

# Einschätzungen von Textunterschieden durch Lehramtsstudent\*innen: Eine Studie am Beispiel von Texten zum Umgang mit (Leistungs-)Heterogenität

Fabio Nagele  · Michaela Katstaller · Ulrike Greiner

Eingegangen: 12. August 2020 / Überarbeitet: 16. April 2021 / Angenommen: 31. Mai 2021 / Online publiziert: 30. September 2021  
© Der/die Autor(en) 2021

**Zusammenfassung** Bildungs- und erziehungswissenschaftliche Fragestellungen, die sich z. B. mit dem Umgang mit (Leistungs-)Heterogenität von Schüler\*innen im Klassenzimmer befassen, sind nicht nur aus der Perspektive des pädagogischen Professionswissens relevant, sondern auch Bestandteil von Diskursen zwischen Wissenschaft, Politik, Gesellschaft und pädagogischer Praxis. Da (angehende) Lehrpersonen als aktiv Beteiligte sowohl produktiv als auch rezeptiv an diesen Diskursen partizipieren, benötigen sie die generische Fähigkeit, Aussagen aus verschiedenen Informationsquellen sowie deren Bedeutsamkeit adäquat einschätzen zu können. Basierend auf dem Multiple-Document-Comprehension-Modell (MDC-Modell) wird in dieser Studie untersucht, ob Lehramtsstudent\*innen des Bachelorstudiums ( $n = 193$ ) an der Paris Lodron Universität Salzburg die aus dem MDC-Modell abgeleiteten kognitiven Anforderungen im Vergleich von Texten beherrschen. Den Student\*innen wurde dazu ein wissenschaftlicher Text und ein journalistischer Kommentar zu der Frage des Umgangs mit (Leistungs-)Heterogenität vorgelegt, auf deren Basis Verifikationsitems bearbeitet werden mussten. Anschließend wurden mithilfe eines kognitiven Diagnosemodells (CDM) die kognitiven Anforderungen des MDC-Modells modelliert, um dessen Mehrdimensionalität empirisch abzubilden. Die Ergebnisse zeigen, dass ca. 93 % der Lehramtsstudent\*innen zumindest einen der vier Anforderungsbereiche des MDC-Modells nicht beherrschen. Der Beitrag zeigt somit neue Möglichkeiten in der Diagnostik von Fertigkeiten zum Vergleich multipler Doku-

---

Dr. Fabio Nagele (✉) · Dr.in Michaela Katstaller · Univ.-Prof.in DDr.in Ulrike Greiner  
School of Education, Paris Lodron Universität Salzburg, Erzabt-Klotz-Straße 1, 5020 Salzburg,  
Österreich  
E-Mail: [fabio.nagele@plus.ac.at](mailto:fabio.nagele@plus.ac.at)

Dr.in Michaela Katstaller  
E-Mail: [michaela.katstaller@plus.ac.at](mailto:michaela.katstaller@plus.ac.at)

Univ.-Prof.in DDr.in Ulrike Greiner  
E-Mail: [ulrike.greiner@plus.ac.at](mailto:ulrike.greiner@plus.ac.at)

mente gemäß des MDC-Ansatzes im Bereich der Lehrer\*innenbildungsforschung auf, welche zukünftig auch für die Ausbildung von angehenden Lehrpersonen genutzt werden können.

**Schlüsselwörter** Multiple-Document-Comprehension · Leistungsheterogenität · Lehrer\*innenbildung · Kognitives Diagnosemodell · CDM

## **On the assessment of student teachers' ability to evaluate different text sources: A study on the example of texts on how to deal with academic heterogeneity**

**Abstract** How to deal with students' academic heterogeneity in the classroom is not only an important question from the perspective of pedagogical professional knowledge, but it is also part of a discourse between science, politics, society, and educational practice. As student teachers actively participate in these discourses, both productively and receptively, they need the general ability to adequately evaluate statements that derive from different sources of information. Based on the Multiple-Document-Comprehension model (MDC), it is examined whether undergraduate student teachers ( $n=193$ ) at the Paris Lodron University of Salzburg master the cognitive procedures that were derived from the MDC model. The student teachers were given a scientific text and a journalistic commentary on the question of how to deal with academic heterogeneity. Based on these texts, verification items had to be processed. Subsequently, the cognitive procedures were statistically modelled by using a Cognitive Diagnosis Model (CDM) in order to examine empirically the MDC model's multi-dimensionality. The results show that approx. 93% of the student teachers have not mastered at least one of the four MDC-related cognitive procedures. Thus, the study offers new possibilities in the diagnosis of skills for the comparison of multiple documents according to the MDC model in the field of teacher education research. Implications of this study can also be used for the training of prospective student teachers.

**Keywords** Multiple-Document-Comprehension · Academic Heterogeneity · Teacher Education · Cognitive Diagnosis Modeling · CDM

## **1 Einleitung**

Bildungswissenschaftliches Wissen bildet neben dem fachlichen und fachdidaktischen Wissen einen zentralen Bestandteil in der Lehrer\*innenbildung (Baumert und Kunter 2006; Shulman 1986). Blömeke et al. (2008) postulieren, bildungswissenschaftliches Wissen u. a. in bildungssoziologische Lerninhalte und Fragestellungen zu unterteilen, welche auf die gesellschaftliche Determiniertheit der Erziehungspraxis durch die gesellschaftliche Gesamtpraxis (Benner 1980) bzw. die strukturellen Erfordernisse der jeweils bestehenden gesellschaftlichen Verhältnisse (Schmied-Kowarik 2018) verweisen. Auf die Relevanz gesellschaftlicher Diskurse innerhalb der Lehrer\*innenbildung verweist auch der Qualitätssicherungsrat (der zur Qua-

litätssicherung für die Pädagog\*innenbildung in Österreich konstituiert wurde) in seiner Definition der mit diesem Beruf verbundenen professionellen Kompetenzen (QSR = Qualitätssicherungsrat für Pädagoginnen- und Pädagogenbildung 2013, S. 2): „Pädagoginnen [*sic*] haben ein umfassendes Verständnis ihrer Aufgabe, die im allgemeinsten Sinn mit ‚Bildung‘ umschrieben werden kann. Sie reicht von der Vermittlung fachlicher Kompetenzen, [...] bis hin zur Beteiligung an der Zukunftsgestaltung der Gesellschaft durch die Mitwirkung an und Initiierung von relevanten Diskursen“. Die Verbindung von Wissenschaft, pädagogischer Praxis und Öffentlichkeit stellt ein pädagogisches Handeln dar, in dem Argumentation und diskursive Teilhabe als explizite Bestandteile des Berufsbildes von Lehrpersonen angesehen werden. Ausdrücklich wird hier auf die Diskurskompetenz von Lehrpersonen verwiesen, der neben der Kompetenz, an gesellschaftlichen Diskursen aktiv teilzunehmen, auch die Textsortenkompetenz zuzuschreiben ist.

Durch die Verwobenheit der Erziehungswissenschaft mit den sie vorstrukturierenden gesellschaftlichen Bedingungen lässt sich schlussfolgern, dass angehende Lehrpersonen bei ihrer Einschätzung von Textquellen, welche erziehungswissenschaftliche Themen adressieren, der generischen Kompetenz bedürfen, sowohl die inhaltlichen Aussagen verschiedener Textquellen als auch die (impliziten) Intentionen und Hintergründe der jeweiligen Autor\*innen korrekt einordnen zu können, um zu einer validen Einschätzung über die dargestellten Inhalte zu gelangen. Die damit verbundenen kognitiven Anforderungen zur Bewertung von verschiedenen Textquellen werden aus dem MDC-Modell (*multiple document comprehension*; Britt und Rouet 2012; vgl. auch Schoor et al. 2020) abgeleitet. Aus Sicht der Lehrer\*innenprofession ergibt sich für Lehramtsstudent\*innen die Relevanz des MDC-Modells in zweifacher Hinsicht: (1) Im Sinne einer theoriegeleiteten und reflexiven Theorie-Praxis-Verzahnung (Burn und Mutton 2015; Neuweg 2011; Rhein 2019) ist die situationsgebundene Relevanz und inhaltliche Validität von unterschiedlichen (erziehungswissenschaftlichen) Textquellen für ein professionelles Handeln adäquat zu beurteilen und (2) dieses Handeln ist vor dem Hintergrund medial geführter politischer und gesellschaftlicher Diskurse zu analysieren, in deren Brennpunkt sich die professionelle Lehrperson als Vertreter\*in institutionalisierter Bildungseinrichtungen befindet, wie am Beispiel der gesellschaftspolitischen Diskurse im Kontext internationaler Leistungsvergleichsstudien deutlich wird (Hoffmann-Ocon und Schmidtke 2012; Kuhlmann 2012). Das MDC-Modell basiert dabei auf der Annahme, dass Informationen zu verschiedenen Textquellen (oder verschiedenen Quellen in einem Text) in Form von mentalen Modellen abgespeichert werden, um sich eine integrierte Gesamtrepräsentation eines Themenbereiches zu erschließen.

Für die vorliegende Studie wurde das MDC-Modell bzw. dessen Adaption durch Schoor et al. (2020) als Grundlage für die Fragestellung herangezogen, inwieweit angehende Lehrpersonen die kognitiven Anforderungen zur Bewertung von unterschiedlichen Textquellen beherrschen. Dazu wurden zwei zum Teil fingierte Texte generiert, welche beide den Umgang mit der (Leistungs-)Heterogenität von Schüler\*innen aus verschiedenen Perspektiven beleuchten. Dabei wurde einer der beiden Texte als wissenschaftlicher Text verfasst, während der andere einen journalistischen Kommentar darstellt, welcher jedoch teilweise auch auf wissenschaftliche Quellen Bezug nimmt. Neben der empirischen Erfassung der MDC-basierten Fähig-

keiten von Lehramtsstudent\*innen – für die erstmals ein kognitives Diagnosemodell (CDM) herangezogen wurde – wird in Anschluss an die Befunde von Schoor et al. (2020) die Frage nach der Dimensionalität des MDC-Modells aufgeworfen und kritisch diskutiert.

