



**Kompetenz und Performanz von (angehenden) Lehrkräften
im Bereich Deutsch als Zweitsprache (DaZ) -
Entwicklung und Validierung eines performanznahen Testinstruments**

Von der Fakultät Bildung
der Leuphana Universität Lüneburg zur Erlangung des Grades

Doktorin der Philosophie

- Dr. phil. -

genehmigte Dissertation von

Svenja Lemmrich

geboren am 14.11.1990 in Hamburg

Eingereicht am: 22.12.2020

Mündliche Verteidigung (Disputation) am: 06.09.2021

Betreuer und Erstgutachter: Prof. Dr. Timo Ehmke
Zweitgutachterin: Prof. Dr. Julia Ricart Brede
Drittgutachter: Prof. Dr. Marcus Pietsch

Die einzelnen Beiträge des kumulativen Dissertationsvorhabens sind oder werden ggf. inkl. des Rahmenpapiers wie folgt veröffentlicht:

Lemmrich, S., Bahls, A. & Ehmke, T. (2020). Effekte von mündlichen versus schriftlichen Antwortformaten bei der performanznahen Messung von Deutsch-als-Zweitsprache (DaZ)-Kompetenz – eine experimentelle Studie mit angehenden Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 1–11. <https://doi.org/10.1024/1010-0652/a000282>

Lemmrich, S., Hecker, S.-L., Klein, S., Ehmke, T., Köker, A., Koch-Priewe, B. & Ohm, U. (2019). Performanznahe und videobasierte Messung von DaZ-Kompetenz bei Lehrkräften: Skalierung und dimensionale Struktur des Testinstruments. In T. Ehmke, P. Kuhl & M. Pietsch (Hg.), *Lehrer. Bildung. Gestalten: Beiträge zur empirischen Forschung in der Lehrerbildung* (1. Aufl., S. 190–204). Beltz.

Lemmrich, S., Hecker, S.-L., Klein, S., Ehmke, T., Koch-Priewe, B., Köker, A. & Ohm, U. (2020). Linguistically Responsive Teaching in Multilingual Classrooms. Development of a Performance-Oriented Test to Assess Teachers' Competence. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper & C. Lautenbach (Hg.), *Student Learning in German Higher Education* (1. Aufl., S. 125–140). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27886-1_7

Lemmrich, S. & Ehmke, T. (eingereicht). Performance-Oriented Measurement of Teachers' Competency in Linguistically Responsive Teaching: Connection of LRT-Competence with Relevant Learning Opportunities, Personality Factors and Beliefs. *Language Awareness*.

Veröffentlichungsjahr: 2021

Danksagung

Ich bedanke mich ganz herzlich bei meinem Erstgutachter Timo Ehmke für die verlässliche und geduldige Unterstützung, die großartige Betreuung und das Vertrauen. Danke für die zahlreichen Möglichkeiten, mich weiterzuentwickeln und mich auf Tagungen und anderen Forschungsreisen auszutauschen. Ein großer Dank gilt meiner Zweitgutachterin Julia Ricart Brede, die das *DaZKom-Video*-Projekt mit ihrer Expertise unterstützt und mir im Rahmen ihres Kolloquiums wertvolle Denkanstöße gegeben hat. Danke an meinen Drittgutachter Marcus Pietsch, der mich von Beginn meiner Promotionsphase an begleitet hat und stets eine hilfreiche Antwort auf methodische Fragen wusste. Ein Dank gilt auch meinen Projektkolleginnen und -kollegen im *DaZKom*-Team. Insbesondere bedanke ich mich bei Sarah-Larissa Hecker und Stephanie Falkenstern. Durch unseren intensiven Austausch, spannende Diskussionen und Teamwork haben wir große Datenerhebungen stemmen können, gemeinsame Artikel geschrieben und viel gelernt. Danke an Sina Spiekermeier, die stets einen Überblick über den Projektablauf behalten hat, insbesondere in der Endphase meines Schreibprozesses. An dieser Stelle sei auch unseren Hilfskräften für ihre Arbeit gedankt und allen, die uns bei unseren Datenerhebungen unterstützt haben. Danke an Svenja Hammer und Nele Fischer, die mich zur Promotion ermutigt haben und deren Büro in meiner Anfangszeit ein fröhlicher Ort zum auftanken und zum Frust teilen war. Danke an Anika Bahls für die intensive Unterstützung im Projekt und die vielen gemeinsamen Tassen Kaffee. Danke an Lisa Berkel-Otto, durch die diese Reise anfang, als sie mir die Stelle als Hilfskraft im Institut vermittelte und die mich seitdem bei Datenerhebungen, mit ihrer Literaturdatenbank oder mit einem freundschaftlichen Rat unterstützt. Danke an Bianka Troll und Herrad Schönborn, die diesen Weg gemeinsam mit mir gegangen sind. Danke für die vielen Stunden gemeinsames Schreiben, fürs Zuhören, Motivieren und fürs Feiern kleiner und großer Meilensteine. Danke an Anna Holstein für die wunderbaren Gespräche und Powertalks. Ein Dank gilt auch dem Team der Graduate School, durch die ich an einigen spannenden Workshops und Programmen teilnehmen durfte. Insbesondere bedanke ich mich bei Nadine Dablé für das intensive Coaching. Ich bedanke mich bei meinen Freunden, die mich beizeiten mit viel Lachen und Herzenswärme auch mal auf andere Gedanken brachten. Ich bedanke mich bei meiner wunderbaren Familie, die immer zuhört, auch mal Kritik übt und mir stets den Rücken stärkt. Ein großer Dank gilt Simon Schüssler und Sören Lemmrich fürs Korrekturlesen. Und ein besonderer Dank gilt Patrick Naber für die unendliche Geduld, die Motivation und die liebevolle Unterstützung.

Inhalt

Zusammenfassung	2
Summary	3
1. Einleitung	4
2. Theoretischer Rahmen und Verortung der Dissertation	5
2.1 Deutsch-als-Zweitsprache-Kompetenz in der Lehrkräfte(aus)bildung	5
2.2 Kompetenz, Performanz und performanznahes Messen	13
3. Forschungsdesiderate und Forschungsfragen.....	19
4. Beiträge der Dissertation	22
4.1 Artikel 1: Effekte von mündlichen versus schriftlichen Antwortformaten bei der performanznahen Messung von DaZ-Kompetenz (Lemmrich, Bahls & Ehmke, 2020)	23
4.2 Artikel 2: Performanznahe und videobasierte Messung von DaZ-Kompetenz bei Lehrkräften - Skalierung und dimensionale Struktur (Lemmrich et al., 2019)	25
4.3 Artikel 3: Linguistically Responsive Teaching in Multilingual Classrooms - Development of a Performance-Oriented Test to Assess Teachers' Competence (Lemmrich et al., 2020).....	26
4.4 Artikel 4: Performance-Oriented Measurement of Teachers' Competency in Linguistically Responsive Teaching: Connection of LRT-Competence with Relevant Learning Opportunities, Personality Factors and Beliefs (Lemmrich & Ehmke, eingereicht).....	27
5. Zusammenfassung und Diskussion	29
5.1 Diskussion der übergeordneten Forschungsfragen des Rahmenpapiers.....	31
5.2 Limitationen und Empfehlungen für die Forschung	39
5.3 Empfehlungen für die Lehrkräftebildung	41
6. Literaturverzeichnis	44
7. Artikel 1: Effekte von mündlichen versus schriftlichen Antwortformaten bei der performanznahen Messung von Deutsch-als-Zweitsprache (DaZ)-Kompetenz–eine experimentelle Studie mit angehenden Lehrkräften	54
8. Artikel 2: Performanznahe und videobasierte Messung von DaZ-Kompetenz bei Lehrkräften: Skalierung und dimensionale Struktur des Testinstruments	66
9. Artikel 3: Linguistically Responsive Teaching in Multilingual Classrooms. Development of a Performance-Oriented Test to Assess Teachers' Competence.....	82
10. Artikel 4: Performance-Oriented Measurement of Teachers' Competence in Linguistically Responsive Teaching: Connection of LRT Competence with Relevant Learning Opportunities, Personality Factors and Beliefs	99

Zusammenfassung

Die Forderung nach sprachbildendem Fachunterricht ist aktueller denn je. Deshalb stellt sich die Frage, welche Kompetenzen (Fach-) Lehrkräfte benötigen, um dies leisten zu können und wie sie diese in ihrer Aus- und Weiterbildung erwerben können (Baumann, 2017; Köker, 2018). Zwar gibt es Empfehlungen für die Lehrkräftebildung, jedoch ist die Ausbildungslandschaft im Bereich Deutsch als Zweitsprache (DaZ) deutschlandweit und universitätsabhängig sehr heterogen (Gantefort & Michalak, 2017). Um Aussagen über benötigte Kompetenzen treffen zu können, müssen diese messbar gemacht werden. Wissensfacetten im Bereich DaZ können mit einem Paper-Pencil-Test erfasst werden. Um Handlungsfacetten erfassen zu können, wird eine Testumgebung benötigt, die der Komplexität von Unterricht möglichst gerecht wird, weshalb sich Paper-Pencil-Tests dafür nicht eignen (Blömeke, König et al., 2015). Dem Kompetenz-als-Kontinuum-Modell (Blömeke, Gustafsson & Shavelson, 2015) folgend, widmet sich diese Arbeit der Entwicklung eines videobasierten Testinstruments mit einem mündlichen Antwortformat, um so performanznah DaZ-Kompetenz zu erfassen. Folgende Forschungsfragen werden in dieser Dissertation untersucht:

- 1.) Inwieweit können eine ausreichende psychometrische Qualität und eine dimensionale Struktur des Testinstruments festgestellt werden?
- 2.) Inwieweit kann performanznahe DaZ-Kompetenz bei (angehenden) Lehrkräften festgestellt werden?
- 3.) Wie hängen individuelle und akademische Hintergrunddaten der Proband*innen, DaZ-bezogene Lerngelegenheiten, Überzeugungen zu sprachlich-kultureller Heterogenität in Schule und Unterricht oder Persönlichkeitsmerkmale mit performanznaher DaZ-Kompetenz zusammen?

Das Rahmenpapier bettet die vier Artikel in den übergeordneten theoretischen Kontext ein, diskutiert die Ergebnisse und leitet abschließend Empfehlungen für weitere Forschung und für die Lehrkräftebildung ab.

Summary

The demand for Linguistically Responsive Teaching (LRT) is more relevant than ever. Therefore the question arises as to what competences teachers need in this field and how they can acquire these in teacher training (Baumann, 2017; Köker, 2018). Although by now there are a number of recommendations, teacher training in this field greatly differs depending on federal states and universities across Germany (Gantefort & Michalak, 2017). In order to be able to make statements about required competences, these must be made measurable. Knowledge facets in the area of LRT can be measured with a paper-pencil-test. To record facets of action-related competence, a test environment is needed that does justice to the complexity of teaching as much as possible, which is why paper-pencil tests cannot be used here (Blömeke, König et al., 2015). Following the competence-as-continuum model (Blömeke, Gustafsson & Shavelson, 2015), this work focusses on the development of a video-based test instrument with an oral response format to assess LRT-competence performance-oriented. The following research questions are examined in this dissertation:

- 1.) To what extent can a sufficient psychometric quality and a dimensional structure of the test instrument be determined?
- 2.) To what extent can (preservice) teachers' performance-oriented LRT-competence be determined?
- 3.) How are individual factors of the participants, such as academic background, LRT-related learning opportunities, beliefs concerning linguistic-cultural heterogeneity in school and teaching or personality traits related to LRT-competence?

The framework paper embeds the four articles in the overall theoretical context, discusses the results and finally derives recommendations for further research and for teacher training.

1. Einleitung

„Die prinzipielle Reformbereitschaft zur Sprachbildung als Aufgabe aller Fächer und Lernbereiche entspricht aktuellen gesellschafts- und bildungspolitischen Forderungen auf internationaler und Landesebene.“ (Thürmann et al., 2017, S. 6)

Die Idee von Sprachbildung als Aufgabe aller Fächer geht nicht nur auf migrationsbedingte Mehrsprachigkeit vieler Schüler*innen zurück. Vielmehr erwerben generell zu wenig Schüler*innen im deutschen Schulsystem ausreichend sprachliche Kompetenzen, obwohl diese grundlegend für eine berufliche und gesellschaftliche Teilhabe sind (Becker-Mrotzek & Roth, 2017a). Deshalb muss eine Professionalisierung der Lehrkräfte dahingehend erfolgen, dass sie ihren (Fach-) Unterricht sprachbildend gestalten können (Baumann, 2017).

Wie Lehrkräfte bestmöglich auf die Praxis vorbereitet werden können, bleibt jedoch weiterhin ungeklärt: Findet die Ausbildung hauptsächlich an der Hochschule statt oder erfolgt das ‚wahre‘ Erlernen des Berufs erst in der Praxis? (Röbe et al., 2019). Auf diese Frage gibt es aktuell weder theoretisch noch empirisch eindeutige Antworten und auch in dieser Arbeit wird diese Frage nicht beantwortet werden. Dennoch ist die Frage Grundlage für die theoretische Rahmung und empirische Gestaltung. In dieser Arbeit geht es um die Kompetenz der (angehenden) Lehrkräfte im Bereich Deutsch als Zweitsprache (DaZ). Diese soll performanznah erfasst werden, sodass die Komplexität von Unterricht bestmöglich berücksichtigt werden kann. Dabei wird auch untersucht, welche DaZ-bezogenen Lerngelegenheiten die Testteilnehmer*innen bereits hatten, um so ggf. Rückschlüsse auf besonders effektive DaZ-bezogene Lerngelegenheiten ziehen und die Lehrkräftebildung auf lange Sicht entsprechend verändern zu können. Dafür wird ein Kompetenzbegriff zugrunde gelegt, auf den dann Erkenntnisse aus dem Testentwicklungsprozess und den Datenerhebungen bezogen werden. So soll eine Annäherung an die Frage erarbeitet werden, wie kompetente (DaZ-)Lehrkräfte bestmöglich ausgebildet werden können.

Die vorliegende Dissertation ist im Rahmen des *DaZKom-Video*¹-Projekts entstanden. Ziel des Projekts war es, ein innovatives Testformat zu entwickeln, welches eine performanznahe Messung von DaZ-Kompetenz ermöglicht. Dazu werden Videovignetten und ein mündliches

¹ *DaZKom-Video*: „Performanznahe Messung von Deutsch-als-Zweitsprache-Kompetenz bei (angehenden) Lehrkräften“ ist BMBF-gefördert und Teil des Clusters KoKoHs II (Laufzeit: 2017-2020). Antragsteller*innen sind Barbara Koch-Priewe, Anne Köker, Udo Ohm und Timo Ehmke.

Antwortformat genutzt. Das Forschungsvorhaben bedient sich unterschiedlicher theoretischer Konzepte und Modelle, an denen die theoretische Ausarbeitung dieser Arbeit angelehnt ist. Sie werden als Ausgangspunkt genutzt und dann spezifisch weiter ausgearbeitet und aufeinander bezogen. Für drei der vier Artikel dieser Arbeit werden außerdem Daten genutzt, die im Rahmen des Projekts erhoben wurden. Artikel 1 präsentiert eine Vorstudie zu dem Projekt, welche sich dem Effekt von mündlichen versus schriftlichen Antwortformaten widmet. Artikel 2 und 3 präsentieren jeweils Ergebnisse der ersten und zweiten Pilotierungsstudien des Projekts. Artikel 4 stellt die Ergebnisse der Normierungsstudie vor und legt noch einmal ein besonderes Augenmerk auf das Antwortformat des Testinstruments. Zudem sind die Studien 2, 3 und 4 Teil des Validierungsprozesses des entwickelten Testverfahrens.

2. Theoretischer Rahmen und Verortung der Dissertation

Der folgende Abschnitt stellt der Dissertation zugrundeliegende Theorien, Konzepte und Modelle vor und ordnet die relevanten Begrifflichkeiten ein. So wird ein übergreifender, theoretischer Rahmen geschaffen, aus dem Forschungsdesiderate abgeleitet werden können, die die Artikel dieser kumulativen Dissertation verbinden. Zunächst wird auf die Bedeutung von Deutsch als Zweitsprache in der Lehrkräftebildung eingegangen und in dem Rahmen auch verwandte Konzepte und Begrifflichkeiten erläutert. Es folgt ein Abschnitt zum Kompetenzverständnis und im Zuge dessen auch die Klärung des Begriffs der Performanz sowie die theoretischen Annahmen für entsprechende Messverfahren.

2.1 Deutsch-als-Zweitsprache-Kompetenz in der Lehrkräfte(aus)bildung

„Die Lehrkräfte sind [...] aufgefordert, die sprachlichen Kompetenzen der Schülerinnen und Schüler auszubauen und die Vermittlung (bildungs-)sprachlicher Kompetenzen in jedem Fachunterricht systematisch zu berücksichtigen.“ (Paetsch et al., 2019, S. 52)

Diese Forderung ist mittlerweile nicht mehr neu. Sprachliche Heterogenität in der Schule ist die Normalität, und dem wird die einsprachige Ausrichtung des Bildungssystems, der *monolinguale Habitus*, nicht gerecht (Gogolin, 1994). Damit einher geht die Forderung nach einem Perspektivenwechsel hin zu einem Bildungswesen, das Mehrsprachigkeit als Ressource anerkennt (Geist & Krafft, 2017; Schroeder & Stölting, 2005) und nach einer forschungsgeleiteten Lehrkräfteaus- und fortbildung, die sich der Sprachkompetenz und der Kommunikationsfähigkeit annimmt und auch Reflexion mit einbezieht (Allgäuer-Hackle & Jessner, 2014). Sprachförderung sollte nicht nur als Ergänzung oder Vorbereitung auf

Regelunterricht, z.B. in Deutsch-als-Zweitsprache Kursen stattfinden (Köker, 2018). Vielmehr sind fachliche Inhalte und Sprache untrennbar miteinander verbunden, weshalb Sprachförderung Aufgabe jeder Fachlehrkraft sein sollte (Baumann, 2017; Köker, 2018). In dieser Arbeit steht also die fachunterrichtsrelevante DaZ-Kompetenz von (angehenden) Lehrkräften im Fokus (Ohm, 2018). Auch andere Konzepte fordern den Ausbau der sprachlichen Kompetenzen der Schüler*innen im Fachunterricht, zum Beispiel das der durchgängigen Sprachbildung (Nach FÖRMIG, (Gogolin & Lange, 2011). Demnach sollten die bildungssprachlichen Kompetenzen der Schüler*innen kumulativ aufgebaut werden, institutionen-, stufen- und fächerübergreifend. In den Referenzrahmen für Schulqualität der Bundesländer wird das unterschiedlich berücksichtigt: NRW hat dafür die Dimension „Bildungssprache und sprachsensibler (Fach-)Unterricht“ (Ministerium für Schule und Bildungs des Landes Nordrhein-Westfalen, 2020, S. 44) eingeführt, welche eben diese Aspekte aufgreift. Hier wird eindeutig und differenziert aufgefächert gefordert, dass alle Schüler*innen bildungssprachliche Kompetenzen erwerben und je nach Bedarf individuelle Unterstützung erhalten sollen (Ministerium für Schule und Bildungs des Landes Nordrhein-Westfalen, 2020). In Niedersachsen wiederum wird „Sprachbildung“ (Niedersächsisches Kultusministerium, 2014, S. 8) als ein Teilmerkmal von Kompetenzorientierung aufgegriffen: Schüler*innen sollen im bewussten Umgang mit Sprache und unterschiedlichen Registern gefördert werden. Diese Ausführungen werfen jedoch die Frage auf, wie Lehrkräfte auf sprachliche Förderung in ihrem Fachunterricht vorbereitet werden müssen (Köker, 2018), denn es zeigt sich, dass (angehende) Lehrkräfte sich unvorbereitet fühlen (Ricart Brede, 2019a). Auch Lehrkräfteausbildner*innen sind häufig nicht für die entsprechende Lehrkräftebildung ausgebildet und fühlen sich unvorbereitet (Benholz & Siems, 2016).

Döll und Hägi (2014) fassen Ergebnisse einer Diskussion zu Mehrsprachigkeit in pädagogischen Berufen zusammen, in der eine linguistische Ausbildung gefordert sowie die Notwendigkeit von Kenntnissen über Spracherwerb, Diagnostik aber auch institutioneller Diskriminierung und Überzeugungen befürwortet wird. Dazu betonen sie die Wichtigkeit der Reflexion des schulbezogenen Handelns der Lehrkräfte. Becker-Mrotzek und Roth (2017b) stellen heraus, dass Sprachkompetenz durch Entwicklungs- und Bildungsprozesse beeinflusst wird: „Dispositionell verfügbare Fähigkeiten werden entfaltet und über sprachliche Sozialisation wie über sprachpädagogische Einwirkungen befördert oder auch beeinträchtigt oder gar behindert“ (Becker-Mrotzek & Roth, 2017b, S. 20). Sie plädieren deshalb dafür, Sprachkompetenz möglichst umfassend auszubauen. Dazu gehört demnach ebenfalls

grundlegendes linguistisches Wissen, aber auch die Fähigkeit, angemessen sprachlich zu handeln.

Konkret wurde dies teilweise umgesetzt, indem bspw. in Nordrhein-Westfalen das DaZ-Modul in die Lehrkräftebildung integriert wurde. Es wird als "Deutsch für Schülerinnen und Schüler mit Zuwanderungsgeschichte (DSSZ-Modul)" (Döll et al., 2017, S. 203) bezeichnet und stellt das Konzept der durchgängigen Sprachbildung in den Mittelpunkt, integriert aber u.a. auch Inhalte wie etwa Linguistik und Migrationspädagogik (Döll et al., 2017). Die inhaltliche Umsetzung wiederum obliegt den Universitäten selbst (Gantefort & Michalak, 2017). Nordrhein-Westfalen integriert auch in die zweite Phase der Lehramtsausbildung Themen, die sich sprachbildendem Unterricht widmen (Benholz & Siems, 2016). Welche konkreten Inhalte und Fokussierungen jedoch notwendig sind, ist weiterhin unklar und damit ein wichtiges Forschungsdesiderat (Baumann, 2017). Welche Kompetenzen brauchen Lehrkräfte also, um die bildungssprachlichen Kompetenzen ihrer Schüler*innen auszubauen und wie können wirksame Lerngelegenheiten für Lehrkräfte gestaltet werden?

Mit diesen Fragen beschäftigen sich aktuell verschiedene Forschungsgruppen sowohl im Bereich DaZ, als auch in den Fachdidaktiken und der Bildungsforschung mit dem Ziel der Lehrkräfteprofessionalisierung (Becker-Mrotzek et al., 2017). Auch die *DaZKom*-Projekte widmen sich diesem Desiderat mit einem Strukturmodell für DaZ-Kompetenz (Köker et al., 2015) und mit nunmehr zwei Kompetenztests (*DaZKom*-Paper-Pencil-Test, Ehmke & Hammer, 2018; *DaZKom*-Video-Test, Lemmrich et al., 2019), mit denen die Kompetenz der Lehrkräfte in diesem Bereich erfasst werden kann. Das Projekt geht auch der Frage der DaZ-bezogenen Lerngelegenheiten nach und wie diese zur Professionalisierung der Lehrkräfte im Bereich DaZ beitragen können.

Deutsch als Zweitsprache

Deutsch als Zweitsprache (DaZ) bezeichnet Deutsch als eine weitere Sprache, die neben der Erstsprache erworben wird. Der Erwerbkontext ist, neben anderen Faktoren, entscheidend für die Abgrenzung von Deutsch als Fremdsprache, denn Deutsch wird als Zweitsprache in Deutschland erworben (Kniffka & Siebert-Ott, 2012). Erwerb einer Zweitsprache erfolgt also im „Land der Zielsprache“ (Ahrenholz, 2011, S. 9), während der Erwerb der Sprache in der „Ausgangskultur“ (Henrici & Vollmer, 2001, S. 8) als Erwerb einer Fremdsprache bezeichnet wird. Deutsch als Zweitsprache ist meist eine aktive, kommunikative Sprache im Leben der Schüler*innen, was für Deutsch als Fremdsprache eher nicht zutrifft (Kniffka & Siebert-Ott,

2012). Geist und Krafft (2017) führen die Form des Spracherwerbs als Faktor mit auf: Deutsch als Erst- und Zweitsprache wird demnach ungesteuert erworben, Deutsch als Fremdsprache gesteuert. Eine eindeutige Klassifizierung ist jedoch meist schwierig, denn die Übergänge können fließend sein (Ahrenholz, 2017; Benz, 2019; Kniffka & Siebert-Ott, 2012). So werden im Fremdsprachenlernen vermehrt alltägliche Kommunikationssituationen geschaffen, um den Lernprozess zu unterstützen (z.B. durch Austauschprogramme) (Kniffka & Siebert-Ott, 2012) und neu zugewanderte Schüler*innen erwerben Deutsch als Zweitsprache im Zielland und gleichzeitig gesteuert, z.B. in Vorbereitungsklassen (Benz, 2019).

