Unterschieden der Lehrerausbildung auf das pädagogische Wissen. Dieser wird meist durch einen Vergleich zwischen den Werten des pädagogischen Wissens je Ausbildungsgang in einzelnen Ausbildungsphasen vorgenommen. Darüber hinaus wird der Einfluss der individuellen Nutzung der Lerngelegenheiten auf das pädagogische Wissen seltener und auf unterschiedliche Weise untersucht, sodass einige der oben genannten Studien (LEK, BilWiss-Studie) auf die Anzahl der besuchten Lehrveranstaltungen fokussieren, andere die Einschätzung der Studierenden zu den behandelten Ausbildungsinhalten (EMW, Dissertation von Watson) betrachten, während im Rahmen der Studie (EMW) die schulpraktischen Tätigkeiten analysiert wurden. Es gibt kaum Studien, welche die Entwicklung des pädagogischen Wissens sowohl während der universitären Lehrerausbildung als auch einschließlich des Vorbereitungsdienstes untersuchen (vgl. König, 2013). Auch gibt es außer den Analysen aus EMW (König et al., 2017a) unseres Wissens keine Studien, welche in Bezug auf die individuelle Nutzung der

Lerngelegenheiten sowohl theoretische als auch auf die Praxis bezogene Elemente der Lehrerausbildung in Zusammenhang mit dem pädagogischen Wissen untersuchen. Unsere Studie trägt zur Schließung dieser Forschungslücken bei, indem wir das pädagogische Wissen von Masterstudierenden und Referendare des Faches Englisch in Zusammenhang mit der individuellen Nutzung von theoretischen und praktischen Elementen ihrer Ausbildung in Beziehung setzen.

2.4 Fragestellungen und Hypothesen

Im Fokus dieser Studie stehen zwei Fragen:

- Gibt es Unterschiede bezüglich des pädagogischen Wissens zwischen den Master-Studierenden und den Referendaren des Faches Englisch?

Basierend auf den obigen Ausführungen zur Lehrer-Expertise und zur curricularen Struktur des Lehramtsstudiums (Universitätsstudium, Referendariat) erwarten wir, dass Referendare über ein höheres Niveau an pädagogischem Wissen verfügen als Master-Studierende (Hypothese 1, abgekürzt als H1).

- Welchen Einfluss haben die Merkmale der Lerngelegenheiten (implementiertes Curriculum) auf das Niveau des pädagogischen Wissens der Studierenden und Referendare?

Da die Ausbildungsphase (Master vs. Referendariat) nur einen distalen Indikator der Lerngelegenheiten darstellt, werden auch Skalen zur Erfassung von behandelten pädagogischen Ausbildungsinhalten und lernprozessbezogenen schulpraktischen Lerngelegenheiten (beide basierend auf Selbstberichten der Studierenden) als proximale Indikatoren einbezogen. Während Vorlesungen und Seminare, die sich auf pädagogische Themen beziehen, den Erwerb von hauptsächlich deklarativ-konzeptuellem pädagogischen Wissen unterstützen dürften, bieten lernprozessbezogene schulpraktische Tätigkeiten den Studierenden die Möglichkeit, ihr pädagogisches Wissen weiter zu strukturieren und zu prozeduralisieren (König, 2013). Somit vermuten wir mit unserem in Abbildung 3 dargestellten Untersuchungsmodell, dass mit größerem Umfang von pädagogischen Ausbildungsinhalten und lernprozessbezogenen schulpraktischen Tätigkeiten auch ein höheres Niveau des pädagogischen Wissens einhergeht (Hypothese 2a, nachfolgend abgekürzt als H2a). Des Weiteren erwarten wir, dass die Ausbildungsphase die Merkmale der Lerngelegenheiten positiv beeinflusst (Hypothese 2b, nachfolgend abgekürzt als H2b), da aufgrund der curricularen Struktur des Lehramtsstudiums, indem die bereitgestellten Lerngelegenheiten im Laufe der Lehramtsstudiums zunehmen, die Chance zur individuellen Nutzung dieser maßgeblich determiniert wird. Ebenfalls basierend auf dem in Abbildung 3 dargestellten Untersuchungsmodell vermuten wir, dass der Einfluss der Ausbildungsphase auf das pädagogische Wissen durch die Merkmale der Lerngelegenheiten vermittelt wird (Hypothese 2c, nachfolgend abgekürzt als H2c).

3. Untersuchungsmethode

Zur Prüfung der ersten Fragestellung verwenden wir das Verfahren der Varianzanalyse (vgl. Abschnitt 4.1) mithilfe von *SPSS* (Version 24), um auf deskriptiver (univariater) Ebene die Hypothese H1 zu prüfen. Um H1 auf multivariater Ebene zu prüfen, berechnen wir ein regressionsanalytisches Strukturgleichungsmodell (Modell 1) mithilfe von *Mplus* (Version 7, Muthén & Muthén, 1998–2015) mit der Ausbildungsphase als unabhängige und dem pädagogischen Wissen als abhängige Variable unter Kontrolle der die individuellen Eingangsvoraussetzungen betreffenden Variablen (vgl. Abschnitt 3.2.3). Zudem kontrollieren wir den Effekt des Ausbildungsgangs auf das pädagogische Wissen, um Unterschiede in der curricularen Struktur des jeweiligen Ausbildungsgangs zu berücksichtigen. Statistisch signifikante Unterschiede zwischen den beiden Ausbildungsphasen in den Werten des pädagogischen Wissens stellen das Entscheidungskriterium für die Geltung von H1 sowohl auf univariater als auch auf multivariater Ebene dar.

Zur Prüfung der zweiten Fragestellung berechnen wir schrittweise zwei regressionsanalytische Strukturgleichungsmodelle mithilfe von *Mplus*. Zuerst (Modell 2) berechnen wir den Effekt von Merkmalen der Lerngelegenheiten auf das pädagogische Wissen (H2a), sowie den Effekt der Ausbildungsphase auf die Merkmale der Lerngelegenheiten (H2b) unter Kontrolle der individuellen Eingangsvoraussetzungen und des Ausbildungsgangs. Wir betrachten somit neben dem pädagogischen Wissen auch Merkmale der Lerngelegenheiten als abhängige Variablen. Dabei kontrollieren wir auch den Effekt des Ausbildungsgangs auf die Merkmale der Lerngelegenheiten, da davon ausgegangen werden kann, dass je nach Ausbildungsgang die Merkmale der Lerngelegenheiten individuell durch die angehenden Lehrkräfte unterschiedlich wahrgenommen werden (vgl. König et al., 2017a). Statistisch signifikante Koeffizienten in die von H2b und H2c erwartete Richtung, stellen das Entscheidungskriterium für die Geltung der beiden Hypothesen dar. Letztlich berechnen wir ein Modell (Modell 3), das zusätzlich zu Modell 2 dem Einfluss der Ausbildungsphase auf das pädagogische Wissen Rechnung trägt. Falls die mit H2c angenommene Vermittlung zutrifft, dann sollte der Effekt der Ausbildungsphase auf das pädagogische Wissen nicht mehr signifikant sein.

Abbildung 4: ediationsmodell

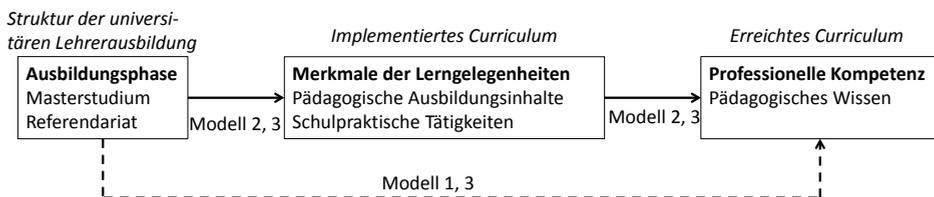
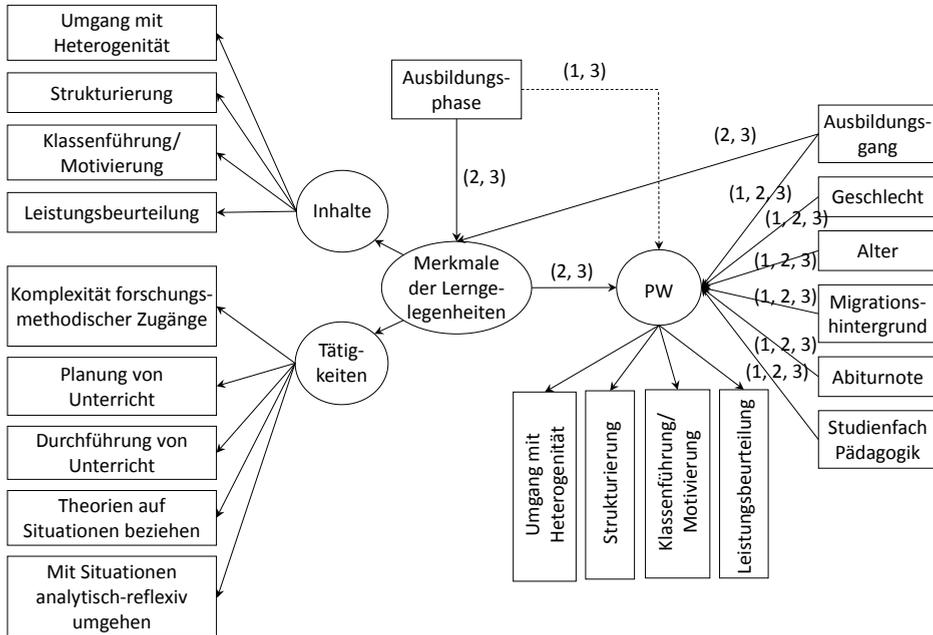


Abbildung 5: Pfadmodelle zu Untersuchung des Einflusses der Ausbildungsphase und der Merkmale der Lerngelegenheiten auf das pädagogische Wissen (die Zahlen in Klammern zeigen die zu berücksichtigenden Pfade pro Modell)



Den Kern unserer Fragestellungen bildet also ein Mediationsmodell, das in Abbildung 4 zu sehen ist, wobei aus Übersichtsgründen die Kontrollvariablen nicht dargestellt werden. Zunächst untersuchen wir den Effekt der Ausbildungsphase auf das pädagogische Wissen (Modell 1). Danach betrachten wir gleichzeitig den Effekt der Ausbildungsphase auf die Merkmale der Lerngelegenheiten und den Effekt der Merkmale der Lerngelegenheiten auf das pädagogische Wissen (Modell 2). Zum Schluss (Modell 3) untersuchen wir sowohl den direkten Effekt der Ausbildungsphase auf das pädagogische Wissen als auch den indirekten Effekt der Ausbildungsphase über die Merkmale der Lerngelegenheiten auf das pädagogische Wissen. Sollte unsere *Mediationshypothese* zutreffen, dann sollte im Modell 3 der direkte Effekt der Ausbildungsphase auf das pädagogische Wissen nicht mehr signifikant sein und im Vergleich zum Modell 1 an Stärke verlieren. Dagegen sollte der indirekte Effekt der Ausbildungsphase auf das pädagogische Wissen über die Merkmale der Lerngelegenheiten statistisch signifikant sein. Darüber hinaus sollten die Modelle 2 und 3 einen besseren Modell-Fit aufweisen als Modell 1, da diese Modelle (2 und 3) das gesamte Zusammenhangsmuster zwischen Lerngelegenheiten und pädagogischem Wissen berücksichtigen. Abbildung 5 fasst die vermuteten und zu analysierenden Beziehungen als Pfadmodell zusammen, wobei, sich die Zahlen in Klammern auf die zu berücksichtigenden Pfade pro Modell beziehen, wie in Abbildung 4. Die Kontrollvariablen befinden sich dabei auf der rechten Seite.

Des Weiteren werden bei den Strukturgleichungsmodellen fehlende Werte mithilfe der Methode der *full-information-maximum-likelihood* mit Mplus modellbasiert geschätzt, d. h. der Behandlung fehlender Werte wird dadurch Rechnung getragen, dass Analyse- und Imputationsmodell identisch sind (vgl. Lüdtke, Robitzsch, Trautwein & Köller, 2007).

3.1 Stichprobe

Im Fokus der PKE-Studie steht die Entwicklung der professionellen Kompetenz von angehenden Englischlehrkräften in NRW. Grundlage für die Ziehung unserer Stichprobe bildet eine Grundgesamtheit von angehenden Englischlehrkräften in NRW der Ausbildungsgänge Haupt-/Real-/Gesamtschule und Gymnasium/Gesamtschule, welche sich am Ende der ersten (Masterstudium) oder der zweiten (Referendariat) Ausbildungsphase befinden. Die Ziehung der Stichprobe in der zweiten Phase bezieht sich auf die Population jener Referendare mit Fach Englisch für die Sekundarstufe, die im Mai 2014 in die zweite Phase eingetreten sind. Diese werden im Falle des Lehramts Gymnasium/Gesamtschule zufällig gezogen und im Falle des Lehramts Haupt-/Real-/Gesamtschule als Vollerhebung durchgeführt, wobei im Falle des Lehramts Gymnasium/Gesamtschule der Regierungsbezirk und die Zentren für schulpraktische Lehrerausbildung (ZfsL) als Strata dienen. Der institutionelle Rücklauf auf Ebene der ZfsLs beträgt 100%, auf der individuellen Ebene der Referendare 78%. Damit entspricht die Stichprobe internationalen Standards ähnlich wie in TEDS-M, d. h., sie ist repräsentativ für eine Population von 544 angehenden Englischlehrkräften der zweiten Phase. Der Rücklauf der Masterstudierenden-Stichprobe beträgt auf Institutionsebene 92%, fehlende Statistiken zu den Studierenden lassen eine präzise Abschätzung des Rücklaufs auf Individualebene aber nicht zu. Alle angehenden Lehrkräfte wurden in Testsessions befragt bzw. getestet, die in Sitzungen von regulär stattfindenden Seminaren durchgeführt werden konnten. Die Testsessions wurden unter einheitlicher Instruktion und unter Beobachtung bzw. in Anwesenheit einer Projektmitarbeiterin bzw. eines Projektmitarbeiters administriert.

Tabelle 1 enthält Basisinformationen zu den Stichproben (216 Studierende im Masterstudium; 228 Referendare).

Tabelle 1: Stichprobenmerkmale

Phase	1. Phase/ Universität <i>n</i> = 216		2. Phase/ Referendariat <i>n</i> = 228	
	Haupt-/Real-/ Gesamtschule <i>n</i> = 35	Gymnasium/ Gesamtschule <i>n</i> = 181	Haupt-/Real-/ Gesamtschule <i>n</i> = 78	Gymnasium/ Gesamtschule <i>n</i> = 150
Ausbildungsgang				
Geschlecht (weiblich)	91.4 %	78.5 %	79.5 %	73.2 %
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Alter	24.4 (3.4)	24.5 (2.5)	29.4 (4.4)	28.6 (2.8)
Abiturnote	2.75 (.43)	2.05 (.52)	2.69 (.57)	2.23 (.56)

3.2 Instrumente

3.2.1 Pädagogisches Wissen

Verwendet werden die geschätzten EAP-Werte der angehenden Lehrkräfte (vgl. Abschnitt 2.2). Um eine bessere Lesbarkeit der varianzanalytischen Ergebnisse zu ermöglichen, werden die EAP-Werte der Referendare auf einen Mittelwert von 500 und einer Standardabweichung von 100 transformiert. Entsprechend dieser Transformation werden dann die EAP-Werte der Masterstudierenden umgerechnet. Für die Überprüfung der restlichen Hypothesen, welche anhand regressionsanalytischer Strukturgleichungsmodelle stattfinden, erfassen wir das pädagogische Wissen auf latenter Ebene anhand der tatsächlichen Testleistung der Studierenden zu den vier Subdimensionen des pädagogischen Wissens, nämlich Umgang mit Heterogenität, Strukturierung von Unterricht, Klassenführung/Motivierung und Leistungsbeurteilung. Die vier Subdimensionen gehen als Item-Parcels (d.h. der Summenwert der Items zu jeder Subdimension) in das Strukturgleichungsmodell ein (Abbildung 5).

3.2.2 Skalen zur Erfassung von Merkmalen der Lerngelegenheiten

Zur Erfassung von Merkmalen der Lerngelegenheiten, die sich auf das implementierte Curriculum beziehen, verwenden wir einerseits ein Messinstrument zur Erfassung der behandelten pädagogischen Ausbildungsinhalte und andererseits ein weiteres Messinstrument zur Erfassung der lernprozessbezogenen schulpraktischen Lerngelegenheiten. Während das erste Messinstrument im Rahmen der EMW-Studie entwickelt wurde (detailliert König et al., 2017a), wurde die Entwicklung und Konzeption des zweiten Messinstruments im Rahmen der PKE-Studie realisiert (König et al., 2016), wobei auch in diesem Fall Erhebungsinstrumente aus der Studie EMW zur Messung der pädagogischen schulpraktischen Lerngelegenheiten wegweisend waren (König et al., 2014).

Tabelle 2: Beispiel-Items zur Messung von Dimensionen pädagogischer Ausbildungsinhalte bei angehenden Lehrkräften und Reliabilität der Skalen (Cronbachs Alpha)

Dimension	Beispiel-Items	Items	Cronbachs Alpha
Umgang mit Heterogenität	Individuelle Förderung im Unterricht	5	0.66
Strukturierung	Unterrichtsmethoden	9	0.78
Klassenführung/Motivierung	Regeln im Unterricht	6	0.77
Leistungsbeurteilung	Lernprozessdiagnostik	9	0.87

Das erste Messinstrument (pädagogische Ausbildungsinhalte) besteht aus insgesamt 29 Items, die sich zu vier Skalen zusammenfassen lassen: (a) Umgang mit Heterogenität (b) Strukturierung von Unterricht (c) Klassenführung/Motivierung und (d) Leistungsbeurteilung. Die Skalen fokussieren somit auf Inhalte, die im erziehungswissenschaftlichen Studium Gegenstand sind und einen unmittelbaren Bezug zum pädagogischen Wissen aufweisen. Die Items besitzen ein dichotomes Antwortformat (0 = nein/ 1 = ja) und werden im Fragebogen mit der folgenden Frage eingeleitet: „Wurden folgende pädagogische Ausbildungsinhalte bisher in Ihrer Lehramtsausbildung behandelt?“. Die einzelnen Dimensionen weisen eine Item-Anzahl zwischen 5 und 9 auf (Tabelle 2). Die gebildeten Skalen bestehen aus dem Durchschnittswert der einzelnen Items.

Das zweite Messinstrument (schulpraktische Tätigkeiten) fokussiert auf die Prozessqualität schulpraktischer Lerngelegenheiten in der Lehrerausbildung und besteht aus insgesamt 99 Items, die sich zu fünf Skalen zusammenfassen lassen (vgl. Tabelle 3). Die erste Skala (Komplexität über forschungsmethodische Zugänge erkunden) bezieht sich auf die Erfassung von reflexiven Kompetenzen, bei denen die angehenden Lehrkräfte einschlägige wissenschaftsbezogene Prinzipien und Denkweisen in Bezug auf ihre spätere praktische Tätigkeit erwerben (König et al., 2014). Die zweite und dritte Skala (Pädagogische Handlungssituationen planen und durchführen) hat mit der Gestaltung von pädagogischen Handlungssituationen zu tun, die die Planung sowie die Durchführung von Unterricht erfordern. Die beiden letzten Skalen (Theorien auf Situationen beziehen/ Mit Situationen analytisch-reflexiv umgehen) fokussieren auf die Verknüpfung zwischen Theorie und praktischen Situationen. Mit der Skala *Theorien auf Situationen beziehen* wird auf theoretische Ansätze und Konzepte Bezug genommen, welche u.a. im Rahmen der akademischen Lehrerausbildung erworben wurden und die in der praktischen Situation genutzt und angewendet werden. Bei der Skala *Mit Situationen analytisch-reflexiv umgehen* geht es um einen analytisch-reflexiven Umgang mit schulpraktischen Situationen im Sinne eines reflektierten Praktikers (Schön, 1983), der Reflexion von Unterrichtserfahrungen und der zunehmenden Kontext- und Situationseinbettung von professionellem Wissen (Berliner, 2004). Die Items zu den fünf Skalen weisen ein dichotomes Antwortformat (0 = nein/ 1 = ja) auf und werden im Fragebogen mit der folgenden Frage eingeleitet: „Haben Sie in Ihrer bisherigen Schulpraxis/ in Ihren bisherigen Schulpraktika die folgenden

Tätigkeiten durchgeführt?“. Die einzelnen Dimensionen weisen eine Item-Anzahl zwischen 9 und 28 auf (Tabelle 3). Die berechneten Skalen bestehen ebenfalls aus dem Durchschnittswert der einzelnen Items.

Das Messinstrument zur Erfassung der schulpraktischen Tätigkeiten ist sprachlich für die fachdidaktischen schulpraktischen Lerngelegenheiten konkretisiert, da die Zielgruppe der PKE-Studie aus angehenden Englischlehrkräften besteht. Aus zeitlichen Gründen wurden aber die pädagogischen, d.h. fächerübergreifenden schulpraktischen Tätigkeiten, wie sie z.B. in der EMW-Studie erfasst werden (König et al., 2014), nicht zusätzlich erhoben. Da das Instrument zur Messung der fachdidaktischen schulpraktischen Lerngelegenheiten auf Basis des EMW-Instruments zur Messung der pädagogischen schulpraktischen Lerngelegenheiten entwickelt wurde, und da sich pädagogische Tätigkeiten immer in einem Fach konkretisieren, dürfte dies die Analyseergebnisse nicht gravierend beeinflussen. In unserer Analyse dürfte sogar der vermutete direkte Effekt auf das pädagogische Wissen eher unter- als überschätzt werden, was letztlich eine konservative Prüfung unserer Hypothesen darstellt. Darüber hinaus lässt sich argumentieren, dass die Lerngelegenheiten mithilfe von latenten Variablen in den Analysen (vgl. Abschnitt 4.2) erfasst werden, Messfehler explizit berücksichtigt werden (Geiser, 2010, S. 42). D. h. die Analysemethode sollte die suboptimale Operationalisierung der schulpraktischen Tätigkeiten zum Teil kompensieren.

Tabelle 3: Beispiel-Items zur Messung von Dimensionen schulpraktischer Tätigkeiten bei angehenden Lehrkräften und Reliabilität der Skalen (Cronbachs Alpha)

Dimension	Beispiel-Items	Items	Cronbachs Alpha
Komplexität über forschungsmethodische Zugänge erkunden	Ich habe die Häufigkeit der mündlichen Schüleräußerungen protokolliert.	9	0.70
Pädagogische Handlungssituationen planen	Ich habe die Planung einer weiteren Unterrichtsstunde auf den bisherigen Unterrichtsverlauf aufgebaut.	19	0.92
Pädagogische Handlungssituationen durchführen	Ich habe Kriterien zur Bewertung mündlicher Leistungen im Unterricht angewandt.	28	0.94
Theorien auf Situationen beziehen	Ich habe mein theoretisches Wissen über Schule und Unterricht auf Englischunterricht bezogen.	24	0.90
Mit Situationen analytisch-reflexiv umgehen	Ich habe reflektiert, ob ich mein Unterrichtsziel erreicht habe.	19	0.90

Insgesamt besteht somit die vorliegende Erfassung von Merkmalen der Lerngelegenheiten aus den beiden latenten Konstrukten *pädagogische Ausbildungsinhalte* und *schulpraktische Tätigkeiten*, d.h. es wird anhand eines faktoriellen Messmodells erster Ordnung operationalisiert (vgl. Abbildung 5). Auf eine differenzierte Analyse des direkten Einflusses beider Konstrukte auf das pädagogische Wissen wird an dieser Stelle verzichtet, weil in der PKE-Studie die Erfassung des pädagogischen Wissens nicht mit dem vollständigen TEDS-M Test,

sondern mit einer Kurzfassung vorgenommen wurde, die auf einem Matrixdesign basiert, das gemeinsam sowohl deklarative als auch prozedurale Aspekte des pädagogischen Wissens erhebt. Insofern ist eine differenzierte Analyse mithilfe von Subskalen nicht möglich. Eine Untersuchung, inwiefern pädagogische Ausbildungsinhalte, die eher auf die Vermittlung von deklarativen Komponenten des pädagogischen Wissens konzipiert sind, oder schulpraktische Tätigkeiten, die auf die Vermittlung von prozeduralen Komponenten des pädagogischen Wissens angelegt sind, werden daher in der vorliegenden Studie nicht vorgenommen.

3.2.3 Kontrollvariablen

Als Kontrollvariablen auf individueller Ebene der angehenden Lehrkräfte beziehen wir das Geschlecht, das Alter, den Migrationshintergrund, die Abiturnote, die Zugehörigkeit der angehenden Lehrkräfte neben dem Englischfach zum Studienfach Pädagogik und den Ausbildungsgang ein. Während Geschlecht, Alter und Migrationshintergrund den Einfluss von soziökonomischen Variablen berücksichtigen, bezieht sich die Abiturnote auf die kognitiven Fähigkeiten der Studierenden. Diese Persönlichkeitsmerkmale werden meist von Studien erhoben (vgl. Kunina-Habenicht et al., 2013), die sich mit dem Einfluss von individuellen Eingangsvoraussetzungen auf den Erwerb von pädagogischem Wissen befassen. Des Weiteren kontrollieren wir den Einfluss des Ausbildungsgangs, um (wie schon in Abschnitt 3 erwähnt wurde) Unterschiede in der curricularen Struktur des jeweiligen Ausbildungsgangs zu berücksichtigen. Der Migrationshintergrund wird über die Angabe zur Muttersprache der Studierenden erfasst. Dabei betrachten wir neben den Antwortkategorien *Deutsch* und *nicht Deutsch* auch die Kategorie *keine Antwort*, um die Anzahl der fehlenden Werte zu reduzieren. Es kann nämlich davon ausgegangen werden, dass es sich bei dieser Kategorie zum größten Teil um ausländische Studierende oder Studierende mit Migrationshintergrund handelt, die aufgrund von z.B. sozialer Erwünschtheit die Auskunft zu dieser Frage verweigern.

4. Ergebnisse

4.1 Deskriptive Befunde

Tabelle 4 sind die deskriptiven Kennwerte wie Mittelwert, Standardabweichung und Standardfehler des Mittelwertes für die Skalen zur Erfassung der behandelten pädagogischen Ausbildungsinhalte und lernprozessbezogenen schulpraktischen Tätigkeiten zu entnehmen. In Bezug auf die Ausbildungsinhalte verdeutlichen die Mittelwerte den Anteil der angehenden Lehrkräfte, die angaben, dass die jeweiligen Inhalte während der von ihnen bisher durchlaufenen Ausbildung behandelt worden waren. Bezogen auf die schulpraktischen Tätigkeiten zeigen die Mittelwerte in Tabelle 4 entsprechend den Anteil der Studierenden auf, die be-

richteten, dass die jeweiligen Tätigkeiten in ihrer bisherigen Schulpraxis bzw. in ihren bisherigen Schulpraktika durchgeführt worden waren. Der Anteil variiert für die Skalen zur Erfassung der Inhalte zwischen 67% (Klassenführung/Motivierung) und 88% (Strukturierung), für die Skalen zu Tätigkeiten zwischen 21% (Komplexität über forschungsmethodische Zugänge erkunden) und 64% (pädagogische Handlungssituationen durchführen).

Tabelle 4: Deskriptive Kennwerte für die Skalen zur Erfassung von Merkmalen der Lerngelegenheiten und standardisierte Faktorladungen

Variable	Mittelwert	Standardabweichung	Standardfehler	standardisierte Faktorladung
<i>Pädagogische Ausbildungsinhalte</i>				
Umgang mit Heterogenität	0.74	0.29	0.01	0.61
Strukturierung	0.88	0.19	0.00	0.59
Klassenführung/Motivierung	0.67	0.32	0.02	0.76
Leistungsbeurteilung	0.70	0.30	0.01	0.70
<i>Schulpraktische Tätigkeiten</i>				
Komplexität über forschungsmethodische Zugänge erkunden	0.21	0.21	0.01	0.67
Pädagogische Handlungssituationen planen	0.63	0.28	0.01	0.94
Pädagogische Handlungssituationen durchführen	0.64	0.28	0.01	0.97
Theorien auf Situationen beziehen	0.50	0.26	0.01	0.73
Mit Situationen analytisch-reflexiv umgehen	0.62	0.27	0.01	0.90

In der letzten Spalte der Tabelle 4 sind die Faktorladungen der entsprechenden manifesten Skalen zu sehen, welche anhand einer konfirmatorischen Faktorenanalyse (zweifaktorielle Lösung) in Mplus berechnet wurden. In Bezug auf die latente Variable *pädagogische Ausbildungsinhalte* beträgt die niedrigste Faktorladung 0.59 (für die Skala *Strukturierung*) und die höchste 0.76 (für die Skala *Klassenführung/Motivierung*). In Bezug auf die latente Variable *schulpraktische Tätigkeiten* fallen die entsprechenden Faktorladungen noch höher aus. Insgesamt weist die konfirmatorische Faktorenanalyse eine gute Modell-Fit-Statistik auf ($\chi^2/df = 2.57^{**}$, CFI = 0.98, RMSEA = 0.06, SRMR 0.04). Die Korrelation zwischen den beiden latenten Variablen beträgt 0.64 und ist hoch genug, um die Messung der Merkmale der Lerngelegenheiten mithilfe von pädagogischen Ausbildungsinhalten und schulpraktischen Tätigkeiten zu rechtfertigen (Abbildung 5). Insgesamt sprechen diese Analyseergebnisse für die Eignung der jeweiligen manifesten Skalen zur Erfassung der beiden latenten Konstrukte.

Abbildung 6: Mittelwerte im pädagogischen Wissen nach Ausbildungsphase (mit 95 %-igem Konfidenzintervall)

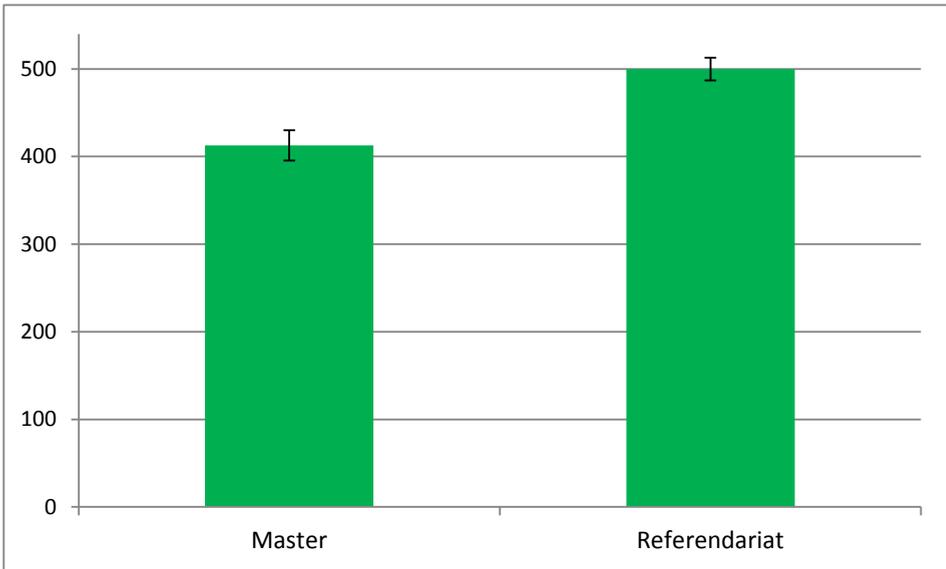


Abbildung 6 enthält die Mittelwerte des pädagogischen Wissens nach Ausbildungsphase. Wie die Ergebnisse der Varianzanalyse zeigen, ist der Unterschied zwischen dem pädagogischen Wissen von Master-Studierenden (412.65) und Referendaren (500) statistisch signifikant und beträgt fast eine Standardabweichung. Die Effektstärke² (η^2) beträgt 0.13 und deutet auf eine mittlere praktische Bedeutsamkeit hin.

4.2 Befunde aus regressionsanalytischen Strukturgleichungsmodellen

Die in diesem Abschnitt vorgestellten multivariaten Analysen basieren auf dem in Abbildung 5 dargestellten Pfadmodell. Die pädagogischen Ausbildungsinhalte werden über vier, die schulpraktischen Tätigkeiten über fünf Skalen jeweils als latente Variablen erfasst. Diese zwei latenten Variablen werden wiederum zu einer latenten Variable zusammengefasst, um die Merkmale der Lerngelegenheiten zu operationalisieren. Das pädagogische Wissen wird ebenfalls auf latenter Ebene erfasst. Bei den folgenden Analysen wurde als Stratifikationsvariable der Standort (Ausbildungsgang in Zentren für schulpraktische Lehrerausbildung

2 η^2 -Werte ab 0.01 und kleiner 0.10 deuten auf geringe praktische Bedeutsamkeit hin, Werte ab 0,10 und kleiner 0.25 deuten auf mittlere praktische Bedeutsamkeit hin, während Werte ab 0.25 auf große praktische Bedeutsamkeit der Zusammenhänge hindeuten (Bortz & Döring 2006).

bzw. Universitäten) berücksichtigt, um Verzerrungen bei der Schätzung der Standardfehler und Teststatistiken durch die Stichprobencluster zu vermeiden.

Tabelle 5: Regressionsanalytische Strukturgleichungsmodelle zur Klärung des Zusammenhangsmusters zwischen Ausbildungsphase, Merkmalen der Lerngelegenheiten und pädagogischen Wissens (standardisierte Regressionsgewichte und Standardfehler)

	Modell 1		Modell 2		Modell 2		Modell 3		Modell 3	
	Abhängige Variable		Abhängige Variablen		Abhängige Variablen		Abhängige Variablen		Abhängige Variablen	
	Pädagogisches Wissen ^a		Merkmale der Lerngeleg. ^a		Pädagogisches Wissen ^a		Merkmale der Lerngeleg. ^a		Pädagogisches Wissen ^a	
Abhängige Variablen:	β	SE	β	SE	β	SE	β	SE	β	SE
Merkmale der Lerngelegenheiten ^a					.69**	.07			.55**	.14
Ausbildungsphase ^b	.62**	.07	.88**	.02			.87**	.02	.15	.14
<i>Kontrollvariablen</i>										
Ausbildungsgang ^{b, c}	.09	.07	.04	.03	.08	.07	.04	.03	.07	.07
Studienfach Pädagogik ^b	-.05	.06			-.05	.06			-.05	.06
Alter ^b	-.04	.09			-.04	.06			-.05	.09
Geschlecht ^{b, d}	-.00	.06			-.01	.06			-.01	.06
Muttersprache ^{b, e} : nicht deutsch ^{b, c}		.05			-.20**	.06			-.20**	.06
keine Antwort ^{b, e}	-.13*	.06			-.13*	.06			-.13*	.06
Abiturnote ^b	-.17*	.07			-.15*	.07			-.15**	.07
<i>R</i> ²	.37		.75		.46		.74		.44	

Anmerkungen: ** $p < 0.01$, * $p < 0.05$, ^a) latente Variable, ^b) manifeste Variable, ^c) der Ausbildungsgang Haupt-/Real-/Gesamtschule bildet die Referenzkategorie ^d) weibliche Studierende bilden die Referenzkategorie, ^e) Muttersprache Deutsch bildet die Referenzkategorie.