## 2 Theoretischer Hintergrund und Forschungsstand

### 2.1 Erziehungswissenschaftliche Diskurse in der Lehrer\*innenprofession

Erziehungswissenschaftliches Wissen als Bestandteil der Lehrer\*innenprofession (Baumert und Kunter 2006) kann in unterschiedliche Teilbereiche gegliedert werden. Neben der Einteilung in allgemeinpädagogische, pädagogisch-psychologische und bildungssoziologische Lerninhalte und Fragestellungen (Blömeke et al. 2008) kann es auch in die Inhaltsbereiche „Erziehung und Bildung“, „Unterricht und Allgemeine Didaktik“ und „Schulentwicklung und Gesellschaft“ (Seifert et al. 2010) unterteilt werden. Während Wissen zu den allgemeinpädagogischen, pädagogisch-psychologischen oder didaktischen Teilbereichen im Rahmen der Lehramtsausbildung zumeist aus didaktischen Lehrbüchern bzw. wissenschaftlich fundierten Quellen (wie bspw. empirischen Studien) bezogen wird, artikulieren sich insbesondere die bildungssoziologischen und gesellschaftsbezogenen Teilbereiche des erziehungswissenschaftlichen Wissens über offen geführte gesellschaftliche Diskurse (Benner 1980; Schmied-Kowarzik 2018). Dies bedeutet, nicht nur das unmittelbar für das professionelle Handeln relevante Wissen (wie z. B. in Form von pädagogischem Wissen), sondern auch gesellschaftspolitische Diskurse gleichermaßen zu berücksichtigen, welche das Handeln von Lehrpersonen in der Praxis zu einem wesentlichen Anteil rahmen.

Gesellschaftliche Diskurse spielen insbesondere im Umgang mit der schüler\*innenbezogenen (Leistungs-)Heterogenität eine besondere Rolle, wie z. B. anhand von Diskursen über die Etablierung von Ganztagschulen – angestoßen durch die PISA-Studien – deutlich wird (Kuhlmann 2012). Das Beispiel schüler\*innenbezogene (Leistungs-)Heterogenität ist daher nicht zufällig gewählt: Wie Schule mit (Leistungs-)Heterogenität umgeht bzw. im Verdacht steht, diese Entwicklung zu verstärken, ist ein gesellschaftliches Brennpunkthema ersten Ranges, denn die Folgen für Bildungsverläufe ihrer Schüler\*innen sind evident. Lehrer\*innen werden hier als rechenschaftspflichtig gesehen und ihre differentielle Diskurskompetenz kann ihre Mitwirkung an gesellschaftlichen Diskursen und damit den gesellschaftlichen Stellenwert der Profession erhöhen. Gleichzeitig wird der Umgang mit (Leistungs-)Heterogenität als eine Subdimension des pädagogischen Professionswissens (König und Blömeke 2009) konzeptualisiert, wodurch die unmittelbare Handlungsrelevanz dieser Thematik für den Lehrberuf hervorgehoben wird. Die notwendige Verschränkung von gesellschaftlich geführten Diskursen mit jener Subdimension pädagogischen Professionswissen, welche den Umgang mit der schüler\*innenbezogenen (Leistungs-)Heterogenität adressiert, ergibt sich aus zweierlei Gründen: (1) (Leistungs-)Heterogenität wird als Produkt eines intersubjektiven Vergleichs anhand eines *tertium comparationis* verstanden, das wiederum

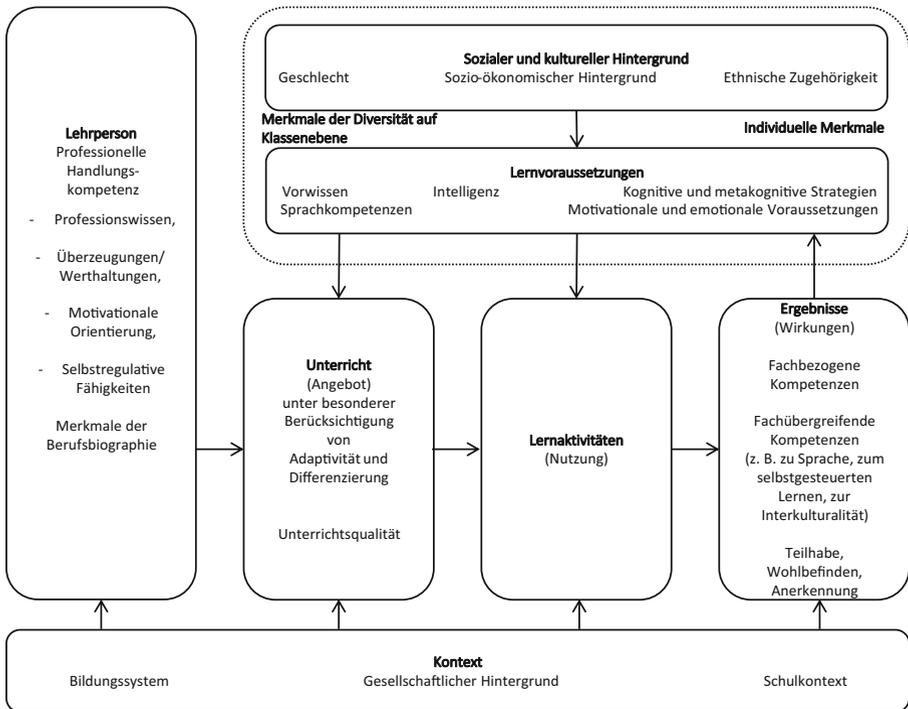


Abb. 1 Heterogenität im Schulsystem. (Gräsel et al. 2017, S. 197)

Ergebnis einer sozialen Konstruktion ist, welche reale gesellschaftliche Folgen mit sich bringt (Budde 2012; Trautmann und Wischer 2011) und (2) der Umgang mit (Leistungs-)Heterogenität ist auch in gesellschaftlich relevante Fragestellungen eingebettet, welche die Frage nach Bildungsgerechtigkeit und Teilhabegerechtigkeit adressieren (Heinrich 2013; Stojanov 2011). Somit ist der Umgang mit (Leistungs-)Heterogenität im Gegensatz zu anderen Subdimensionen des pädagogischen Professionswissens angehender Lehrkräfte (wie z.B. die Strukturierung von Unterricht oder Klassenführung; König und Blömeke 2009) am deutlichsten mit gesellschaftlichen Diskursen verbunden. Dieser Umstand wird auch im Rahmenmodell zum Umgang mit Heterogenität von Gräsel et al. (2017) zum Ausdruck gebracht (s. Abb. 1).

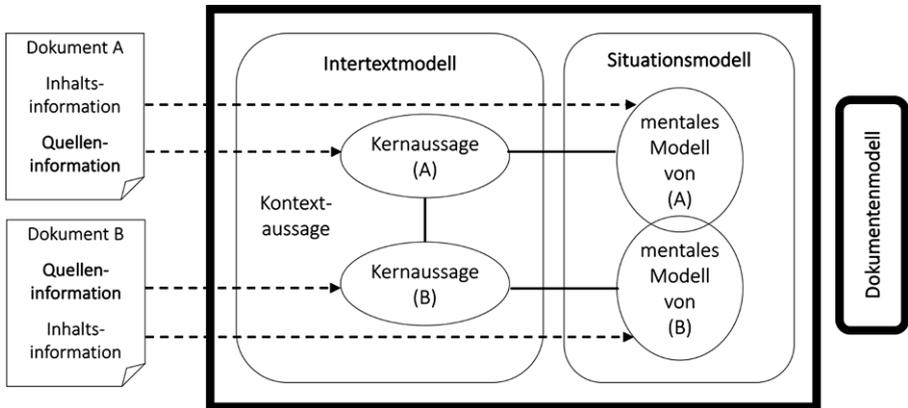
Wie dem Modell von Gräsel et al. (2017) zu entnehmen ist, sind die Heterogenitätsmerkmale einerseits auf der Klassenebene (z.B. Geschlecht, sozio-ökonomischer Hintergrund, ethnische Zugehörigkeit) und andererseits auf der Individual- bzw. Mikroebene (z.B. durch Lernvoraussetzungen, Vorwissen, motivationale und emotionale Voraussetzungen) zu verorten, wobei die Klassen- bzw. Makroebene und die Individual- bzw. Mikroebene stets miteinander verwoben sind (Budde 2012). Für den Unterricht bedeutet dies, dass eine Lehrperson professionelles Wissen benötigt, um den heterogenen Lernvoraussetzungen und der daraus resultierenden (Leistungs-)Heterogenität von Schüler\*innen unter der Berücksichtigung von Adaptivität und Differenzierung gerecht zu werden (Brühwiler 2014; Buholzer und Kummer Wyss 2010). In dem

Modell von Gräsel et al. (2017) ist auch ersichtlich, dass der Kontext (z. B. das Bildungssystem, der gesellschaftliche Hintergrund und der Schulkontext) einen nicht unwesentlichen Einfluss darauf haben, wie sich das Professionswissen von Lehrpersonen definiert und wie (adaptiver und differenzierter) Unterricht gestaltet wird. Vor allem auch die Auswirkung auf die zu erzielenden Ergebnisse spiegeln die normative Komponente wider, in der Fragen nach Teilhabe und Gerechtigkeit thematisiert werden (Stojanov 2011).

