

Zur Erfassung fragwürdiger Überzeugungen zu Bildungsthemen: Entwicklung und erste Überprüfung des Questionable Beliefs in Education-Inventars (QUEBEC)

Jana Asberger, Eva Thomm & Johannes Bauer
Universität Erfurt

Zusammenfassung: Studierende des Lehramts und anderer bildungsbezogener Studiengänge haben häufig fragwürdige oder falsche Überzeugungen zu Bildungsthemen, die zwar subjektiv plausibel erscheinen, jedoch nicht dem Stand der Forschung entsprechen. Diese bildungsbiografisch erworbenen und teilweise fest verwurzelten Fehlkonzepte können die Auseinandersetzung mit wissenschaftlichem Wissen im Studium behindern. Der vorliegende Artikel berichtet Ergebnisse zur ersten Überprüfung eines Fragebogens, der fragwürdige Überzeugungen zu vier exemplarischen Bildungsthemen erfasst (Effekte der Klassengröße, Klassenwiederholung, direkten Instruktion und der Feminisierung des Elementar-/Primarbereichs). Hierfür wurden Studierende verschiedener Studiengänge befragt ($N=217$). Ergebnisse von Faktorenanalysen fielen konform mit der themenbezogenen Struktur des Fragebogens aus und belegten skalare Messinvarianz über Studierende des Lehramts und anderer Studiengänge. Im Vergleich zu Pädagogik/Psychologie-Studierenden wiesen Lehramtsstudierende unter Kontrolle des Studiensemesters geringer ausgeprägte Fehlkonzepte hinsichtlich direkter Instruktion, jedoch höhere hinsichtlich Klassenwiederholungen auf. Im Vergleich zu Studierenden nicht-bildungsbezogener Studiengänge (Wirtschafts- und Sozialwissenschaften/Sonstige) waren Fehlkonzepte zur Klassengröße und direkten Instruktion bei Lehramtsstudierenden geringer ausgeprägt.

Schlüsselbegriffe: Überzeugungen, Fehlkonzepte, Lehramt, Studierende, Diagnostik

What Misconceptions do Preservice Teachers Hold About Core Educational Topics? Development and Psychometric Analysis of the Questionable Beliefs in Education Inventory (QUEBEC)

Summary: Students in teacher education and other educationally relevant disciplines frequently have questionable or false beliefs about school-related topics. Such home-grown and deeply rooted misconceptions may interfere with learning research-based knowledge at university. The present paper delivers results from a first psychometric evaluation of the Questionable Beliefs in Education scale that assesses misconceptions regarding four exemplary educational topics (i. e., effects of class size, grade retention, direct instruction, feminization of elementary/primary education). Data were gathered from $N=217$ students from different disciplines. Results of factor analyses conformed to the scale's four-dimensional topic-related structure. Preservice teachers endorsed misconceptions related to direct instruction to a lesser degree and misconceptions related to grade retention more strongly than educational science and psychology students. As compared to students from non-educational domains, preservice teachers agreed less with misconceptions about class size and direct instruction.

Keywords: Beliefs, misconceptions, preservice teachers, students, assessment

Öffentliche Debatten über bildungsrelevante Themen sind häufig von fragwürdigen Überzeugungen durchdrungen, die nicht dem einschlägigen Stand der Forschung entsprechen (Bromme, Prenzel & Jäger, 2014), und insofern

Fehlkonzepte darstellen (Gardner & Brown, 2013). Dies zeigt sich beispielsweise in Medienberichten, aber auch in öffentlichen Statements von Interessensverbänden oder der Bildungspolitik zu typischen Themen wie der Rolle der

Klassengröße, der Sinnhaftigkeit von Klassenwiederholungen oder einer „neuen Lernkultur“ – (Christodoulou, 2014; Holmes, 2016; Kirschner & De Bruyckere, 2017). Auch bei Studierenden sind häufig fragwürdige Überzeugungen zu Bildungsthemen festzustellen. Solche bildungsbiografisch erworbenen und teilweise fest verwurzelten Fehlkonzepte können die Auseinandersetzung mit wissenschaftlichem Wissen in bildungsbezogenen Studiengängen, wie dem Lehramtsstudium, erheblich behindern (Reusser & Pauli, 2014). Sowohl für die Forschung als auch für die Hochschuldidaktik sind deshalb überprüfte Instrumente zur Diagnose von Fehlkonzepten über bildungsrelevante Themen wünschenswert (Schraw & Olafson, 2015). Sie können genutzt werden, um fragwürdige Überzeugungen bei Studierenden systematisch sichtbar zu machen, sie zu thematisieren und mit wissenschaftlichen Theorien und Befunden zu kontrastieren (Fives & Buehl, 2012; Richardson, 1996). Zudem sind sie im Forschungskontext wertvoll, um Interferenzen durch Fehlkonzepte bei Prozessen der Rezeption wissenschaftlichen Wissens besser untersuchen und verstehen zu können. Daher sind sie auch im breiteren Kontext der Diskussion um Evidenzbasierung im Bildungsbereich relevant. Allerdings gibt es bislang nur wenige empirische Befunde zur Häufigkeit und Ausprägung spezifischer Fehlkonzepte.

Der vorliegende Beitrag berichtet Ergebnisse zur Entwicklung und ersten Überprüfung des *Questionable Beliefs in Education*-Inventars (QUEBEC). Es erfasst typische Fehlkonzepte zu vier exemplarischen Bildungsthemen. Die Fragestellungen der Studie betrafen eine erste psychometrische Evaluation des Instruments mit Fokus auf dessen Dimensionalität und Messinvarianz zum Vergleich Lehramtsstudierender und anderer Studierendengruppen. Darüber hinaus sollte ein explorativer Vergleich Lehramtsstudierender mit Studierenden anderer bildungsbezogener Studiengänge (Pädagogik, Psychologie) und anderer nicht-bildungsbezogener Studiengänge Hinweise auf differenzielle Ausprägungen der untersuchten Fehlkonzepte geben.

Im Folgenden verorten wir zunächst Fehlkonzepte in der Forschung zu Lehrerüberzeugungen und erläutern die Probleme, die sie beim Erwerb wissenschaftlich fundierten Wissens und bei der Kompetenzentwicklung im Studium bereiten können. Zudem beschreiben wir den Stand der Forschung zu bildungsbezogenen Fehlkonzepten und gehen dabei speziell auf die vier exemplarischen Bildungsthemen ein, die im QUEBEC adressiert werden.

Fehlkonzepte und fragwürdige Überzeugungen

Überzeugungen von Lehrkräften gelten als bedeutsame Facette ihrer professionellen Kompetenz (Baumert & Kunter, 2006; Gill & Fives, 2015). Sie sind als subjektive Konzepte, Vorstellungen und Annahmen definiert, die individuell für wahr gehalten werden (Pajares, 1992; Richardson, 1996; Skott, 2015). Im Gegensatz zu Wissen müssen Überzeugungen also kein externes Verifizierungskriterium erfüllen, es genügt ein subjektiver Wahrheitsgehalt. Von außen können Überzeugungen allerdings als fragwürdig oder falsch beurteilt werden, je nachdem wie stark sie mit wissenschaftlich fundiertem Wissen in Konflikt stehen. Gardner und Brown (2013) sprechen in diesem Fall von Fehlkonzepten (vgl. Bensley & Lilienfeld, 2017; Murphy & Mason, 2012). Da wissenschaftliches Wissen im Bildungsbereich grundsätzlich unsicher und vorläufig ist (Bromme et al., 2014; Gollwitzer, Rothmund, Klimmt, Nauroth & Bender, 2014), lassen sich mit dem Forschungsstand konfligierende Überzeugungen jedoch nicht als objektiv falsch klassifizieren. Dies soll der Begriff „fragwürdige Überzeugungen“ zum Ausdruck bringen.

Es wird angenommen, dass sich bildungsbezogene Überzeugungen im Laufe der Bildungsbiografie entwickeln, sodass sie beim Studieneintritt bereits ausgebildet und relativ stabil sind (Kagan, 1992; Kane, Sandretto & Heath, 2002; Pajares, 1992). Lehramtsstudierende, aber auch Studierende verwandter Studiengänge (z. B. Erziehungswissenschaft, Psychologie), beginnen somit ihr Studium als *Insider* auf Basis ihrer Bil-

dungsbiografie (Pajares, 1992; Trautwein, 2013). Dies kann insbesondere dann zu Problemen führen, wenn Vorüberzeugungen mit Studieninhalten konfliktieren, da dies verzerrte Rezeptions- und Bewertungsprozesse begünstigt (Richter & Maier, 2018; Rothmund, Gollwitzer, Nauroth & Bender, 2017). Wenn individuelle Vorüberzeugungen im Widerspruch zu Forschungsbefunden stehen, neigen Personen dazu, wissenschaftliche Evidenz abzuwerten oder gar zu ignorieren anstatt ihre Vorüberzeugungen daran zu reflektieren und ggf. zu korrigieren (Chinn & Brewer, 1998). Die Rezeption bildungswissenschaftlicher Befunde unterliegt vermutlich im besonderen Maße motivierter Informationsverarbeitung, da die Auseinandersetzung mit Bildungsthemen häufig durch eigene (Alltags-)Erfahrungen und Beobachtungen geprägt ist, die eine hohe subjektive Validität haben. Deshalb besteht auch eher das Risiko, die eigenen Auffassungen als widersprüchlich zu wissenschaftlichen Erkenntnissen wahrzunehmen und diese abzuwerten (Gollwitzer et al., 2014). Dies illustrieren u. a. Befunde, dass Lehrkräfte die Nützlichkeit bildungswissenschaftlicher Forschungsbefunde in Abhängigkeit von subjektiver Plausibilität und Übereinstimmung mit eigenen Erfahrungen bewerten und diese bei Diskrepanz auch verwerfen (Hetmanek et al., 2015).