Tabelle 5 enthält die Analyseergebnisse. Anhand der Ergebnisse von Modell 1 ist zunächst wieder der Effekt der Ausbildungsphase auf das pädagogische Wissen erkennbar (0.62). In Modell 2 wird der Effekt der Ausbildungsphase auf die Merkmale der Lerngelegenheiten (0.88) sowie der Effekt der Merkmale der Lerngelegenheiten auf das pädagogische Wissen (0.69) deutlich. Der indirekte Effekt der Ausbildungsphase über die Merkmale der Lerngelegenheiten auf das pädagogische Wissen beträgt 0.61 (in Tabelle 5 nicht dargestellt) und ist statistisch signifikant³ ($p < 0.01$) von Null verschieden. Wird darüber hinaus der di-

3 Die Signifikanztestung der indirekten Effekte fand anhand der Bootstrap-Methode statt, welche eine Verletzung der Annahme der Normalverteilung der Koeffizienten bei indirekten Effekten berücksichtigt (vgl. Geiser 2010).

rekte Einfluss der Ausbildungsphase auf das pädagogische Wissen berücksichtigt (Modell 3), so zeigt sich, dass ihr Effekt 0.15 beträgt und nicht mehr statistisch signifikant ist. Dagegen beträgt der Effekt der Ausbildungsphase auf die Merkmale der Lerngelegenheiten 0.87 und ist statistisch signifikant. Dies gilt auch für den Effekt der Merkmale der Lerngelegenheiten auf das pädagogische Wissen, der 0.55 beträgt. Der indirekte Effekt der Ausbildungsphase über die Merkmale der Lerngelegenheiten auf das pädagogische Wissen weist nun den Wert 0.40 (in Tabelle 5 nicht dargestellt) auf und ist statistisch signifikant³ von Null verschieden. Insgesamt sprechen die Analyseergebnisse im Modell 3 für den Mediationseffekt der individuellen Nutzung der von uns erfassten Merkmale der Lerngelegenheiten auf das pädagogische Wissen. Des Weiteren kann man Tabelle 6 entnehmen, dass bei allen Modellen die Modell-Fit-Statistik gut ist.

Tabelle 6: Model-Fit-Statistiken zu den in Tabelle 5 dargestellten Analysen

	χ^2/df	CFI	RMSEA	SRMR
Modell 1	2.58**	.85	.061	.044
Modell 2	1.80**	.96	.044	.043
Modell 3	1.81**	.96	.044	.043

Anmerkungen: ** $p < 0.01$, * $p < 0.05$

5. Diskussion

Der vorliegende Beitrag hat zum Ziel, die Wirkung der Lerngelegenheiten in der Lehramtsausbildung auf das pädagogische Wissen von angehenden Englischlehrkräften zu untersuchen. Dem Lehrerexpertise-Ansatz (Berliner, 2001, 2004) und seiner Anwendung auf die hiesige Struktur der Lehramtsausbildung (König, 2010b) folgend, erwarteten wir, dass das pädagogische Wissen der Studierenden des Faches Englisch im Laufe ihrer Ausbildung zunimmt. Darüber hinaus vermuteten wir, dass diese Zunahme durch Merkmale der Lerngelegenheiten erklärt werden könnte. Diese Vermutung schließt an derzeitige Modellvorstellungen der Lehrerbildungsforschung an (Kunter et al., 2011; König & Seifert, 2012) und wurde in diesem Beitrag anhand einer für NRW repräsentativen Stichprobe von 216 Masterstudierenden und 228 Referendaren des Faches Englisch für die Sekundarstufe getestet.

Die Ergebnisse unserer Analyse sprechen für die vermutete Lernwirksamkeit der Lehrerausbildung im Fach Englisch: Das pädagogische Wissen von Referendaren ist mit fast einer Standardabweichung deutlich höher als das der Masterstudierenden. Das bedeutet im Allgemeinen, dass sich die mit der jeweiligen *Ausbildungsphase* verbundenen Lerngelegenheiten positiv auf den Erwerb pädagogischen Wissens im Verlauf der Englischlehrerausbildung niederschlagen dürften. Insofern spricht dieses Analyseergebnis für die Geltung der H1 und ist vor

allem im Einklang mit bisherigen Analyseergebnissen zum Zuwachs an pädagogischem Wissen, die im Rahmen echter Längsschnittstudien berichtet werden (König & Seifert, 2012; König, 2013; König & Klemenz, 2015). Auf der multivariaten Ebene kann H1 ebenfalls bestätigt werden. Darüber hinaus bestätigen wir, dass Hypothese 2a ebenfalls gilt, denn die Merkmale der Lerngelegenheiten haben einen signifikanten Effekt auf das pädagogische Wissen. Diese Ergebnisse stehen im Einklang mit Studien wie LEK, EMW und BilWiss (Kunina-Habenicht et al., 2013), die ebenfalls einen positiven Effekt von Merkmalen der Lerngelegenheiten (obwohl anders operationalisiert) zeigen. Ebenfalls als bestätigt kann H2b angesehen werden, denn die Ergebnisse der multivariaten Analysen zeigen, dass es einen signifikanten Effekt der Ausbildungsphase auf die Merkmale der Lerngelegenheiten gibt. Dieser Effekt ist ebenfalls ein Beleg für die Validität der Messinstrumente zur Erfassung von Ausbildungsinhalten und schulpraktischen Tätigkeiten.

Das wichtigste Ergebnis der vorliegenden Analyse stellt allerdings die Bestätigung der Hypothese 2c dar: Der Effekt der Ausbildungsphase auf das pädagogische Wissen wird durch die Merkmale der Lerngelegenheiten im Bereich der Inhalte und schulpraktischen Tätigkeiten vermittelt, da der standardisierte Regressionskoeffizient der Ausbildungsphase auf das pädagogische Wissen nur 0.15 beträgt und nicht signifikant ist (Modell 3 in Tabelle 5), der Effekt der Ausbildungsphase auf die Merkmale der Lerngelegenheiten jedoch signifikant ist und 0.87 beträgt. Ebenfalls signifikant ist in Modell 3 der Effekt der Merkmale der Lerngelegenheiten auf das pädagogische Wissen. Zudem ist der indirekte Effekt der Ausbildungsphase über die Merkmale der Lerngelegenheiten auf das pädagogische Wissen signifikant. Die in diesem Beitrag dargestellten Analyseergebnisse können somit als wichtiger Beitrag zur Frage nach der Wirksamkeit der Lehrerbildung angesehen werden, die in der deutschsprachigen Lehrbildungsforschung seit Jahren wiederkehrend gestellt wird (u.a. Oser & Oelkers, 2001; Terhart, 2012; Voss et al., 2015). Unsere Ergebnisse sprechen für die Wirksamkeit der Lehrerbildung des Faches Englisch in NRW. Darüber hinaus liefert der auf Basis unserer Analyseergebnisse bestätigte Einfluss der Lerngelegenheiten auf das pädagogische Wissen ein zusätzliches Argument für die Validität des von uns eingesetzten Instruments zur Erfassung des pädagogischen Wissens für die Lehramtsstudierenden des Faches Englisch. Da die Zielgruppe angehender Fremdsprachenlehrkräfte bislang kaum in Lehrerbildungsstudien untersucht worden ist, stellen die hier dargelegten Ergebnisse auch eine wichtige Ergänzung zu bisherigen Lehrerbildungsstudien mit Fokus auf Mathematik (u.a. Blömeke et al., 2008, 2010; Kunter et al., 2011) und die naturwissenschaftlichen Fächer dar (z. B. Großschedl et al., 2015).

Des Weiteren sprechen unsere Analyseergebnisse für die Bedeutung schulpraktischer Lerngelegenheiten. Dies ist im Hinblick auf die Debatte über den fehlenden Bezug in Theorie und Praxis in der Lehramtsausbildung (Patry, 2014) von Bedeutung. Die Nutzung von schulpraktischen Tätigkeiten führt zur Anhäufung von pädagogischem Wissen, was bedeutet, dass die schulpraktischen Lerngelegenheiten

als eine Brücke in Bezug auf die *Theorie und Praxis*-Problematik angesehen werden können.

Obwohl unsere Analyseergebnisse vielversprechend sind, ist unsere Studie mit einigen Limitierungen behaftet. Eine Limitierung unserer Studie bezieht sich auf die Erhebung der Lerngelegenheiten insgesamt. Zwar zielt unser Messinstrument auf die Erfassung der Merkmale von formellen Lerngelegenheiten ab, ausgeschlossen werden kann aber nicht, dass informelle Lerngelegenheiten ebenfalls den Erwerb von pädagogischem Wissen vermitteln können. Unter der verschiedentlich genannten Annahme, dass der Erwerb von pädagogischem Wissen Teil eines berufsbiographischen Entwicklungsprozesses ist, der z. B. durch Beobachtung viel früher, vor der Lehrerausbildung während der Schulzeit, beginnt (Borg, 2004) und nach der Lehrerausbildung verschiedentlich, etwa im Rahmen von Fortbildungen oder Maßnahmen der Lehrerkooperation, fortgesetzt wird (Floden, 2015). In diesem Zusammenhang zeigen die Ergebnisse der LEK-Studie (König et al., 2012), dass pädagogische Vorerfahrungen (wie z. B. das Erteilen von Nachhilfeunterricht) neben formellen Lerngelegenheiten die Entwicklung des pädagogischen Wissens determinieren. Im Falle von Fremdsprachenlehrkräften wäre außerdem an weitere Lerngelegenheiten zu denken, die im Zusammenhang mit der Zielsprache stehen, etwa Auslandsaufenthalte und damit verbundene pädagogische Tätigkeiten (z. B. au pair). Insofern sollte die künftige Forschung neben der Erfassung von formellen auch auf die Erfassung von informellen Lerngelegenheiten fokussieren, um den reinen Effekt von beiden Lerngelegenheitstypen auf das pädagogische Wissen zu ermitteln.

Eine weitere Limitierung unserer Studie besteht darin, dass wir den Einfluss der individuellen Eingangsvoraussetzungen auf die Merkmale der Lerngelegenheiten nicht berücksichtigt haben. Beispielsweise zeigen empirische Befunde (König, Rothland, Tachtsoglou & Klemenz, 2017b), dass die Berufswahlmotivation angehender Lehrkräfte, insbesondere die intrinsische Berufswahlmotivation, die Nutzung der schulpraktischen Lerngelegenheiten positiv beeinflusst. Es wäre interessant zu untersuchen, ob der Einfluss von Merkmalen der Lerngelegenheiten auf das pädagogische Wissen unter Berücksichtigung der Berufswahlmotivation bestehen bleibt.

Zukünftige Forschung sollte außerdem unsere Analyseergebnisse für Lehramtsstudierende anderer Fächer replizieren. Denn es wäre auch möglich, dass der Vermittlungseffekt der Merkmale der Lerngelegenheiten je nach Lehramtsfach variiert. Dafür sprechen etwa die Analyseergebnisse von Großschedl et al. (2015), die zeigen, dass es keinen signifikanten Zusammenhang zwischen den formellen Lerngelegenheiten und dem pädagogischen Wissen angehender Biologielehrkräften gibt. Vergleichende Analysen zu den in Lehrerbildungsstudien fokussierten Domänen sind bislang kaum durchgeführt worden (vgl. Blömeke et al., 2011). Ferner sollten unsere Analyseergebnisse mithilfe von Paneldaten repliziert werden. Auf diese Weise können nicht nur Unterschiede in Bezug auf das Niveau vom pädagogischen Wissen oder in Bezug auf den Erklärungsbeitrag der formellen Lerngelegenheiten auf das pädagogische Wissen untersucht werden, sondern es

können auch Entwicklungsverläufe auf intraindividuelle Ebene ermittelt werden. Dies würde eine noch differenziertere als die hier geleistete Analyse zum Einfluss von Merkmalen der Lerngelegenheiten auf den Erwerb professionellen Wissens während der Lehramtsausbildung ermöglichen.

Danksagung

Dr. Sarah Strauß und Dr. Sandra Lammerding sei herzlich gedankt für ihre Unterstützung bei den sprachlichen Vorschlägen zur Verbesserung der Lesbarkeit des Manuskripts.

Literaturverzeichnis

- Anderson, L. W. & Krathwohl, D. R. (Hrsg.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Arnold, K.-H. (2014). Unterrichtsversuche als allgemeindidaktische Lerngelegenheit: Eine vergleichende Curriculumanalyse. In K.-H. Arnold, A. Gröschner & T. Hascher (Hrsg.), *Schulpraktika in der Lehrerbildung. Theoretische Grundlagen, Konzeption, Prozesse und Effekte* (S. 63–86). Münster: Waxmann.
- Bauer, J., Drechsel, B., Retelsdorf, J., Sporer, T., Rösler, L., Prenzel, M. & Möller, J. (2010). Panel zum Lehramtsstudium – PaLea: Entwicklungsverläufe zukünftiger Lehrkräfte im Kontext der Reform der Lehrerbildung. *Beiträge zur Hochschulforschung*, 32, 34–55.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9, 469–520.
- Berliner, D. C (2001). Learning about and learning from expert teachers. *Educational Research*, 35, 463–482.
- Berliner, D. C (2004). Describing the Behavior and Documenting the Accomplishments of Expert Teachers. *Bulletin of Science, Technology & Society*, 24, 200–202.
- Blömeke, S., Bremerich-Vos, A., Haudeck, H., Kaiser, G., Lehmann, R., Nold, G., Schwippert, K. & Willenberg, H. (Hrsg.). (2011). *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT*. Münster: Waxmann.
- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.). (2008). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare – Erste Ergebnisse zur Wirksamkeit der Lehrerbildung*. Münster: Waxmann.
- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.). (2010). *TEDS-M 2008 – Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte im internationalen Vergleich*. Münster: Waxmann.
- Borg, M. (2004). The apprenticeship of observation. *ELT journal*, 58(3), 274–276.
- Bortz, J., & Döring, N. (2006): *Forschungsmethoden und Evaluation für Human und Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Psychologie des Unterrichts und der Schule*, 1 (3), 177–212. Göttingen: Hogrefe.
- Bromme, R. (2008). Lehrerexpertise. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 159–167). Göttingen: Hogrefe.
- Buchholtz, N., Blömeke, S., Kaiser, G., König, J., Lehmann, R., Schwarz, B. & Suhl, U. (2011). Entwicklung von Professionswissen im Lehramtsstudium: Eine

- Längsschnittstudie an fünf deutschen Universitäten. In K. Eilerts, A. Hilligus, G. Kaiser, P. Bender (Hrsg.), *Kompetenzorientierung in Schule und Lehrerausbildung. Perspektiven der bildungspolitischen Diskussion, der Bildungsforschung und der Mathematik-Didaktik. Festschrift für Hans-Dieter Rinkens* (S. 201–214). Paderborn: Lit.
- Cochran-Smith, M. & Zeichner, K. M. (Hrsg.). (2005). *Studying Teacher Education: The report of the AERA Panel on research and teacher education*. Washington: American Educational Research Association.
- Darge, K., Schreiber, M., König, J. & Seifert, A. (2012). Lerngelegenheiten im erziehungswissenschaftlichen Studium. In J. König & A. Seifert (Hrsg.), *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerausbildung* (S. 44–86). Münster: Waxmann.
- De Ayala, R. J., Schafer, W. D. & Sava-Bolesta, M. (1995). An investigation of the standard errors of expected *a posteriori* ability estimates. *British Journal of Mathematical and Statistical Psychology*, 47, 385–405.
- Dreyfus, H. L. & Dreyfus, S. E. (1986). *Mind over Machine. The power of human intuition and expertise in the era of the computer*. New York: The Free Press.
- Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education*, 26(2), 247–273.
- Fend, H. (1980). *Theorie der Schule*. München: Urban und Schwarzenberg.
- Floden, R. (2015). Learning what research says about teacher preparation. In M. J. Feuer, A. I. Berman & R. C. Atkinson (Hrsg.), *Past as prologue: The National Academy of Education at 50. members reflect* (S. 279–284). Washington, DC: National Academy of Education.
- Geiser, C. (2010). *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Grossman, P. L. & Richert, A. E. (1988). Unacknowledged knowledge growth: A re-examination of the effects of teacher education. *Teaching and Teacher Education*, 4(1), 53–62.
- Großschedl, J., Harms, U., Kleickmann, T. & Glowinski, I. (2015). Preservice biology teachers' professional knowledge: Structure and learning opportunities. *Journal of Science Teacher Education*, 26(3), 291–318.
- Gruber, H. (2001). Acquisition of expertise. In N.J. Smelser & P. B. Baltes (Hrsg.), *International encyclopedia of the social and behavioral sciences* (Vol. 8) (S. 5145–5150). Amsterdam: Elsevier.
- Hascher, T. (2014). Forschung zur Wirksamkeit der Lehrbildung. In E. Terhart H. Bennewitz, M. Rothland. (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 542–571). Münster: Waxmann.
- Helmke, A. (2007). *Unterrichtsqualität erfassen, bewerten, verbessern*. (6. Aufl.). Seelze: Klett Kallmeyer.
- Hofmann, J. (2015). Untersuchung des Kompetenzaufbaus von Physiklehrkräften während einer Fortbildungsmaßnahme. In H. Niedderer, H. Fischler & E. Sumfleth (Hrsg.), *Studien zum Physik- und Chemielernen. Band 192*. Berlin: Logos Verlag.
- Kaiser, G., Buchholtz, N., Schwarz, B., Blömeke, S., Lehmann, R., Suhl, U., König, J. & Rinkens, H.-D. (2010). Kompetenzentwicklung in der Mathematik-Gymnasiallehrerausbildung – eine empirische Studie an fünf deutschen Universitäten. *Beiträge zum Mathematikunterricht 2010*, S. 465–468.
- KMK (2004). Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Standards für die Lehrerausbildung: Bildungswissenschaften. Beschluss der Kultusministerkonferenz von 16.12.2004.
- KMK (2015). Schüler, Klassen, Lehrer und Absolventen der Schulen 2004 bis 2013. Statistische Veröffentlichungen der Kultusministerkonferenz. Dokumentation Nr. 2016 – Januar 2015. Bonn: KMK.

- König, J. (2010). Lehrerprofessionalität – Konzepte und Ergebnisse der internationalen und deutschen Forschung am Beispiel fachübergreifender, pädagogischer Kompetenzen. In J. König & B. Hofmann (Hrsg.), *Professionalität von Lehrkräften – Was sollen Lehrkräfte im Lese- und Schreibunterricht wissen und können?* (S. 40–106). Berlin: DGLS.
- König, J. (2012). Die Entwicklung von pädagogischem Unterrichtswissen. In König, J. & Seifert, A. (Hrsg.), *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerbildung* (S. 143–182). Münster: Waxmann.
- König, J. (2013). First comes the theory, then the practice? On the acquisition of general pedagogical knowledge during initial teacher education. *International Journal of Science and Mathematics Education*, 11(4), 999–1028.
- König, J. (2014a). *Designing an international instrument to assess teachers' general pedagogical knowledge (GPK): Review of studies, considerations, and recommendations. Technical paper prepared for the Innovative Teaching for Effective Learning project*. Paris: OECD.
- König, J. (2014b). Forschung zum Erwerb von pädagogischem Wissen angehender Lehrkräfte in der Lehrerbildung. In E. Terhart, H. Bennewitz, M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 571–596). Münster: Waxmann
- König, J. & Blömeke, S. (2009). Pädagogisches Wissen von angehenden Lehrkräften: Erfassung und Struktur von Ergebnissen der fachübergreifenden Lehrerbildung. *Zeitschrift für Erziehungswissenschaft*, 12(3), 499–527.
- König, J. & Blömeke, S. (2010a). *Pädagogisches Unterrichtswissen (PUW). Dokumentation der Kurzfassung des TEDS-M-Testinstruments zur Kompetenzmessung in der ersten Phase der Lehrerbildung*. Berlin: Humboldt-Universität.
- König, J. & Blömeke, S. (2010b). Pädagogisches Wissen angehender Primarstufenlehrkräfte im internationalen Vergleich. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich* (S. 275–296). Münster: Waxmann.
- König, J. & Klemenz, S. (2015). Der Erwerb von pädagogischem Wissen bei angehenden Lehrkräften in unterschiedlichen Ausbildungskontexten: Zur Wirksamkeit der Lehrerbildung in Deutschland und Österreich. *Zeitschrift für Erziehungswissenschaft*, 18(2), 247–277.
- König, J. & Lebens, M. (2012). Classroom Management Expertise (CME) von Lehrkräften messen: Überlegungen zur Testung mithilfe von Videovignetten und erste empirische Befunde. *Lehrerbildung auf dem Prüfstand*, 5(1), 3–29.
- König, J. & Rothland, M. (2012). Motivations for choosing teaching as a career: Effects on general pedagogical knowledge during initial teacher education. *Asia-Pacific Journal of Teacher Education*, 40(3), 291–317.
- König, J. & Rothland, M. (2015). Wirksamkeit der Lehrerbildung in Deutschland, Österreich und der Schweiz. *Journal für LehrerInnenbildung*, 15(4), 17–25.
- König, J. & Seifert, A. (2012). Der Erwerb von pädagogischem Professionswissen: Ziele, Design und zentrale Ergebnisse der LEK-Studie. In J. König & A. Seifert (Hrsg.), *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerbildung* (S. 32–43). Münster: Waxmann.
- König, J. & Tachtsoglou, S. (2012). Pädagogisches Professionswissen und selbsteingeschätzte Kompetenz. In König, J. & Seifert, A. (Hrsg.), *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerbildung* (S. 284–297). Münster: Waxmann.

- König, J., Blömeke, S. & Kaiser, G. (2010). Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich. In S. Blömeke, G. Kaiser & R. Lehmann, (Hrsg.), *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Primarstufenlehrkräfte im internationalen Vergleich* (S. 99–131). Münster: Waxmann.
- König, J., Blömeke, S. & Kaiser, G. (2015). Early career mathematics teachers' general pedagogical knowledge and skills: Do teacher education, teaching experience, and working conditions make a difference? *International Journal of Science and Mathematics Education*, 13(2), 331–350.
- König, J., Buchholtz, C. & Dohmen, D. (2015). Analyse von schriftlichen Unterrichtsplanungen: Empirische Befunde zur didaktischen Adaptivität als Aspekt der Planungskompetenz angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*, 18(2), 375–404.
- König, J., Lammerding, S., Nold, G., Rohde, A., Strauß, S., & Tachtsoglou, S. (2016). Teachers' Professional Knowledge for Teaching English as a Foreign Language: Assessing the Outcomes of Teacher Education. *Journal of Teacher Education*, 67(4), 320–337.
- König, J., Ligtoet, R., Klemenz, S. & Rothland, M. (2017a). Effects of opportunities to learn in teacher preparation on future teachers' general pedagogical knowledge. *Studies in Educational Evaluation*, 53, 122–133.
- König, J., Rothland, M., Tachtsoglou, S. & Klemenz, S. (2017b). Comparing the change of teaching motivations among preservice teachers in Austria, Germany, and Switzerland: Do in-school learning opportunities matter? *International Journal of Higher Education*, 5 (3), 91–103.
- König, J., Tachtsoglou, S., Darge, K. & Lünemann, M. (2014). Zur Nutzung von Praxis: Modellierung und Validierung lernprozessbezogener Tätigkeiten von angehenden Lehrkräften im Rahmen ihrer schulpraktischen Ausbildung. *Zeitschrift für Bildungsforschung*, 4(1), 3–22.
- König, J., Tachtsoglou, S. & Seifert, A. (2012). Individuelle Voraussetzungen, Lerngelegenheiten und der Erwerb von pädagogischem Professionswissen. In J. König, & A. Seifert (Hrsg.), *Lehramtsstudierende erwerben pädagogisches Professionswissen. Ergebnisse der Längsschnittstudie LEK zur Wirksamkeit der erziehungswissenschaftlichen Lehrerausbildung* (S. 234–283). Münster: Waxmann.
- Krauss, S. & Bruckmaier, G. (2014). Das Experten-Paradigma in der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz, M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 241–261). Münster: Waxmann.
- Kunina-Habenicht, O., Schulze-Stocker, F., Kunter, M., Baumert, J., Leutner, D., Förster, D., Lohse-Bossenz, H. & Terhart, E. (2013). Die Bedeutung der Lerngelegenheiten im Lehramtsstudium und deren individuelle Nutzung für den Aufbau des bildungswissenschaftlichen Wissens. *Zeitschrift für Pädagogik*, 59(1), 1–23.
- Kunter, M., Baumert, J., Blum, W., Klusmann U., Krauss, S. & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen. *Psychologische Rundschau*, 58, 103–117.
- McDonnell, L. M. (1995). Opportunity to Learn as a Research Concept and a Policy Instrument. *Educational Evaluation and Policy Analysis*, 17(3), 305–322.
- Muthén, L.K. & Muthén, B.O. (1998–2015). *Mplus User's Guide. Seventh Edition* [Computer software]. Los Angeles, CA: Muthén & Muthén.

- Oser, F & Oelkers, J. (Hrsg.). (2001). *Die Wirksamkeit der Lehrerausbildungssysteme: Von der Allrounderbildung zur Ausbildung professioneller Standards*. Chur: Rüegger.
- Patry, J.-L. (2014). Theoretische Grundlagen des Theorie-Praxis-Problems in der Lehrer/innenausbildung. In K.-H. Arnold, A. Gröschner & T. Hascher (Hrsg.), *Schulpraktika in der Lehrerbildung. Theoretische Grundlagen, Konzeption, Prozesse und Effekte* (S. 63–86). Münster: Waxmann.
- Rost, J. (1996). Lehrbuch Testtheorie, Testkonstruktion. Bern: Huber.
- Schmidt, W.H., Cogan, L. & Houang, R. (2011). The role of opportunity to learn in teacher preparation: An international context. *Journal of Teacher Education*, 62(2), 138–153.
- Schön, D. A. (1983). *The reflective practitioner – how professionals think in action*. New York, NY: Basis Books.
- Seifert, A. & Schaper, N. (2010). Überprüfung eines Kompetenzmodells und Messinstruments zur Strukturierung allgemeiner pädagogischer Kompetenz in der universitären Lehrerausbildung. *Lehrerausbildung auf dem Prüfstand*, 3(2), 179–198.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Havard Educational Review*, 57, 1–22.
- Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R. & Rowley, G. (2008). *Teacher Education and Development Study in Mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics. Conceptual framework*. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.
- Terhart, E. (1993). Pädagogisches Wissen. Überlegungen zu seiner Vielfalt, Funktion und sprachlichen Form am Beispiel des Lehrerwissens. In J. Oelkers & H.-E. Tenorth(Hrsg.), *Pädagogisches Wissen* (S. 129–141). Weinheim: Beltz.
- Terhart, E. (2012). Wie wirkt Lehrbildung. Forschungsprobleme und Gestaltungsfragen. In *Zeitschrift für Bildungsforschung*, 2, 3–21. doi.org/10.1007/s35834-012-0027-3.
- Voss, T., Kunina-Habenicht, O., Hoehne, V. & Kunter, M. (2015). Stichwort Pädagogisches Wissen von Lehrkräften: Empirische Zugänge und Befunde. *Zeitschrift für Erziehungswissenschaft*, 18(2), 187–223.
- Voss, T., Kunter, M. & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, 103(4), 952–969.
- Watson, C. (2015). *Kompetenzentwicklung im bildungswissenschaftlichen Studium: Empirische Befunde zur Wirksamkeit der Lehramtsausbildung. Manteltext zur Dissertation zur Erlangung des akademischen Grades eines Doktors der Philosophie (Dr. phil.) der Fakultät für Kulturwissenschaften der Universität Paderborn*. Zugriff am 03.03.2017 unter <http://digital.ub.uni-paderborn.de/hsx/content/titleinfo/1944807>
- Watson, C., Seifert, A. & Schaper, N. (2017). Der Einfluss von institutionellen Merkmalen und die Entwicklung des bildungswissenschaftlichen Wissens angehender Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*. Online erhältlich unter: <https://doi.org/10.1007/s11618-017-0794-7>.
- Werquin, P. (2010). *Recognition of non-Formal and informal learning: Country practices*. Paris: OECD.
- Wu, M. & Adams, R. (2006). Modelling mathematics problem solving item responses using a multidimensional IRT model. *Mathematics Education Research Journal*, 18(2), 93–113.

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 zu 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 Intraklassenkorelationskoeffizienten 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, Lehrstuhl für Schulpädagogik, 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 fasst 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 schwächer 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: Beispielimens aus dem Instrument zur Erfassung der Einstellungen bzgl. heterogener Schülerschaften

Form der Heterogenität	Fragebogenprompt	Beispielimens	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,

Überzeugungen 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_w die Kovarianzen der Personenmittelwerte und Σ_b 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

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 und 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

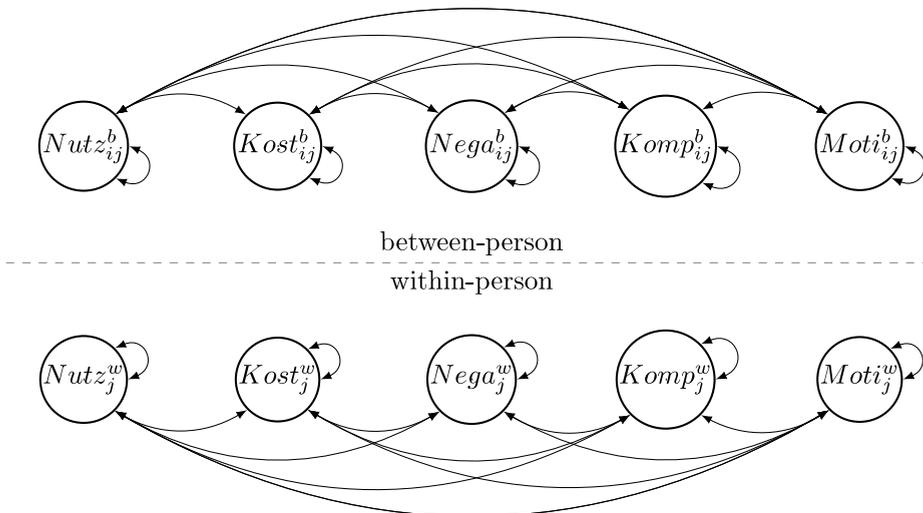
Form	McDonalds ω			
	eigener kompetenter Umgang	Kosten	negative Emotionen	intrinsische motivationale Orientierung
kulturell	.70	.72	.81	.86
leistungsbezogen	.78	.79	.86	.87
sozial	.78	.79	.88	.87

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 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.

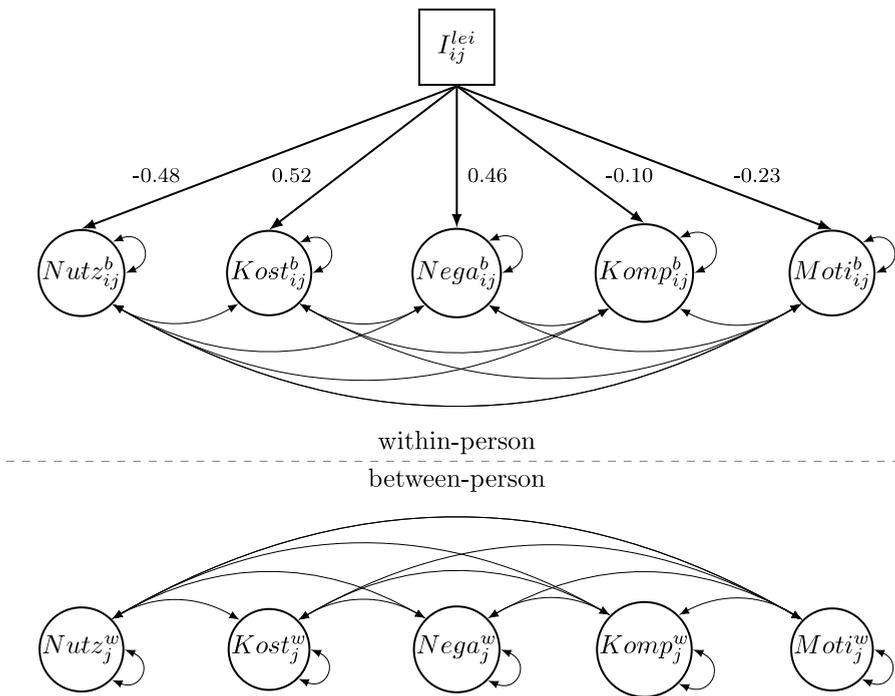
5.4 Differenzierung zwischen Heterogenitätsformen (Forschungsfrage 3)

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 (hier und) 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 geringe-

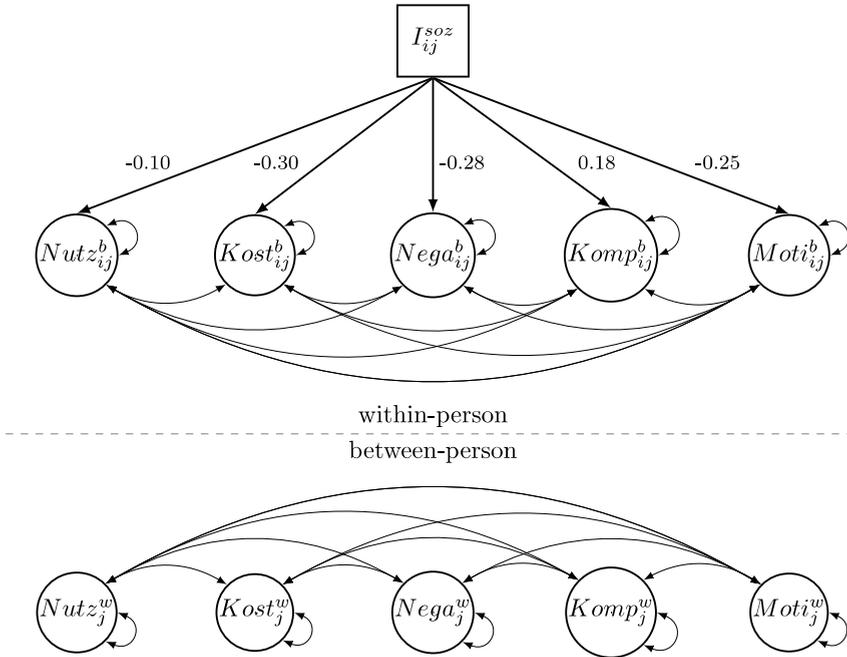
rer Kompetenzselbsteinschätzungen, aber deutlich geringerer Motivation und Wertschätzung einher. Die Indikatorvariable für sozial heterogene Schülergruppen weist ausnahmslos signifikante, kleine Effekte auf: Während die untersuchten Lehramtsstudierenden gegenüber ihrer durchschnittlichen Einstellung (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

Heterogenitätsformen (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 prädiiziert 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 konvergenter/divergenter 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 Beruferschwernis 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: 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. (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. Urdan (Hrsg.), *APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors* (S. 471–499). Washington, DC: APA.
- 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.