Vor dem Hintergrund dieser Ausführungen wird deutlich, dass angehende Lehrpersonen in der Aneignung bildungswissenschaftlichen Wissens bzw. in der Teilhabe an gesellschaftlichen Diskursen über erziehungswissenschaftliche Fragestellungen der generischen Fähigkeit bedürfen, bei der Rezeption von unterschiedlichen Textquellen nicht nur Informationen inhaltlich miteinander zu vergleichen und zu verknüpfen, sondern insbesondere auch die (impliziten) Intentionen der Entstehungskontexte zu berücksichtigen. Wie auch aufgezeigt werden konnte, ist dies vor allem für jene Subdimension des pädagogischen Professionswissens von Relevanz, welche den Umgang mit der (Leistungs-)Heterogenität von Schüler\*innen thematisiert. Zudem konnten Winter-Hözl et al. (2015) empirisch nachweisen, dass Lehramtsstudent\*innen gegenüber Student\*innen der Psychologie und der Erziehungswissenschaft ein geringer ausgeprägtes Wissen über das Genre empirischer Forschungsartikel im Bereich der Bildungswissenschaften verfügen. Daher ist die Frage nach der Modellierung von Kompetenzen im Vergleich von Textquellen für die Lehrer\*innenbildungsforschung von nicht zu unterschätzender Bedeutung. In den folgenden Abschnitten wird auf die kognitiven Anforderungen des MDC-Modells sowie auf den aktuellen Forschungsstand zu deren psychometrischen Erfassung näher eingegangen.

## 2.2 Komponenten des MDC-Modells

Im Rahmen der universitären Ausbildung von Student\*innen wird es als Kernelement eines jeden Studiums betrachtet, dass Student\*innen mit Informationen aus unterschiedlichen Texten umgehen können, um zu einem umfassenden und vertiefenden Verständnis innerhalb ihrer Wissensdomäne zu gelangen (Anmarkrud et al. 2014). Das MDC-Modell von Britt und Rouet (2012), welches sich auf die fach- und genreübergreifenden Prozesse im Vergleich von unterschiedlichen Texten bezieht (s. auch Schoor et al. 2020), unterscheidet hiervon drei mentale Modelle: das Situationsmodell, das Intertextmodell und Dokumentenmodell. Das Situationsmodell erfordert, dass aus den inhaltsbezogenen Informationen zweier Textdokumente neue Aussagen generiert bzw. abgeleitet werden. Das Intertextmodell hingegen zielt darauf ab, dass Textdokumente hinsichtlich der Quelleninformationen (z. B. Autor\*inneninformationen, Genre, Epoche) miteinander in Beziehung gesetzt werden. Im Dokumentenmodell werden das Situations- und Intertextmodell integriert, da Schlussfolgerungen sowohl auf Basis der inhaltsbezogenen Informationen (Situationsmodell) als auch der Quelleninformationen (Intertextmodell) miteinander in Bezug gesetzt werden. In Abb. 2 sind die drei Komponenten des MDC-Modells nach Britt und Rouet (2012) grafisch dargestellt.



**Abb. 2** Das MDC-Modell. (Adaptiert nach Britt und Rouet 2012)

Das MDC-Modell wurde in den jüngeren Forschungsarbeiten von Hahnel et al. (2019) und Schoor et al. (2020) in Anlehnung an Wineburg (1991) um die Strategie ergänzt, bei der Informationen über mehrere Textquellen hinweg miteinander verglichen werden (Corroboration). Der Unterschied zum Situationsmodell besteht bspw. darin, dass bei Corroboration Aussagen lediglich hinsichtlich ihrer *Übereinstimmung* abgeglichen werden, während beim Situationsmodell Aussagen basierend auf den vorliegenden Informationen abgeleitet werden und folglich ein komplexeres *mentales Modell* generiert wird. Für die aktuelle Studie wird an diese Erweiterung des MDC-Modells von Hahnel et al. (2019) und Schoor et al. (2020) um die Strategie „Corroboration“ angeknüpft. Wenngleich die empirischen Befunde von Schoor et al. (2020) aufzeigen, dass „Corroboration“ auf einer eindimensionalen Kompetenzskala als Voraussetzung für den Aufbau des Situations- und Intertextmodells modelliert werden kann, wird im Rahmen der vorliegenden Studie „Corroboration“ als ein qualitativ eigenständiger Anforderungsbereich konzeptualisiert, der in keinem hierarchischen Zusammenhang mit den mentalen Modellen steht. Bräten et al. (2011, S. 63) führen hierzu auf theoretischer Ebene aus, dass sich „Corroboration“ in Anlehnung an Wineburg (1991) darauf bezieht, Inhalte über mehrere Dokumente hinweg zu vergleichen, während in einem mentalen Modell („Situationsmodell“) Inferenzen gebildet werden, die auf einer komplexen Vernetzung zwischen den Dokumenten beruhen. In der vorliegenden Studie wurde insbesondere dieser qualitative Unterschied – Vergleich vs. Inferenz – bei der Itementwicklung berücksichtigt. Bei den Items des Anforderungsbereiches „Corroboration“ wurde auf das (Nicht-)Vorhandensein *identer Inhalte* geachtet, während bei den Items des Anforderungsbereiches „Situationsmodell“ das *Ziehen von Inferenzen* erfasst wurde (s. im Detail Abschn. 4.2).

### 2.3 Empirische Befunde zum MDC-Modell

Wie Schoor et al. (2020) konstatieren, gibt es bisher kaum Ansätze zur systematischen Beforschung von den notwendigen Fähigkeiten im Bereich des Umgangs mit

multiplen Dokumenten, welche mit standardisierten und psychometrischen Modellen durchgeführt werden. Laut den Autor\*innen beziehen sich bisherige Untersuchungen zum Teil auf die Auswertung von Essays, die von Proband\*innen zum Textmaterial verfasst wurden (z. B. Stadtler et al. 2014). Alternativ wurden intertextuelle Inferenz-Verifikationsaufgaben erstellt, welche sich nur auf das Situationsmodell beziehen, das MDC-Modell in seiner Ganzheit jedoch nicht erfassen (z. B. Braasch et al. 2014). Schoor et al. (2020) konnten diese Forschungslücke in ihrer eigenen Untersuchung schließen, indem sie einen standardisierten Test entwickelten, mit dem die vier kognitiven Anforderungen des MDC-Modells basierend auf Verifikations- und Multiple-Choice-Aufgaben psychometrisch abgebildet werden konnten. Methodisch wurden die vier kognitiven Anforderungen mit einem eindimensionalen IRT-Modell (Rasch-Modell) skaliert, wobei nachgewiesen werden konnte, dass die Items des Dokumentenmodells signifikant schwieriger als alle anderen Items waren. Umgekehrt konnte gezeigt werden, dass Items des Typs „Corroboration“ signifikant leichter als alle anderen Items waren, sich im Direktvergleich jedoch nicht signifikant von Items des Intertextmodells unterschieden. Die Autor\*innen führen diese Schwierigkeitsunterschiede darauf zurück, dass Items des Dokumentenmodells allgemein eine höhere Komplexität aufweisen, da diese eine Integration des Situations- und Intertextmodells erfordern, während bei den Corroboration-Items lediglich ein einfacher Vergleich erforderlich ist. Einschränkend halten die Autor\*innen jedoch fest, dass diese Befunde mit anderen Materialien repliziert werden müssten, da nicht ausgeschlossen werden kann, dass diese Effekte auf die ausgewählten Fragen und Themen zurückzuführen sind. Die vorliegende Studie knüpft an dieses Forschungsdesiderat an und unternimmt den Versuch einer alternativen Modellierung unter Verwendung eines kognitiven Diagnosemodells (CDM; George und Robitzsch 2015), mit dessen Hilfe die Annahme der Mehrdimensionalität des MDC-Modells überprüft werden soll.

### 3 Zentrale Fragestellungen

Basierend auf den theoretischen Ausführungen wurden für die vorliegende Studie zwei zentrale Forschungsfragen formuliert:

1. Können die kognitiven Anforderungen des MDC-Modells in ihrer Mehrdimensionalität mithilfe eines kognitiven Diagnosemodells (CDM) empirisch abgebildet werden?
2. Inwieweit werden die kognitiven Anforderungen des MDC-Modells von angehenden Lehrkräften im Bachelorstudium (noch nicht) beherrscht, wenn ihnen ein wissenschaftlicher Text und ein journalistischer Kommentar zum Umgang mit (Leistungs-)Heterogenität von Schüler\*innen vorgelegt wird?

Die erste der beiden Forschungsfragen knüpft vor allem an die Forschungsergebnisse von Schoor et al. (2020) an, welche die kognitiven Anforderungen des MDC-Modells entlang eines eindimensionalen IRT-Modells skalieren konnten. Dabei soll der von Schoor et al. (2020) aufgeworfenen Frage nachgegangen werden, ob das MDC-Modell stets eindimensional ist oder auch mehrdimensional modelliert wer-

den kann. Im Zuge der in dieser Studie vorgenommenen Dimensionalitätsprüfung wird auch der Frage nachgegangen, ob die mit dem Dokumentenmodell verbundene kognitive Anforderung durch das Situations- und Intertextmodell ausreichend erklärt werden kann oder ob das Dokumentenmodell als eine eigenständige Dimension zu modellieren ist.

Die zweite Forschungsfrage verfolgt schließlich in Anschluss an die erste Forschungsfrage das Ziel, zu untersuchen inwiefern angehende Lehrkräfte im Bachelorstudium die kognitiven Anforderungen des MDC-Modells (noch nicht) beherrschen. Hierfür werden die Beherrschungswahrscheinlichkeiten der einzelnen kognitiven Anforderungen sowie die sich daraus ergebenden Profile mithilfe des kognitiven Diagnosemodells (CDM) berechnet.

Zur empirischen Überprüfung beider Forschungsfragen werden in der hier vorliegenden Studie zwei verschiedene Textquellen verwendet, ein wissenschaftlicher Text und ein journalistischer Kommentar, um deutliche Unterschiede in den Textgenres (Intertextmodell) abbilden zu können (s. Abschn. 4). Dazu wurde das Thema „Umgang mit (Leistungs-)Heterogenität“ vor allem deswegen gewählt, da sich dieses Thema im Gegensatz zu anderen Subdimensionen pädagogischen Professionswissens – die weit weniger im öffentlichen Diskurs behandelt werden, wie z. B. die Strukturierung von Unterricht – im Rahmen eines journalistischen Kommentars am ehesten authentisch bzw. ökologisch valide darstellen lässt.