Zentral für diese Arbeit ist, dass die Beherrschung der Unterrichtssprache Deutsch der Schlüssel zu Lerninhalten ist und Schüler*innen ohne ausreichende Sprachkompetenz der Zugang verwehrt wird (Schmölzer-Eibinger, 2013). Das liegt auch daran, dass Schüler*innen in der Schule besonderen sprachlichen Anforderungen begegnen, die sich von alltäglichen unterscheiden und dafür entsprechende weitere sprachliche Kompetenzen benötigen (Schmölzer-Eibinger, 2013). Gogolin (2006) prägte den Begriff der *Bildungssprache* und spricht von *bildungssprachlichen Kompetenzen* in Deutsch (Gogolin, 2006). Das betrifft spezifische Sprachhandlungen (Operatoren), komplexe Satzzusammenhänge und Fachbegriffe (Lange, 2020). Bildungssprache ist das Medium des Wissenstransfers (Lange, 2020) und unzureichende bildungssprachliche Kompetenzen können in der Folge nicht nur die Teilhabe am Unterricht gefährden, sondern auch die gesellschaftliche Teilhabe und das zukünftige Berufsleben negativ beeinflussen (Leisen, 2013). Bildungssprache ist ein sprachliches Register (Feilke, 2012). Als Register werden „Sprachgebrauchsformen in einem bestimmten sozial-funktionalen Kommunikationsfeld“ (Feilke, 2012, S. 6) bezeichnet (Feilke, 2012). Im Kommunikationsfeld Schule hat Sprache eine kommunikative/didaktische Funktion der Wissensvermittlung, aber auch eine sozial-identitätsstiftende und epistemische Funktion (Feilke, 2012; Heine et al., 2018; Morek & Heller, 2012). Neben dem bildungssprachlichen Register gehören auch die spezifischen Besonderheiten des Fachregisters (Schleppegrell, 2007) zu den sprachlichen Anforderungen in der Schule: jedes Fach hat spezifische Merkmale und Funktionen (Heine et al., 2018). Diese Anforderungen können vor allem für Schüler*innen mit Deutsch als Zweitsprache und Deutsch als Fremdsprache eine Schwierigkeit darstellen, aber durchaus auch für Schüler*innen mit Deutsch als Erstsprache (Leisen, 2013), zum Beispiel dann, wenn sie aus bildungsferneren Familien kommen (Meyer & Prediger, 2012). Im Fachunterricht sollten also die im Alltag erworbenen sprachlichen Kompetenzen der

Schüler*innen im Hinblick auf bildungssprachliche Kontexte funktional weiter ausdifferenziert werden, um die Schüler*innen dabei zu unterstützen, fachlich adäquat kommunizieren und handeln zu können und eine gewisse Registerflexibilität zu erreichen (De Oliveira & Schleppegrell, 2015; Marx, 2018; Ohm, 2018).

DaZKom-Modell

Mit dem Ziel, Lehrkräfte dahingehend auszubilden, Schüler*innen mit Deutsch als Zweitsprache im Fachunterricht systematisch fördern zu können, wurde bereits vor einigen Jahren das *DaZKom*²-Strukturmodell für Deutsch als Zweitsprache (DaZ)-Kompetenz (kurz: *DaZKom*-Modell, Köker et al., 2015, Abbildung) entwickelt. Die Dimensionen wurden aus Curricula akademischer Lerngelegenheiten im Bereich DaZ herausgearbeitet (Carlson & Präg, 2018).

Abbildung 1. Strukturmodell für DaZ-Kompetenz (Köker et al. 2015, S.184).

	Dimension	Subdimension	Facetten		
			Stufe I	Stufe II	Stufe III
DaZ-Kompetenz	Fachregister (Fokus auf Sprache)	Grammatische Strukturen und Wortschatz	<i>Morphologie</i>		
			<i>(Lexikalische) Semantik</i>		
			<i>Syntax</i>		
			<i>Textlinguistik</i>		
		Semiotische Systeme	<i>Darstellungsformen</i>		
			<i>Sprachliche Bezüge zwischen Darstellungsformen</i>		
	<i>Mündlichkeit vs. Schriftlichkeit</i>				
	Mehrsprachigkeit (Fokus auf Lernprozess)	Zweitspracherwerb	<i>Interlanguage-Hypothese</i>		
			<i>Meilensteine zweitsprachlicher Entwicklung</i>		
			<i>Gesteuerter vs. Ungesteuerter Zweitspracherwerb</i>		
			<i>Literacy/Bildungssprache</i>		
		Migration	<i>Sprachliche Vielfalt in der Schule</i>		
			<i>Umgang mit Heterogenität</i>		
	Didaktik (Fokus auf Lehrprozess)	Diagnose	<i>Mikro-Scaffolding</i>		
			<i>Makro-Scaffolding</i>		
<i>Umgang mit Fehlern</i>					
Förderung		<i>Mikro-Scaffolding</i>			
		<i>Makro-Scaffolding</i>			
		<i>Umgang mit Fehlern</i>			

² *DaZKom*-Projekt: „Professionelle Kompetenzen angehender Lehrkräfte (Sek I) im Bereich Deutsch als Zweitsprache“ ist BMBF-gefördert und Teil des Clusters KoKoHs I (Laufzeit: 2015-2017). Antragsteller*innen sind Barbara Koch-Priewe, Anne Köker, Udo Ohm und Timo Ehmke.

Im *DaZKom*-Modell wurden drei Hauptdimensionen der generischen DaZ-Kompetenz identifiziert: *Fachregister*, *Mehrsprachigkeit* und *Didaktik*. Diese Dimensionen fächern sich dann in Subdimensionen und einzelne Kompetenzfacetten auf (Köker et al., 2015). In den einzelnen Kompetenzfacetten finden sich die beschriebenen Faktoren wieder, wie z.B. gesteuerter vs. ungesteuerter Zweitspracherwerb oder aber die Bildungssprache, die sich von alltäglichem Sprachgebrauch unterscheidet. Die Dimension *Fachregister* bildet die Spezifität dieses Registers ab, welches auch Teil der Bildungssprache ist. Allgemein beschrieben, fokussiert die Dimension *Fachregister* die Sprache, die im Unterricht nicht nur Lerngegenstand, sondern auch Lernmedium ist. Die Dimension *Mehrsprachigkeit* bezieht sich auf den Lernprozess der mehrsprachigen Schüler*innen. Die dritte Dimension *Didaktik* fokussiert den Lehrprozess und ergänzt damit die beiden erstgenannten Dimensionen. Auf Basis von Diagnostik kann eine angemessene Förderung bereitgestellt werden, in lang- und kurzfristigen Unterrichtssituationen (Köker et al., 2015). Zur letzten Dimension haben sich in den letzten Jahren praktische Konzepte entwickelt, die sich der Frage widmen, wie sich Bildungssprache erfolgreich (im Fachunterricht) vermitteln lässt (Lange, 2020). Dazu zählen neben dem im Modell aufgeführten *Scaffolding* (Gibbons, 2002), ein international bekanntes Konzept für sprachliche Bildung und Zweitspracherwerbsdidaktik (Becker-Mrotzek & Woerfel, 2020, S. 99), *sprachsensibler Fachunterricht* und *durchgängige Sprachbildung*. Beides soll im Folgenden noch einmal erläutert werden.

Sprachsensibler Fachunterricht und durchgängige Sprachbildung

Sprachsensibler Fachunterricht (1) geht u.a. auf Leisen (2013) zurück. Er bezeichnet damit „den bewussten Umgang mit Sprache beim Lehren und Lernen im Fach“ (Leisen, 2013, S. 3). Es sollen die individuellen (sprachlichen) Gegebenheiten der Schüler*innen berücksichtigt und dann systematisch und inhaltsbezogen im Fachunterricht gefördert werden. Leisen unterscheidet dabei explizit zwischen Fachunterricht mit Sprachförderung und Sprachförderunterricht (Leisen, 2013). Sprachsensibler Unterricht wird aber auch als „Sammelbegriff für ein Gestaltungsprinzip des Unterrichts bei sprachlich heterogenen Lerngruppen“ (Becker-Mrotzek & Woerfel, 2020, S. 98) verwendet und bezeichnet dabei Konzepte ähnlich wie das von Leisen (2013) beschriebene. Zentral ist die in den Unterricht integrierte, durchgängige sprachliche Unterstützung (Becker-Mrotzek & Woerfel, 2020).

Durchgängige Sprachbildung ist ein Konzept von Gogolin und Lange (2011) und meint ebenfalls eine „planvolle“ (Gogolin & Lange, 2011, S. 118) Förderung der

bildungssprachlichen Fähigkeiten von Schüler*innen. Dabei legt die durchgängige Sprachbildung ein besonderes Augenmerk auf die gesamte Bildungsbiographie der Schüler*innen. Bestenfalls sollte stationsübergreifend (bspw. Kita-Grundschule) und fächerübergreifend ein Aufbau bildungssprachlicher Kompetenzen stattfinden, aber auch andere Lernbereiche der Schüler*innen sowie alle ihre Sprachen systematisch mit einbezogen werden (schulisch, schulbegleitend, außerschulisch) (Gogolin & Lange, 2011). „Als Ziel der durchgängigen Sprachbildung wird die Fähigkeit gesehen, adressaten-, intentions- und situationsangemessen sowie bildungssprachlich angemessen zu sprechen und zu schreiben“ (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2020, S. 44). Der Begriff *Sprachbildung* ist im Gegensatz zu *Sprachförderung* weniger defizitorientiert und bezieht ganz bewusst auch Schüler*innen mit Deutsch als Erstsprache mit ein (Becker-Mrotzek & Roth, 2017b; Rost-Roth, 2017). Beide Konzepte werden in dieser Arbeit dennoch synonym verwendet. Sie bezeichnen in dieser Arbeit beide jeweils den Prozess der systematischen Förderung der bildungssprachlichen Fähigkeiten aller Schüler*innen im Fachunterricht.

Da auch zwei englische Artikel Teil dieser Arbeit sind und hier sprachsensibler Fachunterricht mit *Linguistically Responsive Teaching* übersetzt wurde, soll dieser Begriff nicht unbeachtet bleiben. *Linguistically Responsive Teaching* steht in der englischsprachigen Forschung für einen konzeptuellen Rahmen, der eine Reihe von Orientierungen und Praktiken in sich vereint, die für das Unterrichten von *English Language Learners (ELL)* (Lucas & Villegas, 2013) fundamental sind. Ausgangspunkt für die Entwicklung dieses Rahmens ist ein erstes, früheres Konzept der Autorinnen, das des *Culturally Responsive Teaching* (Villegas & Lucas, 2002). Das Konzept *Linguistically Responsive Teaching* beinhaltet auch deshalb nicht nur pädagogisches und linguistisches Wissen, sondern greift die Idee einer grundsätzlich wertschätzenden Einstellung gegenüber Diversität und das Verständnis des Zusammenhangs von Sprache, Kultur und Identität mit auf (Lucas & Villegas, 2013). So ist in dem konzeptuellen Rahmen des *Linguistically Responsive Teaching* auch das *Culturally Responsive Teaching* weiterhin mitgedacht (Lucas & Villegas, 2011). Hier zeigt sich auch die Passung zur fachunterrichtsrelevanten DaZ-Kompetenz nach dem DaZKom-Modell: individuelle (migrationsbedingte) Mehrsprachigkeit wird als Teil der sprachlichen Vielfalt in der Schule mitgedacht und die Facette *Umgang mit Heterogenität* bewusst offen verankert (Falkenstern et al., 2021; Ohm, 2018). Insbesondere in Artikel 4 wird also DaZ-Kompetenz mit der Kompetenz mit *Linguistically Responsive Teaching (LRT competence)* übersetzt.

Mehrsprachigkeit und sprachliche Heterogenität

Im Rahmen dieser Arbeit wird, insbesondere zur Beschreibung der Ausgangslage, Bezug genommen auf *mehrsprachige* Schüler*innen (bzw. *multilingual learners*, Studien 3 und 4) und auf *sprachliche Heterogenität*, weshalb beides an dieser Stelle kurz eingeordnet werden soll. Mehrsprachige Schüler*innen sind in dieser Arbeit jene, die zwei oder mehr Sprachen in ihrem Alltag nutzen (Grosjean, 2020). Diese Definition schließt an die der *lebensweltlichen* Mehrsprachigkeit an, die die Sprachen bezeichnet, die in der alltäglichen Sprachpraxis, ob schulisch oder alltäglich, eine zentrale Rolle spielen (Petersen, 2014). Dies bezieht sich sowohl auf Dialekte (*innere Mehrsprachigkeit* nach Wandruszka, 1979) (Busch, 2017) als auch auf mehrere Sprachen im Repertoire der Sprecher*innen (Maak, 2019), deren Kenntnis jedoch stark variieren kann. So wird eine Sprache bspw. nur mündlich genutzt, schriftliche Kompetenzen sind in einer anderen Sprache stärker ausgeprägt (Grosjean, 2020). Für den Kontext dieser Arbeit bedeutet dies, dass das bildungssprachliche Register nicht unbedingt in allen Sprachen, die Schüler*innen nutzen, zur Verfügung steht. Sprachliche Heterogenität beschreibt genau dies: Schüler*innen einer Lerngruppe unterscheiden sich ihrem Sprachrepertoire an sich, in der Kompetenz der Sprachen ihrer Repertoires, in der Sozialisation in diesen Sprachen und in der alltäglichen bzw. bildungssprachlichen Nutzung dieser Sprachen (Lengyel, 2016). Im Gegensatz dazu stehen Lehrkräfte, die (bildungs-)sprachliche Kompetenzen in Deutsch meist (unbewusst) voraussetzen (Putjata & Danilovich, 2019), weshalb die Integration des Aufbaus von DaZ-Kompetenz in die Lehrkräftebildung so zentral ist.

Zusammenfassung

In Anlehnung an das *DaZKom*-Modell umfasst DaZ-Kompetenz also unter anderem Facetten des Zweitspracherwerbs, linguistische Aspekte und schließt mit der Dimension *Didaktik* auch Konzepte für die systematische Sprachförderung im Fachunterricht mit ein. DaZ-Kompetenz bezieht sich in dieser Arbeit somit nicht auf die Kompetenz, additive Sprachförderangebote zu gestalten, wie z.B. Vorbereitungsklassen oder DaZ-Kurse (Becker-Mrotzek & Woerfel, 2020; Rösch, 2011), sondern auf die Kompetenz der Lehrkräfte, integrierte Sprachförderung (Becker-Mrotzek & Woerfel, 2020) in ihrem Fachunterricht zu betreiben, mit dem Ziel einer durchgängigen und systematischen Sprachbildung, unabhängig davon, ob Deutsch für die Schüler*innen die Erstsprache, Zweitsprache oder eine Fremdsprache ist.

2.2 Kompetenz, Performanz und performanznahes Messen

Der Fokus dieser Arbeit liegt auf der Professionalisierung von Lehrkräften im Bereich Deutsch als Zweitsprache. Damit einher geht die Kompetenzmodellierung und -messung. Der Einstieg in dieses Kapitel erfolgt mit einem Abschnitt zur Ausgangslage des *DaZKom-Video*-Projekts, bevor dann das Kompetenzverständnis vorgestellt wird, welches dieser Arbeit zugrunde liegt. In Zuge dessen wird auch auf die Rollen des Professionswissens und des professionellen Handelns eingegangen. Es folgen dann Ausführungen zu Performanz und zu performanznahen Messverfahren.

DaZKom-Video

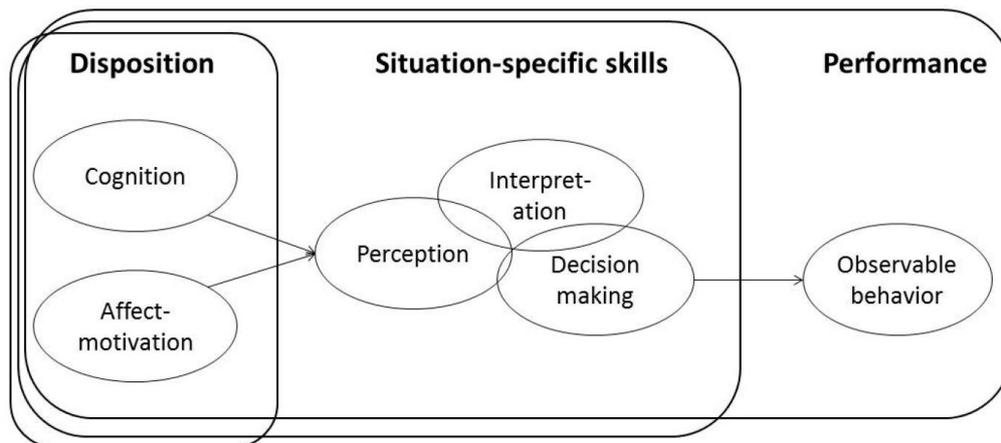
Nachdem im vorherigen Kapitel das DaZ-Modell als Strukturmodell vorgestellt wurde, soll an dieser Stelle die Entwicklungsperspektive mit aufgenommen werden. Es wird davon ausgegangen, dass DaZ-Kompetenz in unterschiedlichen Graden vorliegen kann und diese auch messbar sind (Ohm, 2018). Um die Professionalisierung von Lehrkräften abbilden zu können, wurde auf das fünfstufige Modell des Fähigkeitserwerbs von Erwachsenen (*Five-Stage Model of Adult Skill Acquisition*) Bezug genommen (Dreyfus & Dreyfus, 1986; Ohm, 2018) wonach sich die Kompetenzstufen *Novizen*, *Fortgeschrittene Anfänger*, *Kompetenz*, *Gewandtheit* und schlussendlich in *Expertise* unterscheiden (Dreyfus & Dreyfus, 1986). Die ersten drei Stufen sind kognitiver Natur und können mit dem *DaZKom*-Paper-Pencil-Test valide erfasst werden. Stufe vier und fünf enthalten Handlungsfacetten und müssen daher über spontane Reaktionen operationalisiert werden (Ohm, 2018). Hier setzt das *DaZKom-Video*-Projekt an, welches mithilfe von Videovignetten und mündlichen Antworten der befragten (angehenden) Lehrkräfte eine Erfassung der Handlungsfacetten ermöglichen soll.

Im Vorgängerprojekt *DaZKom* wurden in einem Standardsetting-Verfahren auf Basis der erhobenen Normierungsdaten bereits Kompetenzstufen festgelegt: *unter Mindeststandard*, *Mindeststandard* und *Regelstandard* (Gültekin-Karakoç et al., 2016, S. 139). Im Rahmen des *DaZKom-Video*-Projekts soll mit einem performanznahen Testinstrument (*DaZKom-Video-Test*) DaZ-Kompetenz gemessen werden und im Zuge dessen auch qualitative Stufen von performanznaher DaZ-Kompetenz festgelegt werden.

(Lehrkräfte-) Kompetenz und Performanz

(Professionelle) *Kompetenz* beschreibt in der Forschung zu Lehrkräftebildung eine konkrete, professionsspezifische Fähigkeit (Harms & Riese, 2018), die sich entwickeln und verändern kann (Klieme & Hartig, 2007). Das grenzt sie von persönlichen Merkmalen, wie zum Beispiel Intelligenz, ab (Harms & Riese, 2018). Diese Arbeit (ebenso wie das *DaZKom-Video-Projekt*) basiert auf der Idee, Kompetenz als Kontinuum zu begreifen (Blömeke, Gustafsson & Shavelson, 2015, Abbildung 2). Die Autor*innen des Modells gehen von vorhandenen Dispositionen (kognitiven, affekt-motivationalen) einer Person aus (wie z.B. Wissen und Überzeugungen). Diese Einteilung haben auch Baumert und Kunter (2006) vorgenommen: Sie greifen Shulmans Taxonomie (fachwissenschaftliches, fachdidaktisches, und methodisches Wissen, Shulman, 1986) unter professionellem Wissen auf und nehmen zusätzlich motivationale Orientierungen, selbstregulative Fähigkeiten und Überzeugungen mit auf (Baumert & Kunter, 2006). Blömeke, Gustafsson und Shavelson (2015) beziehen außerdem noch situationsspezifische Fähigkeiten (z.B. professionelle Wahrnehmung) in ihren Kompetenzbegriff mit ein, welche notwendig sind, um Performanz zeigen zu können (Blömeke, Gustafsson & Shavelson, 2015). Das Modell nimmt sich der Konkurrenz zweier Vorstellungen von Kompetenz an: Kompetenz kann einerseits nicht kontextlos und situationsunabhängig nur auf vorhandene Dispositionen zurückgeführt werden. Andererseits kann auch nicht davon ausgegangen werden, dass für kompetentes Verhalten in einer komplexen Situation die Dispositionen einer Person irrelevant sind (König, 2020). Das Modell nimmt also an, dass situationsspezifische Fähigkeiten als Vermittler zwischen Dispositionen und Performanz fungieren (Blömeke, Gustafsson & Shavelson, 2015). Lindmeier (2013) schreibt professioneller Wahrnehmung in ihren Arbeiten eine „Gelenkfunktion“ (Lindmeier, 2013, S. 52) zwischen Wissen und professionellem Handeln zu.

Abbildung 2. Kompetenz-als-Kontinuum-Modell (Blömeke, Gustafsson & Shavelson, 2015, S.9).



Mit Kompetenzentwicklung und Kompetenzmessung geht auch die Frage einher, wann Personen sehr kompetent sind bzw. als Expert*innen bezeichnet werden können. Dafür wurden bereits Kompetenz- bzw. Expertisestufen modelliert, die auch in der Lehrkräfteprofessionsforschung Verwendung finden. Miller (1990) entwickelte Performanztests für (angehende) Mediziner*innen und gliederte die zu erreichende Kompetenz in vier Stufen in eine Pyramiden-Darstellung: *knows (knowledge)*, *knows how (competence)*, *shows how (performance)*, *does (action)* (Miller, 1990, S. 63). Diese Darstellung ähnelt der Idee von Dreyfus und Dreyfus (1986), die ihre Kompetenzstufen von Noviz*innen, über kompetent Handelnde bis hin zu Expert*innen auch in einer Pyramide gliedern (Dreyfus & Dreyfus, 1986). Beide Modelle basieren auf der gleichen Idee: Wissen ist nicht gleichzusetzen mit Können. Jedoch bauen die Stufen aufeinander auf, wobei Wissen die Ausgangsstufe darstellt.

Die Schwierigkeit der Lehrkräfteprofession im Vergleich zu anderen Domänen ist die Vielschichtigkeit und Komplexität, so dass eine eindeutige Definition der Lehrkräfteexpertise kaum möglich ist (Krauss, 2020). In dem Expertise-Paradigma für Lehrkräfte ist keineswegs der Leistungsgedanke zentral, sondern passend zu der eben beschriebenen Komplexität von Unterricht, dass Lehrkräfte über „eine sehr breite und gut vernetzte Wissensbasis in ihrer Expertisedomäne verfügen“ (Krauss, 2020, S. 160) und vor allem, dass sie „dieses Wissen schnell und flexibel einsetzen können“ (Krauss, 2020, S. 160).

Diese Ausführungen zu Lehrkräfteexpertise sind sowohl im Rahmen des Kompetenzverständnisses nach Blömeke, Gustafsson und Shavelson (2015) relevant, als auch in Bezug auf die Kompetenzmessung. In der Konzeption der Kompetenz als Kontinuum sind Dispositionen wie Wissen der Ausgangspunkt, und Performanz wiederum ist nur mit

entsprechenden situationsspezifischen Fähigkeiten möglich. Die *situationsspezifischen Fähigkeiten* (Blömeke, Gustafsson & Shavelson, 2015) umfassen unter anderem *Wahrnehmung (perception)* und *Handlungsentscheidung treffen (decision making)*. Wahrnehmung wird in der (Lehrkräfte-)Professionsforschung als *professionelle Wahrnehmung (professional vision)*, Goodwin, 1994) aufgegriffen. Präzise entscheidende Unterrichtssituationen wahrzunehmen und zu interpretieren ist eine Fähigkeit, die sich zunächst entwickeln muss (van Es & Sherin, 2002). Bezogen auf Blömeke, Gustafsson und Shavelson (2015) bedeutet dies, dass wahrgenommene Situationen nur dann interpretiert werden können, wenn entsprechendes Wissen vorhanden ist. Beides ist aber entscheidend für die Performanz: Die professionelle, wissensbasierte Wahrnehmung ist grundlegend für kompetentes Handeln (Performanz) (Hoth, 2016). Bezogen auf Erkenntnisse aus der Expertiseforschung heißt das, dass es einen entscheidenden Unterschied gibt in der Wahrnehmung zwischen Noviz*innen und erfahrenen Lehrkräften: Erfahrene Lehrkräfte nehmen Unterrichtssituationen genauer und holistischer wahr als Noviz*innen (König & Lebens, 2012). Noviz*innen beschreiben die Situation lediglich, während erfahrene Lehrkräfte Wahrgenommenes mit ihnen bekannten Theorien verknüpfen und so auch komplexe Zusammenhänge identifizieren können (Berliner, 2001). Annahmen aus der kognitionspsychologischen Forschung unterstützen die Überlegungen zu situationsspezifischen Fähigkeiten: Wahrnehmung ist nicht rein reizgesteuert, vielmehr lenkt Wissen die Wahrnehmung (Heins & Zabka, 2019). Dafür ist jedoch auch Erfahrungswissen bzw. Handlungswissen relevant (Heins & Zabka, 2019), welches beim Handeln erworben wird (Grießhaber, 2017). Grießhaber (2017) betont, dass für professionelles Lehrkräftehandeln (Performanz) das Verhältnis beider Wissensarten zentral ist (Grießhaber, 2017, S.93). In der Kompetenzdefinition nach Blömeke, Gustafsson und Shavelson (2015) wird dieses Verhältnis über die situationsspezifischen Fähigkeiten definiert. Diesem Modell entsprechend, folgt nach der professionellen Wahrnehmung einer Situation die Performanz. Unmittelbar und spontan können aber nur erfahrene Lehrkräfte Performanz zeigen, denn nur sie müssen nicht darüber nachdenken, sondern reagieren intuitiv (Gawronski & Creighton, 2013). Dies liegt darin begründet, dass sie auf Erfahrungswerte zurückgreifen können und deshalb nicht abwägen müssen, welche Entscheidung sie treffen und wie diese sich auswirkt (Kahneman, 2013).