Infolgedessen können Studierende aufgrund ihrer bereits vorhandenen Überzeugungen falsches oder fragwürdiges Fachwissen über bildungsrelevante Themen entwickeln, welches sich ungünstig auf die spätere Unterrichtspraxis auswirken kann. Zum Beispiel zeigten König, Darge und Schreiber (2012), dass Lehrkräfte, welche Klassenwiederholungen für eine effektive Maßnahme halten, ihren Schülerinnen und Schülern weniger individuelle Unterstützung und differenzierte Methoden im Unterricht boten. Insgesamt weisen aktuelle Befunde darauf hin, dass korrektes Fachwissen über Bildungsthemen von hoher Bedeutung für die Unterrichtspraxis von angehenden Lehrerinnen und Lehrern ist, z. B. hinsichtlich der Klassenführung, Unterrichtsqualität oder emotionalen Erschöpfung (König & Pflanzl, 2016; Kunter

et al., 2017; Voss, Kunter, Seiz, Hoehne & Baumert, 2014). Dieses Fachwissen wird maßgeblich im Lehramtsstudium erworben.

Sowohl für den Wissens- und Kompetenzerwerb im Studium als auch mit Blick auf potenzielle Effekte auf die spätere Berufspraxis von Lehrkräften (Fives & Buehl, 2012; König et al., 2012; Reusser & Pauli, 2014) ist es daher erforderlich, Fehlkonzepte im Rahmen der Hochschullehre angemessen adressieren zu können. Voraussetzung für einen möglichen Konzeptwechsel ist, dass Studierende ihre bestehenden Überzeugungen evaluieren und diese mit neuen, verständlichen und angemessenen Konzepten vergleichen (Murphy & Mason, 2012). Hierzu können diagnostische Instrumente, die Fehlkonzepte systematisch herausarbeiten und transparent machen, eine Grundlage bieten.

Typische Fehlkonzepte zu Bildungsthemen

Zu Fragen des Lehrens und Lernens, aber auch zu anderen bildungsbezogenen Themen existiert eine große Vielfalt *urbaner Mythen*, die Fehlkonzepte im o. g. Sinn darstellen und die vermutlich weit verbreitet sind (Christodoulou, 2014; Dekker, Lee, Howard-Jones & Jolles, 2012; Holmes, 2016; Kirschner & De Bruyckere, 2017; Renkl, 2015). Neben Konflikten mit dem empirischen Forschungsstand ist für solche Fehlkonzepte auch charakteristisch, dass sie bereits aus theoretischen Gründen fragwürdige subjektive Theorien darstellen (Groeben, Wahl, Schlee & Scheele, 1988; Keil, 2010). Sie basieren häufig auf Übersimplifizierungen der komplexen Zusammenhänge von Voraussetzungen, Prozessen und Ergebnissen des Lehrens und Lernens, etwa in Form pauschaler Annahmen über große, direkte und teilweise monokausal gedachte Einflüsse einzelner Variablen auf bildungsrelevante Outcomes.

Für die Entwicklung des QUEBEC wurden vier einschlägige Themen ausgewählt, die inhaltliche Bereiche der Bildungsforschung adressieren, zu denen häufig fragwürdige Überzeugungen vorliegen: (a) Effekte der *Klassengröße* auf relevante Outcomes wie Unterrichtsqualität,

Lernerfolg von Schülerinnen und Schülern, sowie die Belastung von Lehrkräften; (b) Effekte der *Klassenwiederholung* auf leistungsbezogene und motivationale Outcomes auf Individual-ebene sowie ihre Bedeutung auf der Schulsystemebene; (c) die Wirksamkeit *direkter Instruktion* als Unterrichtsmethode; sowie (d) Effekte des relativen Anteils an Erzieherinnen bzw. Grundschullehrerinnen für die Schulleistung und -laufbahn von Jungen (*Lehrerinnenanteil*). Grund für diese Auswahl war erstens, dass diese Themen in den letzten Jahren kontroverser Gegenstand des öffentlichen Diskurses waren. Unter entsprechenden Schlagworten finden sich hierzu im Internet zahlreiche Medienbeiträge, aber auch Verlautbarungen von Interessensgruppierungen und aus der Bildungspolitik. Zweitens repräsentieren diese Themen zentrale Forschungsfelder der empirischen Bildungsforschung, etwa die Lehr-Lernforschung, die Forschung zu pädagogischen Interventionen, zur Rolle individueller Lernvoraussetzungen und schulstruktureller Merkmale (Gräsel, 2015; Spinath, 2014). Drittens existiert zu ihnen eine relativ solide Befundlage, teilweise auch auf Basis von Metaanalysen, anhand derer sich die Klassifikation bestimmter Annahmen und Positionen als fragwürdig oder gar falsch gut begründen lässt. Im Folgenden skizzieren wir in aller gebotenen Kürze die öffentliche Debatte und den Stand der Forschung zu diesen vier Themenbereichen.

1. Bezüglich der Rolle der *Klassengröße* für Prozessmerkmale und Ergebnisse von Unterricht existieren in der Öffentlichkeit „stabile, aber unzutreffende Erwartungen“ (Bromme et al., 2014, S. 33) zu positiven Effekten kleinerer Klassen. Sie werden mit besserer Unterrichtsqualität und Lernbegleitung, reduzierter Belastung der Lehrkräfte und größeren Lerngewinnen in Verbindung gebracht. Der Stand der Forschung widerlegt solche Erwartungen jedoch weitgehend (Bromme et al., 2014). Demzufolge hat die Klassengröße allenfalls geringe durchschnittliche Effekte auf Schulleistung und dies nur für spezifische Gruppen (Bowne, Magnuson, Schindler, Duncan & Yoshikawa, 2017; Hattie, 2016;

Lankes & Carstensen, 2010; Li & Konstantopoulos, 2017; Watson, Handal & Maher, 2016). Auch in experimentellen Interventionsstudien gefundene Effekte, die als stärkste Evidenz für die Relevanz der Klassengröße gesehen werden könnten, fallen relativ gering aus (Hattie, 2016; Mosteller, 1995). Zudem zeigt sich, dass Lehrkräfte kleinere Klassen kaum für eine bessere Unterrichtsqualität nutzen und dass die Klassengröße weitgehend unabhängig von ihrem beruflichen Belastungserleben ist (Lankes & Carstensen, 2010).

2. *Klassenwiederholungen* – also die angeordnete Wiederholung einer Jahrgangsstufe bei Leistungsdefiziten – werden in der Öffentlichkeit, aber auch von Lehrkräften und Eltern als pädagogische Maßnahme mehrheitlich befürwortet. Hierfür existieren Befunde sowohl im deutschsprachigen als auch im internationalen Raum (Bali, Anagnostopoulos & Roberts, 2005; Ehmke, Drechsel & Carstensen, 2008; Roßbach & Tietze, 2010). Mit Klassenwiederholungen als Maßnahme werden Erwartungen auf unterschiedlichen Ebenen verknüpft: die betroffenen Schülerinnen und Schüler sollen über die zusätzliche Lernzeit leistungsbezogene, sozial-emotionale und motivationale Defizite aufholen können sowie von dem Druck befreit werden, zu den Leistungsschwächsten zu gehören (Ehmke et al., 2008). Auf Systemebene zielt die Klassenwiederholung auf die Verringerung der Leistungsheterogenität in Schulklassen ab (Roßbach & Tietze, 2010) und wird als Mittel zur Sicherung der Leistungsfähigkeit und Gerechtigkeit des Schulsystems gesehen (OECD, 2011). Die verfügbare Evidenz weist jedoch darauf hin, dass positive Effekte von Klassenwiederholung für die Betroffenen eher geringer und kurzfristiger Natur sind; teilweise wurden auch keine oder negative Effekte gefunden (Allen, Chen, Willson & Hughes, 2009; Ehmke, Sälzer, Pietsch, Drechsel & Müller, 2017; Jimerson, 2001; Jimerson & Brown, 2016; Klapproth et al., 2016). Auf der Systemebene scheinen Klassenwiederholungsquoten negativ mit erzielten Kompetenzniveaus und Bildungsgerechtigkeit assoziiert zu sein (OECD, 2011).

3. Unter dem Stichwort *Frontalunterricht* finden sich in der öffentlichen, aber auch in der wissenschaftlichen Debatte, häufig Positionen, die *traditionelle* Lehr-Lernformen der *direkten Instruktion* (z. B. Vorlesung, lehrerzentriertes Unterrichtsgespräch; Wiechmann, 2009) diskreditieren und im Gegensatz zu *aktiven* Lehr-Lernformen als weniger effektiv darstellen (Holmes, 2016; Renkl, 2015). In diesem Kontext mischen sich vielerlei theoretisch und empirisch fragwürdige Auffassungen. Dazu gehören pyramidenartige Vorstellungen der Effektivität von Lehr-Lernformen, für die keinerlei Evidenz vorliegt („Man behält 5 % aus einer Vorlesung, 75 % aus einer aktiven Übung, ...“; De Bruyckere, Kirschner & Hulshof, 2015); aber auch aus dem Konstruktivismus oder der Hirnforschung begründete Annahmen, die konstruktive und aktive Natur menschlicher Informationsverarbeitung erfordere Lernarrangements wie entdeckendes oder problemorientiertes Lernen. Grundlegend ist dabei die Fehlannahme, es gäbe Lehr-Lernformen, die *generell* (in-)effektiver seien als andere, ohne relevante Moderatorvariablen zu berücksichtigen (Renkl, 2015). Gegenüber solchen abwertenden Positionen existiert „massive empirische Evidenz“ (Renkl, 2015, S. 212), dass Formen der direkten Instruktion sehr effektiv für den Wissenserwerb sind, auch bei ungünstigen Lernvoraussetzungen (Adams & Engelmann, 1996; Borman, Hewes, Overman & Brown, 2003; Chodura, Kuhn & Holling, 2015; Hattie, 2009; Liem & Martin, 2016).