## 4 Methodologie

### 4.1 Generierung der Textdokumente

Wie in Abschn. 2.1 dargelegt werden konnte, greifen bildungs- und erziehungswissenschaftliche Diskurse nicht nur auf (bildungs-)wissenschaftliche Befunde zurück, sondern beziehen sich auch auf gesellschaftlich bedeutsame Themen. Dadurch erhalten bildungs- und erziehungswissenschaftliche Diskurse auch eine explizit normative Komponente, welche in rein (bildungs-)wissenschaftlichen Diskursen zwar auch gegeben sind, jedoch zumeist eher implizit.

Um für die vorliegende Studie zwei verschiedene Textdokumente zum Umgang mit (Leistungs-)Heterogenität zu generieren, in denen die unterschiedlichen Intentionen und Hintergründe von Autor\*innen auch erkennbar sind, wurde ein bildungswissenschaftlicher und journalistischer Text verfasst. Der bildungswissenschaftliche Text wurde einer real existierenden wissenschaftlichen Quelle entnommen (Scharenberg 2012) und für den Rahmen der vorliegenden Studie modifiziert. Das so verfasste Dokument umfasste 678 Wörter (Text 1) und erfüllte die gängigen Kriterien eines wissenschaftlichen Textes: eine stringente Gliederung, ein vollständiges Literatur- und Quellenverzeichnis sowie korrekte Zitation im Text (Retting 2017; Ruhl et al. 2010). Zudem wurden Informationen zur Autorin „Katja Scharenberg“ angegeben, u. a. die forschungsbezogenen Tätigkeiten und Schwerpunkte der Autorin. In dem verfassten Text bezieht die Autorin nicht explizit Stellung, ob sie leistungshomogene oder leistungsheterogene Lerngruppen befürwortet, allerdings lassen die angeführ-

ten empirischen Befunde zu den Nachteilen leistungshomogener Lerngruppen eine kritische Haltung der Autorin gegenüber diesen (zumindest implizit) vermuten.

Das zweite Textdokument (Text 2) basiert auf einem journalistischen Kommentar des fiktiven Autors „Horst Brauckmann“. Es umfasst 553 Wörter und wurde entlang der Kriterien für journalistische Texte verfasst, d.h., meinungsbetont, bewertend und auf die Beeinflussung der Leser\*innenschaft orientiert (Jungová 2006; Lenk und Vesalainen 2012; von la Roche 2013). Inhaltlich greift zwar auch der journalistische Kommentar von Horst Brauckmann wissenschaftliche Befunde auf, nutzt diese jedoch im Sinne der Meinungsbildung (z.B.: „[...] Im Gegenteil, es demotiviert schwächere Schülerinnen und Schüler viel mehr, als wenn sie sich mit denjenigen Peers vergleichen können, die ungefähr auf ihrem Niveau sind. Das behaupten die Forscher Holtappels und Rollet, die darüber 2008 ein spannendes Buch publizierten. [...]“). Während der Text von Katja Scharenberg in Übereinstimmung mit wissenschaftlichen Kriterien keine meinungsbildende Stellung zum Thema bezieht, wird dies im journalistischen Kommentar von Horst Brauckmann expliziert. Wie auch beim wissenschaftlichen Text wurden Informationen zu Horst Brauckmann (z.B. seine Tätigkeit als Journalist für die Bereiche Bildung, Schule und Lernen) angegeben. Zudem wurden in einer Einleitung übergreifend für beide Texte die Begriffe „homogene Lerngruppe“ und „heterogene Lerngruppe“ definiert.

## 4.2 Erstellung der Items

Basierend auf den beiden generierten Textdokumenten wurden dichotome Items (Verifikationsitems) für jede der vier kognitiven Anforderungen des MDC-Modells in Anlehnung an Schoor et al. (2020) generiert: Corroboration=6 Items, Situationsmodell= 7 Items, Intertextmodell= 7 Items, Dokumentenmodell= 7 Items. Die insgesamt 27 Items konnten im Format Ja/Nein bzw. korrekt/nicht korrekt beantwortet werden. In Tab. 1 wird jeweils ein Beispielitem für jede kognitive Anforderung angeführt, im ESM 1 (Electronical Supplement) können alle Items inkl. der Texte eingesehen werden:

Zur korrekten Lösung des in Tab. 1 angeführten Beispielitems für „Corroboration“ war es erforderlich, beide Texte hinsichtlich der Bezugnahme auf soziale Unterschiede abzugleichen. Hier nahm nur der wissenschaftliche Text auf dieses

**Tab. 1** Beispielitems zu den kognitiven Anforderungen des MDC-Modells

Kognitive Anforderung	Beispielitem
Corroboration	Bitte geben Sie an, ob in beiden Texten auf das genannte Kriterium eingegangen wird: Ausgleich sozialer Unterschiede. (C1)
Situationsmodell	Der Leistungszuwachs von Schüler*innen ist in leistungshomogenen Lerngruppen am größten. (S1)
Intertextmodell	Dass in den Texten unterschiedlich argumentiert wird, hängt vermutlich damit zusammen, dass die beiden Texte an eine unterschiedliche Leserschaft adressiert sind. (I5)
Dokumentenmodell	Während sich Horst Brauckmann u. a. auf schulische Alltagserfahrungen bezieht, wenn es um die Vorteile von leistungshomogenen Lerngruppen geht, sind die Argumente von Katja Scharenberg stets mit wissenschaftlicher Literatur belegt. (D5)

Thema Bezug, allerdings nicht der journalistische Kommentar. Schwierigkeitsbestimmend war hier, in beiden Texten auf die explizite Bezugnahme zu achten und keine weiterführenden Interpretationen oder Inferenzen zu tätigen.

Für das Item zum Situationsmodell mussten die Student\*innen die Aussagen der beiden Texte kombinieren, um zu einer Synthese zu gelangen. Die Aussage von Item S1 ist hier falsch, da der wissenschaftliche Text einschränkende Faktoren zum Vorteil leistungshomogener Lerngruppen nennt, welche eine eindeutige Bewertung erschweren, wenngleich sich der journalistische Kommentar ebenfalls auf wissenschaftliche Studien beruft, die eindeutig zum Vorteil leistungshomogener Gruppen ausgelegt werden.

Das Beispielimitem zum Intertextmodell greift explizit die Hintergründe der Autor\*innen auf. Hier galt es einzuschätzen, ob die Autor\*innen unterschiedliche Leser\*innenschaften adressieren, was im vorliegenden Fall korrekt ist: der wissenschaftliche Text ist eher an wissenschaftlich interessierten bzw. an in diesem Bereich tätigen Personen adressiert, während der journalistische Text sich an ein breiteres, nicht-wissenschaftliches Publikum richtet.

Das Item zum Dokumentenmodell erfordert schließlich die Integration des Intertext- und des Situationsmodells: neben dem Fokus auf die Verknüpfung der inhaltsbezogenen Aussagen beider Texte (die Vorteile von leistungshomogenen Lerngruppen) müssen auch die Stilmittel der Autor\*innen im Sinne der Textgenres (Bezugnahme auf schulische Alltagserfahrungen vs. wissenschaftlich belegte Argumente) in Hinblick auf die inhaltsbezogenen Aussagen betrachtet werden. Dieses Item musste ebenfalls als „korrekt“ beurteilt werden, um es richtig zu lösen.

### 4.3 Auswertungsmethode

Um die Annahme der Mehrdimensionalität des MDC-Modells zu überprüfen und die Beherrschungswahrscheinlichkeiten der Lehramtsstudent\*innen hinsichtlich der einzelnen kognitiven Anforderungen zu schätzen, wurde ein kognitives Diagnosemodell (CDM) verwendet (Bley und George 2017; George und Robitzsch 2015). Zentrales Element einer CDM-basierten Modellierung ist die dichotom strukturierte Q-Matrix, anhand derer a priori definiert wird, welche kognitiven Anforderungen (folgend auch „Skills“ genannt) zur Lösung eines jeden Items benötigt werden, wobei eine „1“ in der Q-Matrix bedeutet, dass der betreffende Skill zur Lösung des Items benötigt wird, während eine „0“ bedeutet, dass der Skill nicht benötigt wird. Diese a priori-Setzung der erforderlichen Skills pro Item ermöglicht in der Folge eine konfirmatorische Prüfung des zugrunde gelegten mehrdimensionalen Modells bzw. eine Schätzung sowohl (1) der Beherrschungswahrscheinlichkeiten der einzelnen kognitiven Anforderungen bzw. Skills in der Population als auch (2) der daraus abgeleiteten Auftretenswahrscheinlichkeiten von sogenannten „Skill-Klassen“ in der Population (Bley und George 2017). Die Skill-Klassen umfassen dabei die unterschiedlichen Kombinationsmöglichkeiten von beherrschten und nicht beherrschten Skills.