Neben DaZ-bezogenem Wissen sind ebenso Überzeugungen zu sprachlich-kultureller Heterogenität in Schule und Unterricht (Fischer & Ehmke, 2020) Teil der Dispositionen. In Kombination mit den situationsspezifischen Fähigkeiten, DaZ-relevante Unterrichtssituationen wahrzunehmen und entsprechen Handlungsentscheidungen treffen zu können, zeigt sich in der

Performanz die Kompetenz. Das Konzept der Kompetenz als Kontinuum zwischen Dispositionen, über situationsspezifische Fähigkeiten hin zur Performanz kann um eine Zwei-Wege-Modellierung erweitert werden: So erwerben Lehrkräfte im Handeln Erfahrungswissen, welches dann bewusst reflektiert wird und so die Dispositionen (Professionelles Wissen) erweitern bzw. verändern kann (Blömeke & Kaiser, 2017; Santagata & Yeh, 2016). Da auch professionsbezogene Überzeugungen zu den Dispositionen zählen (Baumert & Kunter, 2006; Blömeke, Gustafsson & Shavelson, 2015; Weinert, 2002), kann also auch davon ausgegangen werden, dass Überzeugungen nicht nur das Handeln der Lehrkräfte beeinflussen, sondern sich Überzeugungen ebenfalls verändern können, wenn Unterrichtserfahrungen gesammelt und kritisch reflektiert werden (Ricart Brede, 2019b; Wischmeier, 2012).

Bei der empirischen Untersuchung von professionellen Kompetenzen werden schon lange Verfahren genutzt, die die kognitiven Facetten der Kompetenz (z.B. professionelles Wissen) erfassen. Das sind zumeist Papier-und-Bleistift-Tests (wie z.B. auch der *DaZKom*-Paper-Pencil-Test, Ehmke & Hammer, 2018) (König & Lebens, 2012). Nun werden immer häufiger die situationsspezifischen Fähigkeiten zur Operationalisierung performanzorientierter Verfahren genutzt (König, 2020).

Performanznahes Messen von DaZ-Kompetenz

Das Erfassen von Kompetenz bezog sich lange Zeit vorrangig auf die Ebene der Schüler*innen und wurde mit dem Fokus auf Wissensfacetten durchgeführt. Erst in den letzten Jahren ist die Kompetenzforschung im Bereich der Lehrkräftebildung vorangeschritten (König, 2020). Im Zuge einer vom Bundesministerium für Bildung und Forschung finanzierten Förderlinie haben sich einige Projekte intensiv mit der Struktur und der Messung von Lehrkräftekompetenz beschäftigt und dabei auch die Messung von kognitiven Facetten und von Handlungsfacetten der Kompetenz unterschieden (König, 2020; Zlatkin-Troitschanskaia et al., 2020). In dieser Förderlinie sind sowohl das *DaZKom* als auch das *DaZKom-Video*-Projekt verortet.

Während Wissenstests bei der Untersuchung von Kompetenz meist in Paper-Pencil-Formaten erfolgen (König & Lebens, 2012), wird die Frage nach der Handlungskompetenz und nach dem Verhältnis von Wissen und Können immer zentraler, auch im Hinblick auf die Kompetenzentwicklung und verschiedener Kompetenz- bzw. Expertisestufen. Demnach ist eine Person erst dann kompetent oder gar ein*e Expert*in, wenn sie in kontextspezifischen Situationen professionell handeln kann (Blömeke, Gustafsson & Shavelson, 2015; Dreyfus &

Dreyfus, 1986). Um dies messbar zu machen, benötigt es performanznahe Verfahren (König, 2015). Die Verwendung von Performanztests ist in der Lehrkräfteprofessionsforschung noch relativ jung. In der medizinischen Ausbildung ist sie bereits ein traditionelles Verfahren. Es werden reale Handlungssituationen als Stimuli verwendet, die unter kontrollierten Testbedingungen eine Handlung von Proband*innen verlangen, welche dann wiederum unter ebenfalls kontrollierten Bedingungen bewertet wird (Miller, 1990). Ziel von Miller (1990) war es, nicht nur zu testen, was die (angehenden) Mediziner wissen, sondern auch, ob sie dies dann in ihren Handlungen zeigen können (Miller, 1990).

In einigen Domänen der Lehrkräfteprofessionsforschung sind in den vergangenen Jahren performanznahe Verfahren entwickelt worden, z.B. in der Physik (Schröder et al., 2020), der Biologie (Kramer et al., 2020) oder der Mathematik (Jeschke et al., 2020). Für die (generische) DaZ-Kompetenz, die in allen Fächern relevant ist (Köker et al., 2015), existiert ein solches Verfahren noch nicht. Alle genannten Forschungsgruppen arbeiten mit Videovignetten, um eine holistische Abbildung der komplexen Unterrichtsinteraktionen zu ermöglichen (Casale et al., 2016; Kersting et al., 2012; König & Lebens, 2012) Videovignetten haben den Vorteil, dass sie authentisches Lehrkräftehandeln und sowohl verbale, als auch nonverbale Kommunikation abbilden können (Gröschel, 2019). Sie können außerdem durch die Fokussierung oder die Fragestellungen dem betreffenden Kontext angepasst werden (Helmke, 2009; Reusser, 2005). Deshalb wird davon ausgegangen, dass Videovignetten sich dazu eignen, Kompetenz performanznah zu erfassen (Kersting, 2008; König, 2015; Lindmeier, 2011). Um eine möglichst handlungsnah Testsituation zu schaffen, wird im *DaZKom-Video*-Projekt ein noch nicht so verbreitetes, mündliches Antwortformat genutzt. So soll sichergestellt werden, dass Proband*innen spontan reagieren, ohne lange über ihre Antwort nachzudenken und die Situation zu analysieren und eine möglichst authentische, intuitive Antwort geben. Das Team um Lindmeier (2011) nutzt in seiner Studie ebenfalls Videovignetten und ein mündliches Antwortformat, um handlungsnah Kompetenz messen zu können. Sie nutzen in ihren Arbeiten den Begriff „action-related competence (AC)“ (Lindmeier et al., 2016, S. 221) und argumentieren, dass „die Reaktion der Lehrperson in natürlicher Sprache“ (Lindmeier, 2013, S. 55) für „Unmittelbarkeit“ (Lindmeier, 2013, S. 55) sorgt.

3. Forschungsdesiderate und Forschungsfragen

Mit dem Ziel, eine performanznahe Messung von DaZ-Kompetenz zu realisieren, lassen sich aus den theoretischen Überlegungen drei Forschungsdesiderate mit entsprechenden Forschungsfragen ableiten:

(1) Psychometrische Qualität und dimensionale Struktur des performanznahen DaZ-Kompetenztests

In der theoretischen Rahmung dieser Arbeit wurde herausgearbeitet, dass Wissenstests nur einen Teil von Lehrkräfte-Kompetenz abdecken: Das Anwenden dieses Wissens in konkreten Unterrichtssituationen, die Performanz, wird so nicht erfasst (König, 2015). Um sich diesem Ziel anzunehmen bedarf es performanznaher Messverfahren, die bereits in anderen Domänen der Lehrkräfteprofessionsforschung genutzt werden (u.a. Schröder et al., 2020). Für die generische Kompetenz *DaZ* jedoch liegen derartige Verfahren noch nicht vor. Die Umsetzung dieses performanznahen Ansatzes erfolgt, wie in anderen performanznahen Messverfahren auch, mit Videovignetten (Casale et al., 2016; Kersting et al., 2012; König, 2015; Lindmeier, 2013) und mündlichen Antworten (Knievel et al., 2015). Im Hinblick auf das Ziel, DaZ-Kompetenz zu messen, werden Videovignetten verwendet, die eine DaZ-relevante Situation zeigen. Die Items zu den Vignetten basieren auf den situationspezifischen Fähigkeiten (*Wahrnehmung* und *Handlungsentscheidung treffen*). Die mündlichen Antworten zu den Items werden aufgezeichnet und mithilfe einer erstellten Kodieranweisung kodiert. Im Hinblick auf den Testentwicklungsprozess ist folgende erste Forschungsfrage zentral:

Inwieweit können eine ausreichende psychometrische Qualität und eine dimensionale Struktur des Testinstruments festgestellt werden?

(2) Performanznahe (DaZ-) Kompetenz von angehenden Lehrkräften

Die Überlegungen zu Kompetenz und Performanz legen nahe, dass kompetente Lehrkräfte über Dispositionen wie professionelles (DaZ)-Wissen verfügen. Dem Kompetenz-als-Kontinuum-Modell von Blömeke, Gustafsson und Shavelson (2015) folgend, bedarf es jedoch auch an situationspezifischen Fähigkeiten, um kompetent handeln zu können, also Performanz zu zeigen. Zwar sind Wissensfacetten von DaZ-Kompetenz bereits im *DaZKom*-Modell (Köker et al., 2015) dargestellt und Erfahrungslernen bzw. Handlungsorientierung sind in Anlehnung an das Modell von Dreyfus und Dreyfus (1986) theoretisch mitgedacht (Ohm, 2018), jedoch bleibt

unklar, wie sich performanznahe DaZ-Kompetenz konkret darstellt. Im Vorgänger-Projekt *DaZKom* wurden bereits die ersten drei Kompetenzstufen der kognitiven Facetten von DaZ-Kompetenz festgelegt: *unter Mindeststandard*, *Mindeststandard* und *Regelstandard* (Gültekin-Karakoç et al., 2016). Über die Hälfte der Studierenden erreichte den Mindeststandard und waren zumindest DaZ-sensibilisiert (Gültekin-Karakoç et al., 2016). Neben der performanznahen Messung von DaZ-Kompetenz ist also das Ziel, weitere qualitative Kompetenzstufen von performanznaher DaZ-Kompetenz festzulegen. Ausgangslage dieser Entwicklungsperspektive für DaZ-Kompetenz ist das fünfstufige Modell des Fähigkeitserwerbs von Erwachsenen (*Novizen, fortgeschrittener Anfänger, Kompetenz, Gewandtheit, Expertise*; Dreyfus & Dreyfus, 1986), nach welchem sich Expert*innen durch flüssiges Handeln auszeichnen. Daraus ergibt sich folgende zweite Forschungsfrage, die im Rahmen dieser Dissertation bearbeitet wird:

Inwieweit kann performanznahe DaZ-Kompetenz bei (angehenden) Lehrkräften festgestellt werden?

(3) Bedingungsfaktoren und Prädiktoren von performanznaher DaZ-Kompetenz

Sowohl DaZ-bezogenes Wissen als auch Überzeugen zu sprachlich-kultureller Heterogenität in Schule und Unterricht (Fischer & Ehmke, 2020) sind Teil der Dispositionen (Blömeke, Gustafsson & Shavelson, 2015) und sind gemeinsam mit situationsspezifischen Fähigkeiten Bedingungsfaktoren für performanznahe DaZ-Kompetenz. Diese Bedingungsfaktoren wiederum werden ggf. durch DaZ-bezogene Lerngelegenheiten sowie durch individuelle und akademische Hintergrunddaten (Angaben der Proband*innen zu ihren soziodemographischen Daten und zu ihrer Ausbildung) der Proband*innen positiv beeinflusst.

Obwohl es bereits eine Reihe allgemeiner Forderungen und Empfehlungen gibt, wie die DaZ-bezogene Aus- und Weiterbildung gestaltet werden sollte (Döll & Hägi, 2014; Gogolin & Lange, 2011; Leisen, 2013), fehlt es an Hinweisen zu konkreten, wirksamen DaZ-bezogenen Lerngelegenheiten und darauf, wie sich diese auf die Kompetenz der angehenden Lehrkräfte auswirken. Im Vorgängerprojekt *DaZKom* wurden bereits Zusammenhänge von individuellen und akademischen Hintergrunddaten sowie DaZ-bezogenen Lerngelegenheiten mit den kognitiven Facetten der DaZ-Kompetenz von Lehramtsstudierenden untersucht (Ehmke et al., 2018). Dabei hat sich unter anderem gezeigt, dass Studierende, die selbst Deutsch als Zweitsprache haben, entgegen den Erwartungen keinen Vorteil im Erwerb von DaZ-Kompetenz haben (Ehmke & Lemmrich, 2018). DaZ-bezogene Lerngelegenheiten (eine Skala

von DaZ-bezogenen Themen, die Studierende im Rahmen ihrer Ausbildung ggf. bereits behandelt haben) weisen erwartungsgemäß einen Vorhersagebeitrag zur DaZ-Kompetenz auf (Ehmke & Lemmrich, 2018). Davon ausgegangen, dass Überzeugungen den Wissenserwerb beeinflussen und umgekehrt (Fischer et al., 2018; Wischmeier, 2012), wurden im *DaZKom*-Projekt auch *Überzeugungen zu Sprache im Fachunterricht* mit erhoben (Fischer et al., 2018): Studierende mit höherer DaZ-Kompetenz weisen positivere Überzeugungen auf (Hammer et al., 2016).

Die Evaluation dieser Faktoren dient auch der Überprüfung der Validität des Testverfahrens. Zur Validierung wurden im Hinblick auf das mündliche Antwortformat zusätzlich Persönlichkeitsmerkmale der Proband*innen (Schupp & Gerlitz, 2014) miterhoben, da es offeneren, extrovertierten Personen möglicherweise leichter fällt, spontan mündlich zu antworten und dadurch die Messung der Kompetenz verzerrt wird. Aus diesem Forschungsdesiderat wird folgende dritte Forschungsfrage für die vorliegende Dissertation abgeleitet:

*Wie hängen individuelle und akademische Hintergrunddaten der Proband*innen, DaZ-bezogene Lerngelegenheiten, Überzeugungen zu sprachlich-kultureller Heterogenität in Schule und Unterricht und Persönlichkeitsmerkmale mit performanznaher DaZ-Kompetenz zusammen?*

4. Beiträge der Dissertation

Ausgehend von der theoretischen Rahmung und den entsprechenden Forschungsdesideraten sind vier Studien (bzw. Artikel) entstanden, die sich den Forschungsfragen dieser Dissertation annehmen. Tabelle 1 zeigt eine Übersicht über die Artikel und deren Verortung im Dissertationsprojekt.

Tabelle 1. Übersicht der Studien (Artikel).

	Studie 1	Studie 2	Studie 3	Studie 4
Verortung in der Dissertation	Vorstudie, Artikel 1	Prä-Pilotierung, Artikel 2	Pilotierung, Artikel 3	Normierung, Artikel 4
N	90	137	176	295
Personen	Studierende	Studierende/ Lehrkräfte	Studierende/ Lehrkräfte	Studierende/Lehrkräfte
Testformat	3 Vignetten, variiertes Antwortformat (mündlich/schriftlich)	20 Vignetten, mündliches Antwortformat	16 Vignetten, mündliches Antwortformat	12 Vignetten, mündliches Antwortformat
Fokus	Effekte eines mündlichen vs. schriftlichen Antwortformats	Dimensionalität, psychometrische Qualität	Dimensionalität, psychometrische Qualität	Validität
Berücksichtigte Kovariaten	DaZ-bezogene Lerngelegenheiten, individuelle und akademische Hintergrunddaten	DaZ-bezogene Lerngelegenheiten, individuelle und akademische Hintergrunddaten	DaZ-bezogene Lerngelegenheiten, individuelle und akademische Hintergrunddaten	Überarbeitete DaZ-bezogene Lerngelegenheiten, individuelle und akademische Hintergrunddaten, Überzeugungen und Persönlichkeitsmerkmale

Der erste Artikel (Lemmrich, Bahls & Ehmke, 2020) ist eine experimentelle Grundlagenforschung mit dem Ziel zu klären, ob ein mündliches Antwortformat Effekte auf die gegebenen Antworten hat und so möglicherweise zu einer Verzerrung der Kompetenzmessung führt (Studie 1, Tabelle 1).

Im *DaZKom-Video*-Projekt wurden in der ersten Phase zunächst Aufgaben (Items) für das Testinstrument entwickelt. Dafür wurden Videos an Schulen gedreht und diese dann zu DaZ-relevanten Vignetten geschnitten. Vignetten werden als *DaZ-relevant* definiert, wenn der Fokus des Lehrkräftehandelns in der entsprechenden Vignette sich einer Facette der Dimensionen des

DaZKom-Modells (*Fachregister, Mehrsprachigkeit, Didaktik*, Köker et al., 2015) zuordnen lässt. Die Videos wurden mit Kontextinformationen und Untertiteln versehen, um möglichst situationsspezifische Stimuli präsentieren zu können. Das entwickelte Testinstrument wurde dann mit einer Stichprobe (N=137, vgl. Tabelle 1) von Lehrkräften und Studierenden erprobt. Der zweite Artikel widmet sich der Skalierung und der dimensional Struktur des Testinstruments nach der ersten Pilotierungsphase (Lemmrich et al., 2019). Die Analyse des Testinstruments ist ein zentraler Schritt in der Testentwicklung, da auf deren Basis Entscheidungen in Bezug auf die Testitems und das Testvorgehen getroffen werden können. Es wurden bspw. Kontextinformationen der Vignetten geschärft oder Vignetten gänzlich aus dem Testinstrument ausgeschlossen, abhängig von der ermittelten psychometrischen Qualität der dazugehörigen Items. Zusätzlich wurde die DaZ-Relevanz der Unterrichtssituationen in den Vignetten mit Expert*innen-Ratings überprüft (Hecker et al., 2020). Der Weiterentwicklung des Testinstruments folgt eine nächste Pilotierungsphase, in der erneut Daten von Studierenden und Lehrkräften (N = 176, vgl. Tabelle 1) erhoben und anhand dieser die Teststruktur und die dimensionale Struktur untersucht wurden. Die Ergebnisse werden in Artikel 3 (Lemmrich, Hecker et al., 2020) festgehalten. Es folgt eine weitere Entwicklungsphase, in der auf Basis der ermittelten psychometrischen Qualität der Items und der festgestellten dimensional Struktur 12 finale Vignetten ausgewählt wurden. In Artikel 4 (Lemmrich & Ehmke, eingereicht) sind die Ergebnisse der finalen Erprobungsstudie dargestellt. Das Ziel ist die Validierung des Testinstruments, insbesondere des Antwortformates. Bereits Artikel 1 legt ein besonderes Augenmerk auf die Testfairness eines mündlichen Antwortformates. Die Erkenntnisse sollen in Studie 4 ergänzt werden: Effekte von Persönlichkeitsmerkmalen der Proband*innen werden geprüft, da offene, extrovertierte Proband*innen eventuell leichter mündlich antworten können als schüchterne und dies ggf. zu einer Verzerrung der Kompetenzmessung führt. Die folgenden Unterkapitel fassen die jeweiligen Artikel knapp zusammen. Eingeordnet und diskutiert werden die Ergebnisse dann in Kapitel 5.

4.1 Artikel 1: Effekte von mündlichen versus schriftlichen Antwortformaten bei der performanznahen Messung von DaZ-Kompetenz (Lemmrich, Bahls & Ehmke, 2020)

Performanznahe Testverfahren kommen in der aktuellen Lehrkräfteprofessionsforschung vermehrt zum Einsatz. Dafür haben sich Videovignetten bewährt (u.a. König, 2015). Als Antwortformate werden häufig schriftliche oder Multiple-Choice-Formate genutzt (u.a. Kersting et al., 2012). Um authentische, spontane Antworten von Proband*innen zu generieren

und somit möglichst performanznah zu operieren, werden bislang selten mündliche Antwortformate verwendet. Inwiefern ein mündliches Antwortformat einen Effekt auf die Korrektheit und Ausführlichkeit der Antworten hat, ist noch unklar. In dieser Studie werden in einem experimentellen Design mit randomisierter Stichprobenzuweisung zwei Antwortformate (schriftlich versus mündlich) in einer computerbasierten Testumgebung gegeneinander getestet. Beide Varianten sollen DaZ-Kompetenz bei angehenden Lehrkräften performanznah erfassen. Dafür werden Videovignetten genutzt. Ziel ist zu untersuchen, inwieweit das Antwortformat Einfluss auf die Korrektheit und die Ausführlichkeit der Antworten hat. Daraus ergeben sich folgende Forschungsfragen:

1. Welche Effekte hat das Antwortformat, schriftlich versus mündlich, bei der Messung von DaZ-Kompetenz bei angehenden Lehrkräften in Bezug auf die Korrektheit und die Ausführlichkeit (gemessen an der Wörteranzahl) der Antworten?
2. Welche Effekte hat das Antwortformat, schriftlich versus mündlich, in Bezug auf die Korrektheit und die Ausführlichkeit der Antworten, wenn das Ausmaß an DaZ-Lerngelegenheiten der Probandinnen und Probanden und deren Interaktion mit dem Antwortformat kontrolliert wird?

Untersucht wurde die Stichprobe (N = 90) von Lehramtsstudierenden der Leuphana Universität Lüneburg (B. A. / M. Ed., alle Semester). Neben drei Videovignetten mit jeweils zwei auf den situationsspezifischen Fähigkeiten basierenden Items (1. Was nehmen Sie wahr? 2. Wie würden Sie handeln?), wurde ein Fragebogen zu individuellen und akademischen Hintergrunddaten und zu DaZ-bezogenen Lerngelegenheiten eingesetzt. Die Ergebnisse zeigen, dass der Antwortmodus keinen statistisch signifikanten Einfluss auf die Richtigkeit, jedoch auf die Ausführlichkeit der Antworten hat. Im mündlichen Antwortformat wurden statistisch bedeutsam ausführlichere Antworten gegeben. Zudem zeigt sich, dass Testpersonen, die mehr DaZ-bezogene Lerngelegenheiten hatten, die Aufgaben eher richtig lösen, unabhängig vom Antwortmodus. Interaktionseffekte zeigten sich nicht. Für die Testentwicklung kann damit die Schlussfolgerung gezogen werden, dass ein mündliches Antwortformat zumindest bei Studierenden nicht zu unfairen Testergebnissen führt, verglichen mit einem traditionellen schriftlichen Antwortformat.

4.2 Artikel 2: Performanznahe und videobasierte Messung von DaZ-Kompetenz bei Lehrkräften - Skalierung und dimensionale Struktur (Lemmrich et al., 2019)

Diese zweite Studie des vorliegenden Dissertationsprojekts widmet sich der ersten Erprobung des performanznahen Testinstruments. Der Beitrag prüft im Rahmen des Testentwicklungsprozesses die psychometrische Qualität des Testinstruments und untersucht die Dimensionalität des Konstruktes *DaZ-Kompetenz* sowie Zusammenhänge zwischen der DaZ-Kompetenz mit individuellen und akademischen Hintergrunddaten sowie DaZ-bezogenen Lerngelegenheiten. Getestet wurden Lehramtsstudierende sowie Lehrkräfte unterschiedlicher Klassenstufen (N = 137). Folgende Fragestellungen werden untersucht:

1. Inwieweit kann eine ausreichende psychometrische Qualität der Items des *DaZKom-Video-Testinstruments* festgestellt werden?
2. Welche dimensionale Struktur weist das Testinstrument auf?
3. Welche Zusammenhänge bestehen zwischen Personenmerkmalen, der individuellen DaZ-Erfahrung, der Nutzung von DaZ-bezogenen Lerngelegenheiten und der DaZ-Kompetenz von Testpersonen?

Die Ergebnisse der Item- und Skalenanalyse zeigen eine befriedigende bis sehr gute psychometrische Qualität. Die Items haben eine sehr gute Passung zum Rasch-Modell (Rost, 2004) und erreichen eine zufriedenstellende Trennschärfe, jedoch ist die Itemschwierigkeit zu hoch. Das bedeutet, dass der Test zu schwer bzw. die DaZ-Kompetenz der Proband*innen zu gering ist. Bei der Untersuchung der dimensional Struktur des Konstrukts wurden drei Modelle gegeneinander getestet: das eindimensionale, das zweidimensionale (Unterscheidung nach den situationsspezifischen Fähigkeiten *Wahrnehmung (perception)* und *Handlung (decision making)*) und das dreidimensionale Modell (Unterscheidung nach den Dimensionen des *DaZKom-Modells Fachregister, Mehrsprachigkeit* und *Didaktik*). Die Ergebnisse geben Hinweise darauf, dass es sich um das eindimensionale oder das dreidimensionale Konstrukt handelt. Die zweidimensionale Struktur ist auf Basis dieser Analyse nicht wahrscheinlich, zumal diese zwei Dimensionen (*Wahrnehmung* und *Handlung*) sehr hoch ($r=0.97$) korrelieren. *Wahrnehmung* der Situation und *Handlungsentscheidung treffen* hängen also eng zusammen. Die Berechnungen wurden auf Basis der eindimensionalen Struktur durchgeführt. In Bezug auf individuellen und akademischen Angaben sowie DaZ-bezogener Lerngelegenheiten zeigt sich, dass Personen mit der Erstsprache Deutsch, Zusatzqualifikationen wie das DaZ-Zertifikat oder Forschungserfahrung im Bereich DaZ, statistisch signifikant höhere Testwerte erzielen. Der

Umfang der DaZ-bezogenen Lerngelegenheiten korreliert niedrig mit dem Testergebnis, sie sind in dieser Studie ein schwacher Prädiktor für vorhandene DaZ-Kompetenz. Die Befunde dieser Erprobungsstudie wurden genutzt, um das Testinstrument weiterzuentwickeln.