- Gasterstätt, 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.
- 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.
- 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/>
- 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: 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.

Anita Sandmeier, Martin Gubler & Silvio Herzog

Berufliche Mobilität von Lehrpersonen – Ein strukturierter Überblick über das Forschungsfeld

Zusammenfassung

Die berufliche Mobilität von Lehrpersonen – innerhalb des Lehrberufs und aus dem Lehrberuf hinaus – ist in den vergangenen Jahrzehnten zunehmend in den Brennpunkt der Bildungspolitik geraten. Der Bedarf an wissenschaftlich fundierten Erkenntnissen zum Thema ist vor dem Hintergrund eines sich abzeichnenden Lehrkräftemangels gestiegen. Doch die Forschungslage dazu ist aktuell unsystematisch, insbesondere in Europa. Zwar forschen zahlreiche Akteure zum Thema, jedoch fehlt Forschung über Disziplingrenzen hinweg. Interdisziplinäre Zugänge könnten zu einem vertieften Verständnis und praxisrelevanten Erkenntnissen führen, es finden jedoch kaum je statt, weil jede Forschungsdisziplin ihre eigenen Theorien, Konzepte und Methoden nutzt. Der vorliegende Beitrag identifiziert theoriegeleitet Analyse Kriterien, die relevant sind für die Erforschung von beruflicher Mobilität von Lehrpersonen, bietet einen strukturierten Überblick über das Forschungsfeld und bilanziert kritisch die theoretischen und forschungsmethodischen Grundlagen der aktuellen Forschung. Der Beitrag zeigt auf, dass (a) theoretische Modelle fehlen, (b) im deutschsprachigen Europa die Datenlage zur Mobilität der Lehrkräfte verbessert werden müsste und (c) die zentralen Erkenntnisse zum Thema insbesondere aus qualitativen, professionstheoretischen Studien stammen. Die quantitativen Forschungsdesigns haben bis dahin das Zusammenspiel zwischen kontextuellen und subjektiven Faktoren in ihrer Wirkung auf die berufliche Mobilität über die Zeit zu wenig analysiert.

Schlagworte

Fluktuation, Personalerhaltung, Laufbahn, Berufswechsel

Prof. Dr. Anita Sandmeier (Korrespondenzautorin) · Prof. Dr. Martin Gubler · Prof. Dr. Silvio Herzog, Pädagogische Hochschule Schwyz, Zaystrasse 42, 6410 Goldau, Schweiz
E-Mail: anita.sandmeier@phsz.ch
martin.gubler@phsz.ch
silvio.herzog@phsz.ch

Teachers' career mobility – a structural overview

Abstract

Over the last decades, teachers' occupational mobility has gained in importance for schools and educational policy alike. Although there is an increased practical need for a more in-depth understanding of teacher mobility, the field – particularly in Europe – currently lacks coherence and systematic coverage. Despite many research initiatives, there is a dearth of interdisciplinary approaches. Such studies would lead to a deeper understanding of teacher mobility and to new, practically relevant insights. However, hardly any such project has been conducted because each academic discipline focuses in its own models, concepts and methods. This article identifies relevant criteria to explore teachers' occupational mobility, it provides a structured overview of research on teacher mobility, and it critically assesses the theoretical and methodological approaches of this field. It is argued that (a) there is a lack of conceptual models on teachers' occupational mobility, (b) educational research in Europe would benefit greatly from gathering consistent and coherent data on teacher mobility, and (c) the core insights into teacher mobility have predominantly been based on qualitative studies. In contrast, quantitative research designs to date have only provided limited insights into the interplay of contextual and subjective factors over time.

Keywords

Teacher attrition, Teacher retention, Teacher workforce, Teaching career

1. Einleitung

Die Erforschung von beruflicher Mobilität von Lehrkräften hat in den vergangenen Jahren an Bedeutung gewonnen. Die Fragen, wie viele Lehrpersonen die Stelle wechseln oder den Beruf verlassen, wohin sie wechseln und welche Gründe zu dieser Mobilität führen, hat für die Steuerung von Bildungssystemen, die Bildungspolitik und die Leitung von Schulen eine hohe und unmittelbare Relevanz (OECD, 2006). Eine gewisse Fluktuation ist für die Schule als Organisation und das Schulsystem als Ganzes durchaus normal und sogar förderlich (ebd.). Allerdings kann Fluktuation auch als Indikator für problematische Entwicklungen an Schulen verstanden werden (Ingersoll & May, 2012). In Bildungsinstitutionen ist die Arbeit stark personengebunden und der Faktor Personal spielt eine entscheidende Rolle (Terhart, 2010). Daher können personale Abgänge große Konsequenzen für Qualität und Effizienz an Schulen haben (Guarino, Santibanez & Daley, 2006; Hanushek, Rivkin & Schiman, 2016). Für die Planung von entsprechenden Maßnahmen braucht es evidenzbasiertes Wissen darüber, welche Faktoren die berufliche Mobilität von Lehrkräften beeinflussen.

Das Feld der beruflichen Verbleibs- und Mobilitätsforschung im Bildungsbereich ist aufgrund unterschiedlicher Forschungsparadigmen und disziplinärer Bezugssysteme heterogen (Herzog, 2014) und theoretisch inkonsistent (Rinke, 2014). Im angloamerikanischen Raum war der Ausgangspunkt der Forschung ein dringender Lehrpersonenmangel und somit die Frage, wie man den Personalbedarf im Bildungssystem langfristig und ohne Gefährdung der Schulqualität sichern kann. Die Fragen von *Turnover* (d.h. personeller Fluktuation) und *Retention* (d.h. Personalerhaltung) wurden basierend auf nationalen Bildungsdaten intensiv wissenschaftlich erforscht. Die Ergebnisse dieser quantitativ ausgerichteten Forschung wurde bereits in diversen Überblicksartikeln zusammengefasst (Borman & Dowling, 2008; Guarino et al., 2006; Johnson, Berg, Donaldson & National Retired Teachers Association, 2005). Die Übertragbarkeit dieser Ergebnisse auf andere Bildungssysteme ist jedoch nur bedingt möglich, da berufliche Mobilität geprägt ist von kulturellen und strukturellen Bedingungen in der Ausbildung der Lehrkräfte und deren Anstellungsbedingungen.

In Europa liegen für die Ebene der nationalen Bildungssysteme und die Fragen, wie hoch die berufliche Fluktuation des Lehrpersonals ist und was diese beeinflusst, erst vereinzelte Befunde mit teilweise begrenzter Aussagekraft vor. Dies ist u.a. darauf zurückzuführen, dass berufliche Mobilität von Lehrpersonen traditionellerweise aus der Perspektive der einzelnen Lehrkraft und deren professionellen Entwicklung innerhalb des Lehrberufs diskutiert und erforscht worden ist (Herzog, 2014; Kunze & Stelmaszyk, 2008; Terhart, Altrichter, Schley & Schratz, 1998; Kelchtermans, 2009). Fragen der Sicherung des Bedarfs an Lehrpersonal und die Erforschung beruflicher Wege aus dem Lehrberuf hinaus haben hingegen lange eine eher untergeordnete Rolle gespielt. Dies beginnt sich jedoch aktuell zu ändern. Etliche von internationalen Organisationen und nationalen Bildungsträgern initiierte Expertenberichte (z. B. European Commission, 2013, 2015) gehen der Frage nach, wie man dem vorhandenen bzw. sich abzeichnenden Lehrkräftemangel begegnen kann. Die Prognosen, ob der Lehrerbedarf in den kommenden Jahren gedeckt werden kann, sind dabei teilweise widersprüchlich. In Deutschland sagt eine aktuelle Studie eine Personalnot auf der Primarstufe voraus (Klemm & Zorn, 2018), während die Kultusministerkonferenz ein Überangebot an Lehrerinnen und Lehrern prognostiziert (KMK, 2011). In der Schweiz sind – wie in anderen Ländern auch – große regionale Unterschiede zu beobachten. Aufgrund der steigenden Schülerzahlen rechnet der aktuelle Bildungsbericht auf nationaler Ebene damit, dass nur die Hälfte des prognostizierten jährlichen Mehrbedarfs an Lehrpersonen durch neu diplomierte Lehrerinnen und Lehrer abgedeckt werden kann (SKBF, 2018).

Wenn man die berufliche Mobilität des Lehrpersonals besser verstehen und steuern will, braucht es neben aussagekräftigen Daten auch systematische wissenschaftliche Analysen. Das Forschungsgebiet zu beruflicher Mobilität ist bislang jedoch insgesamt unübersichtlich. Es besteht Bedarf nach einem strukturierenden Überblick, um die Datenlage für systemische Interventionen zu verbessern und die Forschungsdisziplin zur beruflichen Mobilität weiterzuentwickeln. Dazu

leistet der vorliegende Artikel einen Beitrag. Ausgehend von einer theoretischen Präzisierung des Forschungsgegenstandes werden Analysekriterien identifiziert, die für das Verständnis von beruflicher Mobilität zentral sind (Abschnitt 2). Danach wird eine Strukturierung des Forschungsfeldes mittels eines Rasters vorgeschlagen. Entlang dieses Rasters werden exemplarische Studien, deren theoretische Fundierung, untersuchten Faktoren und verwendeten Methoden vorgestellt (Abschnitt 3). Abschließend zieht der Artikel Bilanz, in welchem Ausmaß in den einzelnen Forschungstraditionen die für das Verständnis von beruflicher Mobilität relevanten Analysekriterien berücksichtigt werden (Abschnitt 4), und zeigt in einem Ausblick auf, in welche Richtung sich die Erforschung des Themas entwickeln könnte (Abschnitt 5).

Dieser Überblick kann als Ausgangspunkt dienen, die zukünftigen Forschungskapazitäten gezielter aufeinander abzustimmen und Projekte disziplinübergreifend zu organisieren. Dabei ist die Annahme grundlegend, dass eine Verbindung von systemorientiertem und personenorientiertem Zugang ein besseres Verständnis für Gründe von beruflicher Mobilität von Lehrpersonen gewährleistet (Rinke, 2008; Kelchtermans, 2017). Der Beitrag ist zudem verfasst im Wissen, dass die Bildungsforschung von Erkenntnissen aus anderen Disziplinen profitieren kann. Aus diesem Grund werden auch Überlegungen aus arbeits- und organisationstheoretischen Ansätzen berücksichtigt. Die Differenzen zwischen privatwirtschaftlichem Kontext und Bildungskontext (Terhart, 2016) werden dabei, wo relevant, thematisiert. Ganz grundsätzlich basiert der Beitrag auf der Überzeugung, dass es möglich ist, unabhängig von den zahlreichen länderspezifischen, kulturellen und strukturellen Unterschieden in Ausbildung und Anstellungsbedingungen des Lehrpersonals, Analysekriterien zu identifizieren, die übergreifend relevant sind, um berufliche Mobilität zu verstehen und einen Überblick über verschiedene Forschungszugänge dazu zu geben.

Das Ziel des Beitrages ist es, eine theoretische Systematisierung des Forschungsfeldes vorzunehmen. Eine Bilanzierung der bisherigen Forschungsergebnisse wird damit nicht angestrebt. Eine solche Zusammenfassung des Forschungsstandes findet sich in den Überblicksartikeln zum US-amerikanischen Bildungssystem (Borman & Dowling, 2008; Guarino et al., 2006; Johnson et al., 2005).

2. Berufliche Mobilität von Lehrpersonen aus zwei Perspektiven: Begriffsbestimmung und grundlegende Analysekriterien

Übergeordnet kann berufliche Mobilität in Anlehnung an Vardi (1980) als berufsbezogene Übergänge und die damit zusammenhängenden individuellen Wahrnehmungen, Ziele und Einstellungen verstanden werden. Nachfolgend wird ausgehend von zwei grundlegenden Perspektiven – Bildungssystem und Person

– der Begriff weiter spezifiziert und dargelegt, welche Kriterien in den beiden Zugängen zentral sind, um berufliche Mobilität differenziert zu analysieren.

2.1 Berufliche Mobilität mit Fokus auf das Bildungssystem

Aus der kontextuellen Perspektive wird die berufliche Mobilität als Wechsel, Fluktuation oder Ausfall von Lehrpersonen bezeichnet, die quantitativ gemessen werden kann. Mobilität wird dabei in aggregierter Form in einer Prozentquote ausgedrückt, die aussagt, wie viele Mitarbeitende im Verhältnis zum Personalbestand innerhalb eines Jahres die Organisation oder den Beruf verlassen (Ingersoll, 2001). Eine zusätzliche, für den Lehrberuf spezifische Form des Ausfalls ist die Reduktion des Pensums, die nicht zu beruflicher Mobilität im engeren Sinn gezählt werden kann, jedoch hohe Relevanz besitzt für den Bedarf neuer Lehrpersonen (SKBF, 2018).

In der klassischen *Turnover*-Forschung wird berufliche Mobilität mit den Kriterien Funktionalität, Vermeidbarkeit und Ziel charakterisiert (Hom, Mitchell, Lee & Griffeth, 2012). Das *Kriterium der Funktionalität* ist eine Einschätzung des Wechsels auf der Basis der Qualität des wechselnden Arbeitnehmers, die durch die Organisation vorgenommen wird. Kündigungen von „leistungsstarken“ oder nur schwer zu ersetzenden Personen werden als dysfunktionale Wechsel bezeichnet, während Abgänge von „leistungsschwachen“ Angestellten als funktional für das System verstanden werden (ebd., S. 834; vgl. auch Kelchtermans, 2017). Das *Kriterium der Vermeidbarkeit* ist insbesondere in der Praxis von hoher Relevanz. Es hilft zu unterscheiden, welche Wechsel man durch eine Verbesserung der Arbeitsbedingungen vermeiden könnte und welche Fluktuation unvermeidbar ist, da sie auf Faktoren außerhalb des Einflussbereiches der Leitung beruht. Das *Kriterium des Ziels* von Mobilität fragt danach, wohin die Mobilität führt. Grundsätzlich wird unterschieden zwischen innerbetrieblicher Mobilität durch Funktionswechsel, Wechsel zu einem anderen Arbeitgeber (*organizational turnover*), Berufswechsel (*occupational turnover*), Wechsel zu unbezahlter Arbeit (z. B. Betreuung von Kindern oder pflegebedürftigen Eltern) oder zu anderen Tätigkeiten (z. B. Weiterbildungen). Für den Lehrberuf spezifisch ist, dass innerbetriebliche Mobilität durch Funktionswechsel, aufgrund der geringen Ausdifferenzierung des Berufs, eher selten vorkommen (Herzog & Leutwyler, 2010). Ebenfalls typisch für das Bildungssystem ist die Struktur mit drei Handlungs- und Gestaltungsebenen: Makro- (Bildungsadministration), Meso- (Schule im lokalen Umfeld) und Mikroebene (Unterricht) und ihren verschiedenen Akteuren und Steuerungsinstrumenten (Fend, 2008). Die Fluktuation des Personals kann auf der Makroebene des gesamten Bildungssystems oder auf der Mesoebene der Einzelschule gemessen und beeinflusst werden.

2.2 Berufliche Mobilität mit Fokus auf das Individuum

Berufliche Mobilität aus individueller Sicht fokussiert auf die eigenen beruflichen Wechsel und deren subjektiven Bedeutungen für die einzelne Person. Berufliche Wechsel werden als Übergänge verstanden, die nach ihrem Grad an Institutionalisierung differenziert werden (Herzog, Herzog, Brunner & Müller, 2007). Es gibt normative Übergänge (z. B. Berufseinstieg und Pensionierung) und nicht-normative Transitionen (z. B. Berufs-, Stellen- und Funktionswechsel), aber auch Unterbruch oder Reduktion der Tätigkeit (ebd., S. 70 ff.). Damit ist der *Zielort der beruflichen Mobilität* auch aus dieser Sicht ein zentrales Kriterium für ein differenziertes Verständnis von Mobilität.

Diese Transitionen interessieren aus individueller Perspektive nicht nur isoliert als einzelne Ereignisse. Sie konstituieren in ihrer Gesamtheit die berufliche Laufbahn als einen Aspekt des individuellen Lebenslaufs und entwickeln sich immer über eine zeitliche Sequenz hinweg (Arthur, Hall & Lawrence, 1989). Für ein vertieftes Verständnis individueller beruflicher Mobilität ist daher zwingend das Kriterium der *Zeit* in die Betrachtung mit einzubeziehen (Herzog et al., 2007). Die Zeit ist in dreifacher Weise relevant. Erstens strukturiert sie Mobilitätswege, indem sie berufliche Stationen chronologisch ordnet und die jeweilige Verweildauer pro Station erfasst (ordnende Zeit). Zweitens situiert die Zeit einzelne Übergänge im Kontext des Lebensalters der einzelnen Person und im Kontext der jeweils aktuellen wirtschaftlichen, gesellschaftlichen und politischen Gegebenheiten, z. B. zur jeweiligen Situation auf dem Arbeitsmarkt (situierende Zeit). Drittens ist die Zeit relevant für die individuelle Handlung in der Gegenwart vor dem Hintergrund der eigenen Vergangenheit und der persönlichen Absichten für die Zukunft (modale Zeit) (Herzog, 1988). Berufliche Transitionen sind ausgehend von handlungstheoretischen Modellen zu verstehen als biographische Tatsachen, die sich in der Auseinandersetzung des Individuums mit seinen wechselnden Lebensverhältnissen ergeben. Menschliches Verhalten wird nur dann verstanden, wenn die dynamische *Interaktion zwischen innerpsychischen und kontextuellen Faktoren* berücksichtigt wird unter dem Prinzip der reziproken Determination (Herzog et al., 2007). Konkret heißt dies, dass dieselben kontextuellen Bedingungen je nach Individuum, dessen Zielen, Eigenschaften oder Überzeugungen, unterschiedliche Auswirkungen auf die berufliche Mobilität haben und das Individuum wiederum mit seinen Entscheidungen zurückwirkt auf den Kontext (vgl. auch Gunz & Mayrhofer, 2017).

Um das individuelle Mobilitätsverhalten zu verstehen, sind demnach drei Kriterien grundlegend: *Zielort der beruflichen Mobilität*, *Zeit* und *Interaktion zwischen Individuum und Kontext*.

3. Raster zur strukturierten Beschreibung des Forschungsfeldes

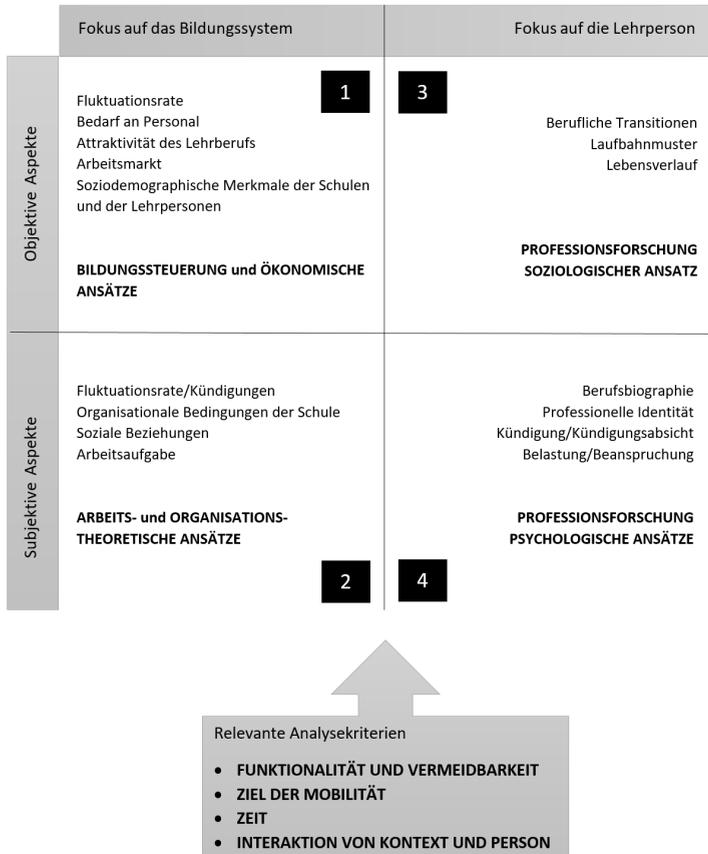
Die Grundlage für den Überblick über das Forschungsfeld bildet ein Klassifikationssystem, das in Anlehnung an Vardi (1980) die verschiedenen Forschungstraditionen zu beruflicher Mobilität von Lehrpersonen entlang von übergeordneten Kriterien aufzeigt (vgl. Abbildung 1). Die x-Achse spiegelt den Fokus der Forschung: Stärkere Konzentration auf berufliche Mobilität in ihrer Bedeutung für das Bildungssystem versus Fokussierung auf die individuelle *Lehrperson*. Auf der y-Achse wird differenziert nach Art der Daten: objektive oder objektivierbare Aspekte versus subjektive Aspekte. Objektive Aspekte können erfasst werden unabhängig vom Urteil der Lehrpersonen und umfassen u.a. statistische Kennzahlen (z. B. die Höhe des Lohns, Größe der Schule, Zusammensetzung der Schülerschaft). Objektivierbare Erhebungen basieren auf Daten, die unabhängig von persönlichen Einstellungen sein sollten (z. B. die retrospektive Schilderung von beruflichen Transitionen). Die subjektiven Aspekte basieren auf Erhebungsmethoden, welche die individuelle Interpretation und Einschätzung der Lehrperson erfassen.

Aus der Kombination dieser zwei Achsen ergeben sich vier Forschungsfelder zu beruflicher Mobilität. Dieses Raster ermöglicht es, die verschiedenen Zugänge zum Thema und ihre zentralen Konzepte systematisch abzubilden.

In Feld 1 (Bildungssystem-Objektiv) sind Studien verortet, die Mobilität in aggregierter Form messen (als Fluktuationsrate) und diese mit objektiven Aspekten in Verbindung bringen. Der Schwerpunkt liegt auf dem Monitoring und der Vorhersage der beruflichen Mobilität des Personals innerhalb und aus dem Bildungssystem. In Feld 2 (Bildungssystem-Subjektiv) wird die berufliche Mobilität von der einzelnen Organisation her untersucht unter Einbezug der subjektiv wahrgenommenen beruflichen Bedingungen. In Feld 3 (Lehrperson-Objektiv) sind Studien kategorisiert, die einen soziologischen Ansatz verfolgen und berufliche Transitionen und Laufbahnmuster mit objektivierbaren Verfahren analysieren im Zusammenhang mit beruflichen und privaten Faktoren. Feld 4 (Lehrperson-Subjektiv) schließlich repräsentiert den psychologischen Ansatz, der berufliche Mobilität als internalen Prozess versteht und diese u. a. vor dem Hintergrund der professionellen Entwicklung und im Zusammenhang mit beruflicher Beanspruchung untersucht.

Zentrale Analyse Kriterien, um berufliche Mobilität zu verstehen, sind die Funktionalität und Vermeidbarkeit von Fluktuation, das Ziel der Mobilität und das Zusammenspiel von kontextuellen und personalen Faktoren über die Zeit.

Abbildung 1: Raster zur Strukturierung des Forschungsfeldes zur beruflichen Mobilität von Lehrpersonen.



3.1 Zentrale Ansätze, typische Fragestellungen, Forschungsmethoden und prototypische Studien in den vier Forschungsfeldern

Im Folgenden werden die vier Forschungsfelder näher beschrieben. Für jedes Feld werden die grundlegenden Ansätze und typische Fragestellungen benannt, um danach einzelne Studien exemplarisch vorzustellen. Die ausgewählten Studien sollen das Feld theoretisch, von den Forschungsmethoden und von den untersuchten Faktoren möglichst repräsentativ darstellen. In einem ersten Schritt werden, falls vorhanden, Studien berichtet, die auf Daten aus Deutschland, Österreich und der Schweiz basieren und seit dem Jahr 2000 wissenschaftlich publiziert worden sind. Ergänzend werden exemplarische Studien aus anderen nationalen Bildungssystemen herangezogen. Dieses Vorgehen liegt im Ziel begründet, neben dem Aufzeigen von exemplarischen Studien einen ersten Überblick über den Forschungsstand im deutschsprachigen Raum zu gewinnen. Die drei Länder sind

in Bildungsfragen mit „analogen oder gar gleichen Problem- und Fragestellungen konfrontiert“ (Stadelmann & Rhyn, 2011, S. 109) und bilden einen gemeinsamen Arbeitsmarkt für ausgebildete Lehrerinnen und Lehrer, trotz teilweise großen strukturellen Unterschieden in Ausbildung und Anstellungsbedingungen.

Die Studien wurden auf der Grundlage einer umfassenden Literaturrecherche ausgewählt, die ausgehend von Handbüchern (Helsper & Böhme, 2008; Terhart, Bennewitz & Rothland, 2014) in verschiedenen Fachliteraturdatenbanken (FIS Bildung, ERIC, PsycInfo) und *Google Scholar* mit den Begriffen *Lehr**, *berufliche Mobilität*, *Fluktuation*, *Karriere*, *Laufbahn*, *Berufswechsel*, *Stellenwechsel*, *Lehrkräftemangel*, *Lehrkräftebedarf* und den entsprechenden englischen Begriffen durchgeführt wurde.

3.2 Feld 1 – Fokus auf das Bildungssystem, objektive Aspekte

Die maßgebenden Ansätze in Feld 1 sind Bildungssteuerung und -ökonomie, die vornehmlich an Fragen der Effektivität und Qualität von Bildungssystemen interessiert sind. Die übergeordneten Fragen dieses Feldes sind, wie hoch die Mobilitätsrate ist und ob der Bedarf an qualifizierten Lehrpersonen abgedeckt werden kann. Damit zusammenhängend ist, die Frage der Schulqualität und Maßnahmen, um den Bedarf und die Qualität des Lehrpersonals zu sichern (u. a. über die Aus-, Fort- und Weiterbildung von Lehrpersonen).

Einen ersten Zugang zu diesen Fragen bilden nationale Bildungsmonitorings und Bildungsberichte, die das Ziel haben, bildungspolitische Entscheidungen auf einer forschungsbasierten Basis zu gründen (Rürup, Fuchs & Weishaupt, 2016). Bisher enthalten diese Berichte im deutschsprachigen Raum jedoch kaum Daten zur beruflichen Mobilität des Lehrpersonals (Autorengruppe Bildungsberichterstattung, 2016; Bruneforth, Lassnigg, Vogtenhuber, Schreiner & Breit, 2015). In Deutschland werden die Daten zur Mobilität länderspezifisch erhoben und in der Schulstatistik veröffentlicht. Die Qualität dieser Statistik ist nach Weishaupt (2014) unvollständig und unsystematisch. In der Schweiz kann seit 2010 durch eine modernisierte Statistik des Schulpersonals und neuen Sozialversicherungsnummern die berufliche Mobilität von Lehrkräften der obligatorischen Schule differenzierter analysiert werden (Bundesamt für Statistik, 2014).

Neben den nationalen und regionalen Bildungsträgern initiieren internationale Organisationen relevante Untersuchungen. Ein Bericht der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD, 2006) basiert auf Länderhintergrundberichten, die von nationalen Expertenkommissionen erstellt wurden. Im Zentrum stehen dabei Fragen der Quantität und Qualität des Lehrkräfteangebots und Maßnahmen (Vergrößerung der Schulklassen, Erhöhung des Pensums für das bestehende Personal), die ergriffen werden, um den Bedarf an Lehrpersonen abzudecken. Ebenfalls auf Expertenbefragungen basieren die bildungspolitischen Studien der Europäischen Union (2013, 2015), die sich um die Frage drehen, wie man die Attraktivität des Lehrberufs verbessern kann. Ein

Befund dieser Befragungen war, dass es keine Wirksamkeitsstudien gibt, die prüfen, welchen Einfluss bildungs- oder berufspolitische Maßnahmen auf die berufliche Mobilität haben (European Commission, 2013, S. 23). Zusätzlich zu diesen bildungspolitischen Studien analysierte die Bildungsökonomie die Frage, welchen Einfluss die Lohnpolitik und der Arbeitsmarkt auf die Mobilität von Lehrkräften haben (Denzler & Wolter, 2009; Weishaupt, 2014).

Im Vergleich zu Europa besteht in den USA mit dem *Schools and Staffing Survey* (SASS) und dem daran gekoppelten *Teacher Follow-up Survey* (TFS) seit 1987 auf nationaler Ebene eine ungleich breitere Datenbasis, um die berufliche Mobilität von Lehrpersonen und deren Ursachen zu analysieren. Diese Surveys basieren auf umfangreichen, national repräsentativen Stichproben und beinhalten neben Fluktuationsraten auch Befragungen von Bildungsträgern auf Distriktebene, Schulleitungen und Lehrpersonen. Die zahlreichen Studien, die auf der Basis dieser Surveys durchgeführt wurden, sind bereits in Meta-Analysen (Borman & Dowling, 2008; Guarino et al., 2006) zusammengefasst worden. Im Zentrum steht dabei der Zusammenhang zwischen schulstrukturellen Merkmalen von Schulen (z. B. Zusammensetzung der Schülerschaft, Ausstattung, Größe), soziodemographischen Merkmalen der Lehrperson (z. B. Geschlecht, Abstammung, Zertifizierung) und der Mobilitätsrate. Borman und Dowling (2008) stellen auf der Basis von 34 Studien fest, dass die Fluktuation eher dysfunktional ist und diese von verschiedenen persönlichen und professionellen Faktoren beeinflusst wird, die über die Laufbahn hinweg variieren. Die Arbeitsbedingungen spielen dabei eine wichtige moderierende Rolle. Fluktuation kann demnach beeinflusst werden durch Politik und Maßnahmen.

3.3 Feld 2 – Fokus auf den Kontext, subjektive Aspekte

Die Forschung in diesem Feld hat ebenfalls das Bildungssystem im Fokus, allerdings aus der subjektiven Sicht der Lehrpersonen. Hier steht die Frage im Mittelpunkt, wie die Lehrpersonen die beruflichen Rahmenbedingungen wahrnehmen und wie diese die Mobilitätsrate der einzelnen Schule beeinflussen. Dabei liegt der Fokus stark auf den organisationalen Bedingungen der einzelnen Schule, z. B. dem Führungsverhalten der Schulleitung, der Unterstützung im Team und den Mitsprachemöglichkeiten. Theoretisch basiert diese Forschung meist auf der Organisations-, Berufs- und Arbeitssoziologie sowie der interdisziplinären Organisationstheorie. Sie findet aktuell hauptsächlich in den USA in der Forschergruppe rund um Richard M. Ingersoll statt (Ingersoll, 2001, 2003; Ingersoll & May, 2012; Ingersoll, Merrill & May, 2016). Empirisch basieren diese Studien auf den in Feld 1 beschriebenen Surveys SASS und TFS. Die Studien zeigen, dass Schulen mit besseren Arbeitsbedingungen tiefere Fluktuationen aufweisen, wobei ein zentraler organisationaler Faktor die individuelle Autonomie im Klassenzimmer ist.

Im deutschsprachigen Raum gibt es unseres Wissens keine Studien, die mit einem organisationstheoretischen Ansatz subjektiv wahrgenommene schulische Kontextfaktoren in ihrer Auswirkung auf das Mobilitätsverhalten in einem ähnlich komplexen Design untersuchen. Schulische Kontextfaktoren werden mehrheitlich ausgehend von psychologischen Theorien und damit mit dem Fokus auf die einzelne Person untersucht. Sie werden deshalb in Feld 4 klassifiziert und dort näher vorgestellt.

3.4 Feld 3 – Fokus auf das Individuum, objektive Aspekte

Dieses Feld basiert primär auf einem soziologischen Ansatz, der berufliche Transitionen beschreibt und diese in einem sozialen Zusammenhang versteht. Studien in diesem Feld betrachten individuelle Laufbahnen mit einem objektivierten Zugang. Sie gehen der Frage nach, ob es typische Laufbahnverläufe und -muster von Lehrpersonen gibt und wohin die berufliche Mobilität führt. Im Vordergrund dieser Studien steht die Beschreibung der beruflichen Übergänge über die Zeit (Herzog, 2014). Dabei wird der berufliche Verlauf entlang von Übergängen beschrieben, beginnend bei der Berufswahl über die Grundausbildung, die Berufseinführungsphase, definitive Anstellung, Stufen- oder Berufswechsel und Rückzug aus dem aktiven Berufsleben, bei den Frauen meist zusätzlich noch der zeitweise Ausstieg in der Familienphase. Im Gegensatz zur berufsbiographischen Forschung, die im Feld 4 lokalisiert ist, erfasst diese Forschung die Verläufe quantitativ aufgrund objektivierter Daten.

Ein typisches Beispiel für dieses Feld ist die Studie von Herzog und Kollegen (2007), die in einer Analyse von vier Kohorten von Absolventen einer Schweizer Grundschullehrerausbildung berufliche Übergänge analysierten, die Lehrpersonen retrospektiv berichtet hatten. Die Autoren zeigten auf, wie hoch die interpersonelle Varianz dieser Verläufe ist, d. h. wie viele individuelle Varianten der beruflichen Veränderungen es gibt. In einer Re-Analyse dieser Daten mittels Sequenzanalyse wurden sechs typische Laufbahnmuster identifiziert und in ihrer Häufigkeit in den verschiedenen Kohorten analysiert (Gubler, Biemann & Herzog, 2017). Weitere Studien in diesem Feld mit ähnlichen Designs stammen von Troesch und Bauer (2017) und Ingrisani (2014) für weitere spezifische Kohorten von Schweizer Lehrpersonen.