Da die Dimensionalität bei kognitiven Diagnosemodellen nicht explorativ, sondern konfirmatorisch ermittelt wird, wurden für die insgesamt 27 Items a priori zwei Q-Matrizen erstellt: Eine Q-Matrix („Q-Matrix A“), bei der die Items der

Dokumentenmodelle (D1–D7) durch die Kombination der mentalen Modelle bzw. Skills „Situation“ und „Intertext“ erklärt werden („Doppelladungen“), und eine weitere („Q-Matrix B“), bei der diese Items durch einen eigenen Skill („Dokumentenmodell“) erklärt werden. Die beiden vollständigen Q-Matrizen befinden sich im ESM 2. Für die Auswertung der empirischen Daten auf Basis der beiden Q-Matrizen A und B wurde jeweils das DINA-Modell (Deterministic Noisy And Gate) angewandt, wodurch angenommen wird, dass Defizite in einer zur Itembearbeitung erforderlichen kognitiven Anforderung nicht durch das Beherrschen einer anderen kognitiven Anforderung kompensiert werden können (George und Robitzsch 2015). Im Fall von Items des Dokumentenmodells in Q-Matrix A bedeutet dies, dass zu deren korrekten Lösung die *gleichzeitige* Anwendung des Situations- als auch des Intertextmodells erforderlich ist. Das fehlende Beherrschen einer der beiden kognitiven Anforderungen führt somit dazu, dass diese Aufgaben nicht gelöst werden können. Die Verknüpfung der beiden Anforderungen in den Items D1–D7 erfolgt somit *multiplikativ*. Alle Analysen wurden mit dem R-Paket „CDM“, Version 7.4-19 (Robitzsch et al. 2019b), und der R-Version 3.6.0 (R Core Team 2019) durchgeführt.

## 5 Stichprobe

Die insgesamt 27 Items, welche sich jeweils auf den Vergleich der beiden Texte beziehen, wurden Lehramtsstudent\*innen im Bachelorstudium an der Paris Lodron Universität Salzburg im Rahmen der Lehrveranstaltung „Diversität und Inklusion: Vielfalt (in) der Schule“ vorgelegt. Diese Lehrveranstaltung wurde für die Datenerhebung gewählt, da (1) das Thema Heterogenität in seinen vielfältigen Facetten basierend auf dem Curriculum Bestandteil der Lehrveranstaltung ist und (2) dadurch gewährleistet ist, dass das Vorwissen von Student\*innen zum Thema „Umgang mit (Leistungs-)Heterogenität“ als potenzielle Störvariable (zumindest aus formaler Sicht) konstant gehalten werden konnte. Für die Befragung wurden alle Parallelkurse dieser Lehrveranstaltung mit einer Gesamtzahl an Student\*innen von  $N=256$  herangezogen. Die Teilnahme war freiwillig und die Möglichkeit einer personalisierten Rückmeldung wurde als Anreiz eingesetzt. Durch die Freiwilligkeit der Teilnahme und den Umstand, dass manche Student\*innen nicht anwesend waren, konnten insgesamt  $n=193$  Personen befragt werden. Die Student\*innen waren zu 64 % weiblich und zu 36 % männlich und befanden sich durchschnittlich im 4,90 Semester ( $Min=2$ ;  $Max=20$ ;  $SD=2,46$ ).

## 6 Analysen und Ergebnisse

Für die Beantwortung der ersten Forschungsfrage, ob die Mehrdimensionalität des MDC-Modells empirisch überprüft werden kann, werden im Folgenden in einem ersten Schritt die Modell-Fit Werte für das dreidimensionale Modell mit insgesamt drei kognitiven Anforderungen bzw. Q-Matrix A herangezogen. In Tab. 2 ist die finale Q-Matrix A jener neun MDC-Items abgebildet, welche die für CDM-basierte Modellierungen erforderlichen Kriterien erfüllten und für die weiteren Berechnun-

**Tab. 2** Finale Q-Matrix A der MDC-Items

Item	Kognitive Anforderung		
	Corroboration	Situationsmodell	Intertextmodell
C1	1	0	0
C5	1	0	0
S1	0	1	0
S4	0	1	0
I3	0	0	1
I5	0	0	1
I6	0	0	1
D5	0	1	1
D6	0	1	1

C Corroboration, S Situationsmodell, I Intertextmodell, D Dokumentenmodell

**Tab. 3** Korrelationen zwischen den drei kognitiven Anforderungen

	Corroboration	Situationsmodell	Intertextmodell
Corroboration	1,00	–	–
Situationsmodell	0,79	1,00	–
Intertextmodell	0,37	0,86	1,00

gen herangezogen werden konnten. Hinsichtlich der Kriterien wurde darauf geachtet, dass kein Item einen Diskriminationswert  $<0,20$  und jedes Items zumindest einen RMSEA-Fit von  $<0,10$  aufweist. Zudem wurden die Chi-Quadrat Tests der Itempaare herangezogen, um problematische Items anhand signifikanter Testwerte zu identifizieren (George et al. 2019).

Die Gesamtgüte des Modells, das auf Basis der Q-Matrix in Tab. 2 berechnet wurde, kann mit  $\chi^2 = 5,42$ ,  $p = 0,72$  und SRMSR = 0,05 als gut beurteilt werden (George und Robitzsch 2015). Die erste Forschungsfrage, ob die Mehrdimensionalität des MDC-Modells basierend auf den drei zugrunde liegenden kognitiven Anforderungen empirisch abgebildet werden kann, ist mit Verweis auf die Modellwerte vorerst positiv zu bewerten. Für eine differenziertere Einschätzung müssen jedoch auch die tetrachorischen Korrelationen zwischen den drei kognitiven Anforderungen berücksichtigt werden (siehe Tab. 3).

Die Korrelationen zeigen, dass zwischen „Corroboration“ und „Intertextmodell“ ein geringer Zusammenhang besteht (0,37), während das „Situationsmodell“ hoch mit „Corroboration“ (0,79) und „Intertextmodell“ (0,86) korreliert. Allerdings könnten die hohen Korrelationen, die vom „Situationsmodell“ ausgehen, evtl. auch darin begründet liegen, dass das Item S1 einen hohen slipping-Wert (0,56) und das Item S4 einen hohen guessing-Wert (0,51) aufweist, was sich folglich in den geringen bis mäßigen Diskriminationswerten niederschlägt (S1 = 0,21 und S4 = 0,39; s. ESM 3).

Um die Annahme des dreidimensionalen CDM-Modells empirisch weiter abzusichern, wurden auch zwei zweidimensionale Modelle und ein vierdimensionales Modell berechnet. Die beiden zweidimensionalen Modelle basieren zum einen auf der Korrelation zwischen den kognitiven Anforderungen „Corroboration“ und „Situationsmodell“ – kurz: „2-dim-a“ – und zum anderen auf jener zwischen „Situa-

**Tab. 4** Modellwerte der einzelnen CDM-Modelle

Modell	AIC	BIC	Max $\chi^2$	$p(\max(\chi^2))$	SRMSR	Mean item RMSEA	Mean item discrimination
2-dim-a	1992,976	2061,493	6,348	0,423	0,060	<b>0,026</b>	0,448
2-dim-b	<b>1992,866</b>	<b>2061,382</b>	5,692	0,614	0,051	0,028	0,422
3-dim	1992,873	2074,440	<b>5,418</b>	<b>0,717</b>	0,050	0,030	0,473
4-dim	2000,391	2108,060	6,183	0,464	<b>0,048</b>	0,041	<b>0,497</b>

AIC Akaike Information Criterion, BIC Bayesian Information Criterion, SRMSR Standardized Root Mean Square Residual, RMSEA Root Mean Square Error of Approximation

**Tab. 5** Vergleiche zwischen den CDM-Modellen

Modell 1	Modell 2	$\chi^2$	df	p
3-dim	4-dim	8,482	8	0,388
2-dim-b	4-dim	16,475	12	0,170
2-dim-a	4-dim	16,585	12	0,166
2-dim-b	3-dim	7,993	4	0,092
2-dim-a	3-dim	8,103	4	0,088

tionsmodell“ und „Intertextmodell“ – kurz: „2-dim-b“. Die korrelierenden kognitiven Anforderungen wurden dabei jeweils auf eine reduziert. Umgekehrt wurde für das vierdimensionale Modell eine weitere kognitive Anforderung angenommen, nämlich das „Dokumentenmodell“, dem die beiden Items D5 und D6 zugeordnet wurden („1“ in der Q-Matrix), wobei gleichzeitig die Ladungen auf dem Situations- und Intertextmodell jeweils auf „0“ gesetzt wurden. Die Modellierung eines vierdimensionalen Modells basiert auf der Annahme, dass es sich bei dem „Dokumentenmodell“ um eine eigenständige kognitive Anforderung handelt, welche sich qualitativ nicht ausschließlich aus der gleichzeitigen Anwendung des Situations- und Intertextmodells erklären lässt. Die einzelnen Modellwerte sowie die Ergebnisse der Modellvergleiche (zur Methodik s. hierzu George und Robitzsch 2015) werden in den Tab. 4 und 5 dargestellt.

Wie Tab. 4 zu entnehmen ist, weist das zweidimensionale Modell (2-dim-b) die niedrigsten AIC- und BIC-Werte auf. Bezogen auf den niedrigsten Chi-Quadratwert der Itempaare wäre hingegen das dreidimensionale Modell zu präferieren. Basierend auf dem mittleren RMSEA-Wert der Items wäre ein zweidimensionales Modell (2-dim-a) zu bevorzugen, während der SRMSR-Wert und die mittlere Itemdiskrimination das vierdimensionale Modell besser erscheinen lassen. Basierend auf den LRT-basierten Modellvergleichen (Tab. 5) wird allerdings deutlich, dass das vierdimensionale Modell die Daten schlechter als alle anderen Modelle erklärt. Das dreidimensionale Modell erklärt hingegen die Daten mit zumindest  $p < 0,10$  signifikant besser als die zweidimensionalen Modelle, allerdings zeigt auch dieses Ergebnis nicht die gewünschte Eindeutigkeit.