4. Die sogenannte Feminisierung des Bildungssystems – insbesondere der hohe bzw. gestiegene *Frauenanteil* im pädagogischen Personal elementarer Bildungseinrichtungen und Schulen – wird in der öffentlichen Debatte häufig als Ursache für Bildungsbenachteiligungen von Jungen ins Feld geführt (Driessen, 2007; Hannover & Kessels, 2011; Neugebauer, Helbig & Landmann, 2011). Der hohe Frauenanteil, so die Annahme, führe bei Jungen u. a. zu motivationalen und leistungsbezogenen Einbußen, reduzierten Bildungschancen, aber auch zu Problemen in der geschlechtsbezogenen Identitäts-

entwicklung. Disparitäten zu Ungunsten von Jungen sind in mehreren Dimensionen sehr gut empirisch belegt (Fiske, 2012; Hannover & Kessels, 2011; Helbig, 2010; Marsh, Martin & Cheng, 2008; Stanat, Bergann & Taraszow, 2018; Voyer & Voyer, 2014): So erzielen Jungen in Schulleistungsvergleichen konsistent deutlich schlechtere Lesekompetenzen als Mädchen (jedoch etwas bessere in Mathematik und Naturwissenschaften), erhalten bei gleicher Kompetenz schlechtere Noten sowie seltener eine Gymnasialempfehlung und erlangen infolgedessen seltener das Abitur. Allerdings findet sich kein Beleg dafür, dass solche Unterschiede auf den Frauenanteil im pädagogischen Personal zurückzuführen sind. Zwar existieren hierzu unseres Wissens noch keine Metaanalysen, weder international noch für den deutschsprachigen Bereich; keine der bisherigen Studien legt jedoch nennenswerte Effekte des Lehrerinnenanteils auf solche Disparitäten nahe (Besnard & Letarte, 2017; Coenen, Cornelisz, Groot, Maassen van den Brink & Van Klaveren, 2018; Driessen, 2007; Gustavsen, 2019; Helbig, 2010; Heyder & Kessels, 2015; Marsh et al., 2008; Schöps, Walter, Zimmer & Prenzel, 2006; Zusammenfassend: Hannover & Kessels, 2011).

Solche Fehlkonzepte zu Bildungsthemen finden sich nicht nur in der öffentlichen Debatte, sondern sind erfahrungsgemäß auch bei Studierenden verbreitet. Bislang gibt es jedoch nur vereinzelte empirische Studien zu ihrer Ausprägung (z. B. zu *Neuromythen*: Dekker et al., 2012). Um bildungsbezogene Fehlkonzepte und ihre potenziellen Auswirkungen auf Lernprozesse im Studium oder professionelles Handeln untersuchen zu können, bedarf es deshalb der Entwicklung themenbezogener Instrumente.

Die vorliegende Studie

Ziele der vorliegenden Studie waren die Entwicklung und erste psychometrische Evaluation des QUEBEC-Inventars sowie die Identifikation von Optimierungsbedarf für eine Weiterentwicklung. Im Besonderen sollten die folgenden Fragestellungen adressiert werden:

1. Welche dimensionale Struktur hat das Inventar und bildet es die vier themenspezifischen Skalen reliabel ab?
2. Ist hinreichende Messinvarianz gegeben, um valide Vergleiche zwischen Lehramtsstudierenden und anderen Studierenden anzustellen?
3. Lassen sich mit dem Inventar Unterschiede zwischen Lehramtsstudierenden, Studierenden anderer bildungsbezogener Studiengänge (Erziehungswissenschaft, Psychologie) und nicht-bildungsbezogener Studiengänge (z. B. Wirtschaftswissenschaften) erfassen?

Hinsichtlich Fragestellung 1 prüften wir faktoranalytisch, ob sich die interne Struktur des Inventars an den vier abgefragten theoretischen Themenbereichen orientiert. Zu Fragestellung 2 testeten wir das Vorliegen skalarer Messinvarianz als Voraussetzung für valide Vergleiche latenter Mittelwerte zwischen Lehramtsstudierenden und anderen Studierenden (Brown, 2015). Fragestellung 3 zielte auf die explorative Identifikation von Unterschieden in der Ausprägung von Fehlkonzepten zwischen Lehramtsstudierenden, Studierenden anderer bildungsbezogener Studiengänge und nicht-bildungsbezogener Studiengänge ab. Da hierzu bislang kaum empirische Befunde vorliegen, sollte die Untersuchung Hypothesen für weiterführende Studien generieren. In den Analysen wurde für das Studiensemester kontrolliert, um Entwicklungsunterschiede im Studienverlauf zu berücksichtigen.

Methoden

Design und Stichprobe

An der Online-Befragung nahmen $N=247$ Probanden teil, als Teilnahmeanreiz wurden Gutscheine verlost. Um sicherzustellen, dass nur gewissenhaft bearbeitete Fragebögen in die Analysen eingingen, wurden $n=30$ Fälle mit unplausiblen Bearbeitungszeiten ausgeschlossen (<2.5 bzw. >45 min). Insgesamt gingen die Daten von $N=217$ Studierenden in die Analysen ein. Ein Großteil der Befragten studierte ein Lehramt ($n=90$; 41,5 %) oder einen anderen Studiengang mit Bezug zu Bildungsfragen (Erziehungswissenschaft, Sozialpädagogik, Psychologie und soziale Arbeit;

$n=41$; 18,9%; im Folgenden: Päd/Psy). Die restlichen 39,6 % ($n=86$) waren Studierende aus Studiengängen ohne direkten Bezug zu Bildungsthemen, überwiegend aus den Wirtschaftswissenschaften (im Folgenden: WISO/Sonstige). Von den teilnehmenden Lehramtsstudierenden strebte die überwiegende Mehrheit die Lehramtszugänge Grundschule (60 %) und Gymnasium/Sekundarstufe 2 (30 %) an. Dies resultiert aus dem Zuschnitt der einbezogenen Universitäten in der Lehrerbildung. Die Studierenden befanden sich durchschnittlich im sechsten Fachsemester ($SD=3.3$).

Instrumente

Das QUEBEC-Inventar besteht aus insgesamt 20 Zielitems sowie 10 Distraktoritems, die Aussagen zu den vier genannten Bildungsthemen beinhalten (s. Anhang A). Zu jedem Item schätzen die Teilnehmenden ein, inwiefern sie die Aussage für zutreffend halten (*überhaupt nicht zutreffend* [1], *voll und ganz zutreffend* [6], 2–5 = keine Labels). Aufgabe ist also die subjektive Einschätzung der Richtigkeit der Aussagen, nicht eine normative Bewertung. Bei den 20 Zielitems handelt es sich um Aussagen, die im Konflikt zum bisherigen Stand der Forschung stehen und deshalb als Fehlkonzepte einzuschätzen sind. Die Distraktoritems beinhalten Aussagen, die entweder mit dem Stand der Forschung übereinstimmen oder die zumindest nicht als klar widerlegt gelten können (vgl. Dekker et al., 2012). Sie wurden eingefügt, damit Teilnehmende mit Vorwissen nicht unmittelbar alle Items als nicht zutreffend klassifizieren und in einen negativ-akquieszenten Antwortstil verfallen. In die Auswertung gehen ausschließlich die Zielitems ein.

Im Vorfeld der vorliegenden Studie führten wir einen kognitiven Prätest durch, um die Verständlichkeit der Items zu prüfen und das Material zu überarbeiten. Hierzu wurden Think-Aloud-Interviews mit $N=5$ Studierenden unterschiedlicher Fachrichtungen eingesetzt und zusätzlich Einschätzungen zur Verständlichkeit der Items abgefragt. Die Erhebung fand in Form von Einzelinterviews statt (ca. 45 Minuten Dauer). Auf einer sechsstufigen Likert-Skala wurden die Items als sehr verständlich bewertet ($M=5.76$, $SD=0.39$; Skala: *überhaupt nicht verständlich* [1] bis *sehr verständlich* [6]). Die Anmerkungen der Studierenden bestätigten die Verständlichkeit der Formulierungen. Kritische Kommentare hinsichtlich unklarer Formulierungen wurden aufgegriffen und Items entsprechend überarbeitet.