3.5 Feld 4 – Fokus auf das Individuum, subjektive Aspekte

In den Studien in Feld 4 steht ebenfalls das Individuum im Fokus. Doch während in Feld 3 die beruflichen Verläufe objektivierbar beschrieben werden, geht es in Feld 4 um die subjektive Deutung der Hintergründe dieser Übergänge. Daher dominieren in diesem Feld psychologische Theorien, im Fokus stehen individuelle Wahrnehmungen, Ziele, Einstellungen und Erwartungen. Zentrale Konzepte in

diesem Feld sind Berufsbiographie, professionelle Identität sowie Belastung und Beanspruchung. Die theoretisch und methodisch am klarsten fundierte Tradition in diesem Feld ist die Berufsbiographieforschung, die mit qualitativen oder Mixed-method-Ansätzen das Thema erforscht. Die zweite Tradition in diesem Feld sind quantitative Studien, die Kündigungsabsicht und Kündigungsverhalten untersuchen.

Berufsbiographieforschung: Die Studien in dieser Tradition wurden bereits in mehreren Überblicksartikeln zusammengefasst. Rothland (2014) gab einen Überblick über die Forschung zu den Berufs- und Studienwahlmotiven. Kunze und Stelmaszyk (2008) bilanzierten die Forschung zu Biographien innerhalb des Lehrberufs, deren Ziel es ist, typische Phasen der professionellen Entwicklung nach inhaltlichen Kriterien zu identifizieren, ohne aber berufliche Wege aus dem Lehrberuf zu untersuchen (z. B. Huberman, 1991; Terhart, 1994). Herzog (2014) beschrieb zusätzlich auch Studien und Ansätze, die Berufsbiographien von Lehrpersonen aus dem Lehrberuf hinaus betrachten.

Die oben erwähnte Studie von Herzog und Kollegen (2007) ist eine der elaboriertesten Untersuchungen beruflicher Übergänge im und aus dem Beruf. Mit einem Mixed-method-Ansatz wurden die standardisierten Daten der individuellen Berufsverläufe in biographischen Interviews mit Lehrpersonen vertiefend interpretiert (vgl. auch Herzog, 2007). Retrospektiv wurden Verläufe und ihre Begründungen in halbstrukturierten Interviews bis zu 40 Jahren zurück erfasst. Die Gründe für den Ausstieg aus dem Beruf waren vielfältig und geschlechtsspezifisch. Wobei Geschlechtsunterschiede in einem klaren Zusammenhang mit eigenen Kindern standen, denn ohne Kinder unterscheiden sich Männer und Frauen kaum (vgl. auch Abele, Dette & Hermann, 2003).

Ebenfalls auf dem biographischen Ansatz basieren die qualitativen Fallstudien von Rinke (2013, 2014) in den USA, die in einem längsschnittlichen, prospektiven Design den Prozess der Laufbahnentscheide von einzelnen Lehrpersonen über zwölf Monate hinweg verfolgte. Rinke untersuchte die Frage, wie die Pläne und Motive von Lehrpersonen sich auf die berufliche Mobilität auswirken. Ihre zentrale Erkenntnis war, dass die individuellen Ziele des Individuums und die professionelle Identität wichtige moderierende Faktoren sind zwischen den kontextuellen Merkmalen der Schule und dem Verbleib im Beruf. Hatte eine Lehrperson den Beruf bewusst als langfristige Perspektive gewählt, reagierte sie auch bei schwierigen kontextuellen Bedingungen nicht mit Kündigung, sondern mit dem aktiven Bemühen und Engagement, diese Bedingungen zu verbessern.

Quantitative Studien: Es lassen sich zwei Arten von quantitativen Studien unterscheiden, die berufliche Mobilität mit Fokus auf das Individuum und subjektiv wahrgenommenen Faktoren standardisiert untersuchen: Befragungen im Auftrag von Bildungsträgern und wissenschaftlich publizierte Studien. Im Auftrag von lokalen Bildungsträgern werden Studien durchgeführt, die kündigende Lehrpersonen nach dem Entscheid der Kündigung retrospektiv zu ihren Kündigungsgründen befragen. Diese Studien basieren auf querschnittlichen Designs und deskriptiven statistischen Verfahren, die nicht den Zusammenhang von Variablen untersuchen,

sondern personenzentriert nach Unterschieden zwischen verschiedenen Gruppen von Lehrpersonen suchen. Die Ergebnisse werden nicht immer wissenschaftlich publiziert, sondern oft in Form eines Forschungsberichtes online gestellt. Ein Beispiel dafür ist die Studie von Keller-Schneider (2010): Theoretisch basierend auf Belastungs-Beanspruchungsmodellen (u. a. Rudow, 1994) verglich sie kündigende mit bleibenden Lehrpersonen in verschiedenen Laufbahnphasen bezüglich ihrer beruflichen Beanspruchung und identifizierte Muster von Kündigungsgründen und Kündigungszielen.

Wissenschaftlich publizierte Studien in diesem Feld untersuchen mehrheitlich die Kündigungsabsicht und nicht die tatsächliche Mobilität von Lehrpersonen. Ihr Ziel ist es, die Kündigungsabsicht in Abhängigkeit von kontextuellen und personellen Faktoren vorherzusagen. Eine niederländische Studie von Houkes, Janssen, Jonge und Bakker (2003) basiert ebenfalls auf theoretischen Modellen der Belastung und Beanspruchung. An einer Stichprobe von Berufsschulehrkräften wurde der Einfluss von Kontextfaktoren (Arbeitsaufgabe, Überlastung, soziale Unterstützung und unerfüllte Laufbahnpläne) auf die Kündigungsabsicht untersucht. Basis der durchgeführten Strukturgleichungsmodelle waren Paneldaten zu zwei Messzeitpunkten mit einem Intervall von einem Jahr. Zentrale Erkenntnis dieser Längsschnittanalyse war es, dass die Kündigungsabsicht die Wahrnehmung der Kontextfaktoren im Jahr darauf beeinflusst. Hat eine Lehrkraft eine hohe Absicht zu kündigen, führte dies dazu, dass sie weniger Entwicklungsmöglichkeiten wahrnahm oder suchte, bzw. weniger vom Vorgesetzten gefördert wurde.

4. Diskussion

Der vorliegende Artikel hat das Ziel, entlang eines Rasters einen strukturierten Überblick über das Forschungsfeld zur beruflichen Mobilität von Lehrpersonen zu geben. Im Folgenden wird die Forschung aus den beiden zentralen Perspektiven (Bildungssystem vs. Lehrperson) bilanziert mit der Frage, in welchem Ausmaß diese Studien die für das Verständnis von beruflicher Mobilität relevanten Kriterien berücksichtigen (vgl. Abschnitt 2.1. und 2.2.). Davon ausgehend werden Forschungsdesiderate formuliert.

Für die Erforschung beruflicher Mobilität mit dem Fokus auf das Bildungssystem ist es relevant, dass man neben der Mobilitätsrate auch über Informationen verfügt, wohin die Mobilität führt, ob sie funktional oder dysfunktional ist und ob sie durch das System hätte beeinflusst werden können (Hom et al., 2012). Die in Feld 1 und 2 vorgestellte Forschung zeigt, dass in Europa zwar eine Intensivierung der Forschungsbemühungen auf der Makroebene stattgefunden hat, die methodisch jedoch vor allem auf Expertenbefragungen basieren. Es bestehen bisher noch wenige Daten zur Höhe der Mobilitätsrate von Lehrpersonen, wohin diese wechseln und ob der Wechsel funktional ist für das System. Statistiken zu Mobilität werden im deutschsprachigen Raum, falls überhaupt, regional und undifferenziert

erfasst (z. B. ohne Zielort der Mobilität). Es gibt bisher keine Studien, die datenbasiert untersuchen, ob bildungspolitische Maßnahmen sich tatsächlich positiv auswirken (European Commission, 2013). Im Gegensatz dazu werden in den USA auf nationaler Ebene regelmäßig umfangreiche, national repräsentative Surveys durchgeführt, die einerseits wichtige deskriptive Daten generieren für die Bildungspolitik und andererseits eine erstklassige Datenbasis liefern für wissenschaftliche, komplexere Analysen.

Faktoren auf der Mesoebene, d. h. die organisationalen Bedingungen der einzelnen Schule, wurden in deutschsprachigen Studien vor allem explorativ als individuelle Kündigungsgründe erfasst und deskriptiv beschrieben. Wie diese einzelnen Faktoren zusammenspielen und ob sie für alle Lehrpersonen ähnlich relevant sind, wurde bisher hauptsächlich in den USA erforscht (z. B. Ingersoll & May, 2012). Die Replikation dieser auf organisationstheoretischen Konzepten beruhenden Studien in anderen nationalen Bildungssystemen steht bisher noch aus. Die Frage, welche Auswirkungen strukturelle Bedingungen haben, ist auch vor dem Hintergrund zunehmender Unterschiede in der Zusammensetzung der Schülerschaft zwischen den Schulen relevant (Weishaupt, 2014).

Zusammenfassend kann für die Erforschung der beruflichen Mobilität mit Fokus auf das Bildungssystem festgehalten werden, dass im deutschsprachigen Europa vergleichsweise wenige Daten vorliegen. Ebenso besteht wenig gesichertes Wissen darüber, mit welchen Maßnahmen auf der Ebene der Makro-, bzw. Mesoebene man die berufliche Mobilität der für das System „wertvollen“ Personen minimiert.

Um die berufliche Mobilität mit dem Fokus auf das Individuum besser zu verstehen, ist es erstens relevant zu differenzieren, wohin die berufliche Mobilität führt. Zweitens gilt es, die Interaktion von innerpsychischen Faktoren und die Einflüsse im beruflichen und privaten Kontext in ihrer Wirkung auf die berufliche Mobilität über die Zeit zu analysieren. Der zentrale Erkenntnisgewinn von Studien mit objektiver Perspektive auf die individuelle berufliche Mobilität (Feld 3) besteht in ihrem Fazit, dass berufliche Mobilität sehr viel mehr beinhaltet als die traditionell untersuchten Stellen- und Berufswechsel. Diese Studien vermögen berufliche Wege von Lehrpersonen nachzuzeichnen und widerlegen damit das „Sackgassen“-Image des Lehrberufs. So zeigen sie u. a. klar auf, dass die berufliche Mobilität von Lehrkräften selten aus dem Bildungssystem hinausführt (Herzog et al., 2007). Die Studien weisen zudem darauf hin, wie wenig standardisiert berufliche Laufbahnen im Lehrberuf sind. Die beruflichen Entwicklungsmöglichkeiten beruhen mehrheitlich auf individuellen Lösungen. Es bestehen aktuell nur wenige formalisierte, z. B. an Weiterbildungen geknüpfte, berufliche Laufbahnmodelle für Lehrpersonen.

Bezüglich der Frage nach Gründen für berufliche Mobilität konnten insbesondere die qualitativen, berufsbiographisch orientierten Studien (Feld 4) vertiefte Erkenntnisse generieren und aufzeigen, wie individuelle und kontextuelle Faktoren über die Zeit zusammenwirken. Diese Studien zeigen, dass es keine einfachen kausalen Zusammenhänge gibt, sondern dass personale Faktoren mit professionellen und systemischen Faktoren interagieren (vgl. auch Kelchtermans, 2017).

Die kontextuellen beruflichen Bedingungen auf der Ebene des Bildungssystems und der einzelnen Schule wirken sich nicht für alle Personen gleich aus, sondern werden moderiert durch die individuelle Biographie, besonders auch Gründe im außerberuflichen Bereich, die Ziele, Überzeugungen und Erwartungen des Individuums (Rinke, 2014). Inwiefern diese in begrenzten Kohorten gefundenen Zusammenhänge generalisierbar sind, müsste in quantitativen Studien geprüft werden. Die bisherigen quantitativen Studien haben durch ihre mehrheitlich querschnittlichen, retrospektiven Designs besonders dazu beigetragen, Kündigungsgründe zu identifizieren, die Lehrkräfte nach ihrem Entscheid berichten, die Stelle zu verlassen. Diese Studien sind oft explorativ angelegt und werden nicht immer wissenschaftlich publiziert. Der retrospektive Zugang hat den Nachteil, dass die Kündigungsgründe stark kognitiv gefiltert sein können. Insbesondere bei der Frage, welche Rolle negative Beanspruchungen für Kündigungen spielen, könnten nachträgliche Situationsumbewertungen zur Folge haben, dass diese nicht mehr das Gewicht erhalten, das sie während des Entscheidungsprozesses innehat- ten (Lazarus & Folkman, 1984; Herzog et al., 2007).

Wissenschaftliche Studien, die berufliche Mobilität prospektiv längsschnittlich untersuchen, sind selten. Die wenigen vorliegenden Panelstudien (z. B. Houkes et al., 2003; Kelly & Northrop, 2015) zeigen auf, wie grundlegend die Analyse des Zusammenspiels von kontextuellen und personellen Faktoren über die Zeit ist, um den Prozess zu verstehen, der zu beruflicher Mobilität führt.

Zusammenfassend kann für die Erforschung der beruflichen Mobilität mit Fokus auf das Individuum festgehalten werden, dass Studien, die objektive berufliche Übergänge in Kombination mit subjektiven Mobilitätsmotiven erfassen, am meisten Erkenntnisse generiert haben.

5. Fazit und Ausblick

Der Überblick über die aktuelle Forschung in den verschiedenen Feldern hat gezeigt, dass im deutschsprachigen Europa im Vergleich zu den USA wenig empirische Daten vorliegen und die Forschung zu beruflicher Mobilität von Lehrkräften bislang eher unsystematisch erfolgt ist. Forschungsdesiderate lassen sich in drei Bereichen lokalisieren: (1) Theorie, (2) Datenlage und (3) quantitative Auswertungsmethoden.

(1) Aufgrund der verschiedenen disziplinären Ausrichtungen tendiert das Forschungsfeld zu theoretischer Inkonsistenz. You und Conley (2015) gehen davon aus, dass es nicht möglich ist, ein theoretisches Modell zu entwickeln, um berufliche Mobilität zu prognostizieren. Diese Haltung führt in den quantitativen Studien oftmals dazu, dass Zusammenhänge nicht hypothesengeleitet, sondern explorativ untersucht werden, was zur Folge hat, dass bislang wenig gesicherte Erkenntnisse gewonnen werden konnten. Diese theoretische Lücke könnte geschlossen werden mit Modellen der Arbeits- und Organisationspsychologie zu beruflicher Mobilität

(als Überblick vgl. Holtom, Mitchell, Lee & Eberly, 2008; Hom et al., 2012), die bislang in der deutschsprachigen Lehrerforschung kaum rezipiert werden. Weitere theoretische Anknüpfungspunkte bestehen bei Modellen der psychologischen Lebenslauf- und der soziologischen Lebensverlaufsforschung (Baltes, Lindenberger & Staudinger, 1998; Brandtstädter & Rothermund, 2002), die die Interaktion zwischen Kontext und Individuum über die Zeit im Zentrum haben. Ausgehend von diesen könnte versucht werden, ein integrierendes Modell zur Erklärung beruflicher Mobilität zu entwerfen. Ein solches Modell kombiniert die strukturellen Rahmenbedingungen des Bildungssystems mit den individuellen Beweggründen der Lehrpersonen und postuliert Interaktionen von kontextuellen und individuellen Faktoren in ihrer Wirkung auf berufliche Mobilität über die Zeit. Es berücksichtigt kontextuelle Faktoren auf verschiedenen Ebenen des Bildungssystems (Unterricht, Schule, Schulsystem) und Faktoren im privaten Kontext des Individuums.

(2) Als weiteres Forschungsdesiderat kann festgehalten werden, dass für die Erforschung der beruflichen Mobilität von Lehrkräften in Europa die Datenlage verbessert werden muss. Die Bildungsträger sollten die Personalfuktuation systematischer monitorieren und dabei neben Individualdaten (wie z. B. Geschlecht, Berufserfahrung, unterrichtete Klassenstufe, Pensum) zumindest den Zielort der Mobilität und die strukturellen Bedingungen der jeweiligen Schulen differenziert erfassen und für die weiterführende Forschung zur Verfügung stellen (Weishaupt, 2014). Der Aufbau einer nationalen Schulpersonaldatenbank, die es über eine feste Identifikationsmöglichkeit von Einzelpersonen ermöglicht, deren Beschäftigungssituation und Mobilität über die Jahre zu verfolgen und diese mit Merkmalen auf der Schulebene zu verbinden, wäre für die Gewinnung vertiefter Erkenntnisse zu gestaltbaren Gründen für Mobilität grundlegend.

International vergleichende Forschung und ein systematischer Vergleich von Lehrerpopulationen verschiedener Länder und mit Kontrollgruppen anderer Berufe würden helfen zu verstehen, welche Rolle institutionelle und strukturelle Rahmenbedingungen spielen (Denzler & Wolter, 2009; Weishaupt, 2014). Eine Ergänzung dieser Schulstatistiken mit einer quantitativen oder qualitativen Befragung der kündigenden Lehrpersonen könnte einerseits dazu beitragen, relevante Faktoren durch die individuelle Interpretation in ihrer Bedeutung zu kontextualisieren (Rinke, 2008, S. 10). Sie würde andererseits helfen, Widersprüche zwischen objektiven Faktoren und deren individuellen subjektiven Deutung aufzudecken.

(3) Die bisherigen Erhebungs- und Auswertungsmethoden in den quantitativen Studien sind nur bedingt geeignet, um das Zusammenspiel von kontextuellen und subjektiven Faktoren über die Zeit zu analysieren. Bei nicht publizierten Studien, die keinem wissenschaftlichen Review unterworfen waren, stellt sich die Frage der Qualität und der Aussagekraft der Ergebnisse besonders deutlich. Es braucht dringend längsschnittliche Studiendesigns und Auswertungsmethoden, die Moderations- und Mediierungsprozesse modellieren, wie diese in der Entwicklungspsychologie bereits seit längerem üblich sind (Magnusson & Stattin, 1998). Zusätzlich würden Mehrebenenanalysen erlauben, Auswertungen auf

der Individualebene mit Analysen auf der Schulebene zu kombinieren und damit Zusammenhänge auf der Organisationsebene zu erklären (Raudenbush, Bryk, Cheong, Congdon & du Toit, 2011).

Der vorliegende Artikel hat sich das Ziel gesetzt, das Forschungsfeld zu beruflicher Mobilität von Lehrpersonen zu strukturieren und exemplarische Studien näher vorzustellen. Was der Artikel explizit nicht bietet, ist eine systematische Review. Die Auswahl der exemplarischen Studien war nicht immer einfach, speziell wenn mehrere Studien mit einem ähnlichen Forschungsdesign vorlagen. Der Beitrag hat aufgezeigt, dass Studien rund um berufliche Mobilität von Lehrkräften entlang von zwei Dimensionen systematisiert und voneinander abgegrenzt werden können. Eine solche Systematisierung kann dazu beitragen, verschiedene Forschungsansätze zu beruflicher Mobilität von Lehrpersonen besser aufeinander abzustimmen und bislang kaum beachtete Themenfelder zu adressieren. Eine verstärkte interdisziplinäre Zusammenarbeit könnte u. a. neue Impulse geben, um theoretische Modelle zu entwickeln, die perspektivenübergreifend sind, und in der Folge helfen, die empirischen Analysen konsequenter an theoretischen Annahmen auszurichten. Mit einer solchen bereichsübergreifenden Forschung können neue Erkenntnisse gewonnen werden, die unmittelbare praktische Relevanz für Maßnahmen zur besseren Steuerung der Fluktuation des Lehrpersonals haben.

Literatur

- Abele, A. E., Dette, D. E. & Hermann, B. (2003). Lehrerinnen und Lehrer. Vom Examen zum Schuldienst. In A. E. Abele, E. H. Hoff & H. U. Hohner (Hrsg.), *Frauen und Männer in akademischen Professionen. Berufsverläufe und Berufserfolg* (S. 71–88). Heidelberg: Asanger.
- Arthur, M. B., Hall, D. T. & Lawrence, B. S. (1989). Generating new directions in career theory: the case for transdisciplinary approach. In M. B. Arthur, D. T. Hall & B. S. Lawrence (Hrsg.), *Handbook of career theory* (S. 7–25). Cambridge: Cambridge University Press.
- Autorengruppe Bildungsberichterstattung (2016). *Bildung in Deutschland 2016. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*. Berlin: Bundesministerium für Bildung und Forschung.
- Baltes, P. B., Lindenberger, U. & Staudinger, U. M. (1998). Life-span theory in developmental psychology. In W. Damon & R. M. Lerner (Hrsg.), *Handbook of child psychology, Vol. 1: Theoretical models of human development*. (S. 1029–1143). New York, NY: Wiley.
- Borman, G. D. & Dowling, N. M. (2008). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of Educational Research*, 78(3), 367–409.
- Brandtstädter, J. & Rothermund, K. (2002). The life-course dynamics of goal pursuit and goal adjustment: A two-process framework. *Developmental Review*, 22(1), 117–150.
- Bundesamt für Statistik (2014). *Mobilität der Lehrkräfte der obligatorischen Schule 2014*. Neuchâtel: Bundesamt für Statistik.
- Bruneforth, M., Lassnigg, L., Vogtenhuber, S., Schreiner, C. & Breit, S. (2015). *Nationaler Bildungsbericht Österreich*. Wien: Bundesministerium für Bildung und Frauen.

- Denzler, S. & Wolter, S. C. (2009). Laufbahnentscheide im Lehrberuf aus bildungsökonomischer Sicht. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Professionalität – Bedingungen, Genese, Wirkungen und Messung* (S. 641–654). Weinheim: Beltz.
- European Commission (2013). *Study on policy measures to improve the attractiveness of the teaching profession in Europe*. Brussels: European Commission.
- European Commission (2015). *Shaping career-long perspectives on teaching. A guide on policies to improve Initial Teacher Education*. Brussels: European Commission.
- Fend, H. (2008). *Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Guarino, C. M., Santibanez, L. & Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research*, 76(2), 173–208.
- Gubler, M., Biemann, T. & Herzog, S. (2017). An apple doesn't fall far from the tree – Or does it? Occupational inheritance and teachers' career patterns. *Journal of Vocational Behavior*, 100, 1–14.
- Gunz, H. & Mayrhofer, W. (2017). *Rethinking career studies: Facilitating conversation across boundaries with the social chronology framework*. Cambridge: University Press.
- Hanushek, E. A., Rivkin, S. G. & Schiman, J. C. (2016). Dynamic effects of teacher turnover on the quality of instruction. *Economics of Education Review*, 55, 132–148.
- Helsper, W. & Böhme, J. (Hrsg.). (2008). *Handbuch der Schulforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Herzog, S. (2007). *Beanspruchung und Bewältigung im Lehrerberuf*. Münster: Waxmann.
- Herzog, S. (2014). Über den Berufseinstieg hinaus: Berufsbiografien von Lehrerinnen und Lehrern im Blickfeld der Forschung. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 314–338). Münster: Waxmann.
- Herzog, S. & Leutwyler, B. (Hrsg.). (2010). *Entwicklungslandkarte für Lehrpersonen. Notwendigkeit, Konzept und Implikationen eines biografisch orientierten Personalentwicklungsinstruments*. Bern: Haupt.
- Herzog, W. (1988). Das Verständnis der Zeit in psychologischen Theorien der Entwicklung. *Schweizerische Zeitschrift für Psychologie*, 47(2), 135–145.
- Herzog, W., Herzog, S., Brunner, A. & Müller, H. P. (2007). *Einmal Lehrer, immer Lehrer? Eine vergleichende Untersuchung der Berufskarrieren von (ehemaligen) Lehrpersonen*. Bern: Haupt.
- Holtom, B. C., Mitchell, T. R., Lee, T. W. & Eberly, M. B. (2008). Turnover and retention research: A glance at the past, a closer review of the present, and a venture into the future. *The Academy of Management Annals*, 2(1), 231–274.
- Hom, P. W., Mitchell, T. R., Lee, T. W. & Griffeth, R. W. (2012). Reviewing employee turnover: Focusing on proximal withdrawal states and an expanded criterion. *Psychological Bulletin*, 138(5), 831–851.
- Houkes, I., Janssen, P. P. M., Jonge, J. & Bakker, A. B. (2003). Specific determinants of intrinsic work motivation, emotional exhaustion and turnover intention: A multisample longitudinal study. *Journal of Occupational and Organizational Psychology*, 76(4), 427–450.
- Huberman, M. (1991). Der berufliche Lebenszyklus von Lehrern: Ergebnisse einer empirischen Untersuchung. In E. Terhart (Hrsg.), *Unterrichten als Beruf. Neuere amerikanische und englische Arbeiten zur Berufskultur und Berufsbiografie von Lehrerinnen und Lehrern* (S. 249–267). Köln: Böhlau.
- Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, 38(3), 499–534.

- Ingersoll, R. M. (2003). *Is there really a teacher shortage?* University of Pennsylvania, PA: GSE Publications. Zugriff am 06.03. 2018 unter http://repository.upenn.edu/gse_pubs/133
- Ingersoll, R. M. & May, H. (2012). The magnitude, destinations, and determinants of mathematics and science teacher turnover. *Educational Evaluation and Policy Analysis*, 34(4), 435–464.
- Ingersoll, R., Merrill, L. & May, H. (2016). Do accountability policies push teachers out? *Educational Leadership*, 73(8), 44–49.
- Ingrisani, D. (2014). *Die "neuen" Lehrerinnen und Lehrer: eine Befragung der ersten Jahrgänge der neuen Vorschul-und Primarlehrerinnen-und-lehrerausbildung der deutschen Schweiz*. Bern: Haupt.
- Johnson, S. M., Berg, J. H., Donaldson, M. L. & National Retired Teachers Association. (2005). *Who stays in teaching and why?: A review of the literature on teacher retention*. Project on the Next Generation of Teachers, Harvard. Zugriff am 06.03.2018 unter http://projectngt.gse.harvard.edu/files/gse-projectngt/files/harvard_report.pdf
- Kelchtermans, G. (2009). Career stories as gateway to understanding teacher development. In M. Bayer, U. Brinkjaer, H. Plauborg & S. Rolls (Hrsg.), *Teachers' career trajectories and work lives* (S. 29–47). Dordrecht: Springer.
- Kelchtermans, G. (2017). 'Should I stay or should I go?': Unpacking teacher attrition/retention as an educational issue. *Teachers and Teaching*, 23(8), 961–977.
- Keller-Schneider, M. (2010). *Kündigungsmotive von Lehrpersonen – Berufseinstiegende und berufserfahrene, kündigende und nicht-kündigende Lehrpersonen im Vergleich*. Zürich: Pädagogische Hochschule.
- Kelly, S., & Northrop, L. (2015). Early career outcomes for the 'best and the brightest' selectivity, satisfaction, and attrition in the beginning teacher longitudinal survey. *American Educational Research Journal*, 52(4), 624–656.
- Klemm, K. & Zorn, D. (2018). *Lehrkräfte dringend gesucht. Bedarf und Angebot für die Primarstufe*. Gütersloh: Bertelsmann Stiftung. Zugriff am 13. Juni 2018 unter <https://www.bertelsmann-stiftung.de/es/publikationen/publikation/did/lehrkraefte-dringend-gesucht/>.
- KMK – Kultusministerkonferenz. (2011). *Lehrereinstellungsbedarf und Lehrereinstellungsangebot in der Bundesrepublik Deutschland. Modellrechnung 2010 - 2020*. Berlin: KMK Kultusministerkonferenz. Zugriff am 13.06.2018 unter https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/BERICHT_MODELLRECHNUNG_online.pdf
- Kunze, K. & Stelmaszyk, B. (2008). Biographien und Berufskarrieren von Lehrerinnen und Lehrern. In W. Helsper & J. Böhme (Hrsg.), *Handbuch Schulforschung* (S. 821–838). Wiesbaden: Springer.
- Lazarus, R. S. & Folkman, S. (1984). *Stress, appraisal, and coping*. New York, NY: Springer.
- Magnusson, D. & Stattin, H. (1998). Person-context interaction theories. In W. Damon & R. M. Lerner (Hrsg.), *Handbook of child psychology. Fifth edition. Volume 1: Theoretical models of human development* (S. 685–759). New York, NY: John Wiley & Sons.
- OECD – Organisation for Economic Co-operation and Development. (2006). *Stärkere Professionalisierung des Lehrerberufs. Wie gute Lehrer gewonnen, gefördert und gehalten werden können*. Paris: OECD. Zugriff am 06.03.2018 unter <http://www.oecd.org/edu/school/attractingdevelopingandretainingeffectiveteachers-final-reportteachersmatter.htm>
- Raudenbush, S., Bryk, A., Cheong, Y. F., Congdon, R. & du Toit, M. (2011). *HLM 7. Hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software Int.
- Rinke, C. R. (2008). Understanding teachers' careers: Linking professional life to professional path. *Educational Research Review*, 3(1), 1–13.

- Rinke, C. R. (2013). Teaching as exploration? The difficult road out of the classroom. *Teaching and Teacher Education*, 34(6), 98–106.
- Rinke, C. R. (2014). *Why half of teachers leave the classroom. Understanding recruitment and retention in today's schools*. Lanham: Rowman & Littlefield Education.
- Rothland, M. (2014). Warum entscheiden sich Studierende für den Lehrerberuf? Berufswahlmotive und berufsbezogene Überzeugungen von Lehramtsstudierenden. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf (2. überarb. und erweit. Aufl.)* (S. 349–385). Münster: Waxmann.
- Rudow, B. (1994). *Die Arbeit des Lehrers: Zur Psychologie der Lehrertätigkeit, Lehrbelastung und Lehrergesundheit*. Bern: Huber.
- Rürup, M., Fuchs, H.-W. & Weishaupt, H. (2016). Bildungsberichterstattung – Bildungsmonitoring. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 411–437). Wiesbaden: Springer Fachmedien.
- Stadelmann, W. & Rhyn, H. (2011). Zusammenarbeit zwischen Deutschland, Österreich und der Schweiz in Bildungsfragen am Beispiel Lehrerbildung. *Beiträge zur Lehrerbildung*, 29(1), 109–114.
- SKBF – Schweizerische Koordinationsstelle für Bildungsforschung. (2018). *Schweizer Bildungsbericht 2018*. Aarau: Schweizerische Koordinationsstelle für Bildungsforschung.
- Terhart, E. (1994). *Berufsbiographien von Lehrern und Lehrerinnen*. Frankfurt a. M.: Lang.
- Terhart, E. (2010). Personalauswahl, Personaleinsatz und Personalentwicklung an Schulen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 255–275). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Terhart, E. (2016). Personalauswahl, Personaleinsatz und Personalentwicklung an Schulen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 279–299). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Terhart, E., Bennewitz, H. & Rothland, M. (2014). *Handbuch der Forschung zum Lehrerberuf*. Münster: Waxmann.
- Terhart, E., Altrichter, H., Schley, W. & Schratz, M. (1998). Lehrerberuf: Arbeitsplatz, Biographie, Profession. In H. Altrichter, W. Schley & M. Schratz (Hrsg.), *Handbuch zur Schulentwicklung* (S. 560–585). Innsbruck: Studien Verlag.
- Troesch, L. M. & Bauer, C. E. (2017). Gehen oder bleiben? Berufsverbleib und Ausstiegsgründe von Berufswechslern und Berufswechslerrinnen in den Lehrerberuf. In C. E. Bauer, C. Bieri & N. Safi (Hrsg.), *Berufswechsel in den Lehrerberuf. Neue Wege der Professionalisierung*. Bern: hep.
- Vardi, Y. (1980). Organizational career mobility: An integrative model. *Academy of Management Review*, 5(3), 341–355.
- Weishaupt, H. (2014). Lehrerberuf: Arbeitsplatz, Arbeitszeit, Arbeitsmarkt. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf (2. überarb. und erweit. Aufl.)* (S. 103–122). Münster: Waxmann.
- You, S. & Conley, S. (2015). Workplace predictors of secondary school teachers' intention to leave. An exploration of career stages. *Educational Management Administration & Leadership*, 43(4), 561–582.

Eveline Wittmann

Indikatorenentwicklung im Föderalstaat – Governancetheoretisch fundierte Analysen im Rahmen der kanadischen Bildungsberichterstattung

Zusammenfassung

Der nationalen Bildungsberichterstattung wird die Funktion einer rationalen Kontrolle von politischem Steuerungshandeln zugeschrieben. Allerdings ist die Bildungsberichterstattung selbst Gegenstand politischer Aushandlungsprozesse an der Schnittstelle von Wissenschaft und Politik. Diese können die Qualität der Indikatorenentwicklung beeinträchtigen. Empirisch von Interesse ist die Frage nach der Ausgestaltung der Interdependenz, den resultierenden Beschränkungen für die Indikatorenentwicklung und den Strategien der bericht-erstattenden Einrichtungen zur Bewältigung der Interdependenz. Vor allem im Föderalstaat ist mit aus governancetheoretischer Perspektive interessierenden Handlungsrestriktionen für die Bericht erstattenden Einrichtungen bei der Indikatorenformulierung zu rechnen. Der vorliegende Beitrag analysiert anhand governancetheoretischer Kriterien die kanadische Bildungsberichterstattung, da diese wie die deutsche Bildungsberichterstattung in einem föderalen Bildungssystem angesiedelt ist. Die kanadischen Erfahrungen ermöglichen Hinweise darauf, wie mit den Problemen der Interdependenz im Föderalstaat umgegangen werden kann.

Schlagworte

Indikatorenentwicklung, Bildungsberichterstattung, Governance, Mehrebenen-system

Prof. Dr. Eveline Wittmann, TUM School of Education, Lehrstuhl für Berufspädagogik, Technische Universität München, Arcisstr. 21, 80333 München, Deutschland
E-Mail: eveline.wittmann@tum.de

Indicator development in federal states – An analysis from the perspective of governance in the context of Canadian educational reporting

Abstract

One function attributed to national education reporting is to enable a rational control of government regulation. However, educational reporting itself is the subject of political negotiation processes at the interface between science and politics. These can affect the quality of the development of the indicators. Empirically the question is of interest of how interdependence is shaped, which limitations ensue for the development of indicators, and which strategies the reporting institutions apply to deal with interdependence. Particularly in federal states, it is to be expected that the reporting authorities will be affected by restrictions which matter from the perspective of governance theory. The present paper analyzes the Canadian education reporting on the basis of criteria from governance theory, since, like German education, it is located in a federal education system. The Canadian experience provides evidence of how to deal with the problems of interdependence in the federal state.