Da die Modellkennwerte mit Q-Matrix A nicht zufriedenstellend erfüllt wurden, wurde basierend auf Q-Matrix B a priori ein vierdimensionales Modell berechnet und erneut jene Items ausgeschlossen, welche die erforderlichen Kriterien nicht erfüllten, wobei nun anstatt der Items S1 und S4 die Items S5 und S6 am besten zum Modell

**Tab. 6** Finale Q-Matrix B der MDC-Items

Item	Kognitive Anforderung			
	Corroboration	Situationsmodell	Intertextmodell	Dokumentenmodell
C1	1	0	0	0
C5	1	0	0	0
S5	0	1	0	0
S6	0	1	0	0
I3	0	0	1	0
I5	0	0	1	0
D5	0	0	0	1
D6	0	0	0	1

C Corroboration, S Situationsmodell, I Intertextmodell, D Dokumentenmodell

**Tab. 7** Korrelationen zwischen den vier kognitiven Anforderungen

	Corroboration	Situationsmodell	Intertextmodell	Dokumentenmodell
Corroboration	1,00	–	–	–
Situationsmodell	–0,71	1,00	–	–
Intertextmodell	0,19	–0,57	1,00	–
Dokumentenmodell	0,43	–0,78	0,82	1,00

passten, während das Item I6 entfernt werden musste. Die finale Q-Matrix B ist in Tab. 6 dargestellt:

Die Gesamtgüte des Modells, das auf Basis der Q-Matrix in Tab. 6 berechnet wurde, kann mit  $\chi^2 = 2,29$ ,  $p = 1,00$  und  $SRMSR = 0,04$  als sehr gut beurteilt werden und weist damit bessere Werte auf, als das auf Q-Matrix A basierende Modell (George und Robitzsch 2015). In Tab. 7 sind ebenfalls die tetrachorischen Korrelationen angeführt.

Die tetrachorischen Korrelationen liegen für dieses Modell entweder im negativen Bereich zwischen  $-0,79$  und  $-0,57$  bzw. zwischen  $0,19$  und  $0,43$ , lediglich das Intertext- und das Dokumentenmodell korrelieren mit  $0,82$ , weshalb diese beiden Skills im Zuge der Dimensionalitätsprüfung in einem eigenen Modell zusammengelegt werden. Zur erneuten Prüfung der Dimensionalität ergeben sich daher drei Modellvarianten: ein vierdimensionales, ein dreidimensionales (bei dem die Items des Dokumentenmodells jeweils zum „Situations-“ und „Intertextmodell“ zugeordnet werden, kurz „3-dim-a“) und ein weiteres dreidimensionales (Zusammenlegung der Skills „Intertext-“ und „Dokumentenmodell“ auf Basis der Korrelation von  $0,82$ , „3-dim-b“).

Bezogen auf die Werte AIC, BIC und mittlerem RMSEA-Fit der Items weist das Modell „3-dim-b“ die besten Kennzahlen auf, während das vierdimensionale Modell „4-dim“ die besseren Werte im Chi-Quadrat-Test der Itempaare, im SRMSR-Fit des Gesamtmodells und in der mittleren Itemdiskrimination zeigt (Tab. 8). Schlussendlich verdeutlichen jedoch die LR-Tests in Tab. 9, dass das vierdimensionale Modell die Daten signifikant besser erklärt ( $p < 0,000$ ). Um auch ein eindimensionales Modell in Anlehnung an Schoor et al. (2020) auszuschließen, wurde zusätzlich ein eindimensionales Raschmodell mit den R-Paketen *ltm* (Rizopoulos 2018), *TAM*

**Tab. 8** Modellwerte der einzelnen CDM-Modelle

Modell	AIC	BIC	Max $\chi^2$	$p(\max(\chi^2))$	SRMSR	Mean item RMSEA	Mean item discrimination
3-dim-a	1928,257	2003,298	15,517	0,002	0,103	0,054	0,454
3-dim-b	<b>1901,404</b>	<b>1976,446</b>	6,002	0,400	0,054	<b>0,030</b>	0,511
4-dim	1903,817	2004,960	<b>2,290</b>	<b>1,000</b>	<b>0,036</b>	0,045	<b>0,574</b>

AIC Akaike Information Criterion, BIC Bayesian Information Criterion, SRMSR Standardized Root Mean Square Residual, RMSEA Root Mean Square Error of Approximation

**Tab. 9** Vergleiche zwischen den CDM-Modellen

Modell 1	Modell 2	$\chi^2$	df	p
3-dim-a	<b>4-dim</b>	40,440	8	<b>0,000</b>
3-dim-b	<b>4-dim</b>	13,587	8	<b>0,000</b>

**Tab. 10** Zusammenfassung der Itemkennwerte

	Min	Max	M	SD
p-value	0,35	0,83	0,58	0,16
Guessing	0,00	0,60	0,21	0,21
Slipping	0,03	0,43	0,21	0,14
Discrimination	0,22	0,97	0,57	0,25
Easiness	0,33	0,72	0,50	0,13
RMSEA	0,00	0,09	0,04	0,03

(Robitzsch et al. 2019a), *WrightMap* (Torres Irribarra und Freund 2016) und *irtoys* (Partchev et al. 2017) berechnet. Die EAP-Reliabilität des berechneten Raschmodells lag jedoch bei 0,00, zudem ergab der LR-Test nach Andersen zur Prüfung der Eindimensionalität ein signifikantes Ergebnis ( $\chi^2(7) = 16,114, p = 0,024$ ), weshalb die Annahme der Eindimensionalität verworfen werden musste. Die zusammengefassten Itemkennwerte des somit anzunehmenden vierdimensionalen Modells sind in Tab. 10 angeführt (Bley und George 2017; George und Robitzsch 2015). Die einzelnen Werte eines jeden Items können in ESM 3 eingesehen werden.

Die p-values beschreiben die relativen Lösungshäufigkeiten, welche mit einem Mittelwert von 0,58 auf einen Itempool mit mittlerer Schwierigkeit hinweisen. Die guessing-Parameter der Items zeigen einen Mittelwert von 0,21, d.h. die Wahrscheinlichkeit, ein Item trotz nicht vorhandener Fähigkeiten korrekt zu lösen, ist im Mittel gering. Gleiches trifft auf die slipping-Parameter zu ( $M = 0,21$ ), welche angeben, mit welcher Wahrscheinlichkeit ein Item trotz vorhandener Fähigkeiten nicht gelöst wird. Mit einer mittleren Diskriminationsfähigkeit von 0,57 wird im Schnitt gut zwischen kompetenten und nicht-kompetenten Proband\*innen unterschieden. Ähnlich wie die p-values beschreibt die „easiness“ die mittlere Lösungshäufigkeit der Items (basierend auf den Werten von „guessing“ und „slipping“, s. George und Robitzsch 2015), wobei der Mittelwert von 0,50 auf eine ausgewogene Lösungshäufigkeit hinweist. Die RMSEA-Werte stellen mit 0,04 einen guten mittleren Item-Fit dar, weshalb davon ausgegangen werden kann, dass die Items zum postulierten

**Tab. 11** Beherrschungswahrscheinlichkeiten der kognitiven Anforderungsbereiche

Kognitiver Anforderungsbereich	Beherrschungswahrscheinlichkeit	SE
Corroboration	0,63	0,06
Situationsmodell	0,54	0,08
Intertextmodell	0,54	0,05
Dokumentenmodell	0,84	0,07

**Tab. 12** Kombinationen von beherrschten Anforderungsbereichen in der Population

Klasse				Auftretenswahrscheinlichkeit	SE
Corr.	Sit.	Inter.	Dok.		
0	1	0	0	0,09	0,03
1	1	0	0	0,06	0,03
1	0	0	1	0,11	0,04
0	1	0	1	0,11	0,04
0	0	1	1	0,05	0,04
1	1	0	1	0,09	0,04
1	0	1	1	0,29	0,05
0	1	1	1	0,11	0,04
1	1	1	1	0,07	0,08

Klasse „0100“: nur *Situationsmodell* wird beherrscht; Klasse „1100“: nur *Corroboration* und *Situationsmodell* wird beherrscht; usw.

Corr. Corroboration, Sit. Situationsmodell, Inter. Intertextmodell, Dok. Dokumentenmodell

Modell passen, wengleich manche Items den optimalen Schwellenwert von 0,05 überschreiten, jedoch zumindest noch  $<0,10$  liegen.

Zur Beantwortung der zweiten Forschungsfrage, mit welcher Wahrscheinlichkeit die vier aus dem MDC-Modell abgeleiteten kognitiven Anforderungsbereiche in der Population der Lehramtsstudent\*innen beherrscht werden und mit welcher Wahrscheinlichkeit deren Kombinationen (Skill-Klassen) in der Population auftreten, werden im Folgenden die (1) Beherrschungswahrscheinlichkeiten der einzelnen Anforderungsbereiche (siehe Tab. 11) und (2) die Auftretenswahrscheinlichkeiten der unterschiedlichen Kombinationen von beherrschten Anforderungsbereichen in Form von Skill-Klassen (siehe Tab. 12) angeführt.

Wie Tab. 11 entnommen werden kann, werden die einzelnen Anforderungsbereiche „Corroboration“ (ca. 63 %), „Situationsmodell“ (ca. 54 %) und „Intertextmodell“ (ca. 54 %) in der Population der untersuchten Lehramtsstudent\*innen mit mittlerer Wahrscheinlichkeit beherrscht, während das „Dokumentenmodell“ die größte Beherrschungswahrscheinlichkeit mit ca. 84 % aufweist. Dieses Ergebnis überrascht insofern, als basierend auf den Forschungsergebnissen von Schoor et al. (2020) anzunehmen wäre, dass Aufgaben zum Dokumentenmodell schwieriger zu lösen sind als Aufgaben der anderen Anforderungsbereiche. Dieser Befund könnte sich aber auch dadurch erklären lassen, dass die Items D5 und D6 eine Lösungshäufigkeit von  $p=0,81$  und  $p=0,71$  aufweisen, diese Items daher einen zu geringen Schwierigkeitsgrad aufweisen.