4.3 Artikel 3: Linguistically Responsive Teaching in Multilingual Classrooms - Development of a Performance-Oriented Test to Assess Teachers' Competence (Lemmrich et al., 2020)

Die dritte Studie dieser Dissertation präsentiert Ergebnisse der zweiten Pilotierungsstudie des *DaZKom-Video*-Projekts. Nach der ersten Pilotierungsstudie wurden auf Grundlage der Ergebnisse zu psychometrischer Qualität und Dimensionalität Vignetten aus dem Testinstrument entfernt oder die Kontextinformationen geschärft. Außerdem wurden Expert*innen hinzugezogen, um die DaZ-Relevanz und Authentizität der Vignetten zu überprüfen. Die Ergebnisse dieses Expert*innen-Ratings wurden auch im Überarbeitungsprozess berücksichtigt. Die zweite Pilotierungsstudie wurde also mit einem überarbeiteten Testinstrument durchgeführt. Es werden für diese Studie erneut die psychometrische Qualität des *DaZKom-Video*-Tests und die Dimensionalität des Konstrukts der DaZ-Kompetenz evaluiert, um die Ergebnisse für die Finalisierung des Testinstruments zu nutzen. Diese Studie ist in englischer Sprache erschienen. DaZ-Kompetenz wird begrifflich als *German-as-a-Second-Language (GSL)*-Competence übersetzt. Neben der psychometrischen Qualität werden auch Zusammenhänge zwischen DaZ-Kompetenz und individuellen und akademischen Hintergrunddaten sowie DaZ-bezogenen Lerngelegenheiten untersucht. Die Studie untersucht folgende Forschungsfragen:

1. Inwieweit kann eine ausreichende psychometrische Qualität der Items des *DaZKom-Video*-Testinstruments festgestellt werden? (How satisfactory is the *DaZKom-Video* test instrument's psychometrical quality?)
2. Welche dimensionale Struktur weist das Testinstrument auf? (Which dimensional structure can be determined for the test instrument?)
3. Welche Zusammenhänge bestehen zwischen Personenmerkmalen, der individuellen DaZ-Erfahrung, der Nutzung von DaZ-bezogenen Lerngelegenheiten und der DaZ-Kompetenz von Proband*innen? (Which correlations can be identified between the GSL competence of the test persons, the sociodemographic characteristics and learning opportunities?)

Die Stichprobe (N = 184) umfasst Lehramtsstudierende unterschiedlicher Universitäten und Lehrkräfte verschiedener Schulformen und wurde an mehreren Standorten in ganz Deutschland erhoben. Die Items und Skalenanalysen ergeben eine gute psychometrische Qualität: Sowohl der Item-Fit zum Rasch-Modell (Rost, 2004) als auch die Trennschärfe sind zufriedenstellend. Auch in dieser Studie zeigt sich eine sehr hohe Itemschwierigkeit. Wie schon in der ersten Pilotierungsstudie, wird in dieser Studie ein Modellvergleich vorgenommen, um die dimensionale Struktur zu überprüfen: Es wird ein ein-, zwei- und ein dreidimensionales Modell geprüft. Es zeigt sich, dass das eindimensionale und das dreidimensionale Modell die beste Passung aufweisen. Die beiden Dimensionen des zweidimensionalen Modells, Wahrnehmung und Handlung, zeigen, wie auch in der ersten Pilotierungsstudie, eine hohe Korrelation ($r=0.97$). Proband*innen mit DaZ-bezogener Lehrerfahrung oder DaZ-bezogener Zusatzqualifikation (wie zum Beispiel das DaZ-Zertifikat), sowie Proband*innen mit vielen DaZ-bezogenen Lerngelegenheiten (Skala DaZ-bezogener Themen, die im Rahmen der Lehramtsaus- und Weiterbildung der Proband*innen bereits thematisiert wurden) erreichen statistisch signifikant höhere Testergebnisse im *DaZKom-Video*-Test. Die Ergebnisse weisen auf eine zuverlässige und valide performanznahe Messung von DaZ-Kompetenz hin.

4.4 Artikel 4: Performance-Oriented Measurement of Teachers' Competency in Linguistically Responsive Teaching: Connection of LRT-Competence with Relevant Learning Opportunities, Personality Factors and Beliefs (Lemmrich & Ehmke, eingereicht)

Diese letzte Studie der Dissertation präsentiert die Ergebnisse der Normierungsstudie, die mit dem finalen und erprobten *DaZKom-Video*-Testinstrument durchgeführt wurde. Für die Finalisierung des Testinstruments wurden auf Basis der Analysen zur Pilotierungsstudie erneut Vignetten aus dem Test ausgeschlossen. So bleiben nunmehr 12 finale Vignetten im Testinstrument. Neben der performanznahen DaZ-Kompetenz werden Zusammenhänge mit erhobenen Kovariaten erfasst und so eine Konstruktvalidierung vorgenommen. In dem Zuge wird insbesondere auch die Überprüfung der Fairness des mündlichen Antwortformats fokussiert. Damit schließt diese Studie nicht nur an Artikel 2 und 3 an, sondern auch an die Untersuchung aus Artikel 1. Als Kovariaten wurden individuelle und akademische Hintergrunddaten erfasst, wie zum Beispiel Studien-/Unterrichtsfächer, DaZ-bezogene Lehr-/Forschungserfahrung und DaZ-bezogene Lerngelegenheiten. Außerdem werden Überzeugungen zu sprachlich-kultureller Heterogenität in Schule und Unterricht (Fischer &

Ehmke, 2020) sowie Persönlichkeitsmerkmale (Schupp & Gerlitz, 2014) erfasst, um zu überprüfen, inwiefern bestimmte Persönlichkeitsmerkmale mit spontanen, mündlichen Antworten zusammenhängen. Wenn interne Proband*innen bspw. durch das mündliche Antwortformat gehemmt wären, könnte dies zur Verzerrung der Kompetenzmessung führen. Extrovertierte Proband*innen hätten wiederum ggf. einen Vorteil und die Testfairness wäre beeinträchtigt. Auch dieser Artikel wurde in englischer Sprache verfasst. Im Gegensatz zu Studie 3 wurde DaZ-Kompetenz nicht direkt übersetzt, sondern auf das englischsprachige Konstrukt des *Linguistically Responsive Teaching* (LRT, Lucas & Villegas, 2013) übertragen. Aufgrund der großen Überlappung der Konstrukte (DaZ-Kompetenz (Köker et al., 2015) und Linguistically Responsive Teaching (Lucas & Villegas, 2013)) wurde DaZ-Kompetenz hier mit *LRT competence* übersetzt. In englischsprachigen Forschungsarbeiten ist dies ein verbreiteter Begriff und macht diese Studie anschlussfähig an den internationalen Diskurs. Folgende Forschungsfragen wurden in der letzten Studie bearbeitet:

1. Inwiefern hängen die Ergebnisse des *DaZKom-Video-Tests* mit individuellen und akademischen Merkmalen der Proband*innen sowie der Nutzung von DaZ-bezogenen Lerngelegenheiten zusammen? (Is there a correlation between the test results obtained and individual and academic characteristics and the use of LRT-relevant learning opportunities?)
2. Welche Zusammenhänge lassen sich zwischen Persönlichkeitsfaktoren der Teilnehmer*innen und den Ergebnissen des *DaZKom-Video-Testinstruments* feststellen? (Which connections can be identified between personality factors of participants and LRT-test results?)
3. Welcher Zusammenhang lässt sich zwischen den Überzeugungen der Teilnehmer*innen bezüglich sprachlich-kultureller Heterogenität in Schule und Unterricht und den Testergebnissen feststellen? (Which relationship can be identified between beliefs of participants concerning multilingualism in school and teaching and the LRT-test results?)

Die Stichprobe (N = 295) besteht aus Lehramtsstudierenden und Lehrkräften aller Schulformen, erhoben an Universitäten und Schulen in ganz Deutschland. Die Ergebnisse bestätigen einige unserer Hypothesen, wie z.B. DaZ-spezifische Lehrerfahrung als Prädiktor für performanznahe DaZ-Kompetenz. Überraschend war, dass Proband*innen mit DaZ-bezogener Forschungserfahrung höhere Testergebnisse erzielen sowie auch diejenigen, die Englisch

studieren bzw. unterrichten. Unerwartet war auch, dass DaZ-bezogene Lerngelegenheiten keinen signifikanten Zusammenhang mit der DaZ-Kompetenz zeigen. Dies ist ggf. dadurch erklärbar, dass diese Skala der DaZ-bezogenen Lerngelegenheiten Themen abfragt, die im Rahmen von Aus- und Weiterbildungsangeboten bereits behandelt wurden. Damit wären sie Teil des professionellen Wissens und Teil der Dispositionen (Blömeke, Gustafsson & Shavelson, 2015). Um in performanznahen Umgebungen Kompetenz zeigen zu können, sind jedoch auch situationsspezifische Fähigkeiten relevant, die ggf. noch nicht geschult sind. Einen statistisch besonders signifikanten Zusammenhang mit der DaZ-Kompetenz zeigen positive Überzeugungen zu sprachlich-kultureller Heterogenität in Schule und Unterricht. Persönlichkeitsfaktoren dagegen zeigen keinen Zusammenhang mit den Testergebnissen. Proband*innen mit bestimmten Persönlichkeitsmerkmalen werden durch ein mündliches Antwortformat in diesem Testkontext also weder bevorteilt noch benachteiligt.

5. Zusammenfassung und Diskussion

Das vorliegende Rahmenpapier wurde mit dem Ziel verfasst, die vier Studien unter den drei übergreifenden Forschungsfragen zusammenzuführen:

- 1.) Inwieweit können eine ausreichende psychometrische Qualität und eine dimensionale Struktur des Testinstruments festgestellt werden?
- 2.) Inwieweit kann performanznahe DaZ-Kompetenz bei (angehenden) Lehrkräften festgestellt werden?
- 3.) Wie hängen individuelle und akademische Hintergrunddaten der Proband*innen, DaZ-bezogene Lerngelegenheiten, Überzeugungen zu sprachlich-kultureller Heterogenität in Schule und Unterricht oder Persönlichkeitsmerkmale mit performanznaher DaZ-Kompetenz zusammen?

Dafür wurde zunächst die theoretische Rahmung dargelegt: Für die Herleitung des Konzepts *DaZ-Kompetenz* wurde die Definition von Deutsch als Zweitsprache herangezogen und die Abgrenzung zu Deutsch als Fremdsprache verdeutlicht. In einem nächsten Schritt wurde DaZ-Kompetenz anhand des *DaZKom*-Modells (Köker et al., 2015) erläutert, welches sich aus den drei Dimensionen *Fachregister*, *Mehrsprachigkeit* und *Didaktik* in Subdimensionen und Kompetenzfacetten auffächert. Zweitspracherwerb ist eine der Subdimensionen, der wiederum die Art des *Spracherwerbs* (gesteuert versus ungesteuert) und *Bildungssprache* als Facetten zugeordnet werden. *Scaffolding* wird als Kompetenzfacette der Dimension *Didaktik* begriffen.

Die für den Kontext dieser Arbeit zentralen didaktischen Konzepte des *sprachsensiblen Fachunterrichts* und der *durchgängige Sprachbildung* werden als Teil der DaZ-Kompetenz begriffen und ebenfalls der Dimension *Didaktik* zugeordnet. Fachunterrichtsrelevante DaZ-Kompetenz bezeichnet also die Kompetenz von Lehrkräften, sprachbildenden Fachunterricht zu gestalten, der die Schüler*innen angemessen und systematisch bei der Entwicklung bzw. Ausdifferenzierung ihrer sprachlichen Register im Hinblick auf bildungssprachliche bzw. fachsprachliche Kontexte unterstützt, unabhängig davon, ob es sich um Schüler*innen mit Deutsch als Zweit-, Fremd- oder Erstsprache handelt.

Der zweite Teil der theoretischen Rahmung widmet sich der Verortung des zugrundeliegenden Kompetenzbegriffs: Kompetenz wird als Kontinuum zwischen Dispositionen wie Wissen und Überzeugen, situationsspezifischen Fähigkeiten und Performanz begriffen (Blömeke, Gustafsson & Shavelson, 2015). In Bezug auf DaZ-Kompetenz lässt sich diese Konzeption folgendermaßen übertragen: Die im *DaZKom*-Modell aufgefächerten Dimensionen und Facetten bilden die Wissensfacetten ab. Gemeinsam mit Überzeugungen zu sprachlich-kultureller Heterogenität (Fischer & Ehmke, 2020) bilden sie die dispositionale Ausgangslage. Mit situationsspezifischen Fähigkeiten, wie der professionellen Wahrnehmung und der Fähigkeit, kontextabhängig professionelle Handlungsentscheidungen zu treffen, können (angehende) Lehrkräfte dann ggf. Performanz zeigen.

Eine zentrale Frage in der Forschung zur Professionalisierung von Lehrkräften ist die nach der professionellen Kompetenz, die mit professionellem Wissen nicht vollständig erfüllt ist (König, 2020). Es besteht weitestgehend Einigkeit darüber, dass Wissen und Können nicht gleichzusetzen sind (Baumert & Kunter, 2006) und dass lediglich diejenigen Lehrkräfte als sehr kompetent eingestuft werden können, die auch in einer komplexen Unterrichtssituation ihr Wissen spontan anwenden können (Blömeke, Gustafsson & Shavelson, 2015; Dreyfus & Dreyfus, 1986; Ohm, 2018; Weinert, 2002). Deshalb wird vermehrt der Frage nach der Handlungskompetenz bzw. der Performanz nachgegangen und wie diese bestmöglich erfasst wird. Dafür muss eine Testsituation geschaffen werden, die zum einen der Komplexität von Unterricht gerecht wird (König, 2015) und zum anderen spontane Handlungen von Lehrkräften abfragt (Lindmeier, 2013).

Die vier Studien dieser Dissertation untersuchen die psychometrische Qualität, die dimensionale Struktur, performanznahe DaZ-Kompetenz der (angehenden) Lehrkräfte und Zusammenhänge mit erhobenen Kovariaten: individuelle und akademische Merkmale der

Proband*innen (Studien 1-4), DaZ-bezogene Lerngelegenheiten (Studien 1-4), Überzeugungen (Studie 4) und Persönlichkeitsmerkmale (Studie 4). Es folgt nun eine Gesamtdiskussion der vier Beiträge anhand der übergeordneten Forschungsfragen.

5.1 Diskussion der übergeordneten Forschungsfragen des Rahmenpapiers

Der folgende Abschnitt diskutiert die Ergebnisse der vier Studien jeweils in Abschnitten zu den drei Fragestellungen.

Inwieweit können eine ausreichende psychometrische Qualität und eine dimensionale Struktur des Testinstruments festgestellt werden?

In der Annahme, dass Kompetenz ein Kontinuum zwischen Dispositionen und Performanz ist (Blömeke, Gustafsson & Shavelson, 2015), werden DaZ-relevante Videovignetten und mündliche Antworten genutzt, um eine performanznahe Testumgebung zu realisieren. Die inhaltliche Domäne DaZ wird über das *DaZKom*-Modell definiert, welches drei Dimensionen identifiziert: *Fachregister*, *Mehrsprachigkeit* und *Didaktik* (Köker et al., 2015). Zur Operationalisierung werden die situationsspezifischen Fähigkeiten *Wahrnehmung* und *Handlungsentscheidung* den Items zugrunde gelegt. Das Ergebnis ist ein reliables und valides Testinstrument für die performanznahe Messung von DaZ-Kompetenz. Jedoch ist ein zentrales Ergebnis aller Studien, dass die Itemschwierigkeit und der Personenfähigkeitsschätzer (Weighted Likelihood Estimates, WLEs; Wilson, 2004) wenig überlappen. Das bedeutet, dass der performanznahe Test für DaZ-Kompetenz für alle Proband*innengruppen (Lehramtsstudierende und Lehrkräfte) sehr schwer bzw. die DaZ-Kompetenz der Proband*innen gering ist. In Bezug auf die fehlenden Ausbildungs- und Performanzstandards im Bereich DaZ, der deutschlandweit unterschiedlichen Ausgestaltung der Lehrkräftebildung im Bereich DaZ und damit einhergehend geringen vorhandenen DaZ-bezogenen Lerngelegenheiten in Studium und Weiterbildung (vgl. 2.1), ist dieses Ergebnis nicht überraschend. Angehenden Lehrkräften, die im Augenblick vom großflächigen Ausbau universitärer, DaZ-bezogener Lerngelegenheiten profitieren können, fehlt es noch an Handlungserfahrung. Und Lehrkräfte, die schon eine Weile im Beruf sind, hatten meist wenig universitäre, DaZ-bezogene Ausbildungsanteile, weshalb sie in dieser Domäne kein Wissen erwerben konnten. Diese Perspektive mag auch die Erklärung dafür sein, dass unterrichtserfahrene Lehrkräfte nicht unbedingt höhere Testergebnisse erzielen als Lehramtsstudierende. Der Theorie zur Entwicklung von Expertise folgend, müssten DaZ-

erfahrene Lehrkräfte die DaZ-bezogenen Unterrichtssituationen präzise wahrnehmen und entsprechend konkrete Handlungsvorschläge machen können (Berliner, 2001; König & Lebens, 2012). Dafür sprechen die Ergebnisse der Studien jedoch nicht. Ein weiterer Grund dafür könnte sein, dass sich Kompetenz nicht in allen Dimensionen gleichzeitig entwickelt (Blömeke & Kaiser, 2017), also z.B. Wissensfacetten und Handlungsfacetten. Zentral in dem Modell sind auch die situationsspezifischen Fähigkeiten: Wenn zwar ausgeprägtes DaZ-bezogenes Wissen vorhanden ist, jedoch die professionelle Wahrnehmung (noch) nicht geschult ist, können keine Bezüge zu vorhandenen Theorien und Konzepten hergestellt werden, um professionell Handeln zu können. Hinweise dazu finden sich auch in der Theorie zu *Noticing* und *Knowledge-based Reasoning* als zwei Teilkomponenten professioneller Wahrnehmung (Seidel & Stürmer, 2014). Der konkrete Zusammenhang zwischen den situationsspezifischen Fähigkeiten und Performanz ist jedoch immer noch ungeklärt (Weber et al., 2020).

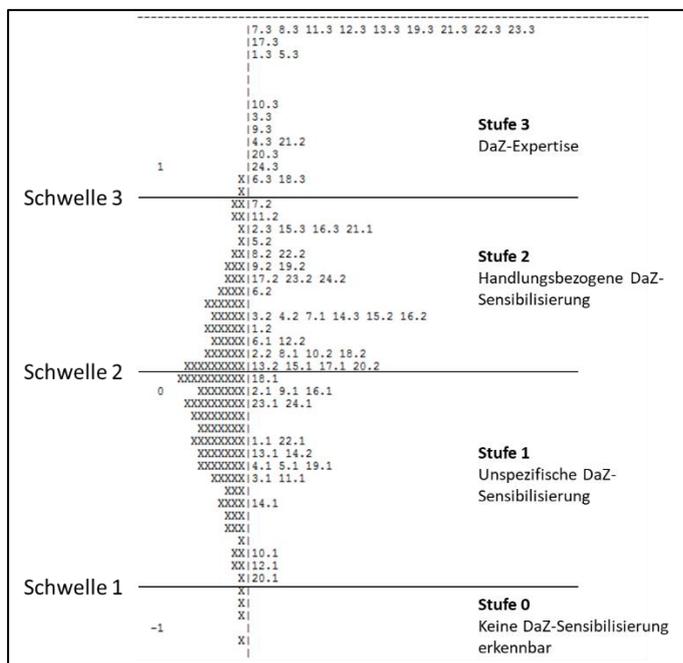
Die Ergebnisse bestätigen, was bereits in den theoretischen Annahmen deutlich wurde: Eine Lehrkraft handelt stets in hochkomplexen und vielschichtigen Situationen, weshalb eine eindeutige Definition ihrer Expertise nur schwer möglich ist (Krauss, 2020). Berliner (2004) betont, dass es Zeit (und entsprechend Erfahrung) braucht, bis sich Expertise entwickelt. Allerdings gibt es auch Hinweise darauf, dass Lehrkräfte nicht automatisch mit einer höheren Anzahl von Berufsjahren eine höhere Expertise entwickeln (Krauss & Bruckmaier, 2014). Neben möglichen, unterschiedlichen Entwicklungsstadien der Dimensionen des Kompetenz-als-Kontinuum-Modells, können ggf. auch die Dimensionen bzw. Facetten des DaZ-Kom-Modells (Köker et al., 2015) unterschiedlich entwickelt sein. Zwar werden die Dimensionen analytisch getrennt, jedoch treten sie in konkreten Unterrichtssituationen meist gleichzeitig auf. Insbesondere die Dimension *Didaktik* wird als eine grundlegende betrachtet und durchgehend mitgedacht (Ohm, 2018). Die Untersuchung der Dimensionalität in Studie 2 und 3 liefert ebenfalls Hinweise darauf. So zeigen alle Dimensionen zumindest eine mittlere Korrelation, wobei die Dimension *Didaktik* mit den beiden anderen Dimensionen *Fachregister* und *Mehrsprachigkeit* jeweils eine etwas höhere Korrelation zeigt. Sollten also (angehende) Lehrkräfte bspw. bereits Wissen im Bereich der Dimension *Fachregister* haben, aber kein Wissen aus der Dimension *Didaktik*, führt dies möglicherweise ebenfalls zu defizitärer Performanz.

Inwieweit kann performanznahe DaZ-Kompetenz bei (angehenden) Lehrkräften festgestellt werden?

Aus den theoretischen Überlegungen geht hervor, dass Wissensfacetten von Kompetenz mit Paper-Pencil-Formaten bereits valide erfasst werden können. Es geht daraus auch hervor, dass dies nicht zwingend etwas über die (Handlungs-)Kompetenz von angehenden Lehrkräften aussagt. Deshalb wurde die Idee des performanznahen Messens prominent, welches bereits für unterschiedliche Domänen umgesetzt wurde, jedoch nicht für die Domäne DaZ. Diese birgt die Besonderheit, dass sie eine fächerübergreifende Kompetenz ist, die in jedem Fachunterricht relevant ist. Ziel der Dissertation war es zu prüfen, inwieweit ein performanznahes Testinstrument die Kompetenz (angehender) Lehrkräfte erfasst, konkrete, DaZ-relevante Situationen wahrzunehmen und professionell Handeln zu können. Obwohl viele Studien in diesem Feld ähnliche theoretische Rahmungen zugrunde legen, ist nicht klar, inwiefern die zu messende Performanz mit tatsächlicher Unterrichtsperformanz gleichgesetzt werden kann. Während einige Studien *Sprechen über Unterricht* und *Unterricht selbst* unter *Performanz* zusammenfassen (Heins & Zabka, 2019), begreifen andere das Erfassen handlungs**naher** Fähigkeiten als Erfassen von Performanz (Performanztest) (Schröder et al., 2020). Die Begründung liegt nahe: Videovignetten zeigen authentische Unterrichtsinteraktionen. Jedoch handelt es sich jeweils nur um kurze Sequenzen. Lehrkräfte sind nicht wirklich in der Situation, sind nicht in komplexen Interaktionen eingebunden und kennen die Hintergründe der jeweiligen Schüler*innen nicht (Kersting et al., 2012). Zudem werden sie in einer Testsituation nicht in Interaktionen mit anderen Beteiligten involviert, wie es im Unterricht Realität wäre und sind deshalb ggf. anders (bzw. weniger) belastet (Sieland & Jordaan, 2019). Es ist deshalb durchaus davon auszugehen, dass *Unterrichten* und *über Unterricht sprechen* nicht gleichzusetzen sind. Einen Hinweis auf diese Unterscheidung liefert Miller (1990) in seiner Forschung mit (angehenden) Mediziner*innen. Er unterscheidet ganz konkret die Kompetenzstufen „shows how (performance)“ und „does (action)“ (Miller, 1990, S. 63). Die Bezeichnung **performanznahe** Messung scheint, auch unter Beachtung der Ergebnisse zur Validierung (siehe Diskussion zu Forschungsfrage 3), in dieser Dissertation weiterhin angemessen, denn obwohl eine größtmögliche Annäherung an tatsächliche Handlungskompetenz versucht wird, bleibt das Verhältnis von Testperformanz und Unterrichtsperformanz unklar. Hinweise dazu liefern auch König und Lebens (2012) oder Weber et al. (2020). Die drei Kompetenzstufen der kognitiven Facetten von DaZ-Kompetenz (*unter Mindeststandard*, *Mindeststandard* und *Regelstandard*) (Gültekin-Karakoç et al., 2016) sollten im *DaZKom-Video*-Projekt um weitere

qualitative Kompetenzstufen von performanznaher DaZ-Kompetenz ergänzt werden. Dafür wurden die Daten der Normierungsstudie des *DaZKom-Video*-Projekts genutzt, um anhand dessen in einem Standardsetting-Verfahren diese Kompetenzstufen festlegen zu können. Mithilfe der Expert*innen konnten vier Kompetenzstufen identifiziert werden: *Keine DaZ-Sensibilisierung erkennbar*, *Unspezifische DaZ-Sensibilisierung*, *Handlungsbezogene DaZ-Sensibilisierung* und *DaZ-Expertise* (Abbildung 3). Etwa die Hälfte der getesteten Lehramtsstudierenden und Lehrkräften weisen eine unspezifische DaZ-Sensibilisierung auf.³

Abbildung 3. Wright Map mit Cut Scores, Normierungsstichprobe.



Diese Ergebnisse der Normierungsstudie zeigen: Nur wenige (angehende) Lehrkräfte erzielen so hohe Testwerte, dass ihnen eine handlungsbezogene DaZ-Sensibilisierung oder gar DaZ-Expertise zugesprochen werden kann. Dies bestätigt Ergebnisse, die zur Forschungsfrage 1 diskutiert wurden: die Itemschwierigkeit ist sehr hoch bzw. die Personenfähigkeit eher gering. Dafür spricht auch der geringe Anteil der Lehrkräfte mit DaZ-Expertise (ca. 4 %). Deskriptive Ergebnisse am Beispiel der Normierungsdaten zeigen, dass Proband*innen unabhängig davon, ob sie Lehramtsstudierende oder Lehrkräfte sind, im Durchschnitt nicht ganz die Hälfte des höchstmöglichen Testwertes erreichen ($Min = 1$, $Max = 54$, $MW_{gesamt} = 21,43$, $SD_{gesamt} =$

³ Eine Publikation zum Standardsetting-Verfahren im *DaZKom-Video*-Projekt ist in Vorbereitung: Hecker et al., Bestimmung von Standards und Stufen handlungsbezogener Kompetenz bei Lehrkräften aller Fächer im Bereich Deutsch als Zweitsprache (DaZ) [Arbeitstitel].