Keywords

Education indicators, Educational Reporting, Governance, Multi level

1. Problemstellung

Mit dem Begriff *Bildungsbericht* angesprochen sind „regelmäßig erscheinende [...] Publikationen, die eine datengestützte, evaluative Gesamtschau über das Bildungswesen ermöglichen sollen“ (Rürup, Fuchs & Weishaupt 2016, S. 412). Bildungsberichterstattung ist grundsätzlich an administrative Entscheidungsträger und die Öffentlichkeit adressiert. Dabei geht es um die Zusammenfassung von Befunden aus unterschiedlichen Erhebungsprogrammen und deren Aufbereitung unter spezifischen übergreifenden Fragestellungen (Rürup et al., 2016); statistische Einzelkennziffern werden dazu mathematisch oder textlich zu übergreifenden Indikatoren integriert, die die Beantwortung dieser Fragestellungen erlauben (ebd.).

Neben der Darstellung von Steuerungswissen wird der Bildungsberichterstattung die Funktion der Systemkontrolle und der Legitimation von Steuerungshandeln zugeschrieben (Rürup et al., 2016). Voßkuhle betont hierbei die Notwendigkeit von Wissen für die Aufrechterhaltung staatlicher Steuerungsgewalt und stellt heraus: „nur der informierte Staat ist in der Lage, seine Gewährleistungsverantwortung für das Gemeinwohl effektiv wahrzunehmen.“ (Voßkuhle, 2005, § 43). Mit der Bildungsberichterstattung intendiert ist eine „evidenzbasierte ‚Rationalisierung‘ von Bildungspolitik“ (Rürup et al., 2016, S. 432;

vgl. Brauckmann et al., 2010; zur Evidenzbasierung von Bildungspolitik und -praxis z. B. auch Bromme, Prenzel & Jäger, 2014; Wurster, Richter & Lenski, 2017).

Empirisch betrachtet ist die Wirksamkeit von Bildungsberichten bislang kaum untersucht. Allerdings stellt sich neben der Frage der Wirkung von Bildungsberichten auf die Bildungspolitik sowie als Voraussetzung für deren Beantwortung die Frage nach der Qualität von in Form von Indikatoren im Rahmen der Bildungsberichterstattung bereitgestellten Informationen (Rürup et al., 2016). Numerische Indikatoren stellen ein „Governanceinstrument [dar], das [...] unbezweifelnd daher kommt und den wissenschaftlichen Laien mit Beurteilungskriterien versorgt, deren Verlässlichkeit und Relativität er nicht zu beurteilen vermag“ (Schuppert, 2008, S. 286). In diesem Beitrag wird daher der Blick darauf gerichtet, dass die Erstellung von Indikatoren selbst, die an der Schnittstelle von Wissenschaft und Politik stattfindet und zumeist bei wissenschaftlichen Forschungsinstituten oder statistischen Ämtern angesiedelt ist, Gegenstand politischer Aushandlungsprozesse ist (vgl. auch Niedlich & Brüsemeister, 2016), die die Qualität der Indikatorenentwicklung beeinflussen kann. Bildungsberichterstattung unterliegt einer Reihe offener oder verdeckter politischer Einflüsse und damit auch den Strategien und Interessen der betreffenden Akteure.

In zentralstaatlichen Systemen finden Aushandlungsprozesse grundsätzlich direkt zwischen berichterstattender Einrichtung und dem zuständigen zentralstaatlichen Ministerium statt. Im Föderalstaat treten weitere mögliche Akteure hinzu: Zum einen sind dies die einzelnen föderalen Einheiten, zum anderen können institutionalisierte Kooperationsarrangements zwischen den föderalen Einheiten mit eigener Akteursqualität, wie die Kultusministerkonferenz, hinzukommen. In solchen politischen Mehrebenensystemen ist Interdependenz zwischen den Ebenen und ihren Akteuren über formale Zuständigkeiten und die Verfügung über Ressourcen für die Aufgabebearbeitung institutionalisiert (Kussau & Brüsemeister, 2007). Neben der Möglichkeit, aktiv Interessen einzubringen, kann bei den genannten Akteuren auch eine Verhinderungsfähigkeit für eine transparente Bildungsberichterstattung liegen (vgl. Benz, 2009). Vor diesem Hintergrund ist mit aus governancetheoretischer Perspektive interessierenden Handlungsrestriktionen für die berichterstattenden Einrichtungen bei der Indikatorenformulierung zu rechnen (Mayntz, 2005). Die Erforschung solcher Restriktionen stellt in der Bildungspolitik grundsätzlich ein Forschungsdefizit dar (Kussau & Brüsemeister, 2007). Empirisch von Interesse ist vor diesem Hintergrund die Frage nach der Ausgestaltung der Interdependenz, den hieraus resultierenden Beschränkungen für die Indikatorenentwicklung und den Strategien der berichterstattenden Einrichtungen zur Bewältigung der Interdependenz.

Aus deutscher Sicht ist es vor diesem Hintergrund besonders von Interesse, die kanadische nationale Bildungsberichterstattung zu betrachten, weil diese der deutschen vergleichbar in einem föderal organisierten Bildungssystem erfolgt. Dadurch lassen sich entsprechende Probleme nachzeichnen und Fragen bezüglich der deutschen Bildungsberichterstattung aufwerfen, die nicht zuletzt die Effektivität nationaler Bildungsberichterstattung als Steuerungsinstrument tangie-

ren. Gleichzeitig ermöglichen die kanadischen Erfahrungen Hinweise darauf, wie mit den Problemen der Bildungsberichterstattung im Föderalstaat umgegangen werden kann.

Der kanadischen Bildungsberichterstattung liegt das sogenannte *Pan-Canadian Education Indicators Program* (PCEIP) zugrunde, bei dem zwischen einem nationalen und einem internationalen Strang zu unterscheiden ist. Im Folgenden werden zunächst die Analysekriterien dargelegt und governancetheoretisch begründet. Sodann werden basierend auf Analysen, die in Wittmann (2011) ausführlich dargelegt sind, die inhaltliche Ausrichtung der publizierten schriftlichen Berichte sowie die Ausgestaltung und Veränderungen der berichteten Indikatoren beschrieben, wobei zwischen dem nationalen und dem internationalen Strang des PCEIP differenziert wird. Hieran schließen die governancetheoretisch begründeten Analysen zu Interdependenzen bei der Indikatorenformulierung und zu Mechanismen der Interdependenzbewältigung an.

2. Kriterien der Analyse und methodisches Vorgehen

Die qualitativen Analysen sind orientiert am Verfahren der strukturierenden Inhaltsanalyse (Mayring, 2002). Es wurde in zwei Durchgängen verfahren: Zunächst wurden vorgängig festgelegte, im Kontext der Indikatorenforschung theoretisch begründete Auswertungskategorien an die vorliegenden Berichte sowie ergänzende Dokumente herangetragen. Im Kontext eines umfassenderen Projekts zur „Indikatorenentwicklung im Rahmen der nationalen Bildungsberichterstattung“ wurden entlang von Berichtssystemen anderer Länder sowie inter- bzw. supranationaler Organisationen umfassend untersucht, wie die Indikatorenentwicklung in das Gesamtmonitoringsystem eingebunden war, außerdem Aufbau und Struktur der Berichte sowie die Kriterien der Indikatorenauswahl und die Entwicklung von Indikatoren in den Bereichen outcome- und prozessbezogener Indikatoren im Zeitablauf (vgl. z. B. auch Wittmann & Hardt, 2011). Hierzu wurden unter anderem die folgenden Kriterien zugrunde gelegt, die die Darstellung im anschließenden Abschnitt leiten (zum Vorgehen im Folgenden Wittmann, 2011; vgl. Döbert & Klieme, 2009; Jaafar & Earl, 2008):

1. *Systematische inhaltliche Abdeckung*: Auf welche Gebiete des Bildungswesens bezieht sich die Erfassung in den Berichten systematisch, insbesondere auch: Inwieweit wird die berufliche Bildung erfasst?
2. *Aggregationsebenen und Referenzen*: Auf welchen Aggregationsebenen erfolgt die Berichterstattung? Was sind ihre Bezugsgrößen, insbesondere: Inwieweit werden Vergleiche zwischen föderalen Einheiten oder zu anderen Bildungssystemen berichtet?

3. *Indikatorenauswahl:*

- a) Welche Indikatoren wurden im Zeitverlauf hinzugenommen oder weggelassen und weshalb? Welche Kennziffern wurden innerhalb der Indikatoren im Zeitverlauf hinzugenommen und weshalb?
- b) Inwieweit ist die Indikatorenauswahl durch die Verfügbarkeit von Daten oder die Qualität der Datenbasis bestimmt, d. h. durch die Zugriffsmöglichkeit auf repräsentative und aktuelle Daten?

Die Analysen wurden ergänzt durch explizierende Inhaltsanalysen qualitativer Experteninterviews (Mayring, 2002), die mit Berichtshauptverantwortlichen geführt wurden und den Erhalt vertiefender Kontextinformationen mit dem Ziel der Prüfung der Validierung der getätigten Interpretationen zur Zielsetzung hatten. Zur Sprache kamen hier insbesondere die Entstehungssituationen der Berichte und einzelner Indikatoren sowie vorgenommener Veränderungen der Gesamtberichterstattung und der Berichterstattung einzelner Indikatoren. Sowohl die Analyse der verschriftlichten Dokumente als auch die Interviews und darauf bezogene Analysen waren darüber hinaus unter anderem an den folgenden Kriterien orientiert:

4. *Einbeziehung politisch-administrativer Instanzen:* Auf welche Weise werden die unterschiedlichen bildungspolitischen bzw. administrativen Ebenen in die Datenbereitstellung und Indikatorenformulierung eingebunden?
5. *Rechenschaft und Konsequenzen:* Erfolgt die Berichterstattung unter Bezugnahme auf zuvor politisch gesetzte Standards, insbesondere auf der Bundesebene konsentrierte curriculare Mindestanforderungen?

Für den hier vorgelegten Artikel bieten Textstellen zu diesen Kriterien sowie zur Datenverfügbarkeit (Kriterium 3b) nun die Gelegenheit zur Reanalyse der in Wittmann (2011) vorgelegten Auswertungen unter Bezugnahme auf ausgewählte governancetheoretische Kategorien, die besondere Relevanz für Interdependenzen im Mehrebenensystem aufweisen. Nach Kussau und Brüsemeister (2007) sind Interdependenzbeziehungen durch Verschränkungen zwischen Verfügungsrechten verursacht. Altrichter und Heinrich (2007) unterscheiden hierbei zwischen Verfügungsrechten i. e. S. und Verfügungsfähigkeiten. Verfügungsrechte stellen Regeln und Normen dar, die unterschiedlich stark institutionalisiert und rechtlich abgesichert sein können. Auch die in der Einleitung angesprochene Verhinderungsfähigkeit ist durch Verfügungsrechte bedingt (Benz, 2009; Tsebelis, 2002). Verfügungsfähigkeiten bestimmen, inwieweit die Inhaber von Verfügungsrechten diese auch tatsächlich ausüben können. Hierzu gehören materielle Ressourcen ebenso wie immaterielle Ressourcen wie Zeit, Kompetenz und das Vermögen, Handlungen *Sinn* zu verleihen.

Für die Bewältigung von durch Ressourcenverflechtungen gekennzeichneten Interdependenzbeziehungen stehen nach Kussau und Brüsemeister (2007) drei Strategien zur Verfügung:

1. Anpassung i. S. eines Verzichts auf ursprüngliche Handlungsziele,
2. Ausweichen auf andere Ressourcen, sowie
3. Zusammenlegen von Ressourcen mit anderen Akteuren mit dem Effekt, die Gewichtung der Interdependenzbeziehung zu verschieben.

Empirisch zu klären ist vor diesem Hintergrund die Frage nach dem Mischungsverhältnis der Merkmale von und der Intensität von Interdependenzbeziehungen (Kussau & Brüsemeister, 2007). Die Fragen, denen hier empirisch nachgegangen werden, beziehen sich in Orientierung an diesem Raster auf

1. die konkrete Ausgestaltung der Interdependenzbeziehungen, ihre Begründung in vorherrschenden Regelungen und Normen sowie verfügbaren materiellen wie immateriellen Ressourcen und ihre Auswirkungen auf die Indikatorenentwicklung;
2. Strategien der Interdependenzbewältigung (Anpassen, Ausweichen auf andere Ressourcen, Zusammenlegen von Ressourcen) und ihre Auswirkungen auf die Indikatorenentwicklung (s. Altrichter & Heinrich, 2007).

Die Analysen beruhen jeweils auf Angaben aus allen verfügbaren schriftlichen Berichtsausgaben.¹ Die ergänzenden Experteninterviews wurden mit insgesamt vier Personen geführt und ebenfalls entlang der Kriterien der strukturierenden qualitativen Inhaltsanalyse nach Mayring (2002) ausgewertet. Ein Experteninterview wurde zum Gesamtsystem des Bildungsmonitoring in Kanada mit Schwerpunkt auf Leistungsvergleichsstudien geführt. Die befragten Personen waren für die Berichterstattung in den ausgewerteten Berichtssystemen an zentraler Stelle und von Beginn an verantwortlich. Die geringe Anzahl der Befragungspersonen ergibt sich daraus, dass diese Merkmale und damit Aussagekraft mit Blick auf die angelegten Kriterien (3) bis (5) lediglich auf sehr wenige in die Berichterstattung involvierte Personen zutreffen. Gleichzeitig ist bei den befragten Experten ein besonders hohes Maß an Expertise aus den unterschiedlichen Perspektiven der beteiligten Akteure anzunehmen. Als Experten fungierten Mitarbeiter von *Statistics Canada* (kanadisches statistisches Bundesamt; 2 Experten) und CMEC (Council of Ministers of Education, d. h. der Rat der für Bildung verantwortlichen Minister der kanadischen Provinzen und Territorien; 2 Experten). *Statistics Canada* und CMEC sind hierbei verantwortlich für die Berichterstattung im Rahmen des PCEIP. Die teilstrukturierten Interviews waren jeweils auf den Kompetenzbereich der Interviewpartner zugeschnitten, so dass nicht von allen Experten dieselben Gegenstandsbereiche thematisiert wurden. Insgesamt wurden drei Interviews geführt, davon zwei Interviews mit jeweils zwei Teilnehmern und ein Interview mit einer einzelnen Person. Eine Person nahm an drei Interviews teil. Maßgeblich

1 Aus Gründen der Übersichtlichkeit der Darstellung werden die Berichtsquellen nur ausgewiesen, wo sie zitiert oder paraphrasiert werden, nicht jedoch, wo sie das Datenmaterial der vergleichenden Inhaltsanalysen darstellen. Eine Nachvollziehbarkeit ist durch Verweise auf konkrete Indikatoren, Kapitel oder die Inhaltsverzeichnisse der betroffenen Berichte jeweils vollumfänglich gegeben.

für die Gruppierung der Interviewpartner war das Angebot der betreffenden Institutionen. Die Interviews wurden im Regelfall als Face-to-face-Interviews geführt; in zwei Fällen erfolgte die Teilnahme am Interview wegen der räumlichen Distanz per Telefonkonferenz. Die Validierung der Interviewinhalte wurde zum einen über den Abgleich mit der Indikatorenentwicklung im Detail gemäß den Inhaltsanalysen der Berichtstexte vorgenommen (vgl. auch Wittmann 2011), zum anderen über einen Abgleich zwischen den Aussagen der Interviewpartner sowohl innerhalb der einzelnen Gespräche als auch zwischen den Gesprächen.

3. Indikatorenentwicklung im *Pan-Canadian Education Indicators Program* (PCEIP) – nationale Perspektive

3.1 Zur ursprünglichen Struktur der Bildungsberichterstattung im PCEIP

Die kanadischen Bildungsberichte tragen den Titel *Education Indicators in Canada*; ihnen liegt das sogenannte *Pan-Canadian Education Indicators Program* (PCEIP) zugrunde. Gemäß Eigendarstellung verfolgt das PCEIP die Zielstellung, einen Satz statistischer Maße zu den kanadischen Bildungssystemen zu veröffentlichen, das es Politikern, Praktikern und der Öffentlichkeit ermöglichen sollte, die Leistung dieser Bildungssysteme über die föderalen Einheiten hinweg ebenso wie im Zeitverlauf zu überwachen (CESC, 2007a). Umfassende schriftliche Berichtsbände wurden jeweils in den Jahren 1996, 1999, 2003, 2005 sowie 2007 publiziert, im Jahr 2007 zusätzlich ein methodenorientiertes Handbuch (CESC, 2007b). Erarbeitet wurden diese Berichtsbände durch den *Canadian Education Statistics Council* (CESC), in dem das kanadische statistische Bundesamt (*Statistics Canada*) mit dem Rat der Bildungsminister der Provinzen und Territorien (CMEC) kooperiert. CMEC stellt eine intergouvernementale Einrichtung dar, welche analog der Kultusministerkonferenz gesehen werden kann und im Jahr 1967 im gegenseitigen Einvernehmen gegründet wurde angesichts der Zuständigkeit der Provinzen und Territorien als föderale Einheiten für die Bildungspolitik. Diese Einrichtung sollte ein Diskussionsforum, einen Mechanismus für gemeinsame Aktivitäten sowie ein Instrument der Beratung und Zusammenarbeit mit Zentralregierung und nationalen Einheiten dienen; außerdem sollte sie bildungspolitische Interessen der Provinzen und Territorien auch international repräsentieren (CMEC, 2009a; Wittmann, 2011).

Die Berichte basierten ab dem Jahr 1999 auf einem im Beratungsgremium des *Strategic Management Committee* des CESC entwickelten Indikatorenrahmen und bezogen sowohl die nationale Ebene als auch die der föderalen Einheiten ein (CESC, 2007a). Veränderungen der Berichterstattung zwischen 1999 und den Folgejahren sind jedoch bezogen auf die berichteten Indikatoren substan-

ziell und werden bereits an der Kapitelstruktur deutlich (CESC, 2000).² Während die Berichtsgliederung des Berichtsbandes aus dem Jahr 1999 einem Kontext-Input/Prozess-Output-Outcome-Modell für das gesamte Bildungswesens zu folgen scheint, lässt sich die Struktur der Berichte ab 2003 im Sinne separater Input-Prozess-Output-Outcome-Modelle für zwei große Abschnitte des Bildungswesens interpretieren, die schulische und die hochschulische Bildung. In den Jahren 2003 bis 2007 bezieht sich die Berichterstattung jeweils auf die Bereiche primärer, sekundärer und postsekundärer Bildung; hierzu werden Inputindikatoren finanzieller Art sowie Output- und Outcomeindikatoren dargestellt. Die Erwachsenenbildung wird subsumiert unter die postsekundäre Bildung, sie bildet mithin keinen eigenständigen Berichtsaspekt. Aus deutscher Sicht erscheint auffällig, dass seit dem Berichtsjahr 2003 über die berufliche Bildung mit Blick auf berufliche Bildungsgänge und Ausbildungseinrichtungen nicht mehr systematisch berichtet wird (CESC, 2003). Insgesamt stellen die Berichte also vorrangig Berichte über das öffentliche allgemeine, vor allem das schulische und hochschulische Bildungswesen dar (Wittmann, 2011).

Vergleichsanalysen für die Referenzen der Indikatoren in den Berichten im Zeitverlauf verdeutlichen: Die Berichte ermöglichen zunehmend eine Zurechenbarkeit von Ergebnissen zur Ebene der bildungspolitisch zuständigen Provinzen und Territorien; im Berichtsjahr 2007 sind die Indikatoren bezogen auf diese Ebene verstärkt ausdifferenziert. Betroffen sind vor allem Indikatoren zu den Outcomes *Transitions to postsecondary education* (E1) sowie *Transitions to the labour market* (E2). Aussagen der Experteninterviews untermauern, dass das Grundanliegen des PCEIP in der Berichterstattung zu den föderalen Einheiten besteht. Die Ausdifferenzierung der Berichterstattung ist demzufolge der Tatsache zuzuschreiben, dass erst mit dem Youth in Transition Survey (YITS) die erforderlichen Daten vorlagen (I1, 602–609; I2, 319–321).³

Insgesamt finden sich im Zeitablauf Veränderungen auf der Ebene der Gesamtindikatoren ebenso wie der einzelner Kennziffern (ausführlich Wittmann, 2011). Beispielsweise wurde im Jahr 2007 im Kapitel zur postsekundären Bildung Indikator D5 (*Literacy*) hinzugenommen.⁴ Auch erheblich ausgeweitet wurden die Indikatoren zu Übergängen in postsekundäre Bildung und Arbeitsmarkt. So wurde Indikator E1 (*Transitions to postsecondary education and the labour market*) aus den Berichten der Jahre 2003 sowie 2005 zum Indikator *Transitions to postsecondary education* (E1) und zum Indikator *Transitions to the labour market* (E2) ausdifferenziert; auch wurde die Anzahl der berichteten Kennziffern in diesen Bereichen deutlich erhöht. Dies resultiert einerseits aus einer erweiterten Berichterstattung zur Ebene der föderalen Einheiten, andererseits daraus,

2 Eine Übersicht über Konkordanzen der Kennziffern und Indikatoren des Berichtsjahres 1999 mit den folgenden Berichtsjahren ist im Handbook (CESC, 2007b) zu finden.

3 Die Angabe der Interviewquellen erfolgt anonymisiert mit der Nummer des geführten Interviews und der Zeilenangabe des Transkripts.

4 Partiiell erfolgte eine Berichterstattung zu diesem Indikator, die sich auf eine frühere Version der IALS-Studie bezieht, bereits im Bericht des Jahres 1999 (CESC 2007b, S. 157 f.; CESC 2000, S. 86 ff.).

dass im Jahr 2007 auch zu High-School-Dropouts, nachgeholten High-School-Abschlüssen, Arbeitsmarktgliederung und Verdienst Kennziffern berichtet wurden. Zugrunde lag eine Erweiterung der Datenquellen, d. h. der Heranziehung des YITS für Indikator E1 und des *National Graduates Survey* sowie des *Follow-up of Graduate Survey* für Indikator E2, wobei letztere Surveys nicht regelmäßig durchgeführt werden (Statistics Canada, 2009b). Die im Bericht des Jahres 2003 enthaltene Schulebene (Indikator C4) war in den Jahren 2005 und 2007 nicht mehr Teil der Berichterstattung. Diesbezüglich hatte die kanadische PISA 2000-Erhebung eine zentrale Datenquelle dargestellt. Außerdem wurde der Indikator D2 (*Adult education and training*) aus 2003 bzw. 2005 im Jahr 2007 nicht mehr berichtet. Für diesen Indikator war der im Jahr 2003 erschienene *Adult Education and Training Survey* (AETS; Statistics Canada, 2009a) herangezogen worden. Auch hier wurde die Berichterstattung also aufgrund der nicht vorhandenen Datenbasis verändert. Ebenfalls durch die Datenbasis begründet ist eine weitere inhaltliche Modifikation; so bezieht sich der Indikator aus dem Berichtsjahr 2003 auf Erwachsenenbildung unterschiedlicher Anbieter, d. h. von Non-Profit-Organisationen bis zu Arbeitgebern, der Indikator des Jahres 2005 demgegenüber auf arbeitsbezogenes Lernen. Allgemein verdeutlichen diese Veränderungen, dass die Verfügbarkeit von Daten die berichteten Indikatoren, deren Ausprägung und Ausdifferenzierung hochgradig beeinflusst (Wittmann, 2011).

Die Berichterstattung zum beschriebenen PCEIP-Strang wurde nach 2007 nur noch online auf Basis zweimonatlich upgedateter Daten fortgeführt (Statistics Canada, 2009c). Allerdings sollte ein zweiter, im Jahr 2009 hinzugekommener Strang des PCEIP in Form einer neuen Print-Publikationsreihe einen Vergleich zwischen den Bildungssystemen der kanadischen Provinzen und Territorien mit den Bildungssystemen der Länder der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) ermöglichen (CMEC, 2009c).

3.2 Zum internationalen Strang der Bildungsberichterstattung im PCEIP

Im Jahr 2009 erschien in Buchform ein Bericht unter dem Titel *Education Indicators in Canada. An International Perspective 2009*. Dieser folgte einer Kritik des *Canadian Council on Learning* (CCL), der zufolge es der kanadischen Bildungsberichterstattung für 60 % der Indikatoren des jährlich erscheinenden OECD-Berichts *Education at a Glance* nicht möglich war, kanadische Daten zu liefern (CCL, 2007). Diese Lücken seien mit dem neuen Berichtsformat geschlossen worden (CCL, 2009b). Der Bericht des Jahres 2009, der Ausgangspunkt für die seither erfolgende kanadische Bildungsberichterstattung ist (CESC, 2016), ist hier also insofern von besonderem Interesse, als mit ihm mit der OECD eine weitere Akteurebene für die kanadische Bildungsberichterstattung relevant wird. Der Bericht *Education Indicators in Canada. An International Perspective* erfolgt abgestimmt mit den Definitionen sowie Methodologien des jährlich er-

scheinenden OECD-Berichts. Er bezieht sich auf die Primar-, Sekundar- und Postsekundarbildung unter Einschluss der beruflichen und tertiären Bildung; für die Erwachsenenbildung existiert demgegenüber weiterhin kein eigenständiger Berichtszweig, ebenso wenig wie non-formale sowie bzw. informelle Aspekte des Lernens systematisch berichtet werden. Berichtet wird zu Output und Outcome und finanziellen Input-Faktoren sowie zur Bildungsbeteiligung und Übergängen in Beschäftigung (CESC, 2009a). Für den mittlerweile jährlich erscheinenden Bericht werden zum einen Daten aus der OECD-Berichterstattung, zum anderen Daten aus zumindest jährlich erfolgenden Erhebungen zugrunde gelegt. Ausnahmen gelten für Daten, die finanzbezogenen Indikatoren der Primar- und Sekundarstufe zugrunde liegen, wobei die Berichterstattung hier in den Zwischenjahren auf Schätzwerten basiert (Statistics Canada, 2009d), sowie für Daten aus der PISA-Erhebung der OECD, welche im Drei-Jahres-Rhythmus getaktet ist. Die Berichterstattung erfolgt durchgehend vergleichend auf supranationaler Ebene (OECD und EU19) sowie auf den Ebenen von Nationalstaaten innerhalb der OECD und von kanadischen Provinzen. Andere Vergleiche, z. B. in Form von Zeitreihen, werden unter anderem wegen des zeitlichen Aufwandes und der Schwierigkeiten, die Provinzdaten mit den OECD-Erfordernissen kompatibel zu machen, bislang kaum vorgenommen (I1, 957–976; Wittmann, 2011). Die Indikatorenauswahl zum Bericht *Education Indicators in Canada. An International Perspective* erfolgt gemäß CESC (2009a) neben der Indikatorenauswahl der OECD in Orientierung daran, welche Daten für Provinzen und Territorien verfügbar sind. So ist der Wegfall des Kapitels D und seiner schul- und klassenbezogenen Indikatoren im kanadischen Bericht von 2009 durch die Tatsache begründbar, dass Kanada an der TALIS-Studie nicht beteiligt war, welche dem entsprechenden Kapitel des OECD-Berichts zugrunde liegt. Daneben wurden im kanadischen Bericht gegenüber dem OECD-Bericht auch Indikatoren zu ökonomischen Outcomes, sozialen und personenbezogenen Outcomes sowie Kennziffern bezüglich der Finanzierung und zum Zusammenhang von Bildungsstand und Arbeitslosigkeit nicht berichtet. Diesbezüglich bestätigen auch die Aussagen der interviewten Experten, dass die Indikatorenformulierung und -auswahl durch die Verfügbarkeit von Daten oder andere der Datenerhebung zugrunde liegende Definitionen beeinflusst ist (I1, 1021–1037; I2, 109–112; Wittmann, 2011).

Politische Entscheidungen können vor dem Hintergrund dieser Erläuterungen bei der Indikatorenentwicklung im Rahmen der nationalen kanadischen Bildungsberichterstattung zumindest in dreierlei Hinsicht eine Rolle spielen: (a) im Rahmen der expliziten Formulierung politischer Anforderungen durch die Ministerien der Provinzen und Territorien, (b) durch die Bereitstellung oder Nicht-Bereitstellung einheitlicher und der Berichterstattung dienlicher Erfassungsstandards administrativer Daten durch die Provinzen und Territorien sowie (c) durch die Bereitstellung finanzieller Mittel für Datenerhebungen.

Diesbezügliche Interdependenzbeziehungen zeichnen sich in den geführten Interviews ab und werden im Folgenden genauer herausgearbeitet.

4. Interdependenzen und ihr Einfluss auf die Indikatorenentwicklung

4.1 Interdependenzen durch formale Kompetenzverteilung im Föderalstaat

Zu den von CMEC wahrgenommenen Aktivitäten gehört neben der Berichterstattung über Bildungsindikatoren auch deren Entwicklung (CMEC, 2009a). CMECs Funktion bei der Indikatorenentwicklung ergibt sich daraus, dass die föderalen Einheiten gemäß *Section 93 BNA*⁵ wie in der Bundesrepublik Deutschland grundsätzlich für Bildung allein zuständig sind. Eine Bundeszuständigkeit für Schulpolitik ist lediglich gegenüber der indianischen Bevölkerung gegeben (*Section 91 CA*; Avenarius, Brauckmann & von Kopp, 2007). Jedoch ist insgesamt der kanadische Bundesstaat traditionell geprägt vom Bestreben nach möglichst großer Unabhängigkeit der föderalen Einheiten im Inneren (Avenarius et al., 2007); bis heute fehlt ein für Bildungsfragen zuständiges föderales Ministerium (The Prime Minister of Canada, 2018).

Die Bildungsminister der Provinzen und Territorien sind formal über Konsultationen in die Diskussion und Formulierung der Indikatoren eingebunden (T(CESC, 2007b)). Die faktische Arbeitsteilung zwischen den Ministerien der Provinzen und Territorien auf der einen Seite, vermittelt durch CMEC, und *Statistics Canada* im *Canadian Education Statistics Council* (CESC) auf der anderen Seite lässt sich den Angaben aus den geführten Experteninterviews zufolge so beschreiben, dass erstere Informationsbedürfnisse formulieren und über den Inhalt entscheiden, während die methodisch-methodologische Entwicklung der Indikatoren durch *Statistics Canada* verantwortet wird (I1, 41–46; I1, 906–909; I2, 183–195).

Eine Ankoppelung an andere Mechanismen einer wissensorientierten Governance, wie Mindeststandards mit oder ohne rechtliche Legitimation (Schuppert, 2008) ist nicht vorhanden. Trotz entsprechender Willenserklärungen von CMEC sind solche Standards bislang kaum verwirklicht, was die Frage nach der diesbezüglichen Handlungsfähigkeit von CMEC als Kollektivorgan aufwirft: Konkret ist in der sogenannten *Victoria Declaration* von CMEC aus dem Jahr 1993 zwar ein nationaler Aktionsplan intendiert, der auf Testmaßnahmen, Standardsetzung und Harmonisierung der Curricula der Provinzen und Territorien fokussiert sein sollte. Auf die Rechenschaftslegung zu entweder durch ein Bundesministerium oder in kooperativen Gremien wie CMEC formulierten Standards könnte sich auch die nationale Bildungsberichterstattung beziehen. Jedoch ist ein gemeinsamer curricularer Rahmen bis heute in Kanada nur für den naturwissenschaftlichen Bereich

5 Der *British North America Act* (BNA) von 1867 ist neben dem 1982 beschlossenen „Constitution Act, 1982“ (CA) das zentrale Verfassungsdokument Kanadas (Avenarius, Brauckmann & von Kopp, 2007, S. 69).

realisiert (siehe CMEC, 1997).⁶ Wie auch im Rahmen der Experteninterviews bestätigt wurde, erfolgt eine Setzung von Standards für den Schulbereich ansonsten lediglich partiell auf der provinziellen Ebene und ohne Anbindung an die nationale Ebene (I3, 171; I3, 220–228). Von den in der *Victoria Declaration* genannten Maßnahmen sind bislang nur gemeinsame Leistungstests verwirklicht (I3, 274–280). CMEC war dabei eingebunden in eine Reihe von Leistungsvergleichsstudien auf der nationalen und internationalen Ebene. Hierzu gehörten:

- das *Pan-Canadian Assessment Program* (PCAP), welches aus seiner Reihe aufeinanderfolgender Leistungstestungen bei 13jährigen Schülerinnen und Schülern in Mathematik, Lesen und Naturwissenschaften besteht;
- die PISA-Studie (*Programme for International Student Assessment*) der OECD, die einen Provinzvergleich enthält;
- die *Teacher Education and Development Study – Mathematics* (TEDS-M) (CMEC, 2009b); sowie
- die Studie *Progress in International Reading Literacy Study* (PIRLS).

Der nationale Bildungsbericht stützt sich zu erheblichen Teilen auf Ergebnisse solcher Erhebungen, insbesondere auf PISA und die nationale Schulleistungstestung; die Rückgriffsmöglichkeit ist aber durch die problematische provinzielle Beteiligung begrenzt. (I3, 275–285).

Insgesamt deutet sich also ein eher hohes Maß an autonomer Gestaltungsmacht der Provinzen und Territorien an; gleichzeitig spielt CMEC einerseits eine Vermittlerrolle, andererseits aber auch eine federführende Rolle bei der Entwicklung der Leistungsvergleichsstudien, welche eine zentrale Datengrundlage für die Bildungsberichterstattung und die Indikatorenentwicklung darstellen. Eine explizite Harmonisierung curricularer Standards zwischen den Provinzen und Territorien im Sinne gemeinsamer Mindestanforderungen und eine hierauf bezogene Bildungsberichterstattung existieren demgegenüber nicht.