Analysen

Da es sich um eine Neuentwicklung handelt, erfolgte die Dimensionalitätsprüfung zu *Fragstellung 1* mit einer explorativen Faktorenanalyse (EFA) mit robuster FIML-Schätzung und obliquer Geomin-Rotation über die 20 Zielitems.¹ Entscheidungskriterien für die Anzahl der Faktoren waren die Parallelanalyse, der Screeplot sowie Velicer's MAP-Test (Field, 2012). Für die Reliabilitätsprüfung wurden Faktorreliabilitäten (McDonald's Ω) berechnet. McDonald's Ω ist hier Cronbach's α vorzuziehen, da die Ergebnisse der Faktorenanalysen darauf hinweisen, dass die Daten nur das Modell kongenerischer, nicht aber das Modell tau-äquivalenter Tests erfüllen (Eid, Gollwitzer & Schmitt, 2017; Furr, 2017). Zur Vorbereitung für *Fragstellung 2* wurde die EFA-Lösung in eine konfirmatorische Faktorenanalyse (CFA) überführt. Dies erfolgte erstens, um weitere Quellen von Misfit aufzudecken, da die CFA hierfür sensitiver ist als die EFA (Gerbing & Hamilton, 1996; Schumacker & Lomax, 2015); zweitens war dies für die Invarianzanalysen erforderlich. Es ist zu betonen, dass die CFA hier in einem modellgenerierenden Modus verwendet wurde (Brown, 2015) und die Ergebnisse keine konfirmatorische Prüfung des Messmodells darstellen. Für die CFA wurden schlecht passende Items (Hauptladungen $< .4$, Nebenladungen $> .3$, lokaler Misfit in Form mehrerer hoher Modifikationsindizes) sukzessive entfernt. Ziel war es, die besten Items für die weitere Entwicklung des Instruments zu identifizieren. Das finale CFA-Modell wurde auf Invarianz für den Vergleich Lehramts- vs. andere Studierende geprüft. Dies erfolgte durch hierarchische Tests konfiguraler (gleiche Struktur des Messmodells), metrischer (+ Gleichheit von Faktorladungen) und skalarer Invarianz (+ Gleichheit von Itemintercepts; Brown, 2015). Kriterium für die Geltung von Invarianz ist, dass die jeweiligen Parameterrestriktionen im Vergleich zur vorausgehenden Invarianzstufe keine signifikante Verschlechterung des Modellfits bewirken. Dabei sprechen ein signifikanter χ^2 -Differenztest bzw. eine Veränderung im Fit von $\Delta CFI \geq .01$ und $\Delta RMSEA \geq 0.015$ gegen das Vorliegen von Invarianz (Brown, 2015; Chen, 2007). *Fragstellung 3* wurde in einem Strukturgleichungsmodell geprüft. Das spezifizierte Modell enthielt die vier untersuchten Fehlkonzepte als abhängige latente Variablen; als Prädiktoren fungierten das Studiensemester sowie die Studiengangzugehörigkeit (manifeste Dummy-Variablen für die Studierendengruppen *Päd/Py* und *WISO/Sonstige* mit *Lehramt* als Referenzkategorie). Alle Analysen wurden in R bzw. *Mplus* 8.1 (MLR-Schätzung) durchgeführt.

Ergebnisse

Deskriptive Befunde

Tabelle 1 enthält die Mittelwerte (Schwierigkeiten) und Standardabweichungen der Zielitems. Die meisten Items zeigten schiefe Verteilungen. Es waren jedoch keine extremen Boden- bzw. Deckeneffekte oder Varianzeinschränkung vorzufinden, sodass keine Items vorab ausgeschlossen wurden. Gemessen an der Antwortskala lagen alle Items zur Klassengröße und zur direkten Instruktion über der theoretischen Skalenmitte, die zum Lehrerinnenanteil darunter. Die Mittelwerte der Items zur Klassenwiederholung variierten relativ eng um die Mitte des Antwortformats.

Fragstellungen 1 und 2: Dimensionalität, Reliabilität und Messinvarianz

Eine erste Prüfung indizierte eine gute Eignung der Daten für eine Faktorenanalyse (Barlett-Test auf Sphärizität: $\chi^2 [190] = 1526.83, p < .001$; KMO-Koeffizient = $.77$; keine Hinweise auf Multikollinearität). Hinsichtlich der Faktorenanzahl wiesen die Parallelanalyse und der Screeplot auf eine Vier-Faktorenlösung hin, der MAP-Test dagegen auf drei Faktoren. Das Vier-Faktorenmodell passte jedoch statistisch signifikant besser auf die Daten als das Drei-Faktorenmodell ($\Delta \chi^2 [17] = 89.85, p < .001$), hatte einen insgesamt akzeptablen Modell-Fit ($\chi^2 [116] = 223.85, p < .001$; RMSEA = $.065$; CFI = $.913$; SRMR = $.039$) und die Lösung bildete eher eine Einfachstruktur ab. In der Gesamtschau der Ergebnisse entschieden wir uns deshalb für vier Faktoren. Die Primärladungen der Items (Tab. 1) spiegelten die vier Themenbereiche der Fehlkonzepte

¹ Die Distraktoritens wurden in einer zusätzlichen Analyse zusammen mit den Zielitems analysiert. Es resultierte dieselbe Faktorenstruktur. Alle Distraktoren luden primär auf ihre jeweiligen themenspezifischen Faktoren. Die Primärladungen waren hoch ($\lambda .49 < .72$), mit Ausnahme eines Distraktors D3 zur Klassenwiederholung ($\lambda = .29$). Alle Nebenladungen waren $< .3$. Deskriptive Statistiken zu den Distraktoren sind Anhang B zu entnehmen.

Tab. 1: Deskriptive Statistiken und standardisierte Faktorladungen der EFA und der CFA für die finale Itemauswahl

Item	M	SD	Klassengröße		Klassenwiederholung		Direkte Instruktion		Lehrerinnenanteil	
			EFA	CFA	EFA	CFA	EFA	CFA	EFA	CFA
KG 1	4.60	1.25	.84	.78	-.03		-.01		-.01	
KG 2	4.33	1.18	.66	.70	.07		-.02		.01	
KG 4	4.68	1.07	.43	.49	.00		.14		.03	
KG 3	5.21	1.03	.40	.47	.06		.20		-.07	
KW 5	2.92	1.24	-.01		.72	.77	.03		-.03	
KW 1	3.75	1.18	-.04		.67	.55	.00		-.18	
KW 3	2.51	1.26	.04		.58		-.27		.05	
KW 4	3.68	1.23	.05		.48	.57	.08		-.04	
KW 2	2.97	1.17	.09		.43		-.15		-.01	
KW 6	2.87	1.23	-.05		.42		.09		.11	
DI 1	4.65	1.29	.06		-.10		.62	.60	.06	
DI 3	3.81	1.28	.10		.06		.51	.55	.05	
DI 5	4.22	1.44	.00		.08		.51	.55	.02	
DI 2	4.31	1.09	-.07		-.34		.44		-.07	
DI 4	4.76	1.09	.04		.05		.38		.01	
LA 2	1.87	1.07	-.03		.01		-.05		.91	.92
LA 4	1.93	1.02	-.01		.06		.05		.87	.83
LA 1	1.86	1.00	.01		-.01		-.05		.87	.89
LA 5	1.79	0.93	.06		-.07		.02		.83	
LA 3	2.31	1.25	-.05		.01		.05		.67	.67
Ω			.69	.72	.70	.67	.63	.59	.88	.89

Anmerkungen: N=217; M (Itemschwierigkeit); 6-stufige Likert-Skala (überhaupt nicht zutreffend [1] bis voll und ganz zutreffend [6]), Ω = Faktorreliabilität.

wider, sodass die Faktoren hier entsprechend bezeichnet werden. Die Faktorreliabilitäten (Tab. 1, McDonald's Ω) fielen akzeptabel bis gut aus, mit Ausnahme geringer Reliabilität der Subskala *direkte Instruktion*.

Auf Basis der anfänglichen Lösung wurden zwei Items des Faktors direkte Instruktion (DI 2, DI 4) sowie ein Item des Faktors Klassenwiederholung (KW 3) aufgrund niedriger Primärladungen bzw. Nebenladungen ausgeschlossen. Anschließend wurden im Rahmen der modellgenerierenden CFAs zunächst die Messmodelle der einzelnen Faktoren separat und anschließend das gemeinsame Messmodell geprüft (Schumacker & Lomax, 2015). Dabei ergab die Betrachtung der Lösungen lokalen Misfit für die Items KW 2, KW 6 und LA 5, jeweils in Form mehrerer hoher Modifikationsindizes zu korre-

lierten Residuen. Dies stellt eine Verletzung der Annahme der Eindimensionalität dar, weshalb die Items aus den weiteren Analysen ausgeschlossen wurden. Die resultierende Lösung hatte einen sehr guten Fit ($\chi^2 [71] = 74.96, p = .35 [n.s.]$; RMSEA = .016; CFI = .995; SRMR = .043) und durchweg substantielle standardisierte Faktorladungen (Tab. 1). Die Reliabilitäten fielen trotz der Itemkürzungen vergleichbar aus.

Tabelle 2 fasst die deskriptiven Statistiken und Korrelationen der Subskalen zusammen. Dabei stehen die Mittelwerte für die durchschnittliche Zustimmung zu den Fehlkonzepten bezüglich der einzelnen Themenbereiche. Fehlkonzepte zur Klassengröße korrelierten in mittlerer Höhe signifikant mit denen zur Klassenwiederholung und zur direkten Instruktion. Alle anderen Korrelationen waren nicht signifikant.

Tab. 2: Ausprägungen (M , 95 %-KI, SD , min, max) und latente Korrelationen der Fehlkonzepte zu den vier Themenbereichen (finale Itemauswahl)

	M	[95 % KI]	SD	min	max	1	2	3
1 Klassengröße	4.71	[4.60; 4.82]	0.83	2.00	6.00			
2 Klassenwiederholung	3.45	[3.33; 3.58]	0.94	1.33	6.00	.37*		
3 Direkte Instruktion	4.23	[4.10; 4.36]	0.99	1.00	6.00	.31*	.10	
4 Lehrerinnenanteil	1.99	[1.87; 2.12]	0.94	1.00	6.00	.06	.00	.14

Anmerkung: * $p \leq .05$.