10,42). Diese Ergebnisse könnten auch Hinweise sein, die die Annahmen zur Performanznähe unterstützen: Da keine echte Unterrichtshandlung erfasst werden kann, sondern nur eine Annäherung an diese erreicht wird, werden neben handlungsnaher Kompetenz auch Elemente der Wissensfacetten miterfasst.

Auffällig ist, dass Referendar*innen (MW = 24,72, SD = 9,82) überdurchschnittliche Testwerte und Wissenschaftler*innen signifikant höhere Testwerte erzielen (MW = 39,71, SD = 7,10). In Anlehnung an das Modell von Dreyfus und Dreyfus (1986) und an das von Blömeke, Gustafsson und Shavelson (2015) wurde davon ausgegangen, dass Lehrkräfte mit zunehmender Lehrerfahrung eher DaZ-bezogene Performanz zeigen können und ihnen so eine höhere Kompetenz- bzw. Expertisestufe zugesprochen werden kann. Die Ergebnisse der Untersuchungen im Rahmen dieser Dissertation zeigen jedoch, dass diese Herleitung so nicht einfach möglich ist. Vielmehr weist dies auf die Relevanz des Kompetenz-als-Kontinuum-Modells (Blömeke, König et al., 2015) als Zwei-Wege-Modell, also ein Kontinuum in beide Richtungen (Blömeke & Kaiser, 2017; Santagata & Yeh, 2016), hin. Auf diese Wechselwirkungen zwischen Wissen und Können wird in dem Abschnitt zu Forschungsfrage 3 mit Bezug zu untersuchten Kovariaten noch einmal eingegangen.

*Wie hängen individuelle und akademische Hintergrunddaten der Proband*innen, DaZ-bezogene Lerngelegenheiten, Überzeugungen zu sprachlich-kultureller Heterogenität in Schule und Unterricht und Persönlichkeitsmerkmale mit performanznaher DaZ-Kompetenz zusammen?*

Ausgehend von dem Kompetenz-als-Kontinuum-Modell (Blömeke, Gustafsson & Shavelson, 2015) sind professionelles (DaZ-bezogenes) Wissen, positive Überzeugungen (zu sprachlich-kultureller Heterogenität in Schule und Unterricht) und geschulte situationsspezifische Fähigkeiten als Bedingungsfaktoren für performanznahe DaZ-Kompetenz angenommen worden. Weiterhin werden individuelle und akademische Hintergrunddaten sowie DaZ-bezogene Lerngelegenheiten als mögliche Prädiktoren angenommen.

Individuelle und akademische Hintergrunddaten

In Studie 4 zeigt sich, dass Proband*innen, die Englisch studieren oder unterrichten, signifikant höhere Ergebnisse im performanznahen *DaZKom-Video-Test* erzielen. Eine mögliche Begründung lässt sich aus der zentralen Bedeutung von Wertschätzung sprachlich-kultureller

Heterogenität ableiten: Budde (2016) betont die Wichtigkeit von *Language Awareness*, Lucas und Villegas (2013) beziehen sich auf positive „Orientations of Linguistically Responsive Teachers“ (Lucas & Villegas, 2013, S. 101) wie die Wertschätzung von sprachlich-kultureller Heterogenität und Gogolin (2020) sieht die Anerkennung der Mehrsprachigkeit der Schüler*innen als Teil ihres Bildungsprozesses (Gogolin, 2020). Interkulturelle Kompetenz ist ein Kompetenzbereich, der Teil des Englischunterrichts in der Grundschule ist (Niedersächsisches Kultusministerium, 2018). Dementsprechend haben sich (angehende) Englischlehrkräfte mehr als alle anderen in ihrer Ausbildung bereits mit dem Thema sprachlich-kultureller Heterogenität und der Wertschätzung ebendieser auseinandergesetzt.

Ein weiteres Ergebnis der Analyse der erhobenen individuellen und akademischen Hintergrunddaten zeigt, dass Proband*innen mit Forschungserfahrung signifikant höhere Ergebnisse im performanznahen *DaZKom-Video-Test* erzielen. Ganz zu Beginn des Projekts waren die Annahmen, dass nur erfahrene Lehrkräfte hohe Testwerte erreichen können, da sie nicht nur über Wissen verfügen, sondern durch Erfahrungen bereits in der Lage sind, intuitiv und spontan in DaZ-relevanten Unterrichtssituationen zu reagieren (Dreyfus & Dreyfus, 1986; Ohm, 2018). Dies setzt jedoch voraus, dass bereits professionelles Wissen vorhanden ist, in das wahrgenommene Situationen eingeordnet werden können. Wird das Kompetenz-als-Kontinuum-Modell (Blömeke, Gustafsson & Shavelson, 2015) nun als Kontinuum mit Wechselwirkungen in beide Richtungen betrachtet (Blömeke & Kaiser, 2017; Santagata & Yeh, 2016), welches davon ausgeht, dass Erfahrungen in der Praxis durch bewusste Reflexion zu Wissen werden bzw. vorhandenes Wissen (und Überzeugungen) verändert werden können, liegt es nahe, dass Forscher*innen in dem Feld hohe Testwerte erreichen. Sie verfügen sowohl über ein hohes Maß an professionellem Wissen als auch über sehr positive Überzeugungen, und setzen sich wiederholt mit konkreten DaZ-bezogenen Situationen auseinander mit dem Ziel, diese in einen theoretischen Kontext einzuordnen. Sie sind deshalb in der Lage, in dieser performanznahen Testumgebung höhere Testwerte zu erzielen.

DaZ-bezogene Lerngelegenheiten

Ein Ergebnis ist in allen Studien ähnlich und im ersten Moment entgegen den Erwartungen: Die Skala der DaZ-bezogenen Lerngelegenheiten zeigt wenig bis keine statistisch signifikante Korrelation mit dem Ergebnis des *DaZKom-Video-Tests*. Ein Grund dafür könnte sein, dass die DaZ-bezogenen Lerngelegenheiten als eine Skala von Themen (z.B. *Gebiete der Linguistik, Scaffolding*) abgefragt wurden. Dies kann daran liegen, dass Proband*innen generell eher

weniger DaZ-bezogene Lerngelegenheiten angegeben haben. Auf einer fünfstufigen Likert-Skala (1 = gar nicht, 2 = in einer Sitzung, 3 = in mehreren Sitzungen, 4 = in einer ganzen Lehrveranstaltung, 5 = in mehreren Lehrveranstaltungen) haben Proband*innen im Durchschnitt angegeben, die Lerngelegenheiten jeweils *in einer Sitzung/in mehreren Sitzungen* behandelt zu haben (Min = 1, Max = 5, MW = 2,58, SD = 0,46). Dies sagt jedoch noch nichts darüber aus, wie intensiv sie behandelt wurden und wie lange diese zurück liegen. Falls Proband*innen diese Themen in ihrer Aus- bzw. Weiterbildung bereits behandelt haben, sind diese dann zunächst erst einmal Teil der Dispositionen. Das Spannungsfeld zwischen den Dispositionen und der Performanz ist sehr komplex, das zeigt sowohl die theoretische Rahmung als auch der erste Teil dieser Diskussion. Sind also diese DaZ-bezogenen Lerngelegenheiten bereits Teil des professionellen Wissens, sagt dies noch nichts über ihre situationsspezifischen Fähigkeiten aus, die weitere Faktoren des Kompetenz-Kontinuums sind.

Überzeugungen zu sprachlich-kultureller Heterogenität in Schule und Unterricht

Insbesondere die letzte Studie (Artikel 4) hat gezeigt, dass neben Wissen und den situationsspezifischen Fähigkeiten als Bedingungsfaktoren für DaZ-spezifische Performanz auch positive Überzeugungen zu sprachlich-kultureller Heterogenität eine wesentliche Rolle einnehmen: Als Teil der Dispositionen war anzunehmen, dass sie einen Einfluss auf die Performanz, auf das professionelle Handeln von Lehrkräften haben, also maßgeblich mit Kompetenz zusammenhängen (Baumert & Kunter, 2006; Blömeke, Gustafsson & Shavelson, 2015; Ricart Brede, 2019b; Weinert, 2002). Das bestätigt auch die Forschung zu Überzeugungen (engl.: *beliefs*, Merk, 2020). Überzeugungen wird eine Filterfunktion zugeschrieben, die Handlungen bewertet und steuert (Fives & Buehl, 2012; Merk, 2020). Wie sie mit Kompetenz zusammenhängen, hängt wiederum auch von dem anderen Teil der Dispositionen, dem Wissen der Lehrkräfte ab (Buehl & Beck, 2014; Ricart Brede, 2019b). Untersuchungen zeigen, dass es für das Bildungssystem eine Herausforderung ist, Schüler*innen mit mehreren Erst-/Zweitsprachen Chancengleichheit zu gewähren im Vergleich zu den Schüler*innen, „deren erste und einzige *Lebenssprache* die Sprache des Bildungssystems ist“ (Gogolin et al., 2020, S. 4). Das mag auch, trotz entsprechend getroffener politischer Vorkehrungen, am immer noch vorherrschenden monolingualen Habitus (Gogolin, 1994) liegen (Krumm, 2020). Im Fokus stehen in der Schule deshalb eher die Schwierigkeiten des Erwerbs der (bildungssprachlichen) Deutschkompetenz (Budde, 2016), statt die Sprachen der Schüler*innen als Ressource wahrzunehmen (Busch, 2017). Dazu kommt, dass

Lehrkräfteausbildner*innen der zweiten Ausbildungsphase sich unvorbereitet fühlen, angehende Lehrkräfte für sprachbildenden Fachunterricht auszubilden (Benholz & Siems, 2016) und, dass Lehramtsstudierende sich eher weniger gut auf Unterricht mit DaZ-Schüler*innen vorbereitet fühlen (Ricart Brede, 2019a). Dies können Gründe für eher weniger positive Überzeugungen gegenüber sprachlich-kultureller Heterogenität in Schule und Unterricht sein. Das hat wiederum zur Folge, dass durch die genannte Filterfunktion von Überzeugungen, sowohl DaZ-bezogene Lerngelegenheiten nicht genutzt werden/wirksam sind und dass DaZ-relevante Situationen nicht als solche wahrgenommen werden, was wiederum zu defizitärer Performanz führen kann. Überzeugungen sind also nicht nur Bedingungsfaktor für performanznahe DaZ-Kompetenz, sie sind auch ein zuverlässiger Prädiktor für DaZ-Kompetenz.

Persönlichkeitsmerkmale

Die Analyse der erhobenen Kovariaten diente nicht nur der Feststellung der Bedingungsfaktoren und Prädiktoren, sondern auch der Validierung des Testinstruments. In dem Zuge wurden auch Persönlichkeitsmerkmale erhoben. Um eine möglichst performanznahe Messung von DaZ-Kompetenz zu realisieren, wurde ein mündliches Antwortformat genutzt mit dem Ziel, eine spontane, unmittelbare Antwort zu erhalten, die einer Handlungsreaktion gleicht. In den Studien 1 und 4 wurde ein besonderer Fokus auf die Evaluation und Validierung des Formates gesetzt. In Studie 1 wurden schriftliche und mündliche Antworten miteinander verglichen in der Annahme, dass Proband*innen, die weniger DaZ-kompetent sind, eher schriftlich antworten können, da sie so länger Zeit haben, über ihre Antwort nachzudenken, was dann wiederum die erwünschte Unmittelbarkeit (Lindmeier, 2013) verhindern würde. In Studie 4 wurden Persönlichkeitsmerkmale (Schupp & Gerlitz, 2014) von Lehrkräften miterhoben, die *Big Five: Extraversion, Verträglichkeit, Gewissenhaftigkeit, Neurotizismus und Offenheit für Erfahrungen*. Die Erfassung dieser Dimensionen hat sich bereits in zahlreichen Studien unterschiedlicher Zielsetzungen als verlässlicher Prädiktor erwiesen und wurde unter anderem auch in Studien der Bildungsforschung, wie zum Beispiel dem Nationalen Bildungspanel (NEPS) eingesetzt (Rammstedt & Danner, 2017). Die Annahme war, dass eher introvertierte, weniger offene Proband*innen weniger gut mündlich antworten können. In beiden Studien, 1 und 4, wurden jedoch keine Effekte festgestellt. Die Studien geben also Hinweise darauf, dass das mündliche Antwortformat fair ist und so für performanznahe Testumgebungen genutzt werden kann. Gleichzeitig deutet ebendieses Ergebnis darauf hin, dass auch ein

ökonomischeres, weniger aufwendiges Antwortformat genutzt werden könnte, da, zumindest in dieser Arbeit, keine Unterschiede festzustellen sind.

5.2 Limitationen und Empfehlungen für die Forschung

Mit der Diskussion der Ergebnisse dieser Dissertation gehen auch Limitationen einher, die im Folgenden mit Empfehlungen für weitere Forschungen formuliert werden.

Psychometrische Qualität und dimensionale Struktur des performanznahen DaZ-Kompetenztests

Die performanznahe Messung wurde über die situationsspezifischen Fähigkeiten operationalisiert. In den Studien, die die Dimensionalität untersucht haben (2 und 3), zeigt das zweidimensionale Modell zwischen den Dimensionen *Wahrnehmung* und *Handlung* jeweils sehr hohe Korrelationen. Dies ist theoriekonform. Die Wahrnehmung der relevanten Situation ist Grundlage für eine entsprechende Handlungsentscheidung (van Es & Sherin, 2002). Fraglich ist, ob beide situationsspezifischen Fähigkeiten abgefragt werden müssen, um Rückschlüsse auf die Performanz bzw. Kompetenz der Proband*innen ziehen zu können. Generell ist der Zusammenhang zwischen situationsspezifischen Fähigkeiten und der gemessenen Performanz weiterhin unklar. Fraglich ist auch, welche Rolle das *Verbalisierungsdilemma* einnimmt (Hecker et al., 2020): Annahmen zur Folge, dass Wissen eine Transformation durchläuft und so verinnerlicht wird, dass es intuitiv verfügbar, aber nicht mehr zu verbalisieren ist (Hecker et al., 2020; König & Lebens, 2012; Neuweg, 2014), würden erfahrene Proband*innen also einen Handlungsvorschlag machen, der aber ihrer tatsächlichen Handlung eventuell nicht entspricht.

Performanznahe (DaZ-) Kompetenz von angehenden Lehrkräften

In Bezug auf empirisches Erfassen von DaZ-bezogener Performanz wurde bereits herausgestellt, dass es sich um eine performanznahe Messung handelt und nicht um einen Performanztest. Die Performanz im Test entspricht also wahrscheinlich nicht der Performanz der Lehrkräfte im Klassenzimmer. Ein Hinweis darauf könnte auch die höhere Kompetenz von Forscher*innen im performanznahen *DaZKom-Video-Test* sein, denn obwohl sie offenbar sehr gut DaZ-bezogene Unterrichtssituationen identifizieren und konkrete Handlungsvorschläge machen können, haben sie nur in wenigen Fällen selber Lehrerfahrung. Wie also Handeln von Lehrkräften messbar abzubilden ist, ist immer noch unklar. Damit ist auch noch nicht umfassend geklärt, wie Wissen nun schlussendlich mit dem Handeln von Lehrkräften

zusammenhängt, das stellen auch andere Forschungsarbeiten in ähnlicher theoretischer Konzeption fest (Weber et al., 2020). Neben Wissen sind auch Überzeugungen Teil der Dispositionen, welche in der letzten Studie (Artikel 4) einen signifikanten Zusammenhang mit der gemessenen DaZ-Kompetenz zeigten. Wissen, Überzeugungen und Handeln hängen wechselseitig zusammen, es bleibt jedoch unklar, wie genau dieser Zusammenhang sich gestaltet (Ricart Brede, 2019b).

Um mehr über (DaZ-) Kompetenzentwicklung zu erfahren und einen Überblick über Kompetenzstufen in Abhängigkeit von (praktischen oder theoretischen) Lerngelegenheiten und der einzelnen Ausbildungsphasen (1-Studium, 2-Referendariat, 3-Fort- und Weiterbildungen (Benholz & Siems, 2016)) zu erhalten, wären Testungen mit beiden Instrumenten, dem *DaZKom-Paper-Pencil-Test* und dem *DaZKom-Video-Test* in längsschnittlichen Erhebungen hilfreich. Weitere Erprobungen der DaZ-Kompetenztests in Kombination mit der Erhebung konkreter DaZ-bezogener Lerngelegenheiten könnten auch dazu beitragen, mehr über die Lehrkräftebildung im Bereich DaZ zu erfahren und diese weiterzuentwickeln. Dieser Ansatz wird aktuell im *DaZKom-Transfer-Projekt*⁴ verfolgt.

Bedingungsfaktoren und Prädiktoren von performanznaher DaZ-Kompetenz

In dieser Arbeit lag der Fokus auf performanznaher DaZ-Kompetenz (angehender) Lehrkräfte sowie Bedingungsfaktoren und Prädiktoren dieser. In diesem Bereich sind auch weiterhin Forschung und vor allem der Transfer in die Praxis notwendig (vgl. Abschnitt 5.3). Was jedoch auch fehlt, ist der Rückbezug zu den Schüler*innen. Ausgehend von Bedarfen der Schüler*innen wurde ein großes Forschungsdesiderat für die Lehrkräfte formuliert und in den letzten Jahren in verschiedenen Projekten und Kontexten beforscht. Hilfreich wäre nun der Einbezug von Schüler*innen-Kompetenzen in Abhängigkeit der gemessenen (DaZ-) Kompetenz bei Lehrkräften, um mehr über die Relevanz von (DaZ-bezogener) Kompetenz von Lehrkräften in Bezug auf Schüler*innen-Kompetenz (bspw. fachlich und bildungssprachlich und in Bezug auf Unterrichtsqualität, Classroom-Management, Lehrkräfte-Gesundheit) zu erfahren (Blömeke & Kaiser, 2017). Diese Wirkungskette wurde bereits in einem Modell theoretisch abgebildet: *The Cascade Model* integriert *COACTIV* in das *Competence as a continuum*-Modell und ergänzt es um zwei weitere Säulen, die sich der (Teachers‘)

⁴ *DaZKom-Transfer*: Transfervorhaben zu „*DaZKom-Video* - Performanznahe Messung von Deutsch-als-Zweitsprache-Kompetenz bei (angehenden) Lehrkräften“, gefördert vom BMBF. Antragsteller*innen sind Andrea Daase, Timo Ehmke, Barbara Koch-Priewe, Anne Köker und Udo Ohm.

performance anschließen: Unterrichtsinteraktion/Lerngelegenheiten und (Lern-) Ziele (Krauss et al., 2020). In der Forschung gibt es bereits zumindest Hinweise darauf, dass eine höhere Wahrnehmungsfähigkeit bei Lehrkräften mit höheren Schüler*innen-Leistungen zusammenhängt (Kersting et al., 2012).

In der Theorie wird gerne auf Videovignetten verwiesen, um die professionelle Wahrnehmung zu schulen, denn theoretisch bieten sie viele Vorteile (Kleinknecht & Gröschner, 2016): Sie bilden z.B. komplexe Unterrichtssituationen authentisch ab, ermöglichen Analysen auf verschiedenen Ebenen und regen Reflexionsprozesse an. Vieles spricht dafür, sie in der Aus- und Weiterbildung von Lehrkräften gezielt einzusetzen. Dennoch ist unklar, inwiefern und welche Kompetenzentwicklung sie unterstützen, weshalb dies ein Forschungsdesiderat darstellt (Kang & van Es, 2019; Krammer, 2020).

5.3 Empfehlungen für die Lehrkräftebildung

„Die Wissenschaft konfrontiert Lehrende fortwährend mit Ideen, Konzepten oder Empfehlungen. [...] Lehrerinnen und Lehrer reagieren häufig mit großer Skepsis.“ (Schart & Legutke, 2012, S. 149)

Das an der Universität erworbene Wissen in der Unterrichtspraxis zu nutzen, stellt Lehrkräfte vor Herausforderungen, denn theoretisches Wissen berücksichtigt die Komplexität der Herausforderungen in der Praxis nicht. Durch die Erfahrungen im Klassenzimmer entwickeln Lehrkräfte Handlungswissen, sie verknüpfen das theoretisch erworbene Wissen mit individuellen Erfahrungen. Allerdings erfassen Lehrkräfte die Komplexität des Unterrichts durch die Vielzahl und Gleichzeitigkeit von Ereignissen nicht gänzlich (Schart & Legutke, 2012). Viele unmittelbare Interaktionen können Stress auslösen und führen auch bei erfahrenen Lehrkräften ggf. zu unprofessionellem Handeln (Sieland & Jordaan, 2019). Erkenntnisse der Forschung bieten eine produktive Möglichkeit zur Reflexion der eigenen Handlungen, wodurch dann Erfahrungswissen entsteht (Schart & Legutke, 2012). Eine Empfehlung für die Lehrkräftebildung ist die oft geforderte Theorie-Praxis-Verknüpfung, wobei es sich dabei um ein nicht eindeutig definiertes Konstrukt handelt. Vielmehr wird nur deutlich, dass offenbar gegensätzliche Begriffe in einem bestimmten Verhältnis zueinanderstehen, wobei sowohl *Theorie* als auch *Praxis* auf unterschiedlichste Art besetzt werden können (Rothland, 2020). Im Kontext der Lehrkräftebildung wird mit Theorie in der Regel wissenschaftliches Wissen bezeichnet, Praxis bezieht sich auf Berufspraxis in der Schule und schließt damit auch Handlungen und Erfahrungen mit ein (Rothland, 2020). Dies ist auch das Verständnis, vor

dessen Hintergrund diese Empfehlung ausgesprochen wird. Es ist nicht zielführend, deklaratives Wissen dem Handlungswissen gegenüberzustellen, beides ist grundlegend für professionelles Lehrkräftehandeln (Grießhaber, 2017). Stürmer (2011) empfiehlt in ihrer Arbeit, anwendungsorientiert und kontextualisiert Wissen zu vermitteln und die Lehrkräfteausbildung ganzheitlich zu betrachten.

Proband*innen, die im Zuge der Studien auf der Skala der DaZ-bezogene Lerngelegenheiten angaben, in Lehrveranstaltungen bereits sprachensible Unterrichtsstunden entworfen zu haben ($r = 0,17^*$, Studie 2) und jene, die sich bereits mit der Unterstützung des sprachlichen Lernprozesses durch Scaffolding ($r = 0,12^*$, Studie 4) auseinandergesetzt haben, erzielten in den Studien signifikant höhere Testwerte. Beides sind didaktische Vorgänge bzw. Konzepte, die Unterricht ganzheitlich betrachten. Dies können Hinweise sein, die die Annahme der Theorie-Praxis-Verknüpfung bzw. der kontextualisierten (Aus-) Bildung der Lehrkräfte stützen.

Eine Theorie-Praxis-Verknüpfung gestaltet sich insbesondere in der ersten Phase der Lehramtsausbildung häufig schwierig, wenn es um Unterrichtssituationen und auch die entsprechende Einbettung und Begleitung geht, weshalb hier Unterrichtsvideos (eigene und fremde) genutzt werden können (Steffensky & Kleinknecht, 2016), die eigene Vorteile mitbringen, welche zur Schulung der professionellen Wahrnehmung zentral sind: Die Situation kann wiederholt angeschaut, in Kontext eingeordnet und auf Theorie bezogen werden. Zudem fallen ggf. Details auf, die in einer Unterrichtssituation auf Grund der hohen Komplexität nicht beobachtet worden wären und der Fokus kann variiert auf unterschiedliche Unterrichtsebenen und – phänome gesetzt werden (Helmke, 2009; Krammer, 2020). Vignetten, die aus testtheoretischen Gründen aus dem Test ausgeschlossen wurden, werden teilweise in der Lehre erprobt. Ein Beispiel dafür ist eine Vignette⁵, die der Subdimension *Migration* des *DaZKom*-Modells zugeordnet ist und Kompetenzfacetten abbildet, die über sprachliche Aspekte hinaus gehen (Falkenstern et al., 2021). Die Proband*innen erzielten für dieses Item signifikant niedrigere Testwerte im Vergleich zu anderen Items im Test. Im Hinblick auf eine migrationspädagogische Sensibilisierung kann diese Vignette ggf. die situationsspezifischen Fähigkeiten schulen und einen reflexiven Umgang mit natio-ethno-kulturellen Differenzordnungen fördern (Falkenstern et al., 2021).

⁵ Das Video, aus dem diese Vignette stammt, ist Teil der Dokumentation *Kampf im Klassenzimmer* (2010) von Nicola Graef und Güner Balci, zusehen auf YouTube: <https://www.youtube.com/watch?v=Rz7veKkaSto>. Sie zeigt ein durch die Lehrerin gesteuertes Gespräch mit ihren Schüler*innen über deren Vorstellungen ihres zukünftigen Lebens.

Bezugnehmend auf die Theorie der beidseitigen Öffnung des Kompetenzkontinuums (Santagata & Yeh, 2016), bieten (eigene) Videos die Möglichkeit, Reflexionsprozesse anzustoßen. Durch die Reflexion von Erfahrungen in der Praxis wird Handlungswissen erworben: „Eine (forschungs-) methodisch fundierte Reflexions- und Selbstreflexionsfähigkeit des Handlungsvollzugs ist ein weiterer, wesentlicher Bestandteil professioneller Handlungskompetenz“ (Fröhlich-Gildhoff, 2015, S. 254). Ergebnisse der Forschungen zu Inklusion bspw. betonen, dass es neben einer entsprechenden Ausbildung für angehende Lehrkräfte wichtig ist, Erfahrungen in inklusiven Unterrichtssettings zu machen, um ihre Einstellungen (positiv) weiterzuentwickeln und eine ressourcen- statt defizitorientierte Haltung gegenüber Inklusion zu entwickeln (Heinrich et al., 2013). Einstellungen bzw. Überzeugungen sind Bestandteil der Dispositionen, weshalb ihnen eine grundlegende Rolle im Kompetenz-Kontinuum zukommt: Sie filtern die Wahrnehmung (Merk, 2020; Ricart Brede, 2019b) und auch sie können durch Reflexionsprozesse verändert werden (Fischer & Lahmann, 2020). Deshalb sollten Diskussions- und Reflexionsprozesse in die Lehrkräftebildung integriert werden, welche aktuell noch nicht systematisch berücksichtigt werden (Ricart Brede, 2019b).