4.2 Interdependenzen durch die Verfügungsfähigkeit der Provinzen im Hinblick auf die Datenbereitstellung

Grundsätzlich liegt das Recht zur Erhebung benötigter administrativer Daten gemäß *Statistics Act* grundsätzlich bei *Statistics Canada*. Hierbei können administrative Datenerhebungen auch mit Verpflichtungscharakter versehen werden (Sect. 31 i. V. m. Sect. 22, *Statistics Act*). Allerdings ist nicht nur die Durchsetzung, sondern auch die Durchsetzbarkeit dieses Verfügungsrechts begrenzt (I1, 255–256; I1, 280–282). Für den Bildungsbereich stellt sich die Situation konkret folgendermaßen dar: Zwar haben die Provinzen hier die rechtliche Zuständigkeit für das Bildungswesen. Für den Hochschulsektor und ins-

6 Für Mathematik etwa existiert lediglich ein Zusammenschluss einzelner Provinzen bei der Formulierung gemeinsamer Standards (Government of Newfoundland and Labrador, 2000).

besondere für die Berichterstattung der Universitäten variiert ihr tatsächlicher Einflussbereich jedoch aufgrund der Hochschulautonomie von Provinz zu Provinz, so dass etwa einheitliche Erfassungsstandards administrativer Daten für das der Bildungsberichterstattung zu den Indikatoren zur postsekundären Bildung (vgl. Abschnitt 3.1) zugrunde liegende *Postsecondary Information System* (PSIS) z. T. schwer zu bewerkstelligen sind. Neben technischen Kapazitäten der betroffenen Institutionen (I1, 309) können für eine einheitliche Datenerfassung darüber hinaus auch finanzielle Ressourcen eine Rolle spielen: So erfolgt die Finanzierung für diese Datenerhebung auch durch die Provinzen und nicht durch *Statistics Canada*, so dass auch keine monetären Anreize für eine Datenerhebung entlang des Erfassungsstandards der statistischen Bundesbehörde existieren (I1, 287–291).

Für die berufliche Bildung in Einrichtungen des staatlichen Bildungswesens resultiert eine weitere Komplikation für die Erhebung einschlägiger Daten aus kulturell bedingten Sinn-Konstruktionen, die beinhalten, dass das Konzept von *vocational education* zwischen den Provinzen variiert und sowohl die dominante institutionelle Verortung entweder in der allgemeinen Schule oder im College- und Hochschulsektor als auch – als Konsequenz hieraus – der Abdeckungsgrad der auf *vocational education* bezogenen Berichterstattung sich zwischen den Provinzen unterscheidet. Dies manifestiert sich unter anderem darin, dass Teilnahme und Abschlüsse für berufliche Bildung im Sekundarbereich lediglich von der Provinz Québec berichtet werden (I1, 123–135). Demgegenüber ist die Datensammlung in Kanada ansonsten eher institutionenbezogen und weniger programmbezogen und klammert somit die berufliche Bildung auf Ebene der Schulen des Sekundarbereichs weitgehend aus. Dies erklärt, dass trotz einschlägiger Erfassung entsprechender Daten und Indikatoren in internationalen Datensammlungen und der Bildungsberichterstattung der OECD eine entsprechende Indikatorenentwicklung in der kanadischen Bildungsberichterstattung nicht zu finden ist (I1, 151–163).

4.3 Interdependenzen durch die Verfügungsfähigkeit über finanzielle Ressourcen

Ein wesentlicher Anteil der Interdependenzbeziehungen, die die Indikatorenentwicklung im Rahmen des PCEIP beeinflussen, ist im Bereich der Verfügungsfähigkeit über Ressourcen angesiedelt. Wie beschrieben, ist *Statistics Canada* nicht nur für die Ausformulierung der Indikatoren, sondern auch für die Durchführung der der Indikatorenentwicklung zugrunde liegenden Datenerhebungen zuständig. Interdependenzen bezüglich der Verfügbarkeit von finanziellen Ressourcen, die die Möglichkeiten der Indikatorenformulierung beschränken, können mithin nicht nur durch eine Unterfinanzierung des PCEIP selbst entstehen (I1, 188–193), sondern auch durch die Unmöglichkeit, über die Erhebung der für die Indikatorenentwicklung benötigten Daten unabhängig zu entscheiden.

Statistics Canada wird hierbei grundsätzlich aus zwei Arten von Budgets finanziert: Zunächst erhält die Einrichtung ein durch die Bundesregierung bereitgestelltes Kernbudget. Administrative Datensammlungen werden im Regelfall aus dem Kernbudget heraus finanziert (I1, 237–238). Aus diesem Budget heraus wird unter anderem auch das PCEIP finanziert. Restriktionen entstehen hierdurch insbesondere im Hinblick auf die Personalkapazität für die Datenauswertung; dies führt insbesondere dazu, dass auf eine detaillierte Berichterstattung im Jahr 2005 verzichtet wurde (I1, 784–786).

Eine hiervon getrennte Finanzierung erfolgt für die großen Surveys; dies betrifft etwa Erhebungen zum Erwachsenenlernen und zur Postsekundären Bildung und die diesbezüglich jeweils entwickelten Indikatoren (I1, 229; I4, 124). Auch das im Grundsatz als administrative Datenerhebung angelegte *Postsecondary Information System* (PSIS) ist nicht vollständig durch das Kernbudget abgedeckt. Vielmehr stellt das zuständige Bundesministerium hierfür einen erheblichen Teil der erforderlichen Mittel bereit (I1, 248; I4, 128). Dementsprechend wird die Durchführung dieser Studien durch politische Prioritätensetzung des Ministeriums bestimmt (I1, 247–248), wobei das Bundesministerium auch einem Lobbying der Provinzregierungen ausgesetzt ist (I1, 244–245). Die vorangehenden Erläuterungen zur Abhängigkeit der Berichterstattung von den verfügbaren Datenquellen verdeutlicht, dass hier dementsprechend trotz formeller Nichtzuständigkeit für Bildungspolitik ein erheblicher Einfluss der Zentralregierung auf die Bildungsberichterstattung vorliegt.

4.4 Mechanismen der Interdependenzbewältigung

Wie oben beschrieben, können zumindest drei Mechanismen der Interdependenzbewältigung aufgrund von Ressourcenabhängigkeiten unterschieden werden:

- Anpassung als Anerkennung der Ressourcenüberlegenheit anderer Akteure äußert sich im Verzicht auf bisherige Handlungsweisen;
- das Ausweichen auf andere Ressourcen beinhaltet demgegenüber ein Erschließen neuer, zusätzlicher Machtpotenziale, um vorhandene Abhängigkeiten zu verringern;
- das Zusammenlegen von Ressourcen schließlich beinhaltet die Bündelung von Beeinflussungspotenzialen und führt ebenfalls zu einer Verschiebung von Abhängigkeitsverhältnissen (Kussau & Brüsemeister, 2007).

Bezüglich Strategien der Bewältigung von Interdependenz im Rahmen des PCEIP sind insbesondere die ersten zwei Ansätze hervorzuheben. Zunächst ist hier auf die bereits angesprochene Neudefinition der PCEIP-Produkte zu verweisen, insbesondere die Veränderung der ursprünglichen Berichtslinie des PCEIP nach 2007. Die Berichterstattung über die ursprünglichen Indikatoren erfolgt derzeit lediglich in Form von häufigeren Online-Updates bislang bereits verwendeter Kennziffern auf der Homepage von *Statistics Canada* sowie sogenannten *Fact*

Sheets zu den Indikatoren. Während dies Aktualität und Zugänglichkeit verbessert, impliziert es gleichzeitig, dass in der Darstellung die Berichterstattung ganzheitlicher Indikatoren gegenüber dem separaten Ausweis einzelner Kennziffern in den Hintergrund rückt. Diese Anpassungsleistung steht im Zusammenhang mit politischen Anforderungen der Provinzen, die Berichterstattung zeitnäher zu leisten und das ursprünglich rein innerkanadisch angelegte Berichtsprogramm verstärkt mit der internationalen Berichterstattung zu harmonisieren. Als weiteres Produkt im Rahmen des PCEIP ist in diesem Zuge der internationale Bericht zu sehen, in dem eine entsprechende Harmonisierung kanadischer Berichtsdaten und Indikatoren mit dem OECD-Bericht *Education at a Glance* vorgenommen wird, wohingegen das PCEIP ursprünglich explizit nicht am INES-Programm orientiert war, das der OECD-Berichterstattung zugrunde liegt (I1, 27–37).

Darüber hinaus ist im Zuge der Entwicklung einer neuen Datenstrategie Interdependenzbewältigung durch *Ausweichen auf andere Ressourcen* oder *Neugruppierung von Ressourcen* festzustellen: Verstärkt werden die Bildungsminister der Provinzen und Territorien im Rahmen von an den Prioritätsratings⁷ der OECD orientierten Verfahren in die Priorisierung durchzuführender Surveys einbezogen, was einerseits eine Stärkung der Rolle des CESC gegenüber *Statistics Canada* impliziert (I2, 163–177), andererseits aber auch eine besser legitimierte Prioritätensetzung gegenüber der Bundesebene ermöglicht. Außerdem sind in die Prioritätsratings, die zum einen die Elementar- und Sekundar-, zum anderen die postsekundäre Bildung und das Erwachsenenlernen betreffen, auch Stakeholder jenseits der politischen Ebene einbezogen, d. h. beispielsweise Erzieher und Eltern (I2, 85–94).

Schließlich wurde im Sinne eines Zusammenschlusses von Ressourcen auch diskutiert, Problemen mangelnder Einheitlichkeit und hieraus resultierender Vergleichbarkeit der provinziellen Datenerhebungen dadurch zu begegnen, dass in bestimmten Gebieten kleinere Gruppen von Provinzen und Territorien vergleichbare Daten sammeln. Damit ist die Hoffnung verbunden, dass die entsprechenden Erhebungen auf weitere Provinzen und Territorien ausgeweitet werden können (I2, 376–381). Entsprechende weitere Entwicklungen waren zum Befragungszeitraum noch offen.

5. Fazit

Nationale Bildungsberichterstattung als ein mögliches Steuerungsinstrument von Bildungspolitik auf der nationalen Ebene ist im Föderalstaat eingebunden in ein Geflecht von Einflussnahmen, die Umfang und Kontinuität, aber auch die inhaltliche Ausprägung der Bildungsberichterstattung beeinflussen und damit die Verlässlichkeit und Unbestechlichkeit des Instruments für eine *rationa-*

⁷ Die OECD verwendet sogenannte *priority rating exercises* u. a. im Rahmen der TALIS-Studie, um Erhebungsinhalte und ggf. resultierende neue Indikatoren mit den beteiligten Ländern zu definieren (OECD, 2016).

le Bildungssteuerung begrenzen können. Antagonistisch gegenüberstehen können sich insbesondere die Verhinderungsfähigkeit föderaler Einheiten, die auf ihrer formellen Zuständigkeit und den – auch kulturabhängigen – Konstruktionen des jeweiligen Bildungssystems als Grundlage der Datenbereitstellung beharren; Probleme treten diesbezüglich in Kanada unter anderem mit Blick auf die berufliche Bildung auf. Die Frage der Abhängigkeit der Indikatorenentwicklung von kulturabhängigen Konstruktionen erscheint unter anderem mit Blick auf die unterschiedlichen Entwicklungen zwischen östlichen und westlichen Bundesländern auch im Rahmen der deutschen Bildungsberichterstattung relevant.

Darüber hinaus entscheidet die Finanzierungsbereitschaft über die Qualität der Bildungsberichterstattung; dies gilt nicht nur im direkten Sinne, sondern auch indirekt über eine kontinuierliche Bereitstellung und Absicherung von für die Bildungsberichterstattung verwendbaren Studien. Zwischen den Akteuren konsentrierte Datenstrategien scheinen mithin für eine systematische und kontinuierliche Bildungsberichterstattung zentral, gerade im Föderalstaat aber auch besonders anfällig für Interessenlagen der Akteure und tendenziell ein Hebel für eine verstärkte nationale Einflussnahme auf ein föderal organisiertes Bildungswesen.

Über einen eigenständigen Einfluss verfügen neben den koordinierenden Einrichtungen des föderalen Bildungssystems, wie CMEC oder der KMK in Deutschland, hierbei auch die durchführenden Einrichtungen der Bildungsberichterstattung selbst. In diesem Sinne sind die dargestellten Strategien zur Interdependenzbewältigung im Rahmen der kanadischen Bildungsberichterstattung interpretierbar, die unter anderem die Neugruppierung von Ressourcen mit der Zielsetzung einer stakeholderbezogenen, vor allem aber auch politisch gestützten Datenbereitstellung beinhalten. Über analoge, stakeholderbezogene Konsense im Rahmen einer deutschen Bildungsberichterstattung wäre daher im Hinblick auf eine nachhaltige Governancestruktur im Bildungswesen nachzudenken. Eine weitere mögliche Strategie der Interdependenzbewältigung für die bildungsberichterstattenden Einrichtungen selbst liegt hier im Rekurs auf die Berichterstattung inter- und supranationaler Organisationen.

Inwieweit eine zu Kanada analoge Verhinderungsfähigkeit auch für deutsche Bundesländer zum Tragen kommt und wie sich diese evtl. auf Wirksamkeit und Wirkungen der auf evidenzbasierte Bildungspolitik angelegten Bildungsberichterstattung auswirkt, auch mit Blick auf die Nutzerperspektive (Scheerens & Hendriks, 2004) war hier nicht Gegenstand der Analysen und wäre im Weiteren empirisch zu prüfen (vgl. dazu auch Rürup et al. 2016; z. B. auch Maag Merki, 2014). Die sich für Kanada andeutenden Konsensbildungsprozesse deuten die Möglichkeit an, dass die Einflussnahme unterschiedlicher Akteure auf die Indikatorenentwicklung die Reflexivität der Indikatoren auch erhöhen und damit die Rationalität der Bildungssteuerung verbessern kann. Mit der Frage der Wirkungen von indikatorengestützter Bildungsberichterstattung ist jedoch auch eine andere Wirklichkeitsebene angesprochen (Dietrich, 2014), die sich auf die Rekonstruktion von Wirklichkeit im Rahmen der hier analysie-

ren Aushandlungsprozesse und auch auf dadurch möglicherweise induzierte Nebenwirkungen der Bildungsberichterstattung bezieht.

Literatur

- Altrichter, H. & Heinrich, M. (2007). Kategorien der Governance-Analyse und Transformationen der Systemsteuerung in Österreich. In H. Altrichter, T. Brüsemeister & J. Wissinger (Hrsg.). *Educational Governance – Handlungskoordination und Steuerung im Bildungssystem* (S. 55–103). Wiesbaden: VS.
- Avenarius, H., Brauckmann, S. & von Kopp, B. (2007). Verfassungsrechtliche Grundlagen und politisch-administrative Praxis. In Arbeitsgruppe Internationale Vergleichsstudie (Hrsg.), *Schulleistungen und Steuerung des Schulsystems im Bundesstaat. Kanada und Deutschland im Vergleich* (S. 69–104). Münster: Waxmann.
- Benz, A. (2009). *Politik in Mehrebenensystemen*. Wiesbaden: VS.
- Brauckmann, S., Kühne, S., Stäsche, U., Tarazona, M., Weishaupt, H. & Wittmann, E. (2010). Steuerung und Verwaltung des Bildungswesens. In R. S. Jäger, P. Nenniger, H. Petillon, B. Schwarz & B. Wolf (Hrsg.), *Empirische Pädagogik 1990–2010. Eine Bestandsaufnahme der Forschung in der Bundesrepublik Deutschland* (S. 119–160). Landau: Empirische Pädagogik e. V.
- BNA – *British North America Act, 1867*.
- Bromme, R., Prenzel, M. & Jäger, M. (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. *Zeitschrift für Erziehungswissenschaft*, 17(4), 3–54.
- CCL – Canadian Council on Learning (2007). *Canada's PSE performance. File under "m" for missing*. Ottawa, Sept. 20, 2007. Zugriff am 04.11.2009 unter <http://www.ccl-cca.ca/CCL/Newsroom/Releases/20070920ReleaseCanadasPSEperformance.htm>
- CCL – Canadian Council on Learning (2009a). *Annual report 2008–2009*. Ottawa.
- CCL – Canadian Council on Learning (2009b). *New Statistics Canada, CMEC data reporting initiative "a welcome step forward," says Canadian Council on Learning*. Ottawa, Sept. 8, 2009. Zugriff am 04.11.2009 unter <http://www.ccl-cca.ca/CCL/Newsroom/Releases/20090908ReleaseStatisticCanadaCMEC.htm?Language=EN>
- CECSC – Canadian Education Statistics Council (2000). *Education indicators in Canada. Report of the Pan-Canadian Education Indicators Program 1999*. Ottawa.
- CECSC – Canadian Education Statistics Council (2003). *Education indicators in Canada. Report of the Pan-Canadian Education Indicators Program 2003*. Ottawa.
- CECSC – Canadian Education Statistics Council (2005). *Education indicators in Canada. Report of the Pan-Canadian Education Indicators Program 2005*. Ottawa.
- CECSC – Canadian Education Statistics Council (2007a). *Education indicators in Canada. Report of the Pan-Canadian Education Indicators Program 2007*. Ottawa.
- CECSC – Canadian Education Statistics Council (2007b). *Education indicators in Canada. Handbook for the report on the Pan-Canadian Education Indicators Program 2007*. Ottawa.
- CECSC – Canadian Education Statistics Council (2009a). *Education indicators in Canada. An International Perspective 2009*. Ottawa.
- CECSC – Canadian Education Statistics Council (2009b). *Pan-Canadian Education Indicators Program*. Zugriff am 05.11.2009 <http://www.cesc-csce/peicp2003E.html>
- CECSC – Canadian Education Statistics Council (2016). *Education Indicators in Canada. An International Perspective 2015*. Zugriff am 03.10.2016 unter <http://www.stat->

- can.gc.ca/access_acces/alternative_alternatif.action?l=eng&loc=/pub/81-604-x/81-604-x2015001-eng.pdf
- CMEC – Council of Ministers of Education (1993). *Victoria Declaration 1993*. Joint declaration. Sept. 28, 1993. Future directions for The Council of Ministers of Education, Canada (CMEC). Victoria.
- CMEC – Council of Ministers of Education (1997). *Common framework of science learning outcomes, K-12*. Pan-Canadian protocol for collaboration on school curriculum. Toronto.
- CMEC – Council of Ministers of Education (2009a). *About*. Zugriff am 05.11.2009 unter <http://www.cmec.ca/About/Pages/default.aspx>
- CMEC – Council of Ministers of Education (2009b). *Assessment*. Zugriff am 10.11.2009 unter <http://www.cmec.ca/Programs/assessment/Pages/default.aspx>
- CMEC – Council of Ministers of Education (2009c). *Indicators*. Zugriff am 5.11.2009 unter <http://www.cmec.ca/Programs/data/indicators/Pages/default.aspx>
- Constitution Act, 1982*, Canada.
- Dietrich, F. (2014). Objektiv-hermeneutische Governance-Analysen. In Maag-Merki, K., Langer, R. & Altrichter, H. (Hrsg.). *Educational Governance als Forschungsperspektive* (S. 199–227). Wiesbaden: Springer.
- Döbert, H. & Klieme, E. (2009). Indikatoren gestützte Bildungsberichterstattung. In R. Tippelt, & B. Schmidt (Hrsg.). *Handbuch Bildungsforschung* (2., überarb. u. erw. Aufl., S. 317–336). Wiesbaden: VS.
- Government of Newfoundland and Labrador (2000). *Publication of specific learning outcomes for mathematics from entry to grade eight*. News release. 11. Sept. 2000. Halifax.
- Jaafar, S. B. & Earl, L. (2008). Comparing performance-based accountability models. A Canadian example. *Canadian Journal of Education*, 31, 697–726. Zugriff am 06.07.2018 unter <http://files.eric.ed.gov/fulltext/EJ809266.pdf>
- Kussau, J. & Brüsemeister, T. (2007). Educational Governance. Zur Analyse der Handlungskoordination im Mehrebenensystem der Schule. In H. Altrichter, T. Brüsemeister & J. Wissinger (Hrsg.), *Educational Governance – Handlungskoordination und Steuerung im Bildungssystem* (S. 15–54). Wiesbaden: VS.
- Maag Merki, K. (2014). Herausforderungen, Möglichkeiten und Grenzen am Beispiel der Analyse der Wirksamkeit der Einführung zentraler Abiturprüfungen. In K. Maag-Merki, R. Langer & H. Altrichter (Hrsg.), *Educational Governance als Forschungsperspektive* (S. 51–85). Wiesbaden: Springer.
- Mayntz, R. (2005). Governance Theory als fortentwickelte Steuerungstheorie? In G. G. Schuppert (Hrsg.), *Governance-Forschung. Vergewisserung über Stand und Entwicklungslinien* (S. 11–20). Baden-Baden: Nomos.
- Mayring, P. (2002). *Qualitative Sozialforschung. Anleitung zu qualitativem Denken* (5., überarb. u. neu ausg. Aufl.). Weinheim: Beltz.
- Niedlich, S. & Brüsemeister, T. (2016). Bildungsmonitoring zwischen Berichterstattung und Steuerungsanspruch – Entwicklungslinien und akteurtheoretische Implikationen. In Arbeitsgruppe „Lernen vor Ort“ (Hrsg.), *Kommunales Bildungsmonitoring als sozialer Prozess* (S. 75–96). Wiesbaden: VS.
- OECD – Organization for Economic Cooperation and Development (2006). *New OECD Activity on recognition of non-formal and informal learning*. Paris.
- OECD – Organization for Economic Cooperation and Development (2009). *Education at a Glance 2009*. Paris.
- OECD – Organization for Economic Cooperation and Development (2016). *TALIS 2018 survey*. Paris. Zugriff am 26.06.2017 unter https://www.oecd.org/edu/school/TALIS_2018_brochure_ENG.pdf
- Rürup, M., Fuchs, H.-W. & Weishaupt, H. (2016). Bildungsberichterstattung – Bildungsmonitoring. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulwesen*. (2. Aufl., S. 411–437). Wiesbaden: Springer VS.

- Scheerens, J., & Hendriks, M. (2004). Benchmarking the quality of education. *European Educational Research Journal*, 3, 101–399.
- Schuppert, G. F. (2008). Governance durch Wissen. In G. F. Schuppert & A. Voßkuhle (Hrsg.), *Governance von und durch Wissen* (S. 259–303). Baden-Baden: Nomos.
- Statistics Act, 1985, zuletzt geändert 2005.
- Statistics Canada (2009a). *Adult Education and Training Survey*. Zuletzt am 23.07.2009 veränd. Vers.; Zugriff am 30.06.2011 unter <http://www.statcan.gc.ca/dli-ild/data-donnees/ftp/aets-eeefa-eng.htm>
- Statistics Canada (2009b). *National Graduates Survey*. Toronto. Zugriff am 09.12.2009 unter <http://www.statcan.gc.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SDDS=3701&lang=en&db=imdb&adm=8&dis=2>
- Statistics Canada (2009c). *Education Indicators in Canada. Report of the Canadian Education Indicators Program*. Product Main Page. Zugriff am 09.11.2009 unter <http://www.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=81-582-X&lang=eng>
- Statistics Canada (2009d). *Survey of Financial Statistics of Private Elementary and Secondary Schools*. Toronto. Zugriff am 06.01.2009 unter <http://www.statcan.gc.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SDDS=3120&lang=en&db=imdb&adm=8&dis=2>
- The Prime Minister of Canada (2018). The team. Ottawa. Zugriff am 15.06.2018 unter <https://pm.gc.ca/eng/cabinet>
- Tsebelis, G. (2002). *Veto players. How political institutions work*. Princeton: Princeton University.
- Wittmann, E. (2011). Indikatorenentwicklung im Rahmen der nationalen Bildungsberichterstattung in Kanada. In M. Baethge et al. (Hrsg.), *Vertiefende Studien zu ausgewählten Aspekten der Indikatorenentwicklung* für den Bildungsbericht (S. 229–282). Bonn: BMBF.
- Wittmann, E. & Hardt, K. (2011). Indikatorenentwicklung in den „Schlüsselzahlen zum Bildungswesen in Europa“. In M. Baethge et al. (Hrsg.) *Vertiefende Studien zu ausgewählten Aspekten der Indikatorenentwicklung* für den Bildungsbericht (S. 283–338). Bonn: BMBF.
- Wurster, S., Richter, D. & Lenski, A.E. (2017). Datenbasierte Unterrichtsentwicklung und ihr Zusammenhang zur Schülerleistung. *Zeitschrift für Erziehungswissenschaft*, 20(4), 628–650.
- Voßkuhle, A. (2005). Sachverständige Beratung des Staates. In J. Isensee & P. Kirchhof (Hrsg.), *HStR*. Bd. III (§ 43 (Rdnr.)) (3. Aufl.). Heidelberg: Müller.

Marianne Schübach, Wim Nieuwenboom, Lukas Frei &
Benjamin von Allmen

Offene Tagesschulen als Mittel um Mathematikleistungen zu fördern und sozial oder kulturell bedingte Disparitäten zu verringern?

Zusammenfassung

Tagesschulen werden in der Schweiz als Mittel diskutiert, um Schulleistungen zu fördern und darauf bezogene Herkunftseffekte aufgrund des sozioökonomischen Status oder der Erstsprache zu reduzieren. Der vorliegende Beitrag prüft diese Annahme anhand einer Stichprobe von 1806 Schülerinnen und Schülern aus 53 offenen Tagesschulen in der Deutschschweiz, die im Rahmen einer Längsschnittstudie zwischen Ende der 1. und Ende der 2. Jahrgangsstufe untersucht wurden, wobei nur ein Teil der Schülerschaft die Tagesschulangebote dauerhaft nutzte. Dabei zeigte sich kein genereller Effekt der dauerhaften Angebotsnutzung auf die Entwicklung von Mathematikleistungen und dies unabhängig von der Angebotsqualität. Ende der 1. Jahrgangsstufe wurden Herkunftseffekte festgestellt, wobei sich bei einem höheren sozioökonomischen Status oder bei Deutsch/Schweizerdeutsch als Erstsprache höhere Mathematikleistungen zeigten. Durch eine dauerhafte Nutzung des Tagesschulangebots konnten die sozialen Disparitäten reduziert werden. Hinsichtlich Erstsprache zeigte sich jedoch ein gegenteiliger Effekt, die dauerhafte Teilnahme am Angebot führte zu einer Zunahme der Unterschiede.

Schlüsselwörter

Ganztagsschule, Extended Education, Erstsprache, Mathematikleistung, sozioökonomischer Status

Prof. Dr. Marianne Schübach · Dr. Wim Nieuwenboom (Korrespondenzautor), Lehrstuhl für Grundschulpädagogik und -didaktik, Otto-Friedrich-Universität Bamberg, Markusstraße 8a, 96047 Bamberg, Deutschland.

E-Mail: marianne.schuepbach@uni-bamberg.de
wim.nieuwenboom@uni-bamberg.de

M. Sc. Frei, Lukas · M. Sc. Benjamin von Allmen, Institut für Psychologie, Universität Bern, Fabrikstraße 8, 3012 Bern, Schweiz

E-Mail: lukas.frei@edu.unibe.ch
benjamin.vonallmen@edu.unibe.ch

Open all-day schools as a means to promote mathematical skills and to reduce social or cultural disparities?

Abstract

All-day schools in Switzerland have been discussed as a possible means to promote school achievement and to reduce correspondent disparities based on socioeconomic status or first language. This contribution explores those assumptions based on a sample of 1806 students from 53 open all-day schools located in the German-speaking part of Switzerland, which have been examined as part of a longitudinal study between the end of first and the end of second grade, with some students continuously utilizing all-day school offerings while others did not. Continuous utilization of all-day school offerings did not affect the development of mathematical skills, regardless of the quality of all-day school offerings. At the end of first grade, social and cultural disparities were found, with a higher socioeconomic status and German/Swiss-German as first language being linked to higher mathematical skills. Continuous utilization of all-day school offerings reduced disparities based on socioeconomic status. With regard to first language, opposite results emerged, with continuous utilization resulting in increased disparities.

Keywords

all-day school, extended education, first language, mathematical skills, socioeconomic status

1. Einleitung

Die PISA-Ergebnisse von 2012 zeigen, dass die Mathematikleistungen bei Abschluss der obligatorischen Schulzeit in der Schweiz zwischen 2003 und 2012 auf hohem Niveau stabil geblieben sind (vgl. Nidegger, 2014). Dies ist zwar erfreulich, vergleichbar mit Deutschland und Österreich lässt sich jedoch auch ein stabiler Zusammenhang zwischen sozialer Herkunft, Erstsprache und der Mathematikleistung nachweisen (vgl. OECD, 2014): Schülerinnen und Schüler mit niedrigem sozioökonomischem Status und solche, die zuhause nicht die Unterrichtssprache sprechen (Deutsch als Zweitsprache), schnitten bei beiden Erhebungswellen deutlich schlechter ab. Dieser Befund deckt sich mit weiteren Untersuchungen in der Deutschschweiz (vgl. Übersicht von Coradi Vellacott, 2007). Die Studie von Moser und Bayer (2010) zeigt zudem, dass Leistungsrückstände in Mathematik bei Schülerinnen und Schülern mit Deutsch als Zweitsprache und ungünstiger sozialer Herkunft bereits bei Schuleintritt bestehen und bis zur 3. Klasse einigermaßen stabil bleiben. Somit ist anzunehmen, dass nicht ausschließ-

lich sekundäre, sondern auch primäre Herkunftseffekte (vgl. Boudon, 1974) einen Einfluss auf die Mathematikleistung der Schülerinnen und Schüler haben.

Im Nachgang der ersten Ergebnisse von PISA 2000, die insbesondere bezüglich Leseleistungen und Herkunftseffekten einen Handlungsbedarf aufzeigten, wurde die Einführung von Tagesschulen durch die nationale PISA-Projektleitung als mögliche Reaktion diskutiert (vgl. Coradi Vellacott, Hollenweger, Nicolet & Wolter, 2003) und in einigen der 26 Schweizer Kantone auch umgesetzt. Vergleichbar mit den Ganztagschulen in Deutschland werden darunter Schulen verstanden, die an mehreren Tagen pro Woche über den traditionellen Unterricht hinaus ganztägige Bildungs- und Betreuungsangebote (inklusive Mittagsverpflegung) anbieten (vgl. EDK, 2013). Diese Angebote beinhalten üblicherweise das Mittagessen, ein beaufsichtigtes Programm am Nachmittag – freie und geleitete Aktivitäten, insbesondere auch Hausaufgabenbetreuung – und etwas seltener eine Betreuung vor Schulbeginn. Dabei sind geleitete Angebote in Form von Arbeitsgruppen (AG) eher die Ausnahme. Bei freien Aktivitäten, die während der Phase des Freispiels stattfinden, können die Kinder meistens zwischen einer Vielzahl an pädagogischen Materialien auswählen und damit arbeiten und spielen. Währenddessen der Unterricht und die Angebote an einigen Tagesschulen mit Ausnahme sogenannter Auffangzeiten für alle Schülerinnen und Schüler verpflichtend sind (gebundene Tagesschulen), bieten die meisten Tagesschulen neben dem Unterricht modular aufgebaute Angebote an, die freiwillig genutzt werden können (offene Tagesschule) (vgl. Schüpbach 2010). Tagesschulen in der Schweiz beschränken sich bis heute weitgehend auf den Primarbereich. Von Tagesschulen erhofft man sich einerseits, dass sie den Einfluss der sozialen und kulturellen Herkunft mittels sozialer Integration und einer Erweiterung des sozialen Netzwerks reduzieren können, durch mehr Kontakt und Austausch der diesbezüglich benachteiligten Schülerinnen und Schüler mit Kolleginnen und Kollegen von unterschiedlicher Herkunft (vgl. Coradi Vellacott et al., 2003). Andererseits wird vermutet, dass sich der pädagogisch und zeitlich erweiterte Rahmen auch direkt auf die Schulleistungen der teilnehmenden Schülerinnen und Schüler auswirkt (vgl. Aeberli & Binder, 2005). Aus einer theoretischen Perspektive kann ausgehend von der *time-on-task-Hypothese* von Hopf (2005) davon ausgegangen werden, dass sich die Nutzung eines Tagesschulangebots positiv auf die Schulleistung, insbesondere auf die Mathematikleistung auswirkt, da sich Schülerinnen und Schüler in einem Tagesschulangebot länger in einem schulischen Umfeld bewegen. Das heißt, Schülerinnen und Schüler, die ein Tagesschulangebot besuchen, sind länger mit der Bildungssprache im Kontakt, was für den Kompetenzerwerb in allen Fächern im Besonderen auch im Fach Mathematik relevant ist (Paetsch, Felbrich & Stanat, 2015). Zudem setzen sich die Schülerinnen und Schüler im Rahmen der Aktivitäten im Tagesschulangebot mit mathematischem und schlussfolgerndem Denken auseinander. So beschäftigen sie sich etwa mit dem Spielen von Schach, Monopoly, Eile mit Weile oder mit spezifischen Mathematikspielen (vgl. Schüpbach, Rohrbach-Nussbaum & Grütter, 2017). Ein zentrales und tragendes Element des Tagesschulangebots in der Deutschschweiz stellt im Weiteren die

Hausaufgabenbetreuung dar (vgl. Frei, Schüpbach, von Allmen & Nieuwenboom, 2016; Schüpbach et al., 2017). Damit verbunden sind Erwartungen an eine bessere Förderung aller Schülerinnen und Schüler und im Besonderen solcher mit unvorteilhafter familialer Herkunft. Bezüglich letzterem erhofft man sich eine Verringerung sozial und kulturell bedingter Disparitäten. Ob Tagesschulen diesen Erwartungen insbesondere im Hinblick auf die Leistungsentwicklung in Mathematik gerecht werden, wurde allerdings bis jetzt im deutschsprachigen Raum und insbesondere in der Schweiz kaum untersucht.