Tab. 3: Messinvarianztests: Lehramts- vs. andere Studierende

Modell	χ^2	df	RMSEA	CFI	$\Delta\chi^2$	Δdf	p
M1 Konfigural	162.458	142	.036	.975			
M2 Metrisch	166.886	152	.030	.982			
M1 vs. M2					6.212	10	.80
M3 Skalar	184.188	162	.036	.972			
M2 vs. M3					17.283	10	.07

Die Ergebnisse der Messinvarianztests sind Tabelle 3 zu entnehmen. Alle Modelle wiesen insgesamt einen akzeptablen Modell-Fit auf. Die sukzessive eingeführten Gleichheitsrestriktionen auf die Faktorladungen (metrische Invarianz) und die Itemintercepts (skalare Invarianz) führten jeweils zu keiner statistisch signifikanten Verschlechterung im χ^2 -Wert. Ebenso blieben die jeweiligen Veränderungen im CFI und RMSEA unterhalb der festgelegten Cut-Off-Werte. Die Inspektion von Modifikationsindizes lieferte keine Hinweise auf lokale Fehlspezifikationen mehr. Insgesamt kann damit die Annahme skalarer Messinvarianz beibehalten werden, sodass die Fehlkonzepte über die Gruppen hinweg gleich interpretierbar und Vergleiche latenter Mittelwerte zulässig sind.

Fragestellung 3: Vergleich von Lehramtsstudierenden mit anderen bildungsbezogenen und nicht-bildungsbezogenen Studierendengruppen

Abbildung 1 illustriert Mittelwertsunterschiede in der Zustimmung zu den Fehlkonzepten zwischen den drei untersuchten Studierendengruppen. Deskriptiv sind bei der nicht-bildungsbe-

zogenen Studierendengruppe Fehlkonzepte zur Klassengröße, Klassenwiederholung und zur direkten Instruktion stärker ausgeprägt als bei den anderen beiden Gruppen. Bei den letzten beiden Inhaltsbereichen fanden sich zudem Unterschiede zwischen den Lehramts- und den Päd/Psy-Studierenden: Fehlkonzepte zur Klassenwiederholung sind bei den Lehramtsstudierenden stärker ausgeprägt, solche zur direkten Instruktion bei den Päd/Psy-Studierenden. Hinsichtlich der Einschätzungen zu Effekten des Lehrerinnenanteils liegen alle Gruppen gleichauf.

Die gefundenen deskriptiven Unterschiede wurden im Rahmen eines Strukturgleichungsmodells unter Kontrolle des Semesters getestet. Die Variable zum Lehrerinnenanteil wurde trotz der geringen deskriptiven Unterschiede einbezogen, um eine potenzielle Suppression zu prüfen. Das Modell wies einen guten Fit auf ($\chi^2[101]=126.56$, $p = .04$, RMSEA=.034, CFI=.970, SRMR=.046), die Ergebnisse finden sich in Tabelle 4. Dabei sind die Effekte der Studiengangszugehörigkeit als standardisierte Mittelwertsdifferenzen zwischen der jeweiligen Gruppe und den Lehramtsstudierenden zu interpretieren (analog zu Cohen's d ; Brown, 2015). Die Unterschiede zwischen den Lehramts- und den Päd/Psy-Studierenden fielen bei den Fehlkonzepten zur

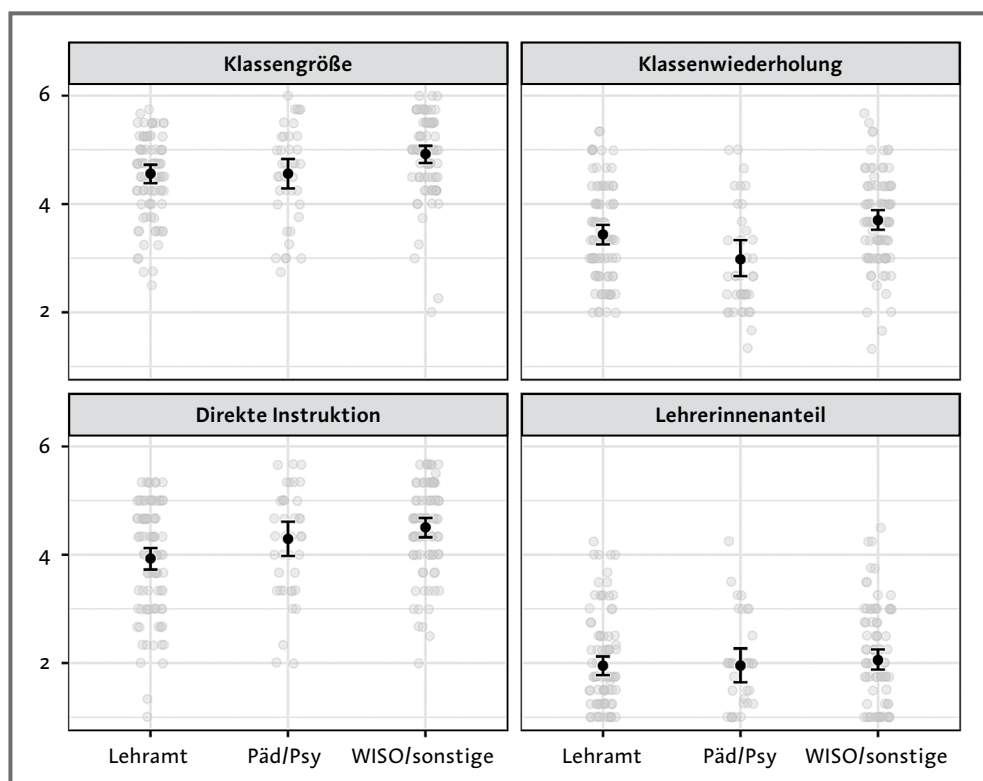


Abb. 1: Fehlkonzepte nach Studiengängen (M und 95 %-KI, höhere Werte = höhere Zustimmung).

Klassenwiederholung und zur direkten Instruktion mit mittleren Effektstärken statistisch signifikant aus. Bei den Studierenden nicht-bildungsbezogener Studiengänge waren die Fehlkonzepte zur Klassengröße und zur direkten Instruktion im Vergleich zur Lehramtsgruppe signifikant höher ausgeprägt, mit einer mittleren bzw. hohen Effektstärke. Für Fehlkonzepte zum Lehre-

rinnenanteil fanden sich keine signifikanten Unterschiede. Die Kovariate *Semester* wurde für die Fehlkonzepte zur direkten Instruktion und zum Lehrerinnenanteil statistisch signifikant, allerdings fielen die entsprechenden Effekte klein aus. Tendenziell scheinen Erstere im Studienverlauf abzunehmen, Letztere werden offenbar leicht verstärkt.

Tab. 4: Latente Regression der Fehlkonzepte auf Studiengang (Lehramt = Referenzgruppe) und Fachsemester

Prädiktor	Klassen- größe		Klassen- wiederholung		Direkte Instruktion		Lehrerinnen- anteil	
	β	p	β	p	β	p	β	p
Päd/Psy ^a	-.09	.70	-.60	.02	.63	.02	-.09	.69
WISO/Sonstige ^a	.48	.01	.29	.10	.84	<.001	.15	.36
Semester	.08	.32	-.06	.50	-.20	.02	.17	.04
R^2	.07		.11		.16		.03	

Anmerkung: ^a = Dummy-kodiert; β = teilstandardisierte Koeffizienten für Dummy-Variablen, vollstandardisierte für Semester.

Diskussion

In der vorliegenden Studie berichteten wir die Entwicklung des QUEBEC-Inventars zur Erfassung von Fehlkonzepten über Bildungsthemen sowie erste explorative Ergebnisse zu seiner psychometrischen Qualität. Bisherige Studien zu bildungsbezogenen Fehlkonzepten fokussierten in der Regel nur einzelne und relativ spezifische Themen (z. B. Neuromythen; Dekker et al., 2012). Mit dem QUEBEC wurde nun erstmalig ein Inventar entwickelt, das die Ausprägung fragwürdiger Überzeugungen zu verschiedenen Bildungsthemen fokussiert, die in der öffentlichen Diskussion häufig thematisiert werden (Bromme et al., 2014; Spinath, 2014). Zudem wurde in der vorliegenden Studie ein erster explorativer Vergleich der Fehlkonzeptausprägungen zwischen Lehramtsstudierenden und anderen Studierenden aus Studiengängen mit und ohne Bildungsbezug vorgenommen.

Zu Fragestellung 1 fielen die Befunde hinsichtlich der Dimensionalität des Instruments konsistent mit den vier untersuchten inhaltlichen Bereichen fragwürdiger Überzeugungen aus. Diese Themenbereiche ließen sich in den Faktorenanalysen empirisch als eigenständige Dimensionen voneinander abgrenzen, auch wenn sie untereinander Korrelationen aufweisen. Daraus lässt sich schließen, dass die Ausprägungen der untersuchten Fehlkonzepte tatsächlich einer themenbezogenen Struktur folgen. Denkbar wäre auch eine stärkere Bündelung im Sinne eines übergeordneten Faktors gewesen. Dabei zeigten sich durchaus Zusammenhänge zwischen den Fehlkonzepten zur Klassengröße, zur direkten Instruktion und zur Klassenwiederholung, also in den Faktoren, die sich im weiteren Sinne mit Voraussetzungen, Methoden und Ergebnissen von Unterricht befassen. Dies deutet darauf hin, dass es möglicherweise Cluster von Fehlkonzepten gibt. Dennoch fallen diese Korrelationen eher moderat aus, sodass die Fehlkonzepte inhaltlich voneinander abgrenzbar sind. Eine zukünftige Erweiterung des QUEBEC um weitere Themen könnte dazu beitragen, solche Cluster besser herauszuarbeiten. Zudem könnten vertiefende Studien auch

die subjektiven Theorien hinter den Fehlkonzepten explizieren und sie auf typische Problem-bereiche hin untersuchen (monokausales und deterministisches Denken etc.).