Tab. 12 gibt Auskunft über die Auftretenswahrscheinlichkeiten der einzelnen Skill-Klassen, wobei aus Gründen der Übersicht nur jene angeführt werden, welche mind. 5% erreichen – die vollständige Auflistung befindet sich in ESM 4:

Bei den Skill-Klassen zeigt sich, dass sich die meisten bei einer Auftretenswahrscheinlichkeit von ca. 5 bis 11% bewegen, weshalb sich hier keine eindeutigen Muster zeigen. Am größten ist jene Skill-Klasse, bei der nur das Situationsmodell nicht beherrscht wird, jedoch alle anderen Anforderungsbereiche (ca. 29%). Alle Anforderungsbereiche beherrschen jedoch nur ca. 7% in der befragten Population, d. h., ca. 93% der Student\*innen beherrschen zumindest einen der vier Anforderungsbereiche des MDC-Modells nicht. Daraus lässt sich ableiten, dass für die meisten Lehramtsstudent\*innen in der befragten Population zumindest einer der vier Anforderungsbereiche noch gezielt gefördert werden müsste. Die übrigen ca. 2%, die in Tab. 12 nicht angeführt sind, verteilen sich auf die hier nicht angeführten Skill-Klassen.

## 7 Diskussion

Die Ergebnisse der empirischen Untersuchung konnten insgesamt aufzeigen, dass die vier kognitiven Anforderungsbereiche, die von Schoor et al. (2020) basierend auf dem MDC-Modell abgeleitet wurden – „Corroboration“, „Situationsmodell“, „Intertextmodell“ und „Dokumentenmodell“ – mithilfe eines kognitiven Diagnosemodells mehrdimensional modellierbar sind und von den Lehramtsstudent\*innen mit mittelhoher bis hoher Wahrscheinlichkeit beherrscht werden. Dabei konnte auch gezeigt werden, dass sich das Dokumentenmodell im Rahmen der Dimensionalitätsprüfungen nicht auf die Kombination des Situation- und Intertextmodells reduzieren ließ, sondern dass es vielmehr einen qualitativ eigenständigen Anforderungsbereich darstellt. Mit einem durchschnittlichen  $p$ -value von 0,58 waren die Items für die Lehramtsstudent\*innen insgesamt von mittelhoher Schwierigkeit. Allerdings waren die Lösungshäufigkeiten der Items D5 und D6 für das Dokumentenmodell insgesamt zu hoch ( $p=0,71$  und  $0,83$ ), weshalb hier von einer systematischen Überschätzung der Lehramtsstudent\*innen in diesem Bereich ausgegangen werden kann (ca. 84% Beherrschungswahrscheinlichkeit in der Population). Zudem korreliert der Anforderungsbereich des Dokumentenmodells sehr hoch mit dem Anforderungsbereich des Intertextmodells (0,82), weshalb diese Items vermutlich zu wenig die Anforderung erfüllen, die Autor\*innen- und Textmerkmale mit den aus den Texten abzuleitenden inhaltlichen Inferenzen zu verknüpfen. In zukünftigen Studien wäre daher insgesamt anzustreben, insbesondere für diesen Anforderungsbereich mehr Items zu generieren, die zu einer etwas niedrigeren Lösungswahrscheinlichkeit und zu einer geringeren Korrelation mit dem Anforderungsbereich des Intertextmodells führen.

Ergänzend zu den CDM-basierten Ergebnissen, welche die vierdimensionale Modellierung als optimale Modellierung des MDC-Modells nahelegen, konnten mit einer eindimensionalen Rasch-Modellierung die Ergebnisse von Schoor et al. (2020) in der vorliegenden Studie nicht repliziert werden. Mit einer EAP-Reliabilität von 0,00 und einem signifikanten Ergebnis des Andersen LR-Tests auf Eindimensionalität ( $p=0,024$ ) konnte nicht nachgewiesen werden, dass den MDC-Aufgaben eine

gemeinsame Dimension zugrunde liegt. Vielmehr sprechen die vorliegenden Ergebnisse dafür, dass für jeden der vier Anforderungsbereiche unterschiedlich schwierige Aufgaben generiert werden können, sich diese jedoch qualitativ voneinander unterscheiden und daher keine schwierigkeitsbestimmenden Merkmale aufzuweisen scheinen, die über alle Anforderungsbereiche hinweg ident sind. Ob sich die Ergebnisse der vorliegenden Studie generalisieren lassen, wäre zukünftig in weiteren Untersuchungen zu überprüfen.

Neben den Ergebnissen zur statistischen Modellierung konnte festgestellt werden, dass jene Skill-Klasse, in der alle vier Anforderungsbereiche beherrscht werden, in der Population lediglich ca. 7% umfasst. D.h. im Umkehrschluss, dass ca. 93% der Lehramtsstudent\*innen zumindest einen Anforderungsbereich nicht beherrschen. Da die Skill-Klasse, in der das Situationsmodell nicht beherrscht wird (jedoch alle anderen Anforderungsbereiche), ca. 29% umfasst, wäre eine mögliche Implikation für die Lehrer\*innenbildung, die Student\*innen stärker darin zu fördern, basierend auf zwei unterschiedlichen Textquellen die richtigen Inferenzen zu ziehen. Positiv kann hingegen der Umstand interpretiert werden, dass jene Skill-Klasse, in der keiner der vier Anforderungsbereiche beherrscht wird, unter 2% liegt.

In Bezug auf die Aussagekraft der Ergebnisse hinsichtlich der Kompetenzeinschätzungen ist als Limitation einschränkend zu erwähnen, dass durch die standardisierte psychometrische Erfassung mithilfe von Verifikationsitems keine Aussage darüber getroffen werden kann, ob und wie die Lehramtsstudent\*innen Textquellen im Rahmen von zu schreibenden Essays miteinander verknüpfen würden (z. B. Stadler et al. 2014). Diese Forschungslücke könnte möglicherweise geschlossen werden, indem Essays hinsichtlich der darin angewandten Strategien zum Vergleich multipler Textdokumente dichotom kodiert würden. Mit der daraus erzeugten Datenmatrix könnten ebenfalls kognitive Diagnosemodelle berechnet werden, ohne dabei auf Antwortvorgaben in Form von Verifikationsitems zurückgreifen zu müssen.

Eine weitere Einschränkung besteht in der Erfassung des Skills „Corroboration“. Diese kognitive Anforderung wurde in Anlehnung an Schoor et al. (2020) bzw. Wineburg (1991) als die Fähigkeit definiert, inhaltliche Aussagen zwischen Texten auf ihre Übereinstimmung hin abzugleichen. In den Corroboration-Items wurde jedoch nicht nach der inhaltlichen Übereinstimmung von Aussagen in beiden Texten gefragt, sondern danach, ob beide Texte ein bestimmtes Thema (wie z. B. den Ausgleich sozialer Unterschiede) *grundsätzlich* thematisieren, unabhängig von inhaltlichen Übereinstimmungen. Zukünftige Forschungen müssten diesem Umstand Rechnung tragen, indem auch Corroboration-Items konzipiert werden, bei denen Aussagen hinsichtlich der Übereinstimmung ihrer Inhalte überprüft werden.

Abschließend bleibt im Rahmen der vorliegenden Ergebnisse die Frage offen, inwiefern die in Anlehnung an Schoor et al. (2020) definierten kognitiven Anforderungen zum Vergleich von Textdokumenten mit Fähigkeiten im Bereich der Lesekompetenz korrelieren bzw. übereinstimmen. Erste Forschungsergebnisse hierzu konnten bereits Mahlow et al. (2020) generieren, die aufzeigen, dass basierend auf dem MDC-Test von Schoor et al. (2020) und dem (ergebnisorientierten) Lesekompetenztest des Nationalen Bildungspanels (NEPS) bei  $n=501$  Student\*innen zweier deutscher Universitäten die MDC-Fähigkeiten von der erfassten Lesekompetenz empirisch getrennt werden können. Die statistischen Modellierungen legen

hierbei nahe, dass die MDC-Anforderungen und die Lesekompetenz weder durch einen gemeinsamen Faktor noch durch die Annahme, dass MDC-Fähigkeiten auf die Lesekompetenz aufbauen, erklärt werden können. Vielmehr scheint es sich basierend auf den empirischen Befunden von Mahlow et al. (2020) bei MDC-Fähigkeiten und der Lesekompetenz um separate Konstrukte zu handeln, wenngleich beide mit  $r=0,84$  korrelieren. Mahlow et al. (2020) gehen davon aus, dass die hohe Überlappung beider Konstrukte auf allgemeine Fähigkeiten in der Dekodierung von Wörtern und Sätzen oder die Intelligenz zurückzuführen sind. Wie die Autor\*innen jedoch aufzeigen, bedürfen zukünftige Studien über den Zusammenhang von MDC-Fähigkeiten und Lesekompetenz jedoch auch des Einbezugs anderer Konzeptualisierungen von Lesekompetenz, wie z. B. prozessorientierte Perspektiven der Kognitionspsychologie.

**Zusatzmaterial online** Zusätzliche Informationen sind in der Online-Version dieses Artikels (<https://doi.org/10.1007/s11618-021-01048-x>) enthalten.

**Funding** Open access funding provided by Paris Lodron University of Salzburg.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

**Interessenkonflikt** F. Nagele, M. Katstaller und U. Greiner geben an, dass kein Interessenkonflikt besteht.

## Literatur

- Anmarkrud, Ø., Bråten, I., & Strømso, H.I. (2014). Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences, 30*, 64–76. <https://doi.org/10.1016/j.lindif.2013.01.007>.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft, 9*(4), 469–520.
- Benner, D. (1980). Das Theorie-Praxis-Problem in der Erziehungswissenschaft und die Frage nach Prinzipien pädagogischen Denkens und Handelns. *Zeitschrift für Pädagogik, 26*(4), 485–497.
- Bley, S., & George, A. C. (2017). Kognitive Diagnosemodelle zur Begleitung individualisierter Lehr- und Lernprozesse. Neue Möglichkeiten aus einem alternativen Ansatz? *Zeitschrift für Berufs- und Wirtschaftspädagogik, 113*(1), 56–85.
- Blömeke, S., Felbrich, A., & Müller, C. (2008). Messung erziehungswissenschaftlichen Wissens angehender Lehrkräfte. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer: Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematik-Studierender und -referendare – Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung* (S. 171–193). Münster: Waxmann.
- Braasch, J.L.G., Bråten, I., Strømso, H.I., & Anmarkrud, Ø. (2014). Incremental theories of intelligence predict multiple document comprehension. *Learning and Individual Differences, 31*, 11–20. <https://doi.org/10.1016/j.lindif.2013.12.012>.