2. Theoretischer Hintergrund und empirischer Forschungsstand

2.1 Theoretischer Hintergrund

Theoretische Wirkungsmodelle zu außerunterrichtlichen Angeboten an Tagesschulen oder Ganztagschulen offenbaren eine Vielfalt an Dimensionen und Strukturen, die sich zum Teil auch in den verwendeten Begrifflichkeiten und deren Bedeutung stark unterscheiden (vgl. Fischer et al., 2011; Holtappels, 2009; Schüpbach, 2010). Gemeinsam ist den meisten Modellen, dass im Sinne des *CIPO-Modells* (Context Input Process Output oder auch CIPP-Modell; vgl. Stofflebeam, 1972) zwischen einer Input- bzw. Kontextebene, einer Prozessebene und einer Output- bzw. Wirkungsebene unterschieden wird. Häufig wird auch dem Umstand Rechnung getragen, dass sich Qualitätsmerkmale auf das Bildungssystem, auf die Einzelschule, auf Klassen oder auf einzelne Kinder beziehen können. Ausgehend von Angebots-Nutzungs-Modellen (z. B. Helmke, 2003) berücksichtigen neuere Modelle zudem stärker, dass die Wirkung der Angebote auch von der Nutzung der einzelnen Schülerinnen und Schüler abhängt. Exemplarisch dafür ist das Modell der Bildungsqualität außerunterrichtlicher Angebote in der Ganztagschule von Stecher, Radisch, Fischer und Klieme (2007), welches für erzieherische Wirkungen sowie Wirkungen auf Leistung und Schulerfolg einen theoretischen Rahmen bildet. Nebst Schulmerkmalen, externem Kontext und individuellem wie auch familiärem Kontext und somit u. a. soziale und kulturelle Herkunft stellt das Modell insbesondere die Prozessqualität der außerunterrichtlichen Angebote und deren Nutzung durch die einzelnen Schülerinnen und Schüler in den Mittelpunkt. Zusätzlich zur Frage nach einer generellen Wirkung der Nutzung von Tagesschulangeboten auf die Schulleistung gilt es entsprechend, die Rolle von Angebotsqualität und die Art der Nutzung zu berücksichtigen. Zudem geht das Modell von der Annahme aus, dass die Wirkung der Angebotsnutzung in Abhängigkeit der familialen Herkunft variiert. So wurde wiederholt die Annahme geäußert, dass insbesondere Schülerinnen und Schüler mit einem erhöhten Risiko für schulischen Misserfolg von außerunterrichtlichen Bildungs- und Betreuungsangeboten profitieren können, indem kulturelle oder soziale Benachteiligungen in der familialen Umwelt kom-

pensiert werden (vgl. Coradi Vellacott, 2007; Miller, 2003). Im Folgenden wird auf den bisherigen Forschungsstand zu diesen Aspekten eingegangen.

2.2 Wirkung der Nutzung des Tagesschulangebots auf die Schulleistung

Die einzige Studie, die Tagesschulen in der Schweiz und zudem spezifisch Mathematikleistungen zum Gegenstand hatte, stellt die *EduCare-Studie* dar (vgl. Schüpbach 2010). Anhand einer Stichprobe von 519 Schülerinnen und Schülern wurde in dieser Studie die Wirkung der Nutzung von offenen und gebundenen Tagesschulangeboten auf die Mathematik- und Leseleistung untersucht. Einzig bei einer intensiven Angebotsnutzung über mindestens 7,5 h pro Woche zeigte sich eine positive Wirkung auf die Entwicklung, sowohl der Mathematik- als auch der Leseleistung (vgl. Schüpbach 2014a; Schüpbach, Herzog & Ignaczewska, 2013).

Weitere Befunde aus dem deutschsprachigen Raum, wo in den letzten Jahren in diesem Zusammenhang vergleichbare Entwicklungen stattgefunden haben, liegen aus Deutschland vor: In einer Sekundäranalyse von repräsentativen Daten der internationalen Grundschul-Leseuntersuchung (IGLU) untersuchten Radisch, Klieme und Bos (2006) für Deutschland, ob Schülerinnen und Schüler in der 4. Jahrgangsstufe ein besseres Leseverständnis aufweisen, wenn sie eine Schule mit ganztägigen Angeboten besuchen. Die Querschnittanalyse zu den 184 untersuchten Schulen zeigte diesbezüglich allerdings keine signifikanten Unterschiede. Weiter kann eine Analyse von Strietholt, Manitius, Berkemeyer und Bos (2015) erwähnt werden, welche neben IGLU-Daten auch Daten der *Trends in International Mathematical and Science Study* (TIMSS) und *Programm for International Student Assessment* (PISA) im Hinblick auf die Ganztagschule untersuchte. Strietholt et al. (2015) stellten fest, dass in Deutschland keine Leistungsunterschiede zwischen Ganztags- und Halbtagschulen vorliegen. Ähnliche Befunde fanden Reinders et al. (2011) in einer Teilstudie des Projekts *Ganztagschule und Integration von Migrantinnen*, in der ein Einfluss der Schulform auf die Entwicklung der Lesegeschwindigkeit zwischen der 1. und 2. Jahrgangsstufe untersucht wurde. In der Stichprobe von 755 Schülerinnen und Schülern aus Bayern und Hamburg konnte kein Vorteil von Schülerinnen und Schülern an offenen oder (teil-)gebundenen Ganztagschulen gegenüber Schülerinnen und Schülern an Halbtagschulen gefunden werden. Weitere Ergebnisse zum Einfluss von Ganztagschulen in Deutschland auf die Entwicklung der Leseleistung stammen von der Langzeitstudie zur *Ganztagsorganisation im Grundschulbereich* (GO), welche die Entwicklung von 604 Schülerinnen und Schülern zwischen Ende 1. bis Ende 2. Jahrgangsstufe untersuchte (vgl. Bellin und Tamke, 2010). Dabei zeigte sich ein leicht stärkerer Leistungszuwachs für Schülerinnen und Schüler, welche den offenen Ganztagsbetrieb nutzten. Wurden zusätzlich die kognitive Leistungsfähigkeit und das Vorwissen der Schülerinnen und Schüler berücksichtigt, wurde dieser Effekt allerdings knapp nicht mehr signifikant.

Während die vorherigen Studien den Primarschulbereich fokussierten, untersuchte die *Studie zur Entwicklung von Ganztagsschulen* (StEG) den Einfluss von offenen und gebundenen Ganztagsschulen auf Schulnoten in der Sekundarstufe (vgl. Kuhn und Fischer, 2011). In dieser Stichprobe von 3047 Schülerinnen und Schülern, welche von der 5. bis zur 9. Jahrgangsstufe untersucht wurden, zeigte sich kein Einfluss der Angebotsnutzung auf die Entwicklung der Schulnoten in Mathematik, Deutsch und der ersten Fremdsprache. Fischer, Sauerwein, Theis und Wolgast (2016) untersuchten eine StEG-Teilstichprobe von 2105 Schülerinnen und Schülern, die aus Ganztagsschulen stammten, in welchen Leseförderangebote im Ganztagsschulangebot freiwillig zum Angebot stehen. Es zeigten sich keine Leseleistungsunterschiede zwischen Schülerinnen und Schülern mit Nutzung von Leseförderangeboten und Schülerinnen und Schülern ohne Nutzung von Leseförderangeboten im Verlaufe der ersten Schuljahreshälfte in der fünften Klasse (Fischer et al., 2016). Ebenfalls anhand einer StEG-Teilstichprobe (2149 Schülerinnen und Schüler) stellten Lossen, Tillmann, Holtappels, Rollett und Hannemann (2016) keine nachweisbaren Effekte der Teilnahme an naturwissenschaftlichen Ganztagsangeboten auf die Kompetenzentwicklung in Naturwissenschaften von Anfang 3. bis Ende 4. Jahrgangsstufe fest, selbst wenn Intensität und Kontinuität der Teilnahme berücksichtigt wurden.

Die bisherigen Studien aus dem deutschsprachigen Raum konnten somit – mit Ausnahme von Schübach (2014a) und Schübach (2013) – keinen Effekt der Teilnahme am Tagesschul- oder Ganztagsangebot auf die Schulleistung ausmachen. Da die Verbreitung von Tages- und Ganztagsschulen primär in den letzten 15 Jahren stark zunahm und das außerunterrichtliche Angebot laufend weiterentwickelt wurde, ist der Forschungsstand auf wenige Studien beschränkt. Daher werden zusätzlich die umfangreichen Forschungsergebnisse zu sogenannten *afterschool* Programmen, welche primär im US-amerikanischen Raum verbreitet sind, herangezogen. Bisherige Metaanalysen zeigten einen positiven Effekt von *afterschool* Programmen auf die Entwicklung der Schulleistungen (vgl. Durlak, Weissberg & Pachan, 2010) und spezifisch auf Mathematikleistungen (vgl. Lauer et al., 2006). Zugleich wird aber auch konstatiert, dass die Mehrzahl der Studien keine Effekte findet (vgl. Roth, Malone & Brooks-Gunn, 2010). Als mögliche Erklärung für diese unterschiedlichen Befunde wird insbesondere die Angebotsqualität diskutiert (vgl. Abschnitt 2.4).

2.3 Wirkung der Nutzung des Tagesschulangebots auf soziale und kulturelle Disparitäten bei der Schulleistung

Noch dürftiger ist die Befundlage im Hinblick auf die Frage, ob Tagesschulangebote herkunftsbedingte Ungleichheiten bei Schulleistungen reduzieren können. Bei der EduCare-Studie aus der Schweiz zeigte sich, dass Herkunftseffekte aufgrund des sozioökonomischen Status oder der familialen Entwicklungsförderung auf die Entwicklung der Mathematik- und Leseleistung selbst durch eine intensi-

ve Angebotsnutzung (mindestens 7,5 h pro Woche) nicht kompensiert werden konnten, sondern teilweise sogar zunahmen (vgl. Schüpbach 2014a; Schüpbach et al. 2013). Für Deutschland wurde der Einfluss ganztägiger Angebote auf soziale oder kulturelle Disparitäten bei der Schulleistung anhand der IGLU-Daten untersucht (vgl. Radisch et al. 2006): Unabhängig davon, ob ganztägige Angebote an einer Schule vorhanden waren oder nicht, zeigten der durchschnittliche sozioökonomische Status und der Migrationsanteil auf Schulebene einen vergleichbaren Einfluss auf das Leseverständnis. Auch die Untersuchungen von Strietholt et al. (2015) zeigen, dass an Ganztagschulen keine geringere Bildungsungleichheit in Bezug auf den Migrationshintergrund und den sozioökonomischen Status vorliegt als an Halbtagschulen. Im Rahmen der GO-Studie wurde ein möglicher Effekt der Angebotsteilnahme auf kulturelle Disparitäten hinsichtlich der Leseleistung zwar nicht direkt untersucht, aber Bellin und Tamke (2010) halten fest, dass „Schüler nichtdeutscher Herkunftssprache auch bei Teilnahme am Ganztagsbetrieb den Leistungsvorsprung von Schülern deutscher Herkunftssprache“ nicht einholen (S. 107). Schliesslich berichten auch Lossen et al. (2016), dass soziale Disparitäten in der naturwissenschaftlichen Kompetenz nicht durch den Besuch von domänenspezifischen Ganztagsangeboten kompensiert werden konnten. Da viele US-amerikanischen *afterschool* Programme spezifisch für schulisch benachteiligte Schülerinnen und Schüler konzipiert sind, stellt sich die Frage, ob in Abhängigkeit der sozialen und kulturellen Herkunft unterschiedliche Effekte zu erwarten sind, häufig gar nicht erst. Es gibt aber zumindest Hinweise, dass Schülerinnen und Schüler mit niedrigerem Ausgangsniveau oder mit niedrigerem sozioökonomischem Status stärker von solchen Programmen profitieren können (vgl. McComb und Scott-Little, 2003; Miller, 2003).

2.4 Die Rolle der Angebotsqualität

Wie die beiden vorangehenden Abschnitte zeigen, konnten die Erwartungen an Tages- und Ganztagschulen hinsichtlich einer Förderung von Schulleistungen und einer Reduktion der darauf bezogenen Herkunftseffekte zumindest für den deutschsprachigen Raum kaum bestätigt werden und bezüglich *afterschool* Programmen zeigten sich heterogene Befunde. Einige der Studien untersuchten daher zusätzlich, ob allenfalls bei einer höheren Qualität der Angebote positive Effekte auftreten: Nebst den zuvor berichteten Ergebnissen wurden im Rahmen der EduCare-Studie mit einer Teilstichprobe von 34 Schülerinnen und Schülern, welche die Tagesschulangebote intensiver und somit mehr mindestens 7,5 h pro Woche nutzten, zusätzliche Analysen gerechnet (vgl. Schüpbach 2014b). Dabei zeigte eine höhere Angebotsqualität, welche mittels standardisierter Beobachtung eingeschätzt wurde, einen positiven Einfluss auf die Entwicklung der Mathematikleistung (vgl. Schüpbach 2014b). Bei der StEG-Studie aus Deutschland ergaben sich in der ersten Erhebungswelle Hinweise auf eine vorteilhafte Wirkung einer höheren Prozessqualität (vgl. Kuhn & Fischer, 2011). Da diese erst im 9. Schuljahr einen

signifikanten Zusammenhang zur Entwicklung der Schulnoten von der 5. bis zur 9. Jahrgangsstufe zeigte und es sich um eine Einschätzung auf Ebene der einzelnen Schülerinnen und Schüler handelt, ist die Kausalität allerdings fraglich. Bei der zweiten Erhebungswelle zeigte die Angebotsqualität aus Schülersicht keinen Effekt auf die Kompetenzentwicklung in Naturwissenschaften (Lossen et al., 2016). In der bereits erwähnten Meta-Analyse von Durlak et al. (2010) konnten positive Wirkungen auf die Schulleistung nur für jene *afterschool* Programme nachgewiesen werden, die über spezifische Ziele mit entsprechenden Programmkomponenten verfügten und in denen aktive Lernformen angeboten wurden, welche miteinander verknüpft und koordiniert waren.

2.5 Zusammenfassung des Forschungsstandes und Fazit

Während die US-amerikanische Forschung zu *afterschool* Programmen aufzeigt, dass außerunterrichtliche Bildungs- und Betreuungsangebote durchaus die Schulleistungen verbessern können, fallen die entsprechenden Befunde zu Tages- und Ganztagschulen im deutschsprachigen Raum eher ernüchternd aus. Auch die Hinweise, dass *afterschool* Programme insbesondere bei Schülerinnen und Schülern mit herkunftsbedingten Nachteilen wirksam sein dürften, konnten für den deutschsprachigen Raum bisher nicht bestätigt werden. Einzelne Befunde deuten aber darauf hin, dass zumindest bei einer hohen Angebotsqualität positive Wirkungen zu erwarten sind. Aufgrund der geringen Anzahl Studien und derer Heterogenität bezüglich Alter der Schülerschaft, Form der Angebote, Operationalisierung der zentralen Variablen und Stichprobengrösse sind allerdings Fragen nach einem Effekt der Angebotsnutzung auf die Mathematikleistung und auf soziale oder kulturelle Disparitäten, als auch nach der Rolle der Angebotsqualität, nicht abschliessend geklärt.

3. Fragestellungen

Der vorliegende Beitrag fokussiert die Entwicklung der Mathematikleistung von Ende 1. bis Ende 2. Jahrgangsstufe an offenen Tagesschulen in der Deutschschweiz, wobei standardisierte Leistungstests eingesetzt werden. Ausgehend von den dargelegten Erwartungen an Tagesschulen, den theoretischen Modellen zu außerunterrichtlichen Angeboten und angesichts der erst marginal vorhandenen Befunde zu Ganztags- und insbesondere Tagesschulen, werden folgende Fragestellungen untersucht: (1) Wird die Entwicklung der Mathematikleistung durch die Nutzung des Tagesschulangebots positiv beeinflusst? (2) Sind bei den Mathematikleistungen Ende der ersten Jahrgangsstufe soziale und kulturelle Disparitäten, basierend auf (a) sozioökonomischem Status oder (b) Erstsprache, vorhanden? (3) Werden die (a) sozialen und (b) kulturellen Disparitäten durch die Nutzung des

Tagesschulangebots verringert? (4) Wirkt sich eine hohe Qualität des Angebots positiv auf den Einfluss der Angebotsnutzung hinsichtlich der Entwicklung von Mathematikleistungen aus?

4. Methode

4.1 Untersuchungsdesign und Stichprobe

Die Daten wurden im Rahmen der Längsschnittstudie *EduCare-TaSe – Tagesschulen und Schulerfolg?* erhoben, die durch den Schweizerischen Nationalfonds finanziert wird. EduCare-TaSe untersucht Kinder im Grundschulalter (Ende 1. bis Ende 2. Jahrgangsstufe) an offenen Tagesschulen, d. h. ein Teil der Schülerinnen und Schüler nutzt neben dem Unterricht freiwillig zusätzlich das Tagesschulangebot. Die Grundgesamtheit umfasst – ausgehend von der Definition der kantonalen Erziehungsdirektoren (vgl. EDK, 2013) – offene Tagesschulen in der Deutschschweiz, die somit nebst dem Unterricht ein modular aufgebautes und freiwilliges außerunterrichtliches Bildungs- und Betreuungsangebot anbieten, das an mindestens drei Tagen die Woche jeweils am Mittag sowie am Nachmittag zur Verfügung steht. Aus forschungsökonomischen Gründen wurden bei den Schulen mindestens zwei parallele Erstjahrgangsklassen vorausgesetzt. Das heißt, kleine ländliche Schulen wurden damit ausgeschlossen. In der vorliegenden Studie wurde eine Vollerhebung angestrebt. Von den insgesamt 251 angefragten und genannten Definition entsprechenden offenen Tagesschulen erklärten sich 53 Schulen mit 1990 Schülerinnen und Schülern aus 13 Deutschschweizer Kantonen bereit, an der Studie teilzunehmen. Um Informationen zur familialen Herkunft der Schülerinnen und Schüler zu erhalten, wurden in einem Folgeschritt deren Eltern angefragt und um die Angabe relevanter Hintergrundvariablen gebeten. Aufgrund der Selbstselektion auf Ebene der Schulen kann nicht von einer repräsentativen Stichprobe ausgegangen werden.

Die Datenerhebung fand zwischen 2014 und 2015 statt. Gegen Ende der 1. Jahrgangsstufe, Mitte 2. Jahrgangsstufe sowie Ende 2. Jahrgangsstufe wurde die Mathematikleistung der Schülerinnen und Schüler erhoben. Kognitive Grundfähigkeiten, Geschlecht und Erstsprache wurden ebenfalls direkt bei den Schülerinnen und Schülern erhoben. Der sozioökonomische Status wurde über die Eltern erfragt. Angaben zur Nutzung der Tagesschulangebote im 1. und 2. Schuljahr wurden bei den Leiterinnen und Leitern der Angebote eingeholt.

Fehlende Werte bei den abhängigen und den unabhängigen Variablen wurden mittels FIML (*full information maximum likelihood*) geschätzt. Einzig die fehlenden Werte bezüglich der Angebotsnutzung konnten nicht geschätzt werden, was zu einem Ausschluss von 184 Fällen führte¹. Dem vorliegenden Beitrag liegt somit

1 Es handelt sich hier um ein statistisch-methodisches Problem, das entsteht, weil die Stärke des Zusammenhangs zwischen Angebotsnutzung und Mathematikleistung zwischen den Schulen variieren sollte. Dadurch kann die FIML-Methode hier nicht angewendet

eine Stichprobe von 1806 Schülerinnen und Schülern aus 53 Tagesschulen zugrunde.

4.2 Instrumente und Skalen

4.2.1 Abhängige Variable

Mathematikleistung: Zu allen drei Messzeitpunkten wurde die Mathematikleistung der Schülerinnen und Schüler mit der Skala *Rechenoperationen* des Heidelberger Rechentests (HRT 1–4, vgl. Haffner, Baro, Parzer & Resch, 2005) gemessen. Dies, um grundlegende Rechenprozesse zu prüfen, „die für die Entwicklung komplexerer mathematischer Kompetenzen eine Voraussetzung darstellen und im Bereich der Mathematik ständig benötigt werden“ (Haffner, Baro, Parzer & Resch, 2005, S. 15). Dazu wurden den Schülerinnen und Schülern Rechenaufgaben aus den Bereichen *Addition*, *Subtraktion*, *Ergänzungsaufgaben* und *Grösser-Kleiner-Vergleiche* mit jeweils 40 Items vorgegeben. Die Testungen erfolgten im Klassenverband und wurden von externen Hilfskräften durchgeführt, die im Vorfeld an einer halbtägigen Schulung teilnahmen. Dem Manual folgend wurde für jeden Bereich ein Mittelwert gebildet, der dann im Gesamtwert verrechnet wurde, wobei die interne Konsistenz zu allen drei Messzeitpunkten gute Werte erzielte ($\alpha = .83 - .86$). Der HRT 1–4 kann ab Ende 1. Klasse bis Ende Grundschulzeit eingesetzt werden. Gemäß Haffner et al. (2005) weist die Skala *Rechenoperationen* im Abstand von 1 bis 2 Wochen eine hohe Messzuverlässigkeit auf ($r_{tt} = .93$) und eine hohe Korrelation mit Mathematiknoten ($r = -.67$). In der vorliegenden Untersuchung wurde mit T-Werten gerechnet (*Min.* = 0, *Max.* = 100; *M* = 50). Deskriptive Angaben zur Stichprobe können der Tabelle 1 entnommen werden.

4.2.2 Unabhängige Variablen

Dauerhafte Angebotsnutzung: In offenen Tagesschulen ist es üblich, dass die Eltern ihre Kinder jeweils zu Beginn eines Schuljahres für die regelmäßige Nutzung bestimmter Angebotsmodule verpflichten. Angaben zur Angebotsnutzung für das 1. und 2. Schuljahr wurden jeweils bei den Leiterinnen und Leitern der Tagesschulangebote eingeholt. Analog zu Kuhn und Fischer (2011) wurde unterschieden zwischen Schülerinnen und Schülern, welche die Angebote höchstens für ein Jahr nutzten (78 %, Angebotsnutzung = 0) und Schülerinnen und Schülern bei denen über beide Schuljahre hinweg eine dauerhafte Angebotsnutzung vorliegt

werden (siehe Muthén & Muthén, 2012 für eine ausführlichere Beschreibung). Weil jedoch zusätzliche Analysen zeigten, dass die Wahrscheinlichkeit für einen fehlenden Wert bei der Angebotsnutzung nicht von den abhängigen Variablen (Ausgangsniveau und Entwicklung der Mathematikleistung) beeinflusst wird ($p > .05$), durfte hier stattdessen ein listenweiser Ausschluss angewendet werden (Urban, Mayerl & Wahl, 2016).

(22 %, Angebotsnutzung = 1). Da die Gruppe ohne dauerhafte Angebotsnutzung sowohl Schülerinnen und Schüler ohne Nutzung (89 %) als auch solche mit einer Nutzung während einem Schuljahr (11 %) umfasst, wurden zusätzliche Analysen mit den Extremgruppen *dauerhafte Angebotsnutzung* und *keine Angebotsnutzung* durchgeführt, wobei die Schülerinnen und Schüler mit einer geringen Nutzungsdauer ausgeschlossen wurden (vgl. Tabelle 1). Dabei zeigten sich identische Ergebnisse hinsichtlich unserer Fragestellungen, weshalb im Folgenden nur die Analysen zur Gesamtstichprobe berichtet werden.

Angebotsqualität: Am Ende der 1. Jahrgangsstufe führten Mitglieder des Forschungsteams Beobachtungen in den Tagesschulangeboten mittels standardisiertem Beobachtungsinstrument *Hort- und Ganztagsangebote-Skala* (HUGS; vgl. Tietze, Roßbach, Stendel & Wellner, 2007), einer *Adaption der School-Age Care Environment Rating Scale* (vgl. Harms, Jacobs & White, 1996), durch. Die Beobachterinnen und Beobachter absolvierten im Vorfeld ein Training. Die Beobachtungen dauerten jeweils mindestens vier Stunden und wurden von einer Person durchgeführt. HUGS erfasst fünfzig Merkmale aus sechs Bereichen (Platz und Ausstattung, Gesundheit und Sicherheit, Aktivitäten, Interaktionen, Strukturierung der pädagogischen Arbeit, Berufliche Entwicklungsmöglichkeiten für das Personal), wobei für jedes Merkmal Werte zwischen 0 (ungenügende Qualität) und 6 (exzellente Qualität) möglich sind. Die Merkmale lassen sich zu einer Gesamtskala zusammenfassen, welche im vorliegenden Beitrag verwendet wird ($\alpha = 0.85$, $M = 3.64$, $SD = 0.57$) (vgl. Tabelle 1).

Sozioökonomischer Status: Um den sozioökonomischen Status der Schülerinnen und Schüler zu berücksichtigen, wurde der *International Socio-Economic Index of Occupational Status* (ISEI; vgl. Ganzeboom & Treiman, 2012; Ganzeboom, de Graaf & Treiman, 1992) berechnet. Dieser Index basiert auf einer Einschätzung von Bildungsabschluss und Einkommen bei einem bestimmten Beruf und wurde zuletzt hinsichtlich der revidierten Fassung der *International Standard Classification of Occupation* (ISCO08) überarbeitet. Für diesen Beitrag wurde der jeweils höchste ISEI-Wert (HISEI) der Eltern verwendet, wobei ein höherer Wert einen höheren sozioökonomischen Status bedeutet ($M = 58.35$, $SD = 21.01$) (vgl. Tabelle 1).

Erstsprache: Mitte 2. Jahrgangsstufe wurden die Schülerinnen und Schüler mittels Fragebogen dazu befragt, ob sie zuhause regelmäßig Deutsch oder Schweizerdeutsch sprechen (74 %, Erstsprache = 1). Sofern dies nicht der Fall ist, kann davon ausgegangen werden, dass die Unterrichtssprache Deutsch eine Zweitsprache darstellt (26 %, Erstsprache = 0) (vgl. Tabelle 1).

4.2.3 Kontrollvariablen

Kognitive Grundfähigkeiten: Ebenfalls Mitte 2. Klasse wurde im Klassenverband die revidierte Fassung des Grundintelligenztests (CFT 1-R; vgl. Weiß und Osterland 2013) eingesetzt um die kognitiven Grundfähigkeiten der Schülerinnen und Schüler

zu erheben, wobei höhere Werte eine höhere Intelligenz bedeuten ($M = 52.08$, $SD = 9.34$) (vgl. Tabelle 1).

Geschlecht: Als zweite Kontrollvariable wurde das Geschlecht der Schülerinnen und Schüler via Klassenlisten sowie einer schriftlichen Befragung der Schülerinnen und Schüler erhoben und als männlich (48 %, Geschlecht = 0) oder weiblich (52 %, Geschlecht = 1) codiert (vgl. Tabelle 1).

Tabelle 1: Deskriptive Statistik

Gesamtstichprobe	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
Schülerin / Schüler					
Mathematikleistung (T1)	1656	28.83	91.10	50.28	8.32
Mathematikleistung (T2)	1632	32.83	95.25	57.83	9.23
Mathematikleistung (T3)	1737	35.26	105.89	62.22	10.21
Kognitive Grundfähigkeiten	1513	21.00	80.00	52.08	9.34
Geschlecht	1778	0.00	1.00	-	-
Angebotsnutzung	1806	0.00	1.00	-	-
Sozioökonomischer Status	1266	11.74	88.70	58.35	21.01
Erstsprache	1535	0.00	1.00	-	-
Schule					
Angebotsqualität	53	2.45	4.80	3.64	0.57

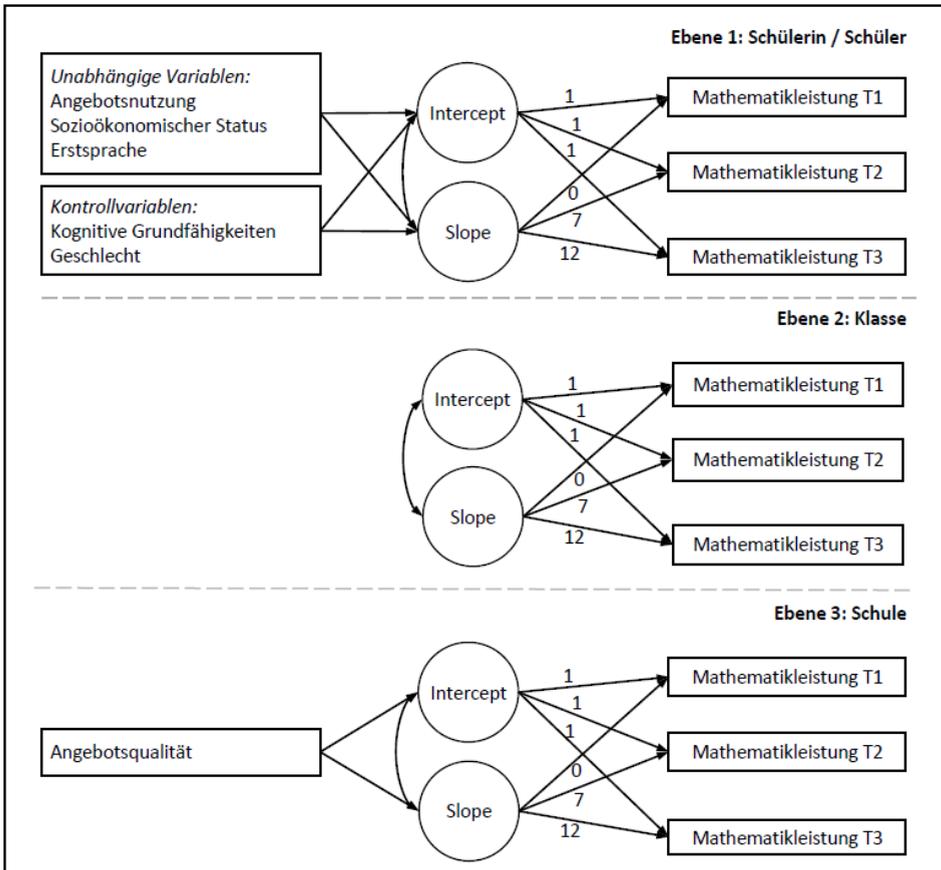
Skalierung: *Geschlecht* (0 = männlich, 1 = weiblich), *Angebotsnutzung* (0 = Nutzung während höchstens einem Schuljahr, 1 = dauerhafte Nutzung über beide Schuljahre), *Erstsprache* (0 = Deutsch als Zweitsprache, 1 = Deutsch als Erstsprache), *Angebotsqualität* (0 = ungenügende Qualität, 6 = exzellente Qualität)

4.3 Statistisches Vorgehen

Die statistischen Analysen wurden mit den Statistikprogrammen SPSS 23.0 (vgl. IBM Corp. 2015) und Mplus Version 7.3 (Muthén und Muthén, 1998–2012) durchgeführt. Um die Fragestellungen zu untersuchen, wurden latente Wachstumskurvenmodelle unter Berücksichtigung von drei Messzeitpunkten gerechnet, wodurch der Entwicklungsstand am Ende der 1. Jahrgangsstufe (Intercept) sowie die Entwicklung der Mathematikleistung bis zum Ende der 2. Jahrgangsstufe (Slope) modelliert wurden. Da die Messungen in Abständen von 7 respektive 5 Monaten erfolgten, wurde die Entwicklung entsprechend mit 0 (T1), 7 (T2) und 12 (T3) codiert. Latente Wachstumskurvenmodelle bieten den Vorteil, dass einerseits der Ausgangswert und andererseits die Entwicklung über drei Messzeitpunkte hinweg klar voneinander getrennt werden können und diese Entwicklung zugleich um den Einfluss des Ausgangswertes bereinigt wird. Um die Klumpenstruktur der Daten angemessen zu berücksichtigen, wurden Mehrebenenmodelle gerechnet, wo-

bei zwischen den Ebenen der Schülerinnen und Schüler (Ebene 1), der Klassen (Ebene 2) und der Schulen (Ebene 3) unterschieden wurde (vgl. Abbildung 1)

Abbildung 1: Mehrebenenmodell mit latenten Wachstumskurven zur Vorhersage der Mathematikleistung Ende 1. Jahrgangsstufe (Intercept) sowie der weiteren Entwicklung bis Ende 2. Jahrgangsstufe (Slope)



Im Modell 1.1 wurde erstens der Einfluss der Angebotsnutzung auf die Entwicklung der Mathematikleistung geschätzt (Tabelle 2, Fragestellung 1) und zweitens untersucht, ob Ende 1. Jahrgangsstufe Disparitäten bezüglich sozioökonomischem Status oder Erstsprache vorhanden sind (Fragestellung 2). Mögliche Interaktionseffekte zwischen familialer Herkunft und Angebotsnutzung auf die weitere Entwicklung der Mathematikleistung wurden im Modell 1.2 getestet (Fragestellung 3). Mit den Modellen 2.1 – 2.4 (Tabelle 3) wurde schließlich untersucht, ob der Einfluss der Angebotsnutzung in Abhängigkeit der Angebotsqualität variiert (Fragestellung 4). Als statistisches Verfahren zur Modellschätzung und Modellbewertung wurde das *Robuste Maximum-Likelihood-Schätzverfahren* (MLR) verwendet. Sämtliche Prädiktoren wurden am Gesamtmittelwert zentriert und – mit Ausnahme der di-

chotomen Variablen – z-standardisiert. Da bei Modellen mit ebenenübergreifenden Interaktionen sowohl R^2 als auch weitere etablierte Fit-Indizes nicht verfügbar sind, werden zusätzlich *Loglikelihood* und AIC angegeben, wobei niedrigere Werte einen besseren Modell-Fit bedeuten (Byrne, 2012). Fehlende Werte wurden mittels *full-information-maximum likelihood* (FIML) geschätzt. Es wurde zweiseitig getestet, mit einem Signifikanzniveau von 5 %.

5. Ergebnisse

5.1 Varianz unterteilt nach den drei Ebenen

Im Durchschnitt hatten die Schülerinnen und Schüler am Ende der 1. Jahrgangsstufe einen Ausgangswert (Intercept) von 50.01, der bis zum Ende der 2. Jahrgangsstufe um 11.96 Skaleneinheiten anstieg (Slope). Der größte Teil der Varianz geht dabei auf die Ebene der Schülerinnen und Schüler zurück (Intercept: 80 %, Slope: 92 %), während durch die Klassenebene (Intercept: 14 %, Slope: 2 %) und die Schulebene (Intercept: 6 %, Slope: 6 %) deutlich weniger Varianz erklärt wurde.

5.2 Generelle Wirkung der Angebotsnutzung

Um zu untersuchen, ob die Angebotsnutzung einen generellen Einfluss auf die Mathematikleistung ausübt und ob damit allfällige soziale oder kulturelle Herkunftseffekte reduziert werden können, wurden die Modelle 1.1 – 1.2 gerechnet (Tabelle 2).

Tabelle 2: Vorhersage von Ausgangswert und Entwicklung der Mathematikleistung von Ende der 1. bis Ende der 2. Jahrgangsstufe.