Ganz zu Beginn dieser Dissertation wurde bereits angedeutet, dass außer der Frage wie Lehrkräfte DaZ-kompetent ausgebildet werden, auch die Frage nach dem wann relevant ist, insbesondere im Hinblick auf *Wissen versus Können* und *Theorie versus Praxis*. Für die konkrete Aus- und Weiterbildung von DaZ-kompetenten Lehrkräften könnte deshalb eine Empfehlung sein, dass fort- und weiterbildende Maßnahmen zu den Themen Sprachbildung/-förderung, DaZ und Mehrsprachigkeit, vor allem auch in der dritten Phase der Lehrerinnen- und Lehrerbildung nicht vernachlässigt, sondern ausgebaut werden sollten, denn hier wird systematisierte Theorie-Praxis-Verknüpfung ermöglicht (Becker-Mrotzek & Woerfel, 2020). Selbst wenn Studierende umfassendes Wissen bezüglich Sprachbildung/DaZ in der ersten Phase der Lehrkräftebildung aufbauen konnten, können sie ihre situationsspezifischen Fähigkeiten nicht ausbauen, wenn in der zweiten oder dritten Phase keine Weiterbildungsmöglichkeiten angeboten werden (Benholz & Siems, 2016), denn Lehrerfahrung alleine reicht zur Professionalisierung im Bereich DaZ nicht aus.

6. Literaturverzeichnis

- Ahrenholz, B. (Hrsg.). (2011). *Deutsch als Zweitsprache: Voraussetzungen und Konzepte für die Förderung von Kindern und Jugendlichen mit Migrationshintergrund* (3., unveränd. Aufl.). Fillibach.
- Ahrenholz, B. (2017). Erstsprache – Zweitsprache – Fremdsprache – Mehrsprachigkeit. In B. Ahrenholz, I. Oomen-Welke & W. Ulrich (Hrsg.), *Deutsch als Zweitsprache* (4. Aufl., S. 3–20). Schneider Verlag Hohengehren.
- Allgäuer-Hackle, E. & Jessner, U. (2014). „Und was sagt die Mehrsprachigkeitsforschung dazu?“ Neue Perspektiven in der Mehrsprachigkeitsforschung und deren Relevanz für Unterricht und Lehrer*innenbildung. In A. Wegner & E. Vetter (Hrsg.), *Mehrsprachigkeit und Professionalisierung in pädagogischen Berufen: Interdisziplinäre Zugänge zu aktuellen Herausforderungen im Bildungsbereich* (S. 125–146). Budrich.
- Baumann, B. (2017). Sprachförderung und Deutsch als Zweitsprache in der Lehrerbildung – ein deutschlandweiter Überblick-ein deutschlandweiter Überblick. In M. Becker-Mrotzek, P. Rosenberg, C. Schroeder & A. Witte (Hrsg.), *Sprachliche Bildung / herausgegeben vom Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache: Band 2. Deutsch als Zweitsprache in der Lehrerbildung* (S. 9–26). Waxmann.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Becker-Mrotzek, M., Rosenberg, P., Schroeder, C. & Witte, A. (Hrsg.). (2017). *Sprachliche Bildung / herausgegeben vom Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache: Band 2. Deutsch als Zweitsprache in der Lehrerbildung*. Waxmann.
- Becker-Mrotzek, M. & Roth, H.-J. (Hrsg.). (2017a). *Sprachliche Bildung: Band 1. Sprachliche Bildung - Grundlagen und Handlungsfelder*. Waxmann.
- Becker-Mrotzek, M. & Roth, H.-J. (2017b). Sprachliche Bildung – Grundlegende Begriffe und Konzepte. In M. Becker-Mrotzek & H.-J. Roth (Hrsg.), *Sprachliche Bildung: Band 1. Sprachliche Bildung - Grundlagen und Handlungsfelder* (S. 11–36). Waxmann.
- Becker-Mrotzek, M. & Woerfel, T. (2020). Sprachsensibler Unterricht und Deutsch als Zweitsprache als Gegenstand der Lehrerinnen- und Lehrerbildung. In C. Cramer, J. König & M. Rothland (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 98–104). Julius Klinkhardt.
- Benholz, C. & Siems, M. (2016). Sprachbildender Unterricht in allen Fächern: Konzepte zur Professionalisierung von Lehrerinnen und Lehrern in den drei Phasen der Lehrerbildung. In B. Koch-Priewe & M. Krüger-Potratz (Hrsg.), *Die deutsche Schule Beiheft: Bd. 13. Qualifizierung für sprachliche Bildung: Programme und Projekte zur Professionalisierung von Lehrkräften und pädagogischen Fachkräften*. Waxmann.
- Benz, J. (2019). *Deutsch als Zweitsprache lehren lernen: Eine Wirksamkeitsstudie zu Lehrerfortbildungen* (1. Auflage).
- Berliner, D. C. (2001). Learning about and learning from expert teachers. *International journal of educational research*, 35(5), 463–482.
- Berliner, D. C. (2004). Describing the Behavior and Documenting the Accomplishments of Expert Teachers. *Bulletin of Science, Technology & Society*, 24(3), 200–212.
<https://doi.org/10.1177/0270467604265535>

- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond Dichotomies. *Zeitschrift für Psychologie*, 223(1), 3–13. <https://doi.org/10.1027/2151-2604/a000194>
- Blömeke, S. & Kaiser, G. (2017). Understanding the development of teachers' professional competencies as personally, situationally and socially determined. In D. J. Clandinin & J. Husu (Hrsg.), *The SAGE handbook of research on teacher education* (S. 783–812). SAGE Publications Ltd.
- Blömeke, S., König, J., Suhl, U., Hoth, J. & Döhrmann, M. (2015). Wie situationsbezogen ist die Kompetenz von Lehrkräften? Zur Generalisierbarkeit der Ergebnisse von videobasierten Performanztests. *Zeitschrift für Pädagogik*(61), 310–327.
- Budde, M. A. (2016). Mehrsprachigkeit - Language Awareness - Sprachbewusstheit. Einführung. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 21(5), 1–4.
- Buehl, M. M. & Beck, J. S. (2014). The Relationship Between Teachers' Beliefs and Teachers' Practices. In H. Fives (Hrsg.), *International Handbook of Research on Teachers' Beliefs* (S. 66–84). Routledge.
- Busch, B. (2017). *Mehrsprachigkeit* (2. Aufl.). UTB GmbH.
- Carlson, S. A. & Präg, D. (2018). Der Prozess der Aufgabenentwicklung im DaZKom-Projekt: von der Rahmenkonzeption bis zur Pilotierung des Testinstruments. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Hrsg.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (S. 93–108). Waxmann.
- Casale, G., Strauß, S., Hennemann, T. & König, J. (2016). Wie lässt sich Klassenführungsexpertise messen? Überprüfung eines videobasierten Erhebungsinstruments für Lehrkräfte unter Anwendung der Generalisierbarkeitstheorie. *Empirische Sonderpädagogik*(2), 119–139.
- De Oliveira, L. C. de & Schleppegrell, M. J. (2015). *Focus on Grammar and Meaning*. Oxford University Press.
- Döll, M. & Hägi, S. (2014). Sprachkompetenzdiagnose und Sprachbildung: Anforderungen an Lehrkräfte. In A. Wegner & E. Vetter (Hrsg.), *Mehrsprachigkeit und Professionalisierung in pädagogischen Berufen: Interdisziplinäre Zugänge zu aktuellen Herausforderungen im Bildungsbereich* (S. 209–228). Budrich.
- Döll, M., Hägi-Mead, S. & Settinieri, J. (2017). Ob ich mich auf eine sprachlich heterogene Klasse vorbereitet fühle? – Etwas! Studentische Perspektiven auf DaZ und das DaZ-Modul (StuPaDaZ) an der Universität Paderborn. In M. Becker-Mrotzek, P. Rosenberg, C. Schroeder & A. Witte (Hrsg.), *Sprachliche Bildung / herausgegeben vom Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache: Band 2. Deutsch als Zweitsprache in der Lehrerbildung* (S. 203–215). Waxmann.
- Dreyfus, H. L. & Dreyfus, S. E. (1986). *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. Oxford: Basil Blackwell Ltd.
- Ehmke, T. & Hammer, S. (2018). Skalierung und dimensionale Struktur des DaZKom-Testinstruments. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Hrsg.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (S. 129–148). Waxmann.
- Ehmke, T., Hammer, S., Köker, A., Ohm, U. & Koch-Priewe, B. (Hrsg.). (2018). *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache*. Waxmann.
- Ehmke, T. & Lemmrich, S. (2018). Bedeutung von Lerngelegenheiten für den Erwerb von DaZ-Kompetenz. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Hrsg.), *Professionelle*

- Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (S. 201–220). Waxmann.
- Falkenstern, S., Hecker, S.-L. & Lemmrich, S. (2021). Lehrkräfteprofessionalisierung im Kontext von Heterogenität. DaZ-kompetenter Umgang mit Differenzordnungen im Unterricht. In I. Dirim & A. Wegner (Hrsg.), *Mehrsprachigkeit und Bildung. Deutsch als Zweitsprache: Inter- und transdisziplinäre Zugänge* (S. 189–202). Verlag Barbara Budrich.
<https://doi.org/10.2307/j.ctv1hm8gfs.13>
- Feilke, H. (2012). Bildungssprachliche Kompetenzen fördern und entwickeln. *Praxis Deutsch*(39 (233), 4–13.
- Fischer, N. & Ehmke, T. (2020). Theoretische Abbildung und empirische Erfassung der Überzeugungen angehender Lehrkräfte zu sprachlich-kultureller Heterogenität in Schule und Unterricht. *DIPF | Leibniz Institute for Research and Information in Education*. Vorab-Onlinepublikation. <https://doi.org/10.7477/395:246:1>
- Fischer, N., Hammer, S. & Ehmke, T. (2018). Überzeugungen zu Sprache im Fachunterricht: Erhebungsinstrument und Skaldokumentation. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Hrsg.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (S. 149–184). Waxmann.
- Fischer, N. & Lahmann, C. (2020). Pre-service teachers' beliefs about multilingualism in school: an evaluation of a course concept for introducing linguistically responsive teaching. *Language Awareness*, 29(2), 114–133. <https://doi.org/10.1080/09658416.2020.1737706>
- 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, S. Graham, J. M. Royer & M. Zeidner (Hrsg.), *APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors* (S. 4279–4499). American Psychological Association.
- Fröhlich-Gildhoff, K. (2015). Kompetenzorientierte Aus- und Weiterbildung – Verankerung von Inklusion als Inhalt und Prinzip. In E. Reichert-Garschhammer, C. Kieferle, M. Wertfein & F. Becker-Stoll (Hrsg.), *V&R Pädagogik. Inklusion und Partizipation: Vielfalt als Chance und Anspruch* (S. 253–262). Vandenhoeck & Ruprecht. <https://doi.org/10.13109/9783666701733.253>
- Gantefort, C. & Michalak, M. (2017). Zwischen Sprache und Fach–Deutsch als Zweitsprache im Lehramtsstudium an der Universität zu Köln. In M. Becker-Mrotzek, P. Rosenberg, C. Schroeder & A. Witte (Hrsg.), *Sprachliche Bildung / herausgegeben vom Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache: Band 2. Deutsch als Zweitsprache in der Lehrerbildung* (S. 61–73). Waxmann.
- Gawronski, B. & Creighton, L. A. (2013). Dual Process Theories. Vorab-Onlinepublikation. <https://doi.org/10.1093/oxfordhb/9780199730018.013.0014>
- Geist, B. & Krafft, A. (2017). *Deutsch als Zweitsprache: Sprachdidaktik für mehrsprachige Klassen. Linguistik und Schule: Bd. 2*. Narr Francke Attempto.
- Gibbons, P. (2002). *Scaffolding Language, Scaffolding Learning*. Portsmouth, NH: Heinemann.
- Gogolin, I. (1994). *Der monolinguale Habitus der multilingualen Schule*. Zugl.: Hamburg, Univ., Habil.-Schr., 1992. *Internationale Hochschulschriften*. Waxmann.

- Gogolin, I. (2006). Bilingualität und die Bildungssprache der Schule. *Die Macht der Sprachen. Englische Perspektiven auf die mehrsprachige Schule*, 79–85.
- Gogolin, I. (2020). Durchgängige Sprachbildung. In I. Gogolin, A. Hansen, S. McMonagle & D. Rauch (Hrsg.), *Handbuch Mehrsprachigkeit und Bildung* (S. 165–173). Springer Fachmedien.
- Gogolin, I., Hansen, A., McMonagle, S. & Rauch, D. (Hrsg.). (2020). *Handbuch Mehrsprachigkeit und Bildung*. Springer Fachmedien. <https://doi.org/10.1007/978-3-658-20285-9>
- Gogolin, I. & Lange, I. (2011). Bildungssprache und durchgängige Sprachbildung. In S. Fürstenau & M. Gomolla (Hrsg.), *Migration und schulischer Wandel: Mehrsprachigkeit* (S. 107–127). VS Verlag für Sozialwissenschaften.
- Goodwin, C. (1994). Professional Vision. *American Anthropologist*, 96(3), 606–633.
- Grießhaber, W. (2017). Lehramtswissen. In M. Becker-Mrotzek, P. Rosenberg, C. Schroeder & A. Witte (Hrsg.), *Sprachliche Bildung / herausgegeben vom Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache: Band 2. Deutsch als Zweitsprache in der Lehrerbildung* (S. 89–106). Waxmann.
- Gröscher, A. (2019). Analyse und Evaluation von Unterricht durch Videographie. In E. Kiel, B. Herzig, U. Maier & U. Sandfuchs (Hrsg.), *Handbuch Unterrichten an allgemeinbildenden Schulen* (S. 486–493). Julius Klinkhardt.
- Grosjean, F. (2020). Individuelle Zwei- und Mehrsprachigkeit. In I. Gogolin, A. Hansen, S. McMonagle & D. Rauch (Hrsg.), *Handbuch Mehrsprachigkeit und Bildung* (S. 13–21). Springer Fachmedien.
- Gültekin-Karakoç, N., Köker, A., Hirsch, D., Ehmke, T., Hammer, S., Koch-Priewe, B. & Ohm, U. (2016). Bestimmung von Standards und Stufen der Kompetenz angehende Lehrerinnen und Lehrer aller Fächer im Bereich Deutsch als Zweitsprache (DaZ). In B. Koch-Priewe & M. Krüger-Potratz (Hrsg.), *Die deutsche Schule Beiheft: Bd. 13. Qualifizierung für sprachliche Bildung: Programme und Projekte zur Professionalisierung von Lehrkräften und pädagogischen Fachkräften* (S. 130–146). Waxmann.
- Hammer, S., Fischer, N. & Koch-Priewe, B. (2016). Überzeugungen von Lehramtsstudierenden zu Mehrsprachigkeit in der Schule. In B. Koch-Priewe & M. Krüger-Potratz (Hrsg.), *Die deutsche Schule Beiheft: Bd. 13. Qualifizierung für sprachliche Bildung: Programme und Projekte zur Professionalisierung von Lehrkräften und pädagogischen Fachkräften* (S. 147–171). Waxmann.
- Harms, U. & Riese, J. (2018). Professionelle Kompetenz und Professionswissen. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Lehrbuch. Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 283–298). Springer. https://doi.org/10.1007/978-3-662-56320-5_17
- Hecker, S.-L., Falkenstern, S., Lemmrich, S. & Ehmke, T. (2020). Zum Verbalisierungsdilemma bei der Erfassung der situationsspezifischen Fähigkeiten von Lehrkräften. *Zeitschrift für Bildungsforschung*, 10(2), 175–190. <https://doi.org/10.1007/s35834-020-00268-1>
- Heine, L., Domenech, M., Otto, L., Neumann, A., Krelle, M., Leiss, D., Höttecke, D., Ehmke, T. & Schwippert, K. (2018). Modellierung sprachlicher Anforderungen in Testaufgaben verschiedener Unterrichtsfächer: Theoretische und empirische Grundlagen. *Zeitschrift für angewandte Linguistik*, 2018(69), 69–96.
- Heinrich, M., Urban, M. & Werning, R. (2013). Grundlagen, Handlungsstrategien und Forschungsperspektiven für die Ausbildung und Professionalisierung von Fachlehrkräften für

- inklusive Schulen. In H. Döbert & H. Weishaupt (Hrsg.), *Inklusive Bildung professionell gestalten: Situationsanalyse und Handlungsempfehlungen* (1. Aufl., S. 69–134). Waxmann.
- Heins, J. & Zabka, T. (2019). „Mentale Prozesse bei der Unterrichtsbeobachtung. Theoretische Klärungen und ein Fallbeispiel zum Literaturunterricht“. *Zeitschrift für Pädagogik*, 65, 903–924.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts; Franz Emanuel Weinert gewidmet* (1. Aufl.). Klett Kallmeyer.
- Henrici, G. & Vollmer, H. J. (2001). Lernen und Lehren von Fremdsprachen: Kognition, Affektion, Interaktion. Ein Forschungsüberblick–Das Spektrum des interdisziplinären Forschungsfeldes: Zielsetzung und Relevanz. *Zeitschrift für Fremdsprachenforschung (ZFF)*, 12(2), 6–17.
- Hoth, J. (2016). *Situationsbezogene Diagnosekompetenz von Mathematiklehrkräften: Eine Vertiefungsstudie zur TEDS-Follow-Up-Studie*. Dissertation. Springer Spektrum.
- Jeschke, C., Lindmeier, A. & Heinze, A. (2020). Vom Wissen zum Handeln: Vermittelt die Kompetenz zur Unterrichtsreflexion zwischen mathematischem Professionswissen und der Kompetenz zum Handeln im Mathematikunterricht? Eine Mediationsanalyse. *Journal für Mathematik-Didaktik*. Vorab-Onlinepublikation. <https://doi.org/10.1007/s13138-020-00171-2>
- Kahneman, D. (2013). Thinking, fast and slow. *New York: Farrar, Straus and Giroux*.
- Kang, H. & van Es, E. A. (2019). Articulating design principles for productive use of video in preservice education. *Journal of Teacher Education*, 70(3), 237–250.
- Kersting, N. (2008). Using Video Clips of Mathematics Classroom Instruction as Item Prompts to Measure Teachers' Knowledge of Teaching Mathematics. *Educational and Psychological Measurement*, 68(5), 845–861. <https://doi.org/10.1177/0013164407313369>
- Kersting, N., Givvin, K. B., Thompson, B. J., Santagata, R. & Stigler, J. W. (2012). Measuring Usable Knowledge. *American Educational Research Journal*, 49(3), 568–589. <https://doi.org/10.3102/0002831212437853>
- Kleinknecht, M. & Gröschner, A. (2016). Fostering preservice teachers' noticing with structured video feedback: Results of an online-and video-based intervention study. *Teaching and Teacher Education*, 59, 45–56.
- Klieme, E. & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In M. Prenzel, I. Gogolin & H.-H. Krüger (Hrsg.), *Zeitschrift für Erziehungswissenschaft, Sonderheft 8. Kompetenzdiagnostik* (S. 11–29). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90865-6_2
- Knievel, I., Lindmeier, A. & Heinze, A. (2015). Beyond Knowledge: Measuring Primary Teachers' Subject-Specific Competences in and for Teaching Mathematics with Items Based on Video Vignettes. *International Journal of Science and Mathematics Education*, 13(2), 309–329. <https://doi.org/10.1007/s10763-014-9608-z>
- Kniffka, G. & Siebert-Ott, G. (2012). *Deutsch als Zweitsprache: Lehren und Lernen* (3. Aufl.). *StandardWissen Lehramt: Bd. 2891*. Schöningh.
- Köker, A. (2018). Zur Relevanz der Ausbildung von Kompetenzen von Lehrkräften im Bereich DaZ. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Hrsg.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (S. 57–71). Waxmann.

- Köker, A., Rosenbrock, S., Ohm, U., Ehmke, T., Hammer, S., Koch-Priewe, B. & Schulze, N. (2015). DaZKom—Ein Modell von Lehrerkompetenz im Bereich Deutsch als Zweitsprache. In B. Koch-Priewe, A. Köker, J. Seifried & E. Wuttke (Hrsg.), *Kompetenzerwerb an Hochschulen: Modellierung und Messung: Zur Professionalisierung angehender Lehrerinnen und Lehrer sowie frühpädagogischer Fachkräfte*. Verlag Julius Klinkhardt.
- König, J. (2015). Measuring classroom management expertise (CME) of teachers: A video-based assessment approach and statistical results. *Cogent Education*, 2(1).
<https://doi.org/10.1080/2331186X.2014.991178>
- König, J. (2020). Kompetenzorientierter Ansatz in der Lehrerinnen- und Lehrerbildung. In C. Cramer, J. König & M. Rothland (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 163–171). Julius Klinkhardt.
- König, J. & Lebens, M. (2012). Classroom Management Expertise (CME) von Lehrkräften messen: Überlegungen zur Testung mithilfe von Videovignetten und erste empirische Befunde. *Lehrerbildung auf dem Prüfstand*, 5(1), 3–28.
- Kramer, M., Förtsch, C., Stürmer, J., Förtsch, S., Seidel, T. & Neuhaus, B. J. (2020). Measuring biology teachers' professional vision: Development and validation of a video-based assessment tool. *Cogent Education*, 7(1), Artikel 1823155. <https://doi.org/10.1080/2331186X.2020.1823155>
- Krammer, K. (2020). Videos in der Lehrerinnen- und Lehrerbildung. In C. Cramer, J. König & M. Rothland (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 691–699). Julius Klinkhardt.
- Krauss, S. (2020). Expertise-Paradigma in der Lehrerinnen- und Lehrerbildung. In C. Cramer, J. König & M. Rothland (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 154–162). Julius Klinkhardt.
- Krauss, S. & Bruckmaier, G. (2014). Das Experten-Paradigma in der Forschung zum Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. Aufl., S. 171–191). Waxmann.
- Krauss, S., Bruckmaier, G., Lindl, A., Hilbert, S., Binder, K., Steib, N. & Blum, W. (2020). Competence as a continuum in the COACTIV study: the “cascade model”. *ZDM*, 52(2), 311–327.
<https://doi.org/10.1007/s11858-020-01151-z>
- Krumm, H.-J. (2020). Mehrsprachigkeit und Identität. In I. Gogolin, A. Hansen, S. McMonagle & D. Rauch (Hrsg.), *Handbuch Mehrsprachigkeit und Bildung* (S. 131–135). Springer Fachmedien.
- Lange, I. (2020). Bildungssprache. In I. Gogolin, A. Hansen, S. McMonagle & D. Rauch (Hrsg.), *Handbuch Mehrsprachigkeit und Bildung* (S. 53–58). Springer Fachmedien.
- Leisen, J. (2013). *Handbuch Sprachförderung im Fach: Sprachsensibler Fachunterricht in der Praxis : Grundlagenwissen, Anregungen und Beispiele für die Unterstützung von sprachschwachen Lernern und Lernern mit Zuwanderungsgeschichte beim Sprechen, Lesen, Schreiben und Üben im Fach*. Ernst Klett Sprachen.
- Lemmrich, S., Bahls, A. & Ehmke, T. (2020). Effekte von mündlichen versus schriftlichen Antwortformaten bei der performanznahen Messung von Deutsch-als-Zweitsprache (DaZ)-Kompetenz – eine experimentelle Studie mit angehenden Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 1–11. <https://doi.org/10.1024/1010-0652/a000282>
- Lemmrich, S. & Ehmke, T. (eingereicht). Preparing Teachers for Linguistically Responsive Teaching: A Test Instrument to Measure Teachers' Competence and Evaluate Learning Opportunities. *Language Awareness*.