- Britt, M., & Rouet, J. (2012). Learning with multiple documents. In J. Kirby & M. Lawson (Hrsg.), *Enhancing the quality of learning: dispositions, instruction, and learning processes* (S. 276–314). Cambridge: Cambridge University Press.
- Brühwiler, C. (2014). *Adaptive Lehrkompetenz und schulisches Lernen: Effekte handlungssteuernder Kognitionen von Lehrpersonen auf Unterrichtsprozesse und Lernergebnisse der Schülerinnen und Schüler. Pädagogische Psychologie und Entwicklungspsychologie*. Münster: Waxmann.
- Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J.-F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist*, 46(1), 48–70. <https://doi.org/10.1080/00461520.2011.538647>.
- Budde, J. (2012). Problematisierende Perspektiven auf Heterogenität als ambivalentes Thema der Schul- und Unterrichtsforschung. *Zeitschrift für Pädagogik*, 58(4), 522–540.
- Buholzer, A., & Wyss, K. A. (Hrsg.). (2010). *Alle gleich – alle unterschiedlich! Zum Umgang mit Heterogenität in Schule und Unterricht*. Seelze-Velber: Klett.
- Burn, K., & Mutton, T. (2015). A review of 'research-informed clinical practice' in initial teacher education. *Oxford Review of Education*, 41(2), 217–233. <https://doi.org/10.1080/03054985.2015.1020104>.
- George, A. C., & Robitzsch, A. (2015). Cognitive diagnosis models in R: a didactic. *The Quantitative Methods for Psychology*, 11(3), 189–205. <https://doi.org/10.20982/tqmp.11.3.p189>.
- George, A. C., Bley, S., & Pellegrino, J. (2019). Characterizing and diagnosing complex professional competencies—an example of intrapreneurship. *Educational Measurement: Issues and Practice*, 38(2), 89–100. <https://doi.org/10.1111/emip.12257>.
- Gräsel, C., Decristan, J., & König, J. (2017). Adaptiver Umgang mit Heterogenität im Unterricht. *Unterrichtswissenschaft*, 45(4), 195–206.
- Hahnel, C., Kroehne, U., Goldhammer, F., Schoor, C., Mahlow, N., & Artelt, C. (2019). Validating process variables of sourcing in an assessment of multiple document comprehension. *British Journal of Educational Psychology*, 89, 524–537. <https://doi.org/10.1111/bjep.12278>.
- Heinrich, M. (2013). Bildungsgerechtigkeit für alle! – aber nicht für jeden? In F. Dietrich, M. Heinrich & N. Thieme (Hrsg.), *Bildungsgerechtigkeit jenseits von Chancengleichheit* (S. 181–194). Wiesbaden: Springer.
- Hoffmann-Ocon, A., & Schmidtke, A. (Hrsg.). (2012). *Reformprozesse im Bildungswesen. Zwischen Bildungspolitik und Bildungswissenschaft*. Wiesbaden: Springer VS.
- Jungová, Z. (2006). Textsorten in der Presse und Publizistik und ihre stilistische Realisierung [Diplomarbeit, Universität Brunn]. <https://is.muni.cz/th/uls6w/final.pdf>. Zugegriffen: 15. Mai 2020.
- König, J., & Blömeke, S. (2009). Pädagogisches Wissen von österreichischen Lehramtsstudierenden. *Erziehung und Unterricht*, 159(1–2), 175–186.
- Kuhlmann, C. (2012). *Bildungspolitik und Leistungsvergleichsstudien. PISA 2000 und die Ganztagschulentwicklung*. Wiesbaden: Springer VS.
- Lenk, H. E. H., & Vesalainen, M. (2012). Der Kommentar als persuasiver Text. Vergleichende Untersuchungen zu einer meinungsbetonten Textsorte in europäischen Massenkommunikationsmedien. In H. E. H. Lenk & M. Vesalainen (Hrsg.), *Persuasionsstile in Europa. Methodologie und Empirie kontrastiver Untersuchungen zur Textsorte Kommentar* (S. 7–32). Hildesheim: Georg Olms.
- Mahlow, N., Hahnel, C., Kroehne, U., Artelt, C., Goldhammer, F., & Schoor, C. (2020). More than (single) text comprehension?—On university students' understanding of multiple documents. *Frontiers in Psychology*, 11, 1–17. <https://doi.org/10.3389/fpsyg.2020.562450>.
- Neuweg, H. G. (2011). Distanz und Einlassung. Skeptische Anmerkungen zum Ideal einer „Theorie-Praxis-Integration“ in der Lehrerbildung. *Erziehungswissenschaft*, 22(43), 33–45.
- Partchev, I., Maris, G., & Hattori, T. (2017). Irtoys. A collection of functions related to item response theory (IRT). *R package version 0.2.1*. <https://cran.r-project.org/web/packages/irtoys/index.html>. Zugegriffen: 24. Sept. 2020.
- QSR = Qualitätssicherungsrat für Pädagoginnen- und Pädagogenbildung (2013). Professionelle Kompetenzen von PädagogInnen. Zielperspektive. [https://www.qsr.or.at/dokumente/1869-20140529-092429-Professionelle\\_Kompetenzen\\_von\\_PaedagogInnen\\_Zielperspektive.pdf](https://www.qsr.or.at/dokumente/1869-20140529-092429-Professionelle_Kompetenzen_von_PaedagogInnen_Zielperspektive.pdf). Zugegriffen: 15. Mai 2020.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>. Zugegriffen: 31. Jan. 2020.
- Retting, H. (2017). *Wissenschaftliche Arbeiten schreiben*. Stuttgart: J. B. Metzler.
- Rhein, R. (2019). Pädagogisches Fallverstehen. Forschendes Lernen im Lehramtsstudium. In M. E. Kaufmann, A. Stilmis & H. A. Mieg (Hrsg.), *Forschendes Lernen in den Geisteswissenschaften. Konzepte, Praktiken und Perspektiven hermeneutischer Fächer* (S. 211–227). Wiesbaden: Springer VS.

- Rizopoulos, D. (2018). ltm. Latent trait models under IRT. R package version 1.1-1. <https://github.com/drizopoulos/ltm>. Zugegriffen: 24. Sept. 2020.
- Robitzsch, A., Kiefer, T., & Wu, M. (2019a). tam. test analysis modules. R package version 3.3-10. <https://github.com/alexanderrobitzsch/TAM>. Zugegriffen: 24. Sept. 2020.
- Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2019b). CDM. R package version 7.4-19. <https://cran.r-project.org/web/packages/CDM/index.html>. Zugegriffen: 31. Jan. 2020.
- Von la Roche, W. (2013). *Einführung in den praktischen Journalismus. Mit genauer Beschreibung aller Ausbildungswege Deutschland – Österreich – Schweiz*. Wiesbaden: Springer VS.
- Ruhl, K., Mahrt, N., & Töbel, J. (2010). *Publizieren während der Promotion*. Wiesbaden: VS.
- Scharenberg, K. (2012). *Leistungsheterogenität und Kompetenzentwicklung. Zur Relevanz klassenbezogener Kompositionsmerkmale im Rahmen der KESS-Studie*. Münster: Waxmann.
- Schmied-Kowarzik, W. (2018). Pädagogik als gesellschaftskritische Wissenschaft. In M. Rühle, A. Bernhard & L. Rothermel (Hrsg.), *Handbuch Kritische Pädagogik* (S. 44–64). Weinheim: Beltz.
- Schoor, C., Hahnel, C., Artelt, C., Reimann, D., Kröhne, U., & Goldhammer, F. (2020). Entwicklung und Skalierung eines Tests zur Erfassung des Verständnisses multipler Dokumente von Studierenden. *Diagnostica*, 66(2), 123–135. <https://doi.org/10.1026/0012-1924/a000231>.
- Seifert, A., Hilligus, A. H., & Schaper, N. (2010). Entwicklung und psychometrische Überprüfung eines Messinstruments zur Erfassung pädagogischer Kompetenzen in der universitären Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, 2(1), 82–103.
- Shulman, L. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher*, 15, 4–14. <https://doi.org/10.3102/0013189X015002004>.
- Stadtler, M., Scharrer, L., Skodzik, T., & Bromme, R. (2014). Comprehending multiple documents on scientific controversies: effects of reading goals and signaling rhetorical relationships. *Discourse Processes*, 51(1–2), 93–116. <https://doi.org/10.1080/0163853x>.
- Stojanov, K. (2011). *Bildungsgerechtigkeit. Rekonstruktionen eines unkämpften Begriffs*. Heidelberg: VS.
- Torres Iribarra, D., & Freund, R. (2016). *WrightMap. IRT item-person map with 'ConQuest' Integration. R package version 1.2.1*. <https://cran.r-project.org/web/packages/WrightMap/index.html>. Zugegriffen: 24. Sept. 2020.
- Trautmann, M., & Wischer, B. (2011). *Heterogenität in der Schule. Eine kritische Einführung*. Wiesbaden: VS.
- Wineburg, S. S. (1991). Historical problem solving: a study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83(1), 73–87. <https://doi.org/10.1037/0022-0663.83.1.73>.
- Winter-Hölzl, A., Wäschle, K., Wittwer, J., Watermann, R., & Nückles, M. (2015). Entwicklung und Validierung eines Tests zur Erfassung des Genrewissens Studierender und Promovierender der Bildungswissenschaften. *Zeitschrift für Pädagogik*, 61, 185–202.