Die Reliabilitäten der Subskalen fielen insgesamt akzeptabel bis gut aus, für die Skala zur direkten Instruktion besteht jedoch Überarbeitungsbedarf. Ein potenzieller Grund für deren geringere Reliabilität ist, dass sie als einzige einen dezidiert fachsprachlichen Begriff („direkte Instruktion“) enthält. Dieser wurde zwar im Fragebogen erläutert, möglicherweise blieben aber dennoch Interpretationsdifferenzen bestehen, die sich reliabilitätsmindernd auswirkten.

Hinsichtlich Fragestellung 2 zeigte sich, dass die Befunde der Messinvarianztestungen konsistent mit der Annahme skalarer Invarianz waren. Insofern ist davon auszugehen, dass die Fehlkonzepte von den verglichenen Studierendengruppen äquivalent interpretiert werden können und die Messeigenschaften der Items hinreichend vergleichbar sind (Brown, 2015).

Bezüglich Fragestellung 3 deuten unsere Analysen darauf hin, dass es tatsächlich systematische Unterschiede in der Ausprägung der Fehlkonzepte zwischen Lehramtsstudierenden, Studierenden anderer bildungsbezogener Studiengänge und solchen nicht-bildungsbezogener Studiengänge gibt. Insofern erwies sich der QUEBEC auch als sensitiv für die Erfassung solcher Differenzen. Aus den explorativen Befunden lassen sich Hypothesen für weiterführende Studien ableiten. So zeigte sich insgesamt, dass die nicht-bildungsbezogenen Studierenden jeweils die höchsten Fehlkonzeptausprägungen aufwiesen. Dies kann als Effekt entsprechenden Wissenserwerbs in den bildungsbezogenen Studiengängen interpretiert werden. Die Unterschiede zwischen den Lehramtsstudierenden und den Päd/Psy-Studierenden liefern aber zudem Hinweise, dass die Ausprägung der Fehlkonzepte neben dem Wissen auch abhängig von Einstellungen und Wahrnehmungen der Bildungspraxis sein könnte. Beispielsweise kann die stärkere Ausprägung der Fehlkonzepte zur Klassenwiederholung bei den Lehramtsstudierenden darauf zurückzuführen sein, dass sie

Klassenwiederholungen stärker normativ befürworten (vgl. Ehmke et al., 2008) und dass sie diese Maßnahme als nützlich für ihren späteren Beruf sowie als konsistent mit der existierenden Bildungspraxis wahrnehmen. Analoges lässt sich für die Unterschiede in den Fehlkonzepten zur direkten Instruktion argumentieren. Solche Hypothesen könnten in weiterführenden Studien unter Einsatz des QUEBEC überprüft werden.

Bei den Fehlkonzepten zum Lehrerinnenanteil fällt auf, dass die Zustimmungen bei allen Studierendengruppen vergleichsweise gering ausfallen. Dies könnte ein Hinweis dafür sein, dass diese Fehlkonzepte bei Studierenden weniger verbreitet sind als in der öffentlichen Debatte. Über Gründe dafür können wir zum jetzigen Stand nur spekulieren. Möglicherweise herrscht aufgrund vieler aktueller Gender-Debatten eine höhere Sensibilität für diesen Themenbereich.

Neben den bereits angesprochenen Punkten sind weitere Limitationen der vorliegenden Studie zu berücksichtigen. Insbesondere handelte es sich um eine größtenteils explorative Querschnittsuntersuchung, deren Ergebnisse mit ausreichend großen Stichproben repliziert werden müssen. Dabei könnten in größeren Studien auch detailliertere Messinvarianztestungen zwischen verschiedenen Studierendengruppen vorgenommen werden, als es hier möglich war. Insbesondere konnte hier aufgrund der Stichprobengröße nicht innerhalb der anderen Studierenden (Studiengänge mit vs. ohne Bezug zu Bildungsthemen) differenziert werden. Weiterhin steht noch eine umfassendere Validierung des Instruments aus. In dieser Studie standen neben der theoriekonformen Testkonstruktion (Testinhalt) die empirische Prüfung der dimensional Struktur, die Sensitivität für Unterschiede sowie die Messinvarianz als Aspekte der Konstruktvalidität im Vordergrund (vgl. AERA, APA & NCME, 2014; Eid et al., 2017). Eine weitere Prüfung des QUEBEC nach einschlägigen Aspekten der Konstruktvalidität soll Gegenstand nachfolgender Studien werden.

Trotz dieser Einschränkungen sind wir der Ansicht, dass der QUEBEC ein sowohl für die Forschung als auch für den Einsatz in der Hochschullehre vielversprechendes Instrument ist. Im

Rahmen der Forschung sehen wir spezifische Einsatzbereiche in der Untersuchung der Rolle von Prädiktoren von Fehlkonzepten, ihrer Entwicklung im Studienverlauf sowie in Interventionsmaßnahmen. Beispielsweise scheint ein unzureichendes Verständnis von Evidenz, Kausalität und Wahrscheinlichkeit die Entwicklung und die Persistenz von Fehlkonzepten zu begünstigen (Bensley & Lilienfeld, 2017; De Bruyckere et al., 2015). Es wäre zu überprüfen, inwiefern dies auch auf bildungsbezogene Fehlkonzepte zutrifft. Darüber hinaus sollte der Einfluss von fragwürdigen Überzeugungen auf die Rezeption, Bewertung, Akzeptanz und Nutzung von bildungswissenschaftlichen Forschungsbefunden weiter untersucht werden (vgl. Bauer, Berthold, Hefter, Prenzel & Renkl, 2017; Gollwitzer et al., 2014; Hetmanek et al., 2015). Hierbei stellt sich insbesondere die Frage, unter welchen Bedingungen solche Fehlannahmen Verzerrungen begünstigen und welche Rolle darauf bezogene Einstellungen spielen. Auf dieser Basis können Ansätze zum *Conceptual Change* bei Lehramtsstudierenden erforscht und erprobt werden. Dabei stellt sich nicht nur die Frage, unter welchen Bedingungen *Conceptual Change* zu erwarten ist, sondern auch wie der Nutzen und die Attraktivität alternativer, evidenzbasierter Konzepte vermittelt werden können (Posner, Strike, Hewson & Gertzog, 1982).

Literatur

- Adams, G. L. & Engelmann, S. (1996). *Research on direct instruction: 25 years beyond DISTAR*. Seattle: Educational Achievement Systems.
- AERA, APA & NCME (2014). *Standards for educational and psychological testing*, 11–31.
- Allen, C. S., Chen, Q., Willson, V. L. & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multilevel analysis. *Educational Evaluation and Policy Analysis*, 20(10), 215–229. <https://dx.doi.org/10.3102/0162373709352239>
- Bali, V. A., Anagnostopoulos, D. & Roberts, R. (2005). Toward a political explanation of grade retention. *Educational Evaluation and Policy Analysis*, 27(2), 133–155. <https://dx.doi.org/10.3102/01623737027002133>
- Bauer, J., Berthold, K., Hefter, M. H., Prenzel, M. & Renkl, A. (2017). Wie können Lehrkräfte und ihre Schülerinnen und Schüler lernen, fragile Evidenz zu verstehen und zu nutzen? *Psychologische Rundschau*, 68(3), 188–192. <https://dx.doi.org/10.1026/0033-3042/a000363>

- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9, 469–520. <https://dx.doi.org/10.1007/s11618-006-0165-2>
- Bensley, D. A. & Lilienfeld, S. O. (2017). Psychological misconceptions: Recent scientific advances and unresolved issues. *Current Directions in Psychological Science*, 26(4), 377–382. <https://dx.doi.org/10.1177/0963721417699026>
- Besnard, T. & Letarte, M. J. (2017). Effect of male and female early childhood education teacher's educational practices on children's social adaptation. *Journal of Research in Childhood Education*, 31(3), 453–464. <https://dx.doi.org/10.1080/02568543.2017.1319445>
- Borman, G. D., Hewes, G. M., Overman, L. T. & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of educational research*, 73(2), 125–230. <https://dx.doi.org/10.3102/00346543073002125>
- Bowne, J. B., Magnuson, K. A., Schindler, H. S., Duncan, G. J. & Yoshikawa, H. (2017). A meta-analysis of class sizes and ratios in early childhood education programs: Are thresholds of quality associated with greater impacts on cognitive, achievement, and socioemotional outcomes? *Educational Evaluation and Policy Analysis*, 39(3), 407–428. <https://dx.doi.org/10.3102/0162373716689489>
- Bromme, R., Prenzel, M. & Jäger, M. (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. In R. Bromme & M. Prenzel (Hrsg.), *Von der Forschung zur evidenzbasierten Entscheidung: die Darstellung und das öffentliche Verständnis der empirischen Bildungsforschung. Zeitschrift für Erziehungswissenschaft*, 27 (Sonderheft), 3–54. Wiesbaden: VS Verlag. <https://dx.doi.org/10.1007/s11618-014-0514-5>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York, NY: Guilford.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504. <https://dx.doi.org/10.1080/10705510701301834>
- Chinn, C. A. & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching*, 35(6), 623–654. [https://dx.doi.org/10.1002/\(SICI\)1098-2736\(199808\)35:6<623::AID-TEA3>3.0.CO;2-O](https://dx.doi.org/10.1002/(SICI)1098-2736(199808)35:6<623::AID-TEA3>3.0.CO;2-O)
- Chodura, S., Kuhn, J. T. & Holling, H. (2015). Interventions for children with mathematical difficulties. *Zeitschrift für Psychologie*, 223(2), 129–144. <https://dx.doi.org/10.1027/2151-2604/a000211>
- Christodoulou, D. (2014). *Seven myths about education*. London: Routledge. <https://dx.doi.org/10.4324/9781315797397>
- Coenen, J., Cornelisz, I., Groot, W., Maassen van den Brink, H. & Van Klaveren, C. (2018). Teacher characteristics and their effects on student test scores: A systematic review. *Journal of Economic Surveys*, 32(3), 848–877. <https://dx.doi.org/10.1111/joes.12210>
- De Bruyckere, P., Kirschner, P. A. & Hulshof, C. D. (2015). *Urban myths about learning and education*. London: Academic Press. <https://dx.doi.org/10.1016/B978-0-12-801537-7.00003-2>
- Dekker, S., Lee, N. C., Howard-Jones, P. & Jolles, J. (2012). Neuromyths in education: Prevalence and predictors of misconceptions among teachers. *Frontiers in Psychology*, 3, 429–436. <https://dx.doi.org/10.3389/fpsyg.2012.00429>
- Driessen, G. (2007). The feminization of primary education: effects of teachers' sex on pupil achievement, attitudes and behavior. *International Review of Education*, 53(2), 183–203. <https://dx.doi.org/10.1007/s11159-007-9039-y>
- Ehmke, T., Drechsel, B. & Carstensen, C. H. (2008). Klassenwiederholen in PISA-I-Plus: Was lernen Sitzbleiber in Mathematik dazu? *Zeitschrift für Erziehungswissenschaft*, 11, 368–387. <https://dx.doi.org/10.1007/s11618-008-0033-3>
- Ehmke, T., Sälzer, C., Pietsch, M., Drechsel, B. & Müller, K. (2017). Kompetenzerwicklung im Schuljahr nach PISA 2012: Effekte von Klassenwiederholungen. *Zeitschrift für Erziehungswissenschaft*, 20(2), 99–124. <https://dx.doi.org/10.1007/s11618-017-0752-4>
- Eid, M., Gollwitzer, M. & Schmitt, M. (2017). *Statistik und Forschungsmethoden*. Weinheim: Beltz.
- Field, A. (2012). *Discovering statistics using R*. London: Sage.
- Fiske, E. B. (2012). *World atlas of gender equality in education*. Paris: UNESCO.
- 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 (Eds.), *APA educational psychology handbook* (Vol. 2, pp. 471–499). APA. <https://dx.doi.org/10.1037/13274-019>
- Furr, R. M. (2017). *Psychometrics: an introduction*. Los Angeles, LA: Sage.
- Gardner, R. M. & Brown, D. L. (2013). A test of contemporary misconceptions in psychology. *Learning and Individual Differences*, 24, 211–215. <https://dx.doi.org/10.1016/j.lindif.2012.12.008>
- Gerbing, D. W. & Hamilton, J. G. (1996). Viability of exploratory factor analysis as a precursor to confirmatory factor analysis. *Structural Equation Modeling*, 3, 62–72. <https://dx.doi.org/10.1080/10705519609540030>
- Gill, M. G. & Fives, H. (2015). Introduction. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teachers' beliefs* (pp. 87–105). New York, NY: Routledge.
- Gollwitzer, M., Rothmund, T., Klimmt, C., Nauroth, P. & Bender, J. (2014). Gründe und Konsequenzen einer verzerrten Darstellung und Wahrnehmung sozialwissenschaftlicher Forschungsbefunde: Das Beispiel der „Killerspiele-Debatte“. *Zeitschrift für Erziehungswissenschaft*, 17(4), 101–117. <https://dx.doi.org/10.1007/s11618-014-0511-8>
- Gräsel, C. (2015). Was ist Empirische Bildungsforschung? In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung: Strukturen und Methoden* (S. 15–30). Wiesbaden: Springer. https://dx.doi.org/10.1007/978-3-531-19992-4_1
- Groeben, N., Wahl, D., Schlee, J. & Scheele, B. (1988). *Das Forschungsprogramm Subjektive Theorien*. Tübingen: Francke.
- Gustavsen, A. M. (2019). Gender differences in academic achievement: A matter of contextual classroom influence? *International Journal of Research*, 8(1), 1–20. <https://dx.doi.org/10.5861/ijrse.2018.2013>
- Hannover, B. & Kessels, U. (2011). Sind Jungen die neuen Bildungsverlierer? Empirische Evidenz für Geschlechterdisparitäten zuungunsten von Jungen und Erklärungsansätze. *Zeitschrift für Pädagogische Psychologie*, 25, 89–103. <https://dx.doi.org/10.1024/1010-0652/a000039>
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge. <https://dx.doi.org/10.4324/9780203887332>

- Hattie, J. (2016). Class size. In J. Hattie E. M. Andermann (Eds.), *International guide to student achievement* (pp. 131–133). London: Routledge.
- Helbig, M. (2010). Sind Lehrerinnen für den geringeren Schulerfolg von Jungen verantwortlich? *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 62(1), 93–111. <https://dx.doi.org/10.1007/s11577-010-0095-0>
- Hetmanek, A., Wecker, C., Gräsel, C., Kiewewetter, J., Trempler, K. & Fischer, M.R. (2015). Wozu nutzen Lehrkräfte welche Ressourcen? Eine Interviewstudie zur Schnittstelle zwischen bildungswissenschaftlicher Forschung und professionellem Handeln im Bildungsbereich. *Unterrichtswissenschaft*, 43(3), 194–210.
- Heyder, A. & Kessels, U. (2015). Do teachers equate male and masculine with lower academic engagement? How students' gender enactment triggers gender stereotypes at school. *Social Psychology of Education*, 18(3), 467–485. <https://dx.doi.org/10.1007/s11218-015-9303-0>
- Holmes, J. D. (2016). *Great Myths of education and learning*. Hoboken: John Wiley & Sons. <https://dx.doi.org/10.1002/9781118760499>
- Jimerson, S.R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30(3), 420–437.
- Jimerson, S.R. & Brown, J.A. (2016). Grade retention. In J. Hattie E. M. Andermann (Eds.), *International guide to student achievement* (pp. 140–142). London: Routledge.
- Kagan, D. M. (1992). Implication of Research on Teacher Belief. *Educational Psychologist*, 27(1), 65–90. https://dx.doi.org/10.1207/s15326985ep2701_6
- Kane, R., Sandretto, S. & Heath, C. (2002). Telling Half the Story: A Critical Review of Research on the Teaching Beliefs and Practices of University Academics. *Review of Educational Research*, 72(2), 177–228. <https://dx.doi.org/10.3102/00346543072002177>
- Keil, F.C. (2010). The feasibility of folk science. *Cognitive Science*, 34(5), 826–862. <https://dx.doi.org/10.1111/j.1551-6709.2010.01108.x>
- Kirschner, P.A. & De Bruyckere, P. (2017). The myths of the digital native and the multitasker. *Teaching and Teacher Education*, 67, 135–142. <https://dx.doi.org/10.1016/j.tate.2017.06.001>
- Klapproth, F., Schaltz, P., Brunner, M., Keller, U., Fischbach, A., Ugen, S. & Martin, R. (2016). Short-term and medium-term effects of grade retention in secondary school on academic achievement and psychosocial outcome variables. *Learning and Individual Differences*, 50, 182–194. <https://dx.doi.org/10.1016/j.lindif.2016.08.014>
- König, J., Darge, K. & Schreiber, M. (2012). Teachers' beliefs about retention: Effects on teaching quality. In J. König (Ed.), *Teachers' pedagogical beliefs* (pp. 191–204). Münster: Waxmann.
- König, J. & Pflanzl, B. (2016). Is teacher knowledge associated with performance? On the relationship between teachers' general pedagogical knowledge and instructional quality. *European Journal of Teacher Education*, 39(4), 419–436. <https://dx.doi.org/10.1080/02619768.2016.1214128>
- Kunter, M., Kunina-Habenicht, O., Baumert, J., Dicke, T., Holzberger, D., Lohse-Bossenz, H., ... Terhart, E. (2017). Bildungswissenschaftliches Wissen und professionelle Kompetenz in der Lehramtsausbildung. In C. Gräsel & K. Trempler (Hrsg.), *Entwicklung von Professionalität pädagogischen Personals* (S. 37–54). Wiesbaden: Springer VS. https://dx.doi.org/10.1007/978-3-658-07274-2_3
- Lankes, E.-M. & Carstensen, C. H. (2010). Kann man große Klassen erfolgreich unterrichten? In W. Bos, S. et al. (Hrsg.), *IGLU 2006: die Grundschule auf dem Prüfstand* (S. 121–142). Münster: Waxmann.
- Li, W. & Konstantopoulos, S. (2017). Does class-size reduction close the achievement gap? Evidence from TIMSS 2011. *School Effectiveness and School Improvement*, 28(2), 292–313. <https://dx.doi.org/10.1080/09243453.2017.1280062>
- Liem, G. A. D. & Martin, A. J. (2016). Direct instruction. In J. Hattie & E. M. Andermann (Eds.), *International guide to student achievement* (pp. 366–368). London: Routledge.
- Marsh, H. W., Martin, A. J. & Cheng, J. H. (2008). A multilevel perspective on gender in classroom motivation and climate: Potential benefits of male teachers for boys? *Journal of Educational Psychology*, 100(1), 78–95. <https://dx.doi.org/10.1037/0022-0663.100.1.78>
- Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children*, 5, 113–127. <https://dx.doi.org/10.2307/1602360>
- Murphy, P. K. & Mason, L. (2012). Changing knowledge and beliefs. In P.A. Alexander & P.H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 305–324). New York, NY: Routledge.
- Neugebauer, M., Helbig, M. & Landmann, A. (2011). Unmasking the myth of the same-sex teacher advantage. *European Sociological Review*, 27(5), 669–689. <https://dx.doi.org/10.1093/esr/jcq038>
- OECD (2011). Wenn Schüler Klassen wiederholen oder die Schule wechseln müssen: Was bedeutet das für die Bildungssysteme? *PISA im Fokus*, 6.
- Pajares, M.F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307–332. <https://dx.doi.org/10.3102/00346543062003307>
- Posner, G. J., Strike, K. A., Hewson, P. W. & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227. <https://dx.doi.org/10.1002/sce.3730660207>
- Renkl, A. (2015). Drei Dogmen guten Lernens und Lehrens: Warum sie falsch sind. *Psychologische Rundschau*, 66, 211–220. <https://dx.doi.org/10.1026/0033-3042/a000274>
- Reusser, K. & Pauli, C. (2014). Berufsbezogene Überzeugungen von Lehrerinnen und Lehrern. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 642–661). Münster: Waxmann.
- Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula (Ed.), *Handbook of research on teacher education* (2nd ed., pp. 102–119). New York, NY: Simon & Schuster Macmillan.
- Richter, T. & Maier, J. (2018). Verstehen kontroverser wissenschaftlicher Themen: Probleme, zugrundeliegende kognitive Prozesse und psychologische Interventionen. *Psychologische Rundschau*, 69, 151–159. <https://dx.doi.org/10.1026/0033-3042/a000371>
- Rößbach, H.-G. & Tietze, W. (2010). Sitzenbleiben. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 781–788). Weinheim: Beltz.
- Rothmund, T., Gollwitzer, M., Nuroth, P. & Bender, J. (2017). Motivierte Wissenschaftsrezeption. *Psychologische Rundschau*, 68, 193–197. <https://dx.doi.org/10.1026/0033-3042/a000364>