- Lemmrich, S., Hecker, S.-L., Klein, S., Ehmke, T., Koch-Priewe, B., Köker, A. & Ohm, U. (2020). Linguistically Responsive Teaching in Multilingual Classrooms: Development of a Performance-Oriented Test to Assess Teachers' Competence. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper & C. Lautenbach (Hrsg.), *Student Learning in German Higher Education* (1. Aufl., Bd. 2020, S. 125–140). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27886-1_7
- Lemmrich, S., Hecker, S.-L., Klein, S., Ehmke, T., Köker, A., Koch-Priewe, B. & Ohm, U. (2019). Performanznahe und videobasierte Messung von DaZ-Kompetenz bei Lehrkräften: Skalierung und dimensionale Struktur des Testinstruments. In T. Ehmke, P. Kuhl & M. Pietsch (Hrsg.), *Lehrer. Bildung. Gestalten: Beiträge zur empirischen Forschung in der Lehrerbildung* (1. Aufl., S. 190–204). Beltz.
- Lengyel, D. (2016). Umgang mit sprachlicher Heterogenität im Klassenzimmer. In J. Kilian, B. Brouër & D. Lüttenberg (Hrsg.), *Handbuch Sprache in der Bildung* (S. 500–522). de Gruyter.
- Lindmeier, A. (2011). *Modeling and measuring knowledge and competencies of teachers: A threefold domain-specific structure model for mathematics*. Zugl.: München, Techn. Univ., Diss., 2010. *Empirische Studien zur Didaktik der Mathematik: Bd. 7*. Waxmann.
- Lindmeier, A. (2013). Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen. In U. Riegel & K. Macha (Hrsg.), *Videobasierte Kompetenzforschung in den Fachdidaktiken* (S. 45–61). Waxmann.
- Lindmeier, A., Hepberger, B., Heinze, A. & Moser Opitz, E. (2016). Modeling cognitive dispositions of educators for early mathematics education, 219–226.
- Lucas, T. & Villegas, A. M. (2011). A Framework for Preparing Linguistically Responsive Teachers. In T. Lucas (Hrsg.), *Teacher preparation for linguistically diverse classrooms: A resource for teacher educators* (S. 55–72). Routledge.
- Lucas, T. & Villegas, A. M. (2013). Preparing Linguistically Responsive Teachers: Laying the Foundation in Preservice Teacher Education. *Theory Into Practice*, 52(2), 98–109. <https://doi.org/10.1080/00405841.2013.770327>
- Maak, D. (2019). „Manchmal ist viel auch besser!": Nutzung von mehrsprachigen Produktbeschreibungen in Schule und Hochschule zur Auseinandersetzung mit Mehrsprachigkeit. In D. Maak & J. Ricart Brede (Hrsg.), *Wissen, Können, Wollen, Sollen?!* (S. 201–233). Waxmann.
- Marx, N. (2018). Förderung, aber welchen Inhalts? Didaktische Perspektiven auf Bildungssprache im DaF-Unterricht. *Informationen Deutsch als Fremdsprache*, 45(4), 401–422.
- Merk, S. (2020). Überzeugungen. In C. Cramer, J. König & M. Rothland (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 825–832). Julius Klinkhardt.
- Meyer, M. & Prediger, S. (2012). Sprachenvielfalt im Mathematikunterricht: Herausforderungen, Chancen und Förderansätze. *Mathematik in der Schule*(54), 2–9.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic medicine : journal of the Association of American Medical Colleges*, 65(9 Suppl), 63-67. <https://doi.org/10.1097/00001888-199009000-00045>
- Ministerium für Schule und Bildungs des Landes Nordrhein-Westfalen. (2020). *Referenzrahmen für Schulqualität NRW: Schule in NRW Nr. 9051*. <https://www.schulentwicklung.nrw.de/referenzrahmen/broschuere.pdf>

- Morek, M. & Heller, V. (2012). Bildungssprache–Kommunikative, epistemische, soziale und interaktive Aspekte ihres Gebrauchs. *Zeitschrift für angewandte Linguistik*, 57(1), 67–101.
- Neuweg, G. H. (2014). Das Wissen der Wissensvermittler. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (2. Aufl., S. 583–614). Waxmann.
- Niedersächsisches Kultusministerium. (2018). *Kerncurriculum für die Grundschule, Schuljahrgänge 3-4, Englisch*.
https://www.cuvo.nibis.de/cuvo.php?skey_lev0_0=Dokumentenart&svalue_lev0_0=Kerncurriculum&skey_lev0_1=Schulbereich&svalue_lev0_1=Primarbereich&skey_lev0_2=Fach&svalue_lev0_2=Englisch&docid=1153&p=detail_view
- Niedersächsisches Kultusministerium. (2014). *Orientierungsrahmen Schulqualität in Niedersachsen*.
https://www.mk.niedersachsen.de/startseite/schule/schulqualitaet/orientierungsrahmen_schulqualitaet_in_niedersachsen/orientierungsrahmen---basis-fuer-schulqualitaet-in-niedersachsen-6339.html
- Ohm, U. (2018). Das Modell von DaZ-Kompetenz bei angehenden Lehrkräften. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Hrsg.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (S. 73–91). Waxmann.
- Paetsch, J., Darsow, A., Wagner, F. S., Hammer, S. & Ehmke, T. (2019). Prädiktoren des Kompetenzzuwachses im Bereich Deutsch als Zweitsprache bei Lehramtsstudierenden. *Unterrichtswissenschaft*, 47(1), 51–77. <https://doi.org/10.1007/s42010-019-00038-7>
- Petersen, I. 1. (2014). *Schreibfähigkeit und Mehrsprachigkeit. DaZ-Forschung 5*. de Gruyter.
http://deposit.d-nb.de/cgi-bin/dokserv?id=4384066&prov=M&dok_var=1&dok_ext=htm
- Putjata, G. & Danilovich, Y. (2019). Sprachliche Vielfalt als regulärer Bestandteil der Lehrerbildung: Zum Bedarf fachlicher und fachdidaktischer Perspektiven. In Y. Danilovich & G. Putjata (Hrsg.), *Edition Fachdidaktiken. Sprachliche Vielfalt im Unterricht: Fachdidaktische Perspektiven auf Lehre und Forschung im DaZ-Modul* (1. Aufl., S. 1–14). Springer Fachmedien Wiesbaden GmbH.
https://doi.org/10.1007/978-3-658-23254-2_1
- Rammstedt, B. & Danner, D. (2017). Die Facettenstruktur des Big Five Inventory (BFI). *Diagnostica*, 63(1), 70–84. <https://doi.org/10.1026/0012-1924/a000161>
- Reusser, K. (2005). Situiertes Lernen mit Unterrichtsvideos. *Journal für Lehrerinnen- und Lehrerbildung*(2), 8–18.
- Ricart Brede, J. (2019a). „Kann man überhaupt vorbereitet sein?“ Zur Vorbereitung auf den Unterricht mit DaZ-SchülerInnen aus Sicht von Lehramtsstudierenden. In D. Maak & J. Ricart Brede (Hrsg.), *Wissen, Können, Wollen, Sollen?!* (S. 161–182). Waxmann.
- Ricart Brede, J. (2019b). Einstellungen – beliefs – Überzeugungen – Orientierungen: Zum theoretischen Konstrukt des Projektes „Einstellungen angehender LehrerInnen zu Deutsch als Zweitsprache in Ausbildung und Unterricht“. In D. Maak & J. Ricart Brede (Hrsg.), *Mehrsprachigkeit: Band 46. Wissen, Können, Wollen - sollen?! (angehende) LehrerInnen und äußere Mehrsprachigkeit* (S. 29–38). Waxmann.
- Röbe, E., Aicher-Jakob, M. & Seifert, A. (2019). *Lehrer werden, Lehrer sein, Lehrer bleiben: Lehrerberuf zwischen Schulalltag und Professionalisierung. utb-studi-e-book*.
- Rösch, H. (2011). *Deutsch als Zweit- und Fremdsprache*. Akademie Verlag.

- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion* (2. Aufl.). *Psychologie Lehrbuch*. Huber.
- Rost-Roth, M. (2017). Lehrprofessionalität (nicht nur) für Deutsch als Zweitsprache – Sprachbezogene und interaktive Kompetenzen für Sprachförderung, Sprachbildung und sprachsensiblen Fachunterricht. In B. Lütke, I. Petersen & T. Tajmel (Hrsg.), *DaZ-Forschung /DaZ-Forj*: v.8. *Fachintegrierte Sprachbildung: Forschung, Theoriebildung und Konzepte für die Unterrichtspraxis* (S. 69–97). de Gruyter. <https://doi.org/10.1515/9783110404166-004>
- Rothland, M. (2020). Theorie-Praxis-Verhältnis in der Lehrerinnen- und Theorie-Praxis-Verhältnis in der Lehrerinnen- und Lehrerbildung. In C. Cramer, J. König & M. Rothland (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 133–140). Julius Klinkhardt.
- Santagata, R. & Yeh, C. (2016). The role of perception, interpretation, and decision making in the development of beginning teachers' competence. *ZDM*, 48(1-2), 153–165. <https://doi.org/10.1007/s11858-015-0737-9>
- Schart, M. & Legutke, M. (2012). *Lehrkompetenz und Unterrichtsgestaltung* (1. Aufl.). *Deutsch als Fremdsprache: / allgemeiner Herausgeber: Goethe-Institut e.V. ; 1.* Langenscheidt; Klett-Langenscheidt; Ernst Klett Sprachen.
- Schleppegrell, M. J. (2007). The Linguistic Challenges of Mathematics Teaching and Learning: A Research Review. *Reading & Writing Quarterly*, 23(2), 139–159. <https://doi.org/10.1080/10573560601158461>
- Schmölzer-Eibinger, S. (2013). Sprache als Medium des Lernens im Fach. In M. 1. Becker-Mrotzek (Hrsg. der Reihe) & M. Becker-Mrotzek, K. Schramm, E. Thürmann & H. J. Vollmer (Hrsg. des Bandes), *Fachdidaktische Forschungen 3. Sprache im Fach Sprachlichkeit und fachliches Lernen* (S. 25–40). Waxmann.
- Schröder, J., Riese, J., Vogelsang, C., Borowski, A., Buschhüter, D., Enkrott, P., Kempin, M., Kulgemeyer, C., Reinhold, P. & Schecker, H. (2020). Die Messung der Fähigkeit zur Unterrichtsplanung im Fach Physik mit Hilfe eines standardisierten Performanztests. *Zeitschrift für Didaktik der Naturwissenschaften*. Vorab-Onlinepublikation. <https://doi.org/10.1007/s40573-020-00115-w>
- Schroeder, C. & Stölting, W. (2005). Mehrsprachig orientierte Sprachstandsfeststellung für Kinder mit Migrationshintergrund. In I. Gogolin (Hrsg.), *FörMig-Edition: Bd. 1. Sprachdiagnostik bei Kindern und Jugendlichen mit Migrationshintergrund: Dokumentation einer Fachtagung am 14. Juli 2004 in Hamburg* (S. 59–74). Waxmann.
- Schupp, J. & Gerlitz, J.-Y. (Hrsg.) (2014). *Big Five Inventory-SOEP (BFI-S)*. ZIS - GESIS Leibniz Institute for the Social Sciences.
- Seidel, T. & Stürmer, K. (2014). Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. *American Educational Research Journal*, 51(4), 739–771. <https://doi.org/10.3102/0002831214531321>
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher*, 15(2), 4–14.
- Sieland, B. & Jordaan, L. (2019). Lehrerinnen-und Lehrerbildung im Umbruch zwischen Auftrag, alten Schwächen und neuen Chancen. *Kompetenzentwicklung im Lehramtsstudium durch professionelles Training*, 7–19.

- Steffensky, M. & Kleinknecht, M. (2016). Wirkungen videobasierter Lernumgebungen auf die professionelle Kompetenz und das Handeln (angehender) Lehrpersonen. Ein Überblick zu Ergebnissen aus aktuellen (quasi-) experimentellen Studien. *Unterrichtswissenschaft*, 44(4), 305–321.
- Stürmer, K. (2011). *Voraussetzungen für die Entwicklung professioneller Unterrichtswahrnehmung im Rahmen universitärer Lehrerbildung* [, Technische Universität München]. EndNote Tagged Import Format.
- Thürmann, E., Krabbe, H., Platz, U. & Schumacher, M. (Hrsg.). (2017). *Ganz In - Materialien für die Praxis. Sprachbildung als Aufgabe aller Fächer und Lernbereiche: Erfahrungen mit Sprachberatung an Ganz-In-Gymnasien*. Waxmann.
- van Es, E. & Sherin, M. (2002). Learning to Notice: Scaffolding New Teachers' Interpretations of Classroom Interactions. *Journal of Information Technology for Teacher Education*, 10(4), 571–596.
- Villegas, A. M. & Lucas, T. (2002). Preparing Culturally Responsive Teachers. *Journal of Teacher Education*, 53(1), 20–32. <https://doi.org/10.1177/0022487102053001003>
- Wandruszka, M. (1979). *Die Mehrsprachigkeit des Menschen*. Piper.
- Weber, K. E., Prilop, C. N., Viehoff, S., Gold, B. & Kleinknecht, M. (2020). Fördert eine videobasierte Intervention im Praktikum die professionelle Wahrnehmung von Klassenführung? – Eine quantitativ-inhaltsanalytische Messung von Subprozessen professioneller Wahrnehmung. *Zeitschrift für Erziehungswissenschaft*, 23(2), 343–365. <https://doi.org/10.1007/s11618-020-00939-9>
- Weinert, F. E. (2002). *Leistungsmessungen in Schulen* (Beltz Pädagogik). Beltz.
- Wilson, M. (2004). *Constructing Measures*. Routledge. <https://doi.org/10.4324/9781410611697>
- Wischmeier, I. (2012). „Teachers’ Beliefs“: Überzeugungen von (Grundschul-) Lehrkräften über Schüler und Schülerinnen mit Migrationshintergrund–Theoretische Konzeption und empirische Überprüfung. In W. Wiater & D. Manschke (Hrsg.), *Verstehen und Kultur: Mentale Modelle und kulturelle Prägungen* (S. 167–189). Springer; Springer VS.
- Zlatkin-Troitschanskaia, O., Pant, H. A., Toepper, M. & Lautenbach, C. (Hrsg.). (2020). *Student Learning in German Higher Education* (1st ed.). Springer Fachmedien Wiesbaden.

7. Artikel 1: Effekte von mündlichen versus schriftlichen Antwortformaten bei der performanznahen Messung von Deutsch-als-Zweitsprache (DaZ)-Kompetenz – eine experimentelle Studie mit angehenden Lehrkräften

Der Beitrag ist wie folgt publiziert:

Lemmrich, S., Bahls, A. & Ehmke, T. (2020). Effekte von mündlichen versus schriftlichen Antwortformaten bei der performanznahen Messung von Deutsch-als-Zweitsprache (DaZ)-Kompetenz – eine experimentelle Studie mit angehenden Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 1–11. <https://doi.org/10.1024/1010-0652/a000282>



Effekte von mündlichen versus schriftlichen Antwortformaten bei der performanznahen Messung von Deutsch-als-Zweitsprache (DaZ)-Kompetenz – eine experimentelle Studie mit angehenden Lehrkräften

Svenja Lemmrich, Anika Bahls und Prof. Dr. Timo Ehmke

Leuphana Universität Lüneburg, Lüneburg, Deutschland

Zusammenfassung: In einer experimentellen Studie mit randomisierter Stichprobenzuweisung werden zwei Antwortformate (schriftlich versus mündlich) in einer computerbasierten Testumgebung gegeneinander getestet. Mit beiden Varianten soll performanznah die Deutsch-als-Zweitsprache (DaZ)-Kompetenz bei angehenden Lehrkräften mit Hilfe von Videovignetten erfasst werden. Die Studie untersucht, inwieweit das Antwortformat Einfluss auf die Korrektheit und die Ausführlichkeit der Antworten hat. Zudem wird geprüft, ob die Effekte auch unter Kontrolle von DaZ-Lerngelegenheiten bestehen bleiben bzw. ob sich ein Interaktionseffekt von DaZ-Lerngelegenheiten und Antwortmodus zeigt. Die Stichprobe (N = 90) wurde aus Lehrveranstaltungen des Lehramtsstudiums gewonnen (B.A./M.Ed., alle Semester). Die Ergebnisse zeigen, dass der Antwortmodus keinen statistisch signifikanten Einfluss auf die Richtigkeit, jedoch auf die Ausführlichkeit der Antworten hat. Im mündlichen Antwortformat wurden statistisch bedeutsam ausführlichere Antworten gegeben. Zudem zeigt sich, dass Testpersonen, die mehr DaZ-Lerngelegenheiten hatten, die Aufgaben eher richtig lösen, unabhängig vom Antwortmodus. Interaktionseffekte zeigten sich nicht. Für die Testentwicklung kann damit die Schlussfolgerung gezogen werden, dass ein mündliches Antwortformat bei Studierenden nicht zu unfairen Testergebnissen führt, verglichen mit einem traditionellen schriftlichen Antwortformat.

Schlüsselwörter: Performanznahe Messung, professionelle Kompetenzen (angehender) Lehrkräfte, Video-Vignetten, mündliche versus schriftliche Antworten, Antwortmodus

Effects of oral vs. written response formats in a performance-oriented assessment of German-as-a-Second-Language (GSL)-competency: An experimental study with pre-service teachers

Abstract: In this experimental study, two response formats (oral vs. written responses) are examined within a computer-based test environment. Both varieties assess German-as-a-Second-Language (GSL) competency of (pre-service) teachers, using video-based items. The study investigates the influence of the response format on correctness and comprehensiveness. Moreover, the study examines if the measured effects maintain when GSL-relevant learning opportunities are controlled or if interaction effects between learning opportunities and the response mode appear. The random sample (N = 90) consists of students of educational studies (pre-service teachers, B.A./M.Ed.). The results of this study show that the response mode has no statistically significant effect on the correctness but on the comprehensiveness of the responses. Oral responses were statistically significant more comprehensive than written responses. Test persons who had more GSL-relevant learning opportunities were more likely to answer correctly, regardless of the response mode. There are no interaction effects. The implication for the test construction therefore is that oral response formats do not lead to unfair test results while testing students, compared to a written format.

Keywords: Performance-oriented testing, professional competences of (pre-service) teachers, video-vignettes, oral versus written responses, response mode

Einleitung

Schülerinnen und Schüler unterschiedlichster Sprachbiographien nehmen am Unterricht teil. Entsprechend viel-

fältig sind die Deutschkompetenzen der Schülerinnen und Schüler. (Paetsch et al., 2019). „Die Lehrkräfte sind deshalb aufgefordert, die sprachlichen Kompetenzen der Schülerinnen und Schüler auszubauen und die Vermitt-

lung (bildungs-)sprachlicher Kompetenzen in jedem Fachunterricht systematisch zu berücksichtigen“ (Paetsch et al., 2019, S.52). Welche Kompetenzen Lehrkräfte wiederum benötigen, diesen Anforderungen in ihrem Unterricht gerecht zu werden, ist unklar. DaZ-bezogene Ausbildungs- und Performanzstandards fehlen bislang (Koch-Priewe, 2018) und die Ausbildung der Lehrkräfte im Bereich Deutsch als Zweitsprache (DaZ) ist deutschlandweit, von Bundesland zu Bundesland und auch von Universität zu Universität, sehr verschieden. Um DaZ-relevante Lerngelegenheiten identifizieren zu können, wird eine performanznahe Messung von DaZ-Kompetenz angestrebt. Dafür wird in dieser Studie ein videobasiertes Testinstrument genutzt: Authentische, videobasierte Testinstrumente sollen neben kognitiven Facetten von Kompetenz auch Handlungsfacetten messbar machen (Gold & Holodynski, 2017). Die Darbietung von Videostimuli und die damit verbundenen Vor- bzw. Nachteile werden daher in der Forschung zu Lehrerinnen- und Lehrerbildung und Professionalisierung bereits untersucht und diskutiert (u.a. Gold & Holodynski, 2017; Borko, 2016; Gaudin & Chaliès, 2015). Weniger häufig thematisiert wird dagegen das Antwortformat. Oft wird derzeit, etwa in computerbasierten Erhebungen, auf Antwortformate zurückgegriffen, wie sie in Papier- und Bleistift-Test eingesetzt werden: geschlossene Formate (z.B. Multiple-Choice), halboffene Formate oder auch offene Antwortformate, die einen Antworttext erfordern (Bauer & Blasius, 2014). Unklar ist jedoch, ob durch ein mündliches Antwortformat, welches eine gewisse Spontanität und Unmittelbarkeit mitbringt, eine performanznähere Messung realisiert werden kann.

In unserer Studie soll geprüft werden, ob ein verändertes Antwortformat Effekte auf die Korrektheit und Ausführlichkeit von Testantworten hat. Dabei wird ein Videostimulus in einer Testaufgabe präsentiert, der eine spontane, mündliche Antwort fordert. Diese mündlichen Antworten werden mit schriftlichen Antworten der ansonsten identischen Testumgebung verglichen. Im Folgenden wird zunächst der theoretische Hintergrund dargelegt, bevor dann die Vorstellung der konkreten Forschungsfrage und der Methode erfolgt und die Ergebnisse präsentiert und diskutiert werden. Zunächst wird auf die zu messende DaZ-Kompetenz und performanznahe Kompetenzmessungen eingegangen. Es wird erläutert, wie diese erfolgt, warum Performanznähe notwendig ist und welche Herausforderungen diese Art von Messung mit sich bringt. Der zweite Teil in diesem Abschnitt widmet sich dem Antwortformat. Es wird herausgearbeitet, welche Formate wie und mit welchen Ergebnissen in Studien verwendet werden

und warum in dieser Studie ein mündliches Antwortformat im Vergleich zu einem schriftlichen erprobt werden soll.

Theoretischer Hintergrund

Performanznahe Messung von DaZ-Kompetenz bei (angehenden) Lehrkräften

Das Konstrukt der DaZ-Kompetenz lässt sich in einem Strukturmodell gliedern: DaZ-Kompetenz umfasst darin die Dimensionen *Fachregister (Fokus auf Sprache)*, *Mehrsprachigkeit (Fokus auf Lernprozess)* und *Didaktik (Fokus auf Lehrprozess)* (*DaZKom*-Strukturmodell, Köker et al., 2015). Alle Dimensionen können sich wiederum in Subdimensionen und Facetten auffächern. Deutsch als Zweitsprache ist kein Unterrichtsfach, sondern ist generisch zu betrachten: Es wird in allen Unterrichtsfächern zum Gegenstand (Köker et al., 2015). So wird DaZ-spezifisches Fachwissen benötigt, welches dann gemeinsam mit dem Fachwissen und dem fachdidaktischen Wissen des jeweiligen Faches (z.B. Mathematik) vermittelt wird. Analog zu dem Strukturmodell wurde auch ein *DaZKom*¹ Paper-Pencil-Test entwickelt, der die kognitiven Facetten von DaZ-Kompetenz valide misst (Ehmke et al., 2018). Jedoch erfasst dieser kognitive Facetten von DaZ-Kompetenz, losgelöst von tatsächlichen Unterrichtssituationen. Mit dem performanznahen Testinstrument wollen wir eine Testsituation schaffen, die nicht explizites DaZ-Fachwissen abfragt, sondern die Probandinnen und Probanden dazu auffordert, ihr DaZ-Fachwissen in Reaktionen auf konkrete Unterrichtssituationen anzuwenden. Dafür eignet sich ein Papier- und Bleistift-Test nicht (Blömeke & König et al., 2015). Vielmehr werden Video-Vignetten als Stimuli genutzt, die eine realitätsnahe Abbildung von Klassenraumsituationen mit ihren komplexen Interaktionen zwischen allen Akteurinnen und Akteuren ermöglichen (Kersting, 2008; Lindmeier, 2013; Casale et al., 2016; Rutsch & Dörfler, 2018).

In der Professionsforschung zu Lehrkräften wird davon ausgegangen, dass professionelle Kompetenzen individuelle, erlernbare Dispositionen sind, die in entsprechenden Unterrichtssituationen aktiviert werden. Wissen ist dabei neben Motivation und Einstellungen eine Teilkomponente der Dispositionen (Blömeke, Gustafson & Shavelson, 2015; Hepberger et al., 2019; Heins & Zabka, 2019). In unserem Verständnis von Kompetenz folgen wir dieser Annahme und nutzen das Modell der *Kompetenz als Kontinuum* (Blömeke, Gustafson & Shavelson, 2015). Die Wissensfacetten der DaZ-Kompetenz (dargestellt im *DaZKom*-Strukturmo-

¹ Professionelle Kompetenzen angehender LehrerInnen (Sek I) im Bereich Deutsch als Zweitsprache (2015 – 2017). Gefördert vom BMBF (Cluster KoKoHs I), Antragstellerinnen und Antragsteller sind Barbara Koch-Priewe, Anne Köker, Udo Ohm und Timo Ehmke

dell) werden (neben u. a. Fachwissen, pädagogischem Wissen und Überzeugungen) als Teil der Dispositionen der Lehrkräfte betrachtet. In einer konkreten Situation entscheiden die vorhandenen Dispositionen und situationspezifischen Fähigkeiten darüber, ob DaZ-relevante Aspekte wahrgenommen werden, wie diese beurteilt werden und welche Handlungsentscheidungen getroffen werden. In der Handlungsfähigkeit einer Person in der konkreten Situation, *in der Performanz*, zeigt sich die Kompetenz einer Person (Blömeke, Gustafson & Shavelson, 2015). Die situationspezifischen Fähigkeiten werden deshalb auch als „Mediatoren“ (Blömeke, Gustafson & Shavelson, 2015, S. 8) zwischen Wissen und Können bezeichnet.

Die situationspezifischen Fähigkeiten nutzen wir als Basis für die Itementwicklung: Die Probandinnen und Probanden müssen in dem Test beantworten, was sie in der gezeigten Unterrichtssituation wahrnehmen und wie sie anstelle der Lehrkraft handeln würden. Die wissensbasierte Wahrnehmung ist die Grundlage für eine erfolgreiche kompetente (Unterrichts-)handlung. Die Testperformanz ist dann Indikator für vorhandene DaZ-Kompetenz. Die Konstrukte der situationspezifischen Fähigkeiten werden als ähnliche Konzepte auch in anderen Studien verwendet, bspw. in solchen, die *professionelle Wahrnehmung* (*professional vision*; Goodwin, 1994) erfassen. Dieses beschreibt wie Personen Situationen, bezogen auf ihre Profession, beobachten und interpretieren (Goodwin, 1994; Seidel et al., 2010). König & Lebens (2012) kommen in ihrer Studie zur professionellen Wahrnehmung zu dem Ergebnis, dass erfahrene Lehrkräfte Unterrichtssituationen genauer und holistischer wahrnehmen als angehende Lehrkräfte. Letztere wiederum schneiden jedoch besser ab, wenn es darum geht, Handlungen zu rechtfertigen. Seidel & Stürmer (2015) betrachten professionelle Wahrnehmung als „anwendungsbezogenen Aspekt professioneller Kompetenz“ (Seidel & Stürmer, 2015, S. 355) und nutzen Video-Vignetten und Rating Items in ihrem Instrument *Observer*, um die Struktur dieser zu erfassen. Sie stützen sich dabei auf die Konzepte des *Noticing/Selective Attention* und des *Knowledge-based Reasoning* (Seidel, Blomberg & Stürmer, 2010; Es & Sherin, 2002). Ersteres beschreibt die Wahrnehmung, welche relevante Unterrichtssituationen herausfiltert, um dann mit letzterem eine theoriebasierte Interpretation der Beobachtung vorzunehmen und Handlungsentscheidungen zu treffen (Es & Sherin, 2002.). Die Fähigkeit, die wahrgenommenen relevanten Situationen in einen breiteren Kontext einzuordnen (und nur so kompetente Handlungsentscheidungen treffen zu können), unterscheidet zwischen Novizinnen- bzw. Novizen-Lehrkräften, welche meist eine reine Beschreibung dessen liefern, was sie beobachten, und erfahrenen Lehrkräften. Diese wiederum stellen eine Verbindung zu ihnen bekannten Konzepten und Theorien über Lehren und Lernen her

und können dadurch schneller und präziser darauf eingehen (u. a. Hoth et al. 2018, Blömeke & König et al., 2015; Lindmeier, 2013; Bromme, 2014).