Prädiktoren	Modell 1.1		Modell 1.2	
	Intercept <i>B (SE)</i>	Slope <i>B (SE)</i>	Intercept <i>B (SE)</i>	Slope <i>B (SE)</i>
Intercept	50.20*** (.35)	1.00*** (.03)	50.20*** (.34)	1.00*** (.03)
Level 1: SchülerIn				
Kognitive Grundfähigkeiten	3.69*** (.16)	.10*** (.01)	3.69*** (.16)	.10*** (.01)
Geschlecht (1 = weiblich)	-2.09*** (.33)	-.05 (.03)	-2.10*** (.33)	-.05 (.03)
SES (sozioökonomischer Status)	.49* (.20)	.01 (.01)	.47* (.20)	.02 (.01)
Erstsprache (1 = Deutsch als Erstsprache)	.90* (.39)	-.06+ (.03)	.94* (.40)	-.06+ (.03)
Angebotsnutzung (1 = dauerhafte Nutzung)	.36 (.45)	.02 (.03)	.27 (.42)	.03 (.03)
SES × Angebotsnutzung			1.23* (.54)	-.06* (.03)
Erstsprache × Angebotsnutzung			-.77 (.83)	.16** (.06)
Kovarianz (Intercept / Slope)	.62*** (.09)		.63*** (.09)	
<i>R</i> ² (Level 1)	.33		.33	
Level 2: Klasse				
<i>R</i> ² Level 2	.00		.00	
Level 3: Schule				
<i>R</i> ² Level 3	.00		.00	
Fit Indizes	Modell 1.1		Modell 1.2	
Loglikelihood	-22049.27		-22260.56	
AIC	44158.54		44597.12	
CFI / TLI	0.97 / 0.96		0.96 / 0.95	
RMSEA	.06		.04	

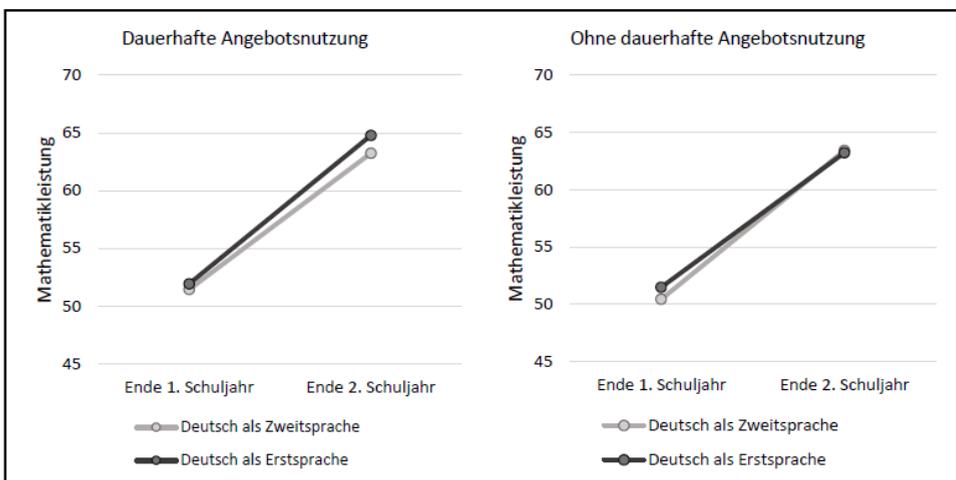
Anmerkungen: *N* = 1806 (Schülerinnen und Schüler) 118 (Klassen) 53 (Schulen); unstandardisierte *B* Koeffizienten (*SE*); * $p \leq .10$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$

In einem ersten Schritt wurden die Variablen *Angebotsnutzung*, *kognitive Grundfähigkeiten*, *Geschlecht*, *sozioökonomischer Status* und *Erstsprache* auf der Ebene der Schülerinnen und Schüler in das Modell aufgenommen (Modell 1.1). Die Angebotsnutzung zeigte keinen signifikanten Einfluss auf die Entwicklung der Mathematikleistung. Mit Blick auf das Ausgangsniveau der Mathematikleistungen ließen sich signifikante Herkunftseffekte ausmachen: Schülerinnen und Schüler mit höherem sozioökonomischem Status wiesen signifikant höhere Ausgangswerte auf als Schülerinnen und Schüler mit niedrigem sozioökonomischem Status ($B = .49$, $p = .013$) und Schülerinnen und Schüler mit Deutsch/Schweizerdeutsch als Erstsprache schnitten signifikant besser ab als jene, welche zuhause nicht regelmäßig Deutsch oder Schweizerdeutsch sprechen ($B = .90$, $p = .023$). Die weitere Entwicklung der Mathematikleistung wurde durch die soziale oder kulturelle

Herkunft nicht signifikant beeinflusst, hinsichtlich der Erstsprache zeigte sich jedoch, dass Schülerinnen und Schüler mit Deutsch als Zweitsprache tendenziell aufholen konnten ($B = -.06, p = .081$). Es lässt sich somit festhalten, dass kein Effekt der Angebotsnutzung auf die Entwicklung der Mathematikleistung gefunden wurde (Fragestellung 1) und dass die Ausgangsleistung Ende 1. Jahrgangsstufe einen signifikanten Zusammenhang mit dem sozioökonomischen Status (Fragestellung 2a) als auch mit der Erstsprache (Fragestellung 2b) zeigte. Mit diesem Modell konnten auf Ebene der Schülerinnen und Schüler 33 % der Varianz im Ausgangsniveau der Mathematikleistung und 12 % in deren Entwicklung erklärt werden.

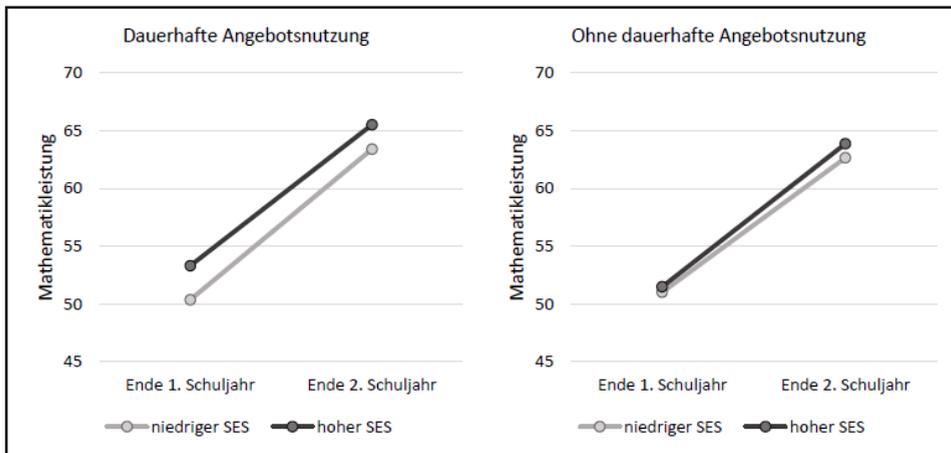
Im zweiten Schritt wurden die *Interaktionen von sozioökonomischem Status und Erstsprache mit der Angebotsnutzung* in das Modell aufgenommen (Modell 1.2), wobei sich mit Blick auf die Entwicklung der Mathematikleistung zwei signifikante Interaktionseffekte zeigten. Die Interaktion von sozioökonomischem Status und Angebotsnutzung ($B = -.06, p = .012$) wird in Abbildung 2 verdeutlicht, welche die Entwicklung der Mathematikleistung für Schülerinnen und Schüler mit hohem sozioökonomischem Status (1 SD über dem Mittelwert) und mit niedrigem sozioökonomischem Status (1 SD unter dem Mittelwert) für Schülerinnen und Schüler mit oder ohne Nutzung von Tagesschulangeboten vergleicht: Während die anfänglichen Unterschiede aufgrund des sozioökonomischen Status bei den Nutzerinnen und Nutzern leicht abnehmen (Schülerinnen und Schüler mit höherem sozioökonomischem Status zeigen einen niedrigeren Anstieg ihrer Mathematikleistung), kommt es bei den anderen Schülerinnen und Schülern, die das Angebot nicht nutzen, zu einer leichten Zunahme

Abbildung 2: Entwicklung der Mathematikleistung für Schülerinnen und Schüler mit hohem (+1 SD) oder niedrigem (-1 SD) sozioökonomischem Status (SES) unterteilt nach Angebotsnutzung und unter Kontrolle des Ausgangsniveaus, der kognitiven Grundfähigkeiten, des Geschlechts und der Erstsprache



Ein gegenteiliger Effekt zeigte sich für die Erstsprache ($B = .16, p = .008$), wie auch in Abbildung 3 ersichtlich ist: Hier kommt es bei den Nutzerinnen und Nutzern zu einer leichten Zunahme der Unterschiede (Schülerinnen und Schüler mit Deutsch als Zweitsprache verzeichnen eine schwächere Zunahme als jene, die zuhause regelmäßig Deutsch oder Schweizerdeutsch sprechen), während es bei den Schülerinnen und Schülern, welche keine Tagesschulangebote nutzen, zu einer Annäherung der beiden Gruppen kommt

Abbildung 3: Entwicklung der Mathematikleistung für Schülerinnen und Schüler mit Deutsch als Erst- oder Zweitsprache unterteilt nach Angebotsnutzung und unter Kontrolle des Ausgangsniveaus, der kognitiven Grundfähigkeiten, des Geschlechts und des sozioökonomischen Status.



Es lässt sich somit festhalten, dass die Nutzung von Tagesschulangeboten zwar den Herkunftseffekt aufgrund des sozioökonomischen Status reduzieren konnte (Fragestellung 3a), jedoch zeigte sich bezüglich Erstsprache eine gegenteilige Wirkung (Fragestellung 3b). Durch dieses zweite Modell konnte zusätzlich 0 % der Varianz im Ausgangsniveau und 3 % in der Entwicklung der Mathematikleistung erklärt werden.

5.3 Bedeutung der Angebotsqualität

Um zu testen, ob allenfalls bei einer höheren Angebotsqualität positive Effekte der Angebotsnutzung auf die Entwicklung der Mathematikleistung auftreten, wurden zusätzlich die Analysen 2.1 – 2.3 durchgeführt (Tabelle 3)

Tabelle 3: Vorhersage von Ausgangswert und Entwicklung der Mathematikleistung von Ende der 1. bis Ende der 2. Jahrgangsstufe unter Berücksichtigung der Angebotsqualität.

Prädiktoren	Modell 2.1		Modell 2.2		Modell 2.3	
	Intercept <i>B</i> (<i>SE</i>)	Slope <i>B</i> (<i>SE</i>)	Intercept <i>B</i> (<i>SE</i>)	Slope <i>B</i> (<i>SE</i>)	Intercept <i>B</i> (<i>SE</i>)	Slope <i>B</i> (<i>SE</i>)
Intercept	50.20*** (.35)	1.00*** (.03)	50.19*** (.35)	1.00*** (.03)	50.19*** (.35)	1.00*** (.03)
Level 1: SchülerIn						
Kognitive Grundfähigkeiten	3.69*** (.16)	.10*** (.01)	3.70*** (.16)	.10*** (.01)	3.69*** (.16)	.10*** (.01)
Geschlecht (1 = weiblich)	-2.09*** (.33)	-.05 (.03)	-2.09*** (.33)	-.05 (.03)	-2.08*** (.33)	-.05 (.03)
SES (sozioökonomischer Status)	.49* (.20)	.01 (.01)	.49* (.20)	.01 (.01)	.47* (.20)	.02 (.01)
Erstsprache (1 = Deutsch als Erstsprache)	.90* (.39)	-.06+ (.03)	.89* (.39)	-.06+ (.03)	.88* (.40)	-.06+ (.03)
Angebotsnutzung (1 = dauerhafte Nutzung)	.36 (.45)	.02 (.03)	.36 (.45)	.02 (.03)	.84 (4.29)	.00 (.01)
Kovarianz (Intercept / Slope)	.62*** (.09)		.62*** (.09)		.62*** (.09)	
<i>R</i> ² (Level 1)	.33	.12	.33	.12	NA	NA
Level 2: Klasse						
<i>R</i> ² Level 2	.00	.00	.00	.00	NA	NA
Level 3: Schule						
Angebotsqualität			.05 (.34)	.03 (.02)	.04 (.34)	.03 (.02)
Angebotsqualität (Level 3) × Angebotsnutzung (Level 1)					.73 (.45)	-.01 (.03)
Kovarianz (Intercept / Slope)			-.03 (.08)		-.03 (.08)	
<i>R</i> ² Level 3	.00	.00	.00	.11	NA	NA
Fit Indizes						
Loglikelihood	Modell 2.1		Modell 2.2		Modell 2.3	
	-22049.27		-22125.85		-22123.96	
AIC	44158.54		44319.70		44327.91	
CFI / TLI	0.97 / 0.96		0.97 / 0.95		NA	
RMSEA	.06		.06		NA	

Anmerkungen: *N* = 1806 (Schülerinnen und Schüler) 118 (Klassen) 53 (Schulen); unstandardisierte *B* Koeffizienten (*SE*); + *p* ≤ .10, * *p* ≤ .05, ** *p* ≤ .01, *** *p* ≤ .001; NA = Fit-Indizes für dieses Modell nicht verfügbar.

Im ersten Schritt wurden erneut *Angebotsnutzung*, *kognitive Grundfähigkeiten*, *Geschlecht*, *sozioökonomischer Status* und *Erstsprache* auf der Ebene der Schülerinnen und Schüler in das Modell aufgenommen (Modell 2.1, vgl. Modell 1.1). Im zweiten Schritt wurde die *Angebotsqualität* auf Ebene der Schule in das Modell

eingefügt (Modell 2.2), wobei sich kein direkter Effekt der Angebotsqualität auf das Ausgangsniveau oder die Entwicklung der Mathematikleistung zeigte und keine zusätzliche Varianz erklärt wurde. Im dritten Schritt wurde die ebenenübergreifende Interaktion zwischen der Angebotsqualität auf Schulebene und der Angebotsnutzung auf Ebene der Schülerinnen und Schüler untersucht (Modell 2.3). Diese zeigte jedoch keine signifikanten Effekte: Selbst bei einer höheren Angebotsqualität scheint die Angebotsnutzung keinen Einfluss auf die Entwicklung der Mathematikleistung zu haben (Fragestellung 4).

6. Diskussion

Dieser Beitrag untersucht erstmals, ob die dauerhafte Nutzung des Tageschulangebots in einer offenen Tagesschule in der Deutschschweiz zu einer Zunahme der Mathematikleistungen führt und ob allfällige sozial oder kulturell bedingte Disparitäten reduziert werden. Zudem wird untersucht, ob der Einfluss der Angebotsnutzung auf die Entwicklung der Mathematikleistung von der Angebotsqualität abhängt. Mit Blick auf unsere Stichprobe zeigte die dauerhafte Nutzung des Tagesschulangebots keinen Einfluss auf die Entwicklung der Mathematikleistung (Fragestellung 1) und dies unabhängig von der Angebotsqualität (Fragestellung 4). Es zeigten sich jedoch Disparitäten Ende 1. Jahrgangsstufe, wobei Schülerinnen und Schüler mit einem höheren sozioökonomischen Status (Fragestellung 2a) oder mit Deutsch/Schweizerdeutsch als Erstsprache (Fragestellung 2b) höhere Mathematikleistungen aufwiesen. Entsprechend unserer Erwartung konnten die Disparitäten aufgrund des sozioökonomischen Status bis Ende 2. Jahrgangsstufe durch eine dauerhafte Angebotsnutzung reduziert werden (Fragestellung 3a), hinsichtlich Erstsprache führte die dauerhafte Angebotsnutzung allerdings zu einer Zunahme der Disparitäten (Fragestellung 3b).

Die dauerhafte Nutzung von außerunterrichtlichen Bildungs- und Betreuungsangeboten an offenen Tagesschulen in der Deutschschweiz wirkte sich nicht positiv auf die Entwicklung der Mathematikleistung aus – dieser ernüchternde Befund widerspiegelt die bisherigen Forschungsergebnisse aus dem deutschsprachigen Raum (vgl. Bellin & Tamke, 2010; Kuhn & Fischer, 2011; Lossen et al., 2016; Radisch et al., 2006; Reinders et al., 2011; Schüpbach 2010) und steht im Widerspruch zur US-amerikanischen Forschung zu *afterschool* Programmen, die zumindest für einzelne Programme positive Wirkungen auf die Schulleistungen nachweisen konnten (vgl. Durlak et al., 2010; Lauer et al. 2006). Die möglichen Gründe für diese Diskrepanz sind vielfältig. So richten sich viele *afterschool* Programme spezifisch an Schülerinnen und Schüler, die hinsichtlich ihrer Bildungslaufbahn als benachteiligt gelten, während Tages- und Ganztagschulen keine spezifische Zielgruppe haben und dadurch mit einer äußerst heterogenen Gruppe konfrontiert sind. Zudem sind die meisten *afterschool* Programme auf eine bestimmte Zeitdauer und spezifische Ziele ausgerichtet. Bei der Einführung von Tagesschulen in der Schweiz

war insbesondere der Bedarf nach Betreuungsangeboten und weniger die zusätzliche Förderung der Schülerinnen und Schüler ausschlaggebend (vgl. Frei et al., 2016), was dazu geführt haben könnte, dass die Förderung weniger zielgerichtet erfolgt.

Unsere Ergebnisse zeigen für die Deutschschweiz aber zumindest in Bezug auf den sozioökonomischen Status, dass eine dauerhafte Angebotsnutzung Disparitäten aufgrund der familialen Herkunft entgegenwirken kann: Während die Unterschiede bei den Schülerinnen und Schülern ohne dauerhafte Angebotsnutzung leicht zunahmen, kam es bei den Schülerinnen und Schülern mit dauerhafter Angebotsnutzung zu einer Annäherung zwischen den Schülerinnen und Schülern mit hohem oder niedrigem sozioökonomischen Status. Allerdings führte die dauerhafte Angebotsnutzung in Bezug auf die Erstsprache gar zu einer Zunahme der Disparitäten: Bei einer dauerhaften Angebotsnutzung zeigten Schülerinnen und Schüler, die zuhause regelmäßig Deutsch oder Schweizerdeutsch sprechen, einen stärkeren Anstieg der Mathematikleistung als Schülerinnen und Schüler mit Deutsch als Zweitsprache. Somit konnten die hoffnungsvollen Befunde zu *after-school* Programmen (vgl. McComb & Scott-Little, 2003; Miller, 2003) nicht vollumfänglich bestätigt werden. In Übereinstimmung mit Bellin und Tamke (2010), Radisch et al. (2006) und Lossen et al. (2016) scheint es den Tagesschulangeboten insbesondere nicht zu gelingen, Schülerinnen und Schüler mit Deutsch als Zweitsprache ausreichend zu fördern. Es bleibt allerdings festzuhalten, dass es sich um sehr geringe Interaktionseffekte handelt, die nur wenig zur Varianzaufklärung in der Entwicklung der Mathematikleistung beitragen.

Abschließend zeigten die Ergebnisse, dass die dauerhafte Angebotsnutzung in der offenen Tagesschule in der Deutschschweiz selbst bei einer höheren Angebotsqualität keinen Effekt auf die Entwicklung der Mathematikleistung hatte. Dieses Ergebnis deckt sich mit den Befunden von Kuhn und Fischer (2011), die ebenfalls keinen Effekt der Prozessqualität auf die nachfolgende Entwicklung der Schulnoten feststellen konnten und den Befunden von Lossen et al. (2016) hinsichtlich naturwissenschaftlicher Kompetenzentwicklung. Die Ergebnisse stimmen jedoch nicht überein mit der Studie von Schübach (2014a), die sowohl offene als auch gebundene Tagesschulen in der Deutschschweiz fokussierte. In der Vorläuferstudie wurde ein positiver Effekt einer höheren Angebotsqualität auf die Entwicklung der Mathematikleistung ausgemacht. Eine mögliche Erklärung für diesen Unterschied bietet die Stichprobe. Da an gebundenen Tagesschulen die Angebote von allen Schülerinnen und Schülern der Schule genutzt werden, ist eine stärkere inhaltliche und zeitliche Verzahnung mit dem Unterricht möglich, was eine gezielte Förderung der Schülerinnen und Schüler begünstigt. Im Gegensatz dazu wechselt die Gruppenzusammensetzung in offenen Tagesschulangeboten ständig, zudem ist auch die Intensität der Teilnahme meist auf wenige Tage und Einheiten beschränkt. Für die Unterschiede könnten aber auch methodische Gründe verantwortlich sein: Die Studie von Schübach (2014a) erstreckte sich über einen Zeitraum von zwei Jahren, wobei insbesondere gegen Ende der Untersuchung Effekte festzustellen waren. Andererseits konnten Klassen- und Schuleffekte in unse-

rer Studie besser kontrolliert werden, da die Angebotsnutzung unabhängig von der Klassenzugehörigkeit erfolgte, was bei Schüpbach (2014a) nicht der Fall war.

Die berichteten Ergebnisse sind mit einigen Einschränkungen verbunden: (1) Der Untersuchungszeitraum war auf eine Dauer von einem Jahr begrenzt, weshalb längerfristige oder verzögerte Effekte nicht untersucht werden konnten. (2) Es kann nicht ausgeschlossen werden, dass sich Schülerinnen und Schüler je nach Angebotsnutzung in der Beanspruchung weiterer Bildungs- und Betreuungsangebote außerhalb der Schule unterscheiden oder über unterschiedlich ausgeprägte familiäre Ressourcen verfügen. (3) Die Teilnahme an Tagesschulangeboten wurde ausschließlich über die Nutzungsdauer operationalisiert. Die Breite der besuchten Angebote sowie das persönliche Engagement der Schülerinnen und Schüler blieben unberücksichtigt, ebenso wie die Intensität der Angebotsnutzung. (4) Die Qualität der Angebote wurde während einer halbtägigen Beobachtung von einer Person eingeschätzt. Dies bietet den Vorteil, dass die externen Einschätzungen unabhängig von der subjektiven Wahrnehmung der Beteiligten stattfanden. Es stellt sich allerdings die Frage, ob eine einmalige Beobachtung ausreichend um die Qualität über ein Schuljahr hinweg ausreichend zu erfassen und ob die für die Entwicklung der Mathematikleistung relevanten Qualitätsaspekte im verwendeten Messinstrument ausreichend Berücksichtigung fanden. Zudem ergeben sich gewisse Einschränkungen bezüglich der Reliabilität. (5) Ein möglicher moderierender Einfluss der Angebotsqualität wurde nur im Hinblick auf die Entwicklung der Mathematikleistung untersucht, nicht aber in Bezug auf soziale oder kulturelle Disparitäten. Eine entsprechende Fragestellung wäre sicher sinnvoll, hätte aber unsere statistischen Modelle überstrapaziert.

Zusammen mit dem zitierten Forschungsstand verdeutlichen die Befunde, dass die hohen Erwartungen an eine Leistungsförderung und an einen Abbau von sozialen oder kulturellen Disparitäten durch Tages- und Ganztagschulen kritisch hinterfragt werden sollten, oder dass eine Verbesserung der Angebote notwendig ist. Insbesondere Schülerinnen und Schüler, für die Deutsch lediglich eine Zweitsprache darstellt, scheinen durch die aktuellen Tagesschulangebote nicht ausreichend gefördert zu werden. Dabei wäre gerade für diese Kinder Potential vorhanden, wie die Forschungsbefunde aus den USA zeigen. Die Ergebnisse verdeutlichen zudem, dass griffige Qualitätskriterien zu außerunterrichtlichen Bildungs- und Betreuungsangeboten fehlen und dass die bloße Nutzung möglicherweise zu kurz greift, wenn die Teilnahme an Tagesschulangeboten erfasst werden soll (vgl. Roth et al., 2010). Stattdessen sollte der Fokus künftiger Studien stärker auf die aktive Rolle der Schülerinnen und Schüler bei der Wahl von Aktivitäten sowie auf ihre kognitive und emotionale Beteiligung an denselben gelenkt werden. Dies gilt auch für die Frage, wie Schülerinnen und Schüler nebst den Tagesschulangeboten ihre unterrichtsfreie Zeit verbringen, damit entsprechende Selektionseffekte ausgeschlossen werden können.

Literatur

- Aeberli, C. & Binder, H.-M. (2005). *Das Einmaleins der Tagesschule. Ein Leitfaden für Gemeinde- und Schulbehörden*. Zürich: Avenir Suisse.
- Bellin, N. & Tamke, F. (2010). Bessere Leistungen durch Teilnahme am offenen Ganztagsbetrieb? *Empirische Pädagogik*, 24(2), 93–112.
- Boudon, R. (1974). *Education, opportunity, and social inequality. Changing prospects in western societies*. New York, NY: Wiley.
- Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming. Multivariate applications series*. New York, NY: Routledge Taylor & Francis Group.
- Coradi Vellacott, M. (2007). *Bildungschancen Jugendlicher in der Schweiz. Eine Untersuchung familiärer, schulischer und sozial-räumlicher Einflüsse auf Leistungsunterschiede am Ende der obligatorischen Schulzeit*. Zürich: Rüegger.
- Coradi Vellacott, M., Hollenweger, J., Nicolet, M. & Wolter, S. C. (Hrsg.). (2003). *Soziale Integration und Leistungsförderung. Thematischer Bericht der Erhebung PISA 2000* (Bildungsmonitoring Schweiz). Neuchâtel: BfS – Bundesamt für Statistik.
- Durlak, J. A., Weissberg, R. P. & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology*, 45(3–4), 294–309.
- EDK – Schweizerische Konferenz der kantonalen Erziehungsdirektoren (2013). *Kantonsumfrage 2012/2013*. Zugriff am 26.09.2016 unter <http://edudoc.ch/record/115194/files/2013.pdf>
- Fischer, N., Holtappels, H. G., Klieme, E., Rauschenbach, T., Stecher, L. & Züchner, I. (Hrsg.). (2011). *Ganztagschule: Entwicklung, Qualität, Wirkungen. Längsschnittliche Befunde der Studie zur Entwicklung von Ganztagschulen (StEG)*. Weinheim: Beltz Juventa.
- Fischer, N., Sauerwein, M. N., Theis, D. & Wolgast, A. (2016). Vom Lesenlernen in der Ganztagschule: Leisten Ganztagsangebote einen Beitrag zur Leseförderung am Beginn der Sekundarstufe I? *Zeitschrift für Pädagogik*, 62(6), 780–796.
- Frei, L., Schübach, M., von Allmen, B. & Nieuwenboom, W. (2016). Bildungsbezogene Erwartungen an Tagesschulen. Förderangebote an offenen Tagesschulen in der Deutschschweiz. *Schweizerische Zeitschrift für Bildungswissenschaften*, 38(3), 549–567.
- Ganzeboom, H. B., Graaf, P. M. de & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1–56.
- Ganzeboom, H. B. G. & Treiman, D. J. (2012). *International stratification and mobility file: Conversion tools*, Department of Social Research Methodology. Zugriff am 26.09.2016 unter <http://www.harryganzeboom.nl/ismf/index.htm>
- Haffner, J., Baro, K., Parzer, P. & Resch, F. (2005). *Heidelberger Rechentest. HRT 1–4; Erfassung mathematischer Basiskompetenzen im Grundschulalter; Manual* (Deutsche Schultests). Göttingen: Hogrefe.
- Harms, T., Jacobs, E. V. & White, D. R. (1996). *School-age care environment rating scale*. New York, NY: Teachers College Press.
- Helmke, A. (2003). *Unterrichtsqualität erfassen, bewerten, verbessern*. Seelze: Kallmeyer.
- Holtappels, H. G. (2009). Qualitätsmodelle – Theorie und Konzeptionen. In I. Kamski, H. G. Holtappels & T. Schnetzer (Hrsg.), *Qualität von Ganztagschule. Konzepte und Orientierungen für die Praxis* (S. 11–25). Münster: Waxmann.
- Hopf, D. (2005). Zweisprachigkeit und Schulleistung bei Migrantenkindern. *Zeitschrift für Pädagogik*, 51(2), 236–251.
- IBM Corp. (2015). *IBM SPSS Statistics for Windows, Version 23.0*. Armonk, NY: IBM Corp.

- Kuhn, H. P. & Fischer, N. (2011). Entwicklung der Schulnoten in der Ganztagschule. Einflüsse der Ganztagsbeteiligung und der Angebotsqualität. In N. Fischer, H. G. Holtappels, E. Klieme, T. Rauschenbach, L. Stecher & I. Züchner (Hrsg.), *Ganztagschule: Entwicklung, Qualität, Wirkungen. Längsschnittliche Befunde der Studie zur Entwicklung von Ganztagschulen (StEG)* (S. 207–226). Weinheim: Beltz Juventa.
- Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D. & Martin-Glenn, M. L. (2006). Out-of-school-time programs. A meta-analysis of effects for at-risk students. *Review of Educational Research*, 76(2), 275–313.
- Lossen, K., Tillmann, K., Holtappels, H. G., Rollett, W. & Hannemann, J. (2016). Entwicklung der naturwissenschaftlichen Kompetenzen und des sachunterrichtsbezogenen Selbstkonzepts bei Schülerinnen und Schülern in Ganztagsgrundschulen – Ergebnisse der Längsschnittstudie StEG-P zu Effekten der Schülerbeteiligung und der Angebotsqualität. *Zeitschrift für Pädagogik*, 62(6), 760–779.
- McComb, E. M. & Scott-Little, C. (2003). *After-school programs. Evaluations and outcomes*. Greenboro, NC: SERVE.
- Miller, B. M. (2003). *Critical hours. Afterschool programs and educational success*. Quincy, MA: Nellie Mae Education Foundation.
- Moser, U. & Bayer, N. (2010). 4. bis 8. *Schlussbericht der summativen Evaluation. Lernfortschritte vom Eintritt in die Eingangsstufe bis zum Ende der 3. Klasse der Primarschule*. Bern: Schulverlag.
- Muthén, L. K. & Muthén, B. O. (1998–2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.
- Nidegger, C. (2014). *PISA 2012. Vertiefende Analysen* (Bildungsmonitoring Schweiz). Neuchâtel: Konsortium PISA.ch.
- OECD – Organisation for Economic Cooperation and Development (2014). *PISA 2012 Ergebnisse: Exzellenz durch Chancengerechtigkeit (Band II). Allen Schülerinnen und Schülern die Voraussetzungen zum Erfolg sichern*. Bielefeld: Bertelsmann.
- Paetsch, J., Felbrich, A. & Stanat, P. (2015). Der Zusammenhang von sprachlichen und mathematischen Kompetenzen bei Kindern mit Deutsch als Zweitsprache. *Zeitschrift für Pädagogische Psychologie*, 29(1), 19–29.
- Radisch, F., Klieme, E. & Bos, W. (2006). Gestaltungsmerkmale und Effekte ganztägiger Angebote im Grundschulbereich. *Zeitschrift für Erziehungswissenschaft*, 9(1), 30–50.
- Reinders, H., Gogolin, I., Gresser, A., Schnurr, S., Böhmer, J. & Bremm, N. (2011). Ganztagsschulbesuch und Integration von Kindern mit Migrationshintergrund im Primarbereich. Erste Näherungen an empirische Befunde einer vergleichenden Untersuchung. *Zeitschrift für Erziehungswissenschaft*, 14(3), 163–183.
- Roth, J. L., Malone, L. M. & Brooks-Gunn, J. (2010). Does the amount of participation in afterschool programs relate to developmental outcomes? A review of the literature. *American Journal of Community Psychology*, 45(3–4), 310–324.
- Schüpbach, M. (2010). *Ganztägige Bildung und Betreuung im Primarschulalter. Qualität und Wirksamkeit verschiedener Schulformen im Vergleich*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schüpbach, M. (2013). Development, quality, and effectiveness of out-of-school time education in Switzerland. In J. Ecarus, E. Klieme, L. Stecher & J. Woods (Hrsg.), *Extended Education - in an International Perspective* (S. 75–98). Opladen: Verlag Barbara Budrich.
- Schüpbach, M. (2014a). Extended education and social inequality in Switzerland: Compensatory effects? An analysis of the development of language achievement with regard to structural and process-related aspects of social background. *Journal for Educational Research Online*, 3, 95–114.

- Schüpbach, M. (2014b). Effects of extracurricular activities and their quality on primary school-age students' achievement in mathematics in Switzerland. *School Effectiveness and School Improvement*, 26(2), 279–295.
- Schüpbach, M., Herzog, W. & Ignaczewska, J. (2013). Entwicklung der Mathematikleistung von Ganztagschulkindern. Kompensatorische Wirkung der Ganztagschule?. *Zeitschrift für Pädagogische Psychologie*, 27(3), 157–167.
- Schüpbach, M., Rohrbach-Nussbaum, R. & Grütter, E. (2017). Pädagogische Qualität: geleitete und freie Aktivitäten in der Tagesschule bzw. im Tagesschulangebot. In Schüpbach, M.; Frei, L.; Nieuwenboom, W. (Hrsg.), *Tagesschulen. Ein Überblick*. Programm VS Research, Springer.
- Stecher, L., Radisch, F., Fischer, N. & Klieme, E. (2007). Bildungsqualität außerunterrichtlicher Angebote in der Ganztagschule. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 27(4), 346–366.
- Strietholt, R., Manitius, V., Berkemeyer, N. & Bos, W. (2015). Bildung und Bildungsungleichheit an Halb- und Ganztagschulen. *Zeitschrift für Erziehungswissenschaft*, 18, 737–761.
- Stufflebeam, D. L. (1972). Evaluation als Entscheidungshilfe. In C. Wulf (Hrsg.), *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen* (S. 113–145). München: Piper; Deutsches Institut für Internationale Pädagogische Forschung.
- Tietze, W., Roßbach, H. G., Stendel, M. & Wellner, B. (2007). *Hort- und Ganztagsangebote-Skala (HUGS). Feststellung und Unterstützung pädagogischer Qualität in Horten und außerunterrichtlichen Angeboten*. Berlin: Cornelsen Scriptor.
- Urban D., Mayerl J., Wahl A. (2016). *Regressionsanalyse bei fehlenden Variablenwerten (missing values): Imputation oder Nicht-Imputation? Eine Anleitung für die Regressionspraxis mit SPSS (2. korrigierte Auflage)*. Stuttgart: SOWI IV; IfS – Institut für Sozialwissenschaften.
- Weiß, R. H. & Osterland, J. (2013). *CFT 1-R. Grundintelligenztest Skala 1*. Göttingen: Hogrefe.