Schöps, K., Walter, O., Zimmer, K. & Prenzel, M. (2006). Disparitäten zwischen Jungen und Mädchen in der mathematischen Kompetenz. In M. Prenzel et al. (Hrsg.), *PISA 2003: Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 209–224). Münster: Waxmann.

Schraw, G. & Olafson, L. (2015). Assessing teachers' beliefs: Challenges and solutions. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teachers' beliefs* (pp. 87–105). New York, NY: Routledge.

Schumacker, R. E. & Lomax, R. G. (2015). *A beginner's guide to structural equation modeling*. Mahwah, NJ: LEA.

Skott, J. (2015). The promises, problems, and prospects of research on teachers' beliefs. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teachers' beliefs* (pp. 13–30). New York, NY: Routledge.

Spinath, B. (Hrsg.) (2014). *Empirische Bildungsforschung*. Berlin: Springer. <https://dx.doi.org/10.1007/978-3-642-41698-9>

Stanat, P., Bergann, S. & Taraszow, T. (2018). Geschlechtsbezogene Disparitäten im deutschen Bildungswesen. In R. Tippelt & B. Schmidt-Hertha (Hrsg.), *Handbuch Bildungsforschung* (S. 1321–1338). Wiesbaden: Springer. https://dx.doi.org/10.1007/978-3-531-19981-8_59

Trautwein, C. R. (2013). Lehrerbezogene Überzeugungen und Konzeptionen: eine konzeptuelle Landkarte. *Zeitschrift für Hochschulentwicklung*, 8(3), 1–14. <https://dx.doi.org/10.3217/zfhe-8-03/02>

Voss, T., Kunter, M., Seiz, J., Hoehne, V. & Baumert, J. (2014). Die Bedeutung des pädagogisch-psychologischen Wissens von angehenden Lehrkräften für die Unterrichtsqualität. *Zeitschrift für Pädagogik*, 60(2), 184–201.

Voyer, D. & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, 140(4), 1174–1204. <https://dx.doi.org/10.1037/a0036620>

Watson, K., Handal, B. & Maher, M. (2016). The influence of class size upon numeracy and literacy performance. *Quality Assurance in Education*, 24(4), 507–527. <https://dx.doi.org/10.1108/QAE-07-2014-0039>

Wiechmann, J. (2009). Direkte Instruktion, Frontalunterricht, Klassenunterricht. In K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (S. 200–203). Bad Heilbrunn: Klinkhardt.

Jana Asberger
Dr. Eva Thomm
Prof. Dr. Johannes Bauer
 Universität Erfurt,
 Bildungsforschung und Methodenlehre
 Nordhäuser Straße 63
 D-99089 Erfurt
 E-Mail: jana.asberger@uni-erfurt.de
eva.thomm@uni-erfurt.de
johannes.bauer@uni-erfurt.de

Anhang A

Items des QUEBEC

Inwiefern halten Sie die folgenden Aussagen zu Themen von Schule und Bildung für zutreffend?

„Die Größe einer Schulklasse beeinflusst entscheidend ...“	
KG 1	... die Qualität des Unterrichts
KG 2	... die Lernergebnisse von Schüler*innen
KG 3	... wie Lehrer*innen im Unterricht auf die individuellen Voraussetzungen und Bedürfnisse der Schüler*innen eingehen
KG 4	... die berufliche Belastung von Lehrer*innen
D 1	... das Auftreten von Unterrichtsstörungen
D 2	... das Klima und den Zusammenhalt in der Klasse
„Schüler*innen mit unzureichenden Leistungen eine Klasse wiederholen zu lassen trägt maßgeblich dazu bei, ...“	
KW 1	... dass sie ihre fachlichen und leistungsmäßigen Defizite aufholen
KW 2	... dass sie Defizite in der sozialen, emotionalen oder verhaltensmäßigen Reife aufholen
KW 3	... ihr Selbstwertgefühl zu stärken (z. B. indem sie von unnötigem Leistungsdruck befreit werden)
KW 4	... ihr Risiko zu reduzieren, die Schule ohne Abschluss zu verlassen
KW 5	... Bildungsgerechtigkeit in der Gesellschaft zu ermöglichen
KW 6	... das Leistungsniveau des Schulsystems sicherzustellen
D 3	... für sie ein pädagogisch wichtiges Signal zu setzen
D 4	... sie zu demotivieren
D 5	... die Leistungsheterogenität von Schulklassen zu verringern

„Ein Unterricht nach der Methode der „Direkten Instruktion“ (d.h. stark lehrerzentriert und mit hohen Anteilen an Lehrervortrag und angeleiteter Übung) ...“	
DI 1	... ist insgesamt weniger effektiv für das Lernen als Lehrmethoden, in denen die Schüler*innen aktiv etwas tun (z. B. problemorientiertes Lernen)
DI 2	... ist schlecht auf die neurologischen/kognitiven Grundlagen menschlichen Lernens ausgerichtet
DI 3	... trägt hauptsächlich dazu bei, nur theoretisches Wissen aufzubauen, das im Alltag nicht angewendet werden kann
DI 4	... eignet sich vor allem für das Erlernen von Faktenwissen, aber nicht für anspruchsvollere Lernziele, wie z. B. Problemlösen
DI 5	... ist insbesondere für Schüler*innen mit Lernschwierigkeiten schlecht geeignet
D 6	... ist eine der effektivsten Lehrmethoden überhaupt
D 7	... führt bei Schüler*innen schnell zu Langeweile
D 8	... ist ein wichtiger Grund, warum Schüler*innen in Deutschland in internationalen Schulleistungsstudien oft nur mittelmäßig abschneiden
„Der deutlich überdurchschnittliche Frauenanteil bei Kita-Erzieher*innen und Grundschullehrer*innen ist eine entscheidende Ursache dafür, dass ...“	
LA 1	... Jungen in der Schule im Durchschnitt schlechtere Leistungen zeigen als Mädchen
LA 2	... Jungen in der Schule im Durchschnitt schlechtere Zensuren erhalten als Mädchen
LA 3	... Jungen in der Schule im Durchschnitt weniger motiviert sind als Mädchen
LA 4	... Jungen in der Schule im Durchschnitt seltener eine Gymnasialempfehlung erhalten als Mädchen
LA 5	... Jungen in der Schule im Durchschnitt häufiger die Schule ohne Abschluss verlassen als Mädchen
D 9	... Jungen in der Schule oft Probleme haben, ein positives Männlichkeitsbild zu entwickeln
D 10	... Jungen in der Schule im Durchschnitt stärkeres Interesse für Schulfächer mit einem höheren Anteil männlicher Lehrkräfte entwickeln als Mädchen (z. B. Physik)

Anmerkungen: KG = Klassengröße; KW = Klassenwiederholung; DI = Direkte Instruktion; LA = Lehrerinnenanteil; D = Distraktoritem.

Anhang B

Deskriptive Statistiken zu den Distraktoren

Item	D 1	D 2	D 3	D 4	D 5	D 6	D 7	D 8	D 9	D 10
<i>M</i>	4.71	3.27	2.60	3.24	2.97	2.20	4.66	3.57	2.73	2.83
<i>SD</i>	1.25	1.35	1.18	1.34	1.22	1.20	1.15	1.31	1.37	1.34

Anmerkungen: *N* = 217; *M* (Itemschwierigkeit); 6-stufige Likert-Skala (überhaupt nicht zutreffend [1] bis voll und ganz zutreffend [6]).