Wahrnehmung wird auch aus der kognitionspsychologischen Perspektive nicht als reiner reizgesteuerter Prozess betrachtet, sondern immer auch als ein wissensbasierter Prozess (Heins & Zabka, 2019) – dies stützt die obigen Annahmen: (Erfahrungs-)wissen, welches in der Wahrnehmung einer Situation aktiviert wird, steuert, was Personen wahrnehmen. Auch in Bezug auf die Handlungsentscheidung (*decision-making*) können Annahmen aus der Kognitionspsychologie herangezogen werden: Der Ansatz der *Dual-Process-Theories* (Gawronski & Creighton, 2013) geht davon aus, dass Entscheidungen entweder kontrolliert oder automatisch getroffen werden. Dabei sind letztere schnelle Entscheidungen, die wenig kognitive Ressourcen benötigen (Gawronski & Creighton, 2013), da es sich um intuitive Prozesse handelt, die auf bereits gesammelte Daten in Form von z. B. Erfahrungen zurückgreifen. Deshalb muss kein bewusstes Abwägen aller Konsequenzen mehr erfolgen (Kahneman, 2013). Um Performanznähe zu schaffen und intuitive Entscheidungen zu fordern, soll in dieser Studie ein mündliches Antwortformat verwendet werden, mit dem Ziel, eine spontane und unmittelbare Reaktion auf eine videobasierte Testsituation zu erreichen. Ein Testformat mit Videos und der Aufzeichnung von mündlichen Antworten macht die Nutzung von Technologie unabdingbar und wird deshalb computerbasiert umgesetzt.

Mündliche und schriftliche Antworten in Testumgebungen und Prüfungsformaten

Obwohl Moduseffekte in verschiedenen Studien-Designs untersucht werden, steht meist der Modus der Darbietung im Vordergrund und weniger die Antwortmodi. Es ist in der Forschung also weitgehend unklar, welche Moduseffekte durch das veränderte Antwortformat *mündlich* versus *schriftlich* bei gleicher Testumgebung erzielt werden. Da die Forschungslage noch sehr begrenzt ist, werden im Folgenden Studien zu mündlichen und schriftlichen Prüfungsformaten sowie auch Studien zu mündlichen und schriftlichen Sprachproduktionen im Allgemeinen herangezogen.

Studien, in denen mündliche im Vergleich zu schriftlichen Formaten vorlagen, kamen zu folgenden Ergebnissen: Pugalee (2004) untersuchte schriftliche und mündliche Dokumentationen von Problemlöseprozessen bei Studentinnen und Studenten. Dabei waren diejenigen, die ihren Problemlöseprozess niederschrieben, signifikant erfolgreicher als diejenigen, die ihre Beschreibung lediglich mündlich äußerten. In seiner theoretischen Rahmung geht Pugalee zum einen darauf ein, dass das Artikulieren von Ideen ein umfassendes Verständnis dieser voraussetzt. Niederschreiben dagegen versteht er als bedeutungsschaffend und als Unterstützung der metakognitiven

Prozesse (Pugalee, 2004). Für mündliche Äußerungen muss also die Idee schon gänzlich durchdrungen und reflektiert sein, wohingegen Schreiben einen konstruierenden Prozess darstellt. Auch Molitor-Lübbert (2002) geht davon aus, dass Denken und Schreiben sich wechselseitig beeinflussen. Scheuermann (2016) schreibt, dass sich durch Schreiben aus implizitem Wissen explizites Wissen entwickeln kann (*Schreibdenken*). Im Rückblick auf die bereits erwähnten kognitionspsychologischen Annahmen aktiviert Sprechen offenbar intuitive, automatische Prozesse. Schreiben hingegen erfolgt kontrolliert und nutzt viele kognitive Ressourcen (Evans, 2008; Samuels, 2009). Dafür spricht auch, dass in gesprochener Sprache im Vergleich zur geschriebenen Sprache weniger Zeit verfügbar ist, um über die genaue Wortwahl und passenden Ausdrücke nachzudenken (Poole & Field, 1976). Dies führt zu weniger elaboriertem Sprachgebrauch, zu Wiederholungen und Dopplungen, aber auch zu mehr Flexibilität (Huxham et al., 2012). Schreiben ist dagegen ein bewusster, analytischer Prozess, in dem es möglich ist, Gedanken strukturiert und überlegt darzulegen (Pugalee, 2004).

Die Auswirkung der mündlichen bzw. schriftlichen Antwortformate unterscheidet sich jedoch offenbar in Abhängigkeit der gemessenen Konstrukte. Pugalee (2004) untersuchte Problemlösefähigkeit, Ryan et al. (2014) hingegen fokussieren Gedächtnistests: Sie verglichen mündliche und schriftliche Antworten in ihrer Studie zu einem Gedächtnistest. Dieser konnte ursprünglich nur mündlich und dadurch nur von Probandinnen und Probanden beantwortet werden, die in der Lage waren, ihre Antworten zu verbalisieren. Um eine Teilnahme auch für Probandinnen und Probanden mit Sprachproblemen zu ermöglichen, wurden Effekte des Antwortformats in einer randomisierten Stichprobe gemessen. Dabei wurden mündliche und schriftliche Antworten verglichen, mit dem Ergebnis, dass sich zwischen beiden Gruppen, mündlich und schriftlich, keine bedeutsamen Unterschiede in den Leistungen ergaben. Thorburn und Collins (2006) kamen in ihrer Studie zu mündlichen und schriftlichen Prüfungsformaten zu dem Ergebnis, dass die meisten Schülerinnen und Schüler ihrer Stichprobe in schriftlichen Prüfungen bessere Leistungen erreichten als in mündlichen Prüfungen. Dies bestätigt die oben beschriebenen Annahmen, dass Schreiben und Denken verwobene Prozesse sind und insbesondere in Bezug auf Prüfungsformate implizites Wissen durch Schreiben aktiviert wird.

In Bezug auf die technische Umsetzung könnte das mündliche Antworten auf Testfragen in einer computerbasierten Testumgebung, ohne direkte verbale Interaktion zwischen Interviewerin bzw. Interviewer und der Testperson, eine ungewohnte Anforderung an die Befragten stellen. Andererseits gehören heutzutage Sprachnachrichten auf neuen Kommunikationsplattformen (z. B.

WhatsApp®), welche ohne Gesprächspartnerin bzw. ohne Gesprächspartner im direkten Gegenüber entstehen, sondern sich diesem über Distanz widmen, zur Alltagskommunikation und sind dadurch nicht mehr so ungewöhnlich, wie noch einige Jahre zuvor (König & Hector, 2017). Dies führt zu der Annahme, dass Audioaufzeichnungen der eigenen Antwort dem Großteil der Befragten vermutlich keine Probleme bereiten dürften.

Der Forschungsüberblick zeigt, dass die Befundlage zu den Effekten von mündlichen vs. schriftlichen Antwortformaten bei Testaufgaben sehr begrenzt ist. Er lässt aber vermuten, dass mündliche versus schriftliche Antwortformate theoretisch unterschiedliche kognitive Prozesse ansprechen und deshalb ein mündliches Antwortformat einen Unterschied machen könnte, der für die Testentwicklung von Bedeutung wäre. Diese Studie soll daher dazu beitragen, die Forschung zu Performanznahmen Messen in videobasierten Testumgebungen, um Ergebnisse zu unterschiedlichen Antwortformaten zu ergänzen.

Forschungsfragen

Diese Studie untersucht zwei Fragestellungen:

1. Welche Effekte hat das Antwortformat, schriftlich versus mündlich, bei der Messung von DaZ-Kompetenz bei angehenden Lehrkräften in Bezug auf die Korrektheit und die Ausführlichkeit (gemessen an der Wörteranzahl) der Antworten?

Anhand der genannten Befunde in der theoretischen Rahmung kann vermutet werden, dass das Antwortformat einen Einfluss auf das Antwortverhalten hat. Testaufgaben mit mündlichem Antwortformat könnten sich demnach für angehende Lehrkräfte (Studierende) als schwieriger erweisen als Testaufgaben mit schriftlichem Antwortformat, da durch die Unmittelbarkeit und Spontanität ein Zeitdruck erzeugt wird, der intuitive kognitive Prozesse anspricht. Dies setzt umfassende DaZ-Kompetenz voraus, damit schnell bereits vorhandenes (Erfahrungs-)wissen aktiviert werden kann. Das schriftliche Format hingegen erlaubt ein tieferes Durchdenken und regt damit kontrollierte kognitive Prozesse an. Schriftliche Antworten können ggf. auch von dem in der Theorie angesprochenen *Schreibdenken* profitieren. Umgekehrt liegt die Vermutung nahe, dass DaZ-erfahrene Testpersonen mündliche Antwortformate leichter lösen können, da sie eben über die nötige Kompetenz verfügen, die intuitives, schnelles Handeln möglich macht. Hinsichtlich der Ausführlichkeit der Antworten ist zu vermuten, dass mündliche Antworten länger bzw. bezogen auf die Wörteranzahl umfangreicher ausfallen als schriftliche Ant-

worten, da die Antwort nicht vorstrukturiert wird, sondern spontan erfolgt.

Mit der zweiten Fragestellung soll geprüft werden, ob sich differentielle Effekte durch Antwortformate in Abhängigkeit von DaZ-Lerngelegenheiten der Studierenden nachweisen lassen:

2. Welche Effekte hat das Antwortformat, schriftlich versus mündlich, in Bezug auf die Korrektheit und die Ausführlichkeit der Antworten, wenn das Ausmaß an DaZ-Lerngelegenheiten der Probandinnen und Probanden und deren Interaktion mit dem Antwortformat kontrolliert wird?

In Bezug auf Unterschiede zwischen Testpersonen mit wenigen DaZ-Lerngelegenheiten und solchen mit vielen DaZ-Lerngelegenheiten sollen drei Hypothesen geprüft werden. (1) Es kann davon ausgegangen werden, dass sich mit zunehmenden Lerngelegenheiten im Bereich DaZ die DaZ-Kompetenz messbar erhöhen sollte. Studentinnen und Studenten mit vielen DaZ-Lerngelegenheiten sollten demnach unabhängig vom Antwortmodus höhere Lösungswahrscheinlichkeiten erzielen als Studentinnen und Studenten mit wenigen DaZ-Lerngelegenheiten. (2) Auch die mittlere Antwortlänge sollte für Studentinnen und Studenten mit vielen DaZ-Lerngelegenheiten höher ausfallen als für jene mit wenigen DaZ-Lerngelegenheiten, da letztere möglicherweise häufiger keine Antwort geben können. (3) Betrachtet man angehende Lehrkräfte wie in dieser Stichprobe, so könnte das mündliche Antwortformat eine spezifische Hürde darstellen. Für diese Personengruppe könnte eine schriftliche Antwort ein Vorteil sein und in einer höheren Lösungsrate resultieren, da die Probandinnen und Probanden Zeit haben, ihre Antwort zu formulieren und ihr Vorwissen abzurufen.

Methoden

Die experimentelle Studie ist als between-subject-design angelegt. Die unabhängige Variable umfasst den Faktor Antwortformat, bei dem zwischen einem mündlichen und einem schriftlichen Antwortformat unterschieden wird. Abhängige Variablen sind die Testleistung in einem performanzorientierten DaZ-Kompetenztest und die Ausführlichkeit der Antworten gemessen an der Wortanzahl. Alle teilnehmenden Probandinnen und Probanden wurden zufällig auf diese beiden Testbedingungen zugewiesen.

Durchführung der Studie und Stichprobe

Die Datenerhebung erfolgte in Lehrveranstaltungen während des Semesters einer deutschen Universität. Die Stich-

Tabelle 1. Deskriptive Stichprobenkennwerte differenziert nach Antwortmodus

Personenmerkmale	Modus	Ausprägung	Häufigkeit	Gesamt
Geschlecht	Mündlich	weiblich	40	44
		männlich	4	
	Schriftlich	weiblich	32	36
		männlich	14	
Erstsprache	Mündlich	Deutsch	37	40
		Andere	3	
	Schriftlich	Deutsch	44	46
		Andere	2	
Auslandsaufenthalt ¹⁾	Mündlich	Ja	26	44
		Nein	18	
	Schriftlich	Ja	22	44
		Nein	24	
Studienfach Deutsch	Mündlich	Ja	29	39
		Nein	10	
	Schriftlich	Ja	30	40
		Nein	10	

¹⁾ Haben Sie bereits länger als drei Monate am Stück im Ausland verbracht?

Tabelle 2. Übersicht über deskriptive Stichprobenkennwerte

Personenmerkmale	Modus	M	SD	N	Gesamt
Alter	Mündlich	24.11	3.47	44	90
	Schriftlich	23.85	3.25	46	
Abiturdurchschnitts- note	Mündlich	2.39	0.62	43	88
	Schriftlich	2.44	0.51	45	
Abiturnote Deutsch	Mündlich	10.44	2.66	41	83
	Schriftlich	10.79	2.19	42	
DaZ- Lerngelegenheiten	Mündlich	1.60	0.47	42	86
	Schriftlich	1.70	0.61	44	

probe (N = 90) umfasst Lehramtsstudierende aller Unterrichtsfächer, sowohl des Bachelorstudiengangs Lehren und Lernen (B.A.) als auch Masterstudierende des Studiengangs Lehramt an Grundschulen (M.Ed.), Lehramt an Haupt- und Realschulen (M.Ed.) oder Lehramt an Berufsbildenden Schulen, mit der Fachrichtung Sozialpädagogik oder Wirtschaftswissenschaften (M.Ed.) im zweiten, vierten oder sechsten Fachsemester. Die Probandinnen und Probanden verteilten sich zufällig auf die beiden Antwortmodi: N = 44 mündlich (48,9%) und N = 46 schriftlich (51,1%). Die Tabellen 1 und 2 zeigen die deskriptiven Stichprobenkennwerte hinsichtlich zentraler Personenmerkmale. Die Stichprobenzuweisung auf die beiden Ex-

perimentalbedingungen erfolgte zufällig. Entsprechend lassen sich keine statistisch bedeutsamen Unterschiede zwischen beiden Gruppen feststellen. Lediglich bezüglich des Geschlechts sind männliche Probanden etwas häufiger in der schriftlichen Bedingung vertreten.

Abhängige Variable: DaZ-Kompetenz und Ausführlichkeit der Antworten

Um die DaZ-Kompetenz der Lehramtsstudierenden zu erfassen, wurde ein computerbasiertes Testinstrument eingesetzt. Der DaZ-Kompetenztest enthält drei Videosequenzen, die jeweils eine DaZ-relevante videographierte Unterrichtssituation zeigen. Zu jeder Videosequenz wurden zwei Testaufgaben formuliert. Der theoretische Hintergrund legt nahe, dass zunächst die Wahrnehmung einer konkreten Situation erfolgt, welche in Kontext eingeordnet wird und dann erst eine Handlungsentscheidung erfolgt. Dieser Theorie folgend, fragte ein Item danach, was die Probandinnen und Probanden in der Videosequenz wahrnahmen (Wahrnehmung/*Perception*, P). Das zweite Item fragte, welche DaZ-relevante Handlungsentscheidung die Testpersonen treffen würden (Handlung/*Decision making*, D). Tabelle 3 zeigt eine Übersicht der drei Videos, ihren Inhalten und den jeweiligen dazugehörigen Testitems. Insgesamt ergeben sich sechs Items, die in einer Variante mit mündlichem Antwortformat und in einer Variante mit schriftlichem Antwortformat vorliegen. Die mündlichen Antworten wurden als Audiodateien aufgezeichnet, die schriftlichen Antworten in Texteingang

befenstern gespeichert. Der Durchführung des Tests ging eine standardisierte Testinstruktion durch die Testleiterinnen und Testleiter voran. Die Beantwortung der Testitems wurde darin konkret erläutert. Die Probandinnen und Probanden sahen jedes Video einmal an, um dann beide Items nacheinander zu bearbeiten. Probandinnen und Probanden durften keine Notizen machen und wurden angehalten, so spontan wie möglich zu antworten. Die mündlichen Antworten der Probandinnen und Probanden wurden automatisch aufgezeichnet. Die Länge der Audioaufzeichnung war nicht zeitlich begrenzt. Die schriftlichen Antworten wurden durch die Probandinnen und Probanden in entsprechende Textfelder des computerbasierten Tests eingetippt. Korrektheit und die Ausführlichkeit der Antworten auf die sechs Items sind die abhängigen Variablen in dieser Studie (Tabelle 3).

Die Korrektheit wurde anhand der Anzahl der durch die Probandinnen und Probanden genannten richtigen Argumente gemessen. Die Kodierung basiert auf Aussagen von Expertinnen und Experten, die mithilfe von Expertinnen- bzw. Experteninterviews zu den Unterrichtsvideos generiert wurden (vgl. dazu auch Klahn, 2016; Hecker & Nimz, 2020). Für jedes Item wurden fünf Argumente notiert, die von Expertinnen und Experten als relevant erachtet wurden. Sobald ein Argument durch die Testperson genannt wurde, wurde die Antwort als richtig gewertet.

Die Ausführlichkeit der schriftlichen Antworten wurde anhand der Wortanzahl erfasst. Für die mündlichen Antworten wurden zuerst Wortlaut-Transkripte erstellt und

Tabelle 3. Übersicht über die Items, deren Inhalt und die Aufgabenstellung

Videosequenz	Inhalt	Aufgabenstellung	Kategorie
Video 1 (Dauer: 1:09 min)	7. Klasse einer Stadtteilschule, Natur & Technik, Einführungsstunde zum Thema Wasser, Schülerinnen und Schüler diskutieren in einer Gruppe das Thema Wasserfall	Bitte beschreiben Sie, welche Aspekte Ihnen in dieser Unterrichtssequenz in Bezug auf sprachlich sensiblen Unterricht auffallen.	Wahrnehmung / Perception (P)
		Bitte beschreiben, wie Sie in dieser Unterrichtssituation als Lehrkraft in Bezug auf die sprachliche Heterogenität im Unterricht handeln würden.	Handlung / Decision making (D)
Video 2 (Dauer: 5:17 min)	7. Klasse einer Stadtteilschule, Natur & Technik, zweiter Teil der Einheit Wasser, eine Gruppe von Schülerinnen und Schüler soll gemeinsam einen Text zum Thema Klärwerk lesen	Denken Sie an die Schwierigkeiten in sprachlich heterogenen Lerngruppen. Was fällt Ihnen zu diesem Thema in diesem Video auf?	Wahrnehmung / Perception (P)
		Angenommen, Sie würden als Lehrkraft in dieser Klasse unterrichten. Wie würden Sie mit der Situation an dem Gruppentisch umgehen?	Handlung / Decision making (D)
Video 3 (Dauer: 1:49 min)	5. Klasse einer integrierten Gesamtschule, die Schülerinnen und Schüler sollen 4er Gruppen bilden und sitzen an 5er Tischen. Sie diskutieren darüber, wer von ihnen ein_e „Ausländer_in“ sei und legen die Zuordnung als Kriterium für die Gruppenbildung fest.	Welche Aspekte fallen Ihnen an dem Verhalten der Lehrkraft auf?	Wahrnehmung / Perception (P)
		Wie würden Sie an Stelle der anwesenden Lehrkraft handeln?	Handlung / Decision making (D)

anschließend die Wortanzahl der Transkripte bestimmt. Als Maß für die Reliabilität der DaZ-Kompetenzskala wurde McDonald's Omega berechnet. Die Reliabilität beträgt $\omega = 0.786$.

Unabhängige Variablen:

DaZ-Lerngelegenheiten und Personenmerkmale

Neben den videobasierten DaZ-Items wurde ein Fragebogen zu soziodemographischen Daten eingesetzt, welcher Fragen zum Geschlecht, zur Muttersprache, zur Anzahl der studierten Semester sowie zum Studienfach beinhaltet. Als unabhängige Variable wurden in dieser Studie außerdem DaZ-relevante Lerngelegenheiten abgefragt. Es wird davon ausgegangen, dass DaZ-Kompetenz gelernt werden kann, zum einen mit wissensbasierten Lerngelegenheiten, zum anderen aber auch mit erfahrungsbasierten Lerngelegenheiten. Es wurden dazu DaZ-relevante Handlungen abgefragt, die im Rahmen von Lehrveranstaltungen bereits durchgeführt wurden (fünfstufige Likert-Skala, 1 = *gar nicht* bis 5 = *in mehreren Lehrveranstaltungen*). Die acht Items umfassende Skala hat nach McDonald's Omega eine Reliabilität von $\omega = 0.921$.

Auswertungsmethode

Für die Beantwortung der ersten Forschungsfrage wurden Mittelwertvergleiche für unabhängige Stichproben mittels T-Testanalysen durchgeführt. Um die zweite Forschungsfrage zu beantworten, wurden zwei latente Strukturgleichungsmodelle mit MPlus 7.3 (Muthén & Muthén, 2014) berechnet. Aufgrund der begrenzten Stichprobengröße wurden für die abhängigen Variablen DaZ-Kompetenz und Wörteranzahl zwei getrennte Modelle geschätzt. Aus testtheoretischen Gründen wurde ein Item aus den Berechnungen ausgeschlossen. Leere Antworten bei den Testaufgaben wurden als falsch kodiert.

Ergebnisse

Mit der ersten Fragestellung soll geprüft werden, inwieweit ein schriftliches versus mündliches Antwortformat, einen Effekt auf die Messung von DaZ-Kompetenz bei an-

gehenden Lehrkräften in Bezug auf die Korrektheit und die Ausführlichkeit der Antworten hat. Die mithilfe eines T-Tests ermittelten Gruppenunterschiede zeigen, dass sich hinsichtlich der Korrektheit der Antworten kein statistisch signifikanter Unterschied zwischen den mündlichen und schriftlichen Antworten festgestellt werden kann. Für beide Varianten des Antwortformats liegt die durchschnittliche Lösungswahrscheinlichkeit über die Items hinweg bei $M = 0.36$. Hinsichtlich der Ausführlichkeit fallen die mündlichen Antworten statistisch signifikant umfangreicher aus als die schriftlichen Antworten (Tab. 4). Während die mündlichen Antworten durchschnittlich 87 Wörter umfassen, sind es in der schriftlichen Variante nur 63 Wörter. Die manifeste Korrelation zwischen Korrektheit und Ausführlichkeit der Antworten beträgt $r = 0.49$.

Um die zweite Forschungsfrage zu beantworten, wurden zwei latente Strukturgleichungsmodelle geschätzt, welche den Einfluss der Faktoren *Antwortmodus*, *DaZ-Lerngelegenheiten* sowie einen Interaktionseffekt (*Antwortmodus* \times *DaZ-Lerngelegenheiten*) auf die gemessene DaZ-Kompetenz (Modell 1, Abb. 1) und die Wortanzahl (Modell 2, Abb. 2) untersuchen. Beide Modelle weisen einen akzeptablen Modellfit auf (Modell 1: $\chi^2(71) = 87,958$, RMSEA = 0.052, CFI = 0.972, TLI = 0.964; Modell 2: $\chi^2(82) = 104.369$, RMSEA = 0.055, CFI = 0.927, TLI = 0.907).

Modell 1 zeigt analog zum Mittelwertvergleich, dass sich kein Effekt des Antwortmodus auf die Testleistung zeigt, Allerdings erzielen Probandinnen und Probanden mit vielen DaZ-Lerngelegenheiten statistisch signifikant bessere Testergebnisse als Personen mit weniger DaZ-Lerngelegenheiten ($\beta = 0.37$, $p < 0.05$). Ein Interaktionseffekt lässt sich wiederum nicht nachweisen.

In Modell 2 zeigt sich, dass der Moduseffekt (mündliches Antwortformat = ausführlichere Antworten) auch bei Kontrolle der DaZ-Lerngelegenheiten bestehen bleibt ($\beta = -0.46$, $p < 0.05$). Ein positiver Zusammenhang von DaZ-bezogenen Lerngelegenheiten auf die Ausführlichkeit der Antwort zeigt sich nicht. Die vermutete Interaktion zwischen beiden Variablen bleibt (knapp) unterhalb der statistischen Signifikanzgrenze (5% beidseitige Testung).

Tabelle 4. Mittelwerte und Standardabweichungen der Korrektheit der Antworten und der Ausführlichkeit der Antworten (Wortanzahl) differenziert nach Modus

	Modus								
	Gesamt		Mündlich		Schriftlich				
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i> (74)	<i>p</i>	<i>Cohen's d</i>
Korrektheit	0.36	0.25	0.36	0.27	0.36	0.23	-0.02	ns	0.00
Wortanzahl	74.24	32.86	87.09	37.23	62.67	23.31	-3.46	< 0.05	-0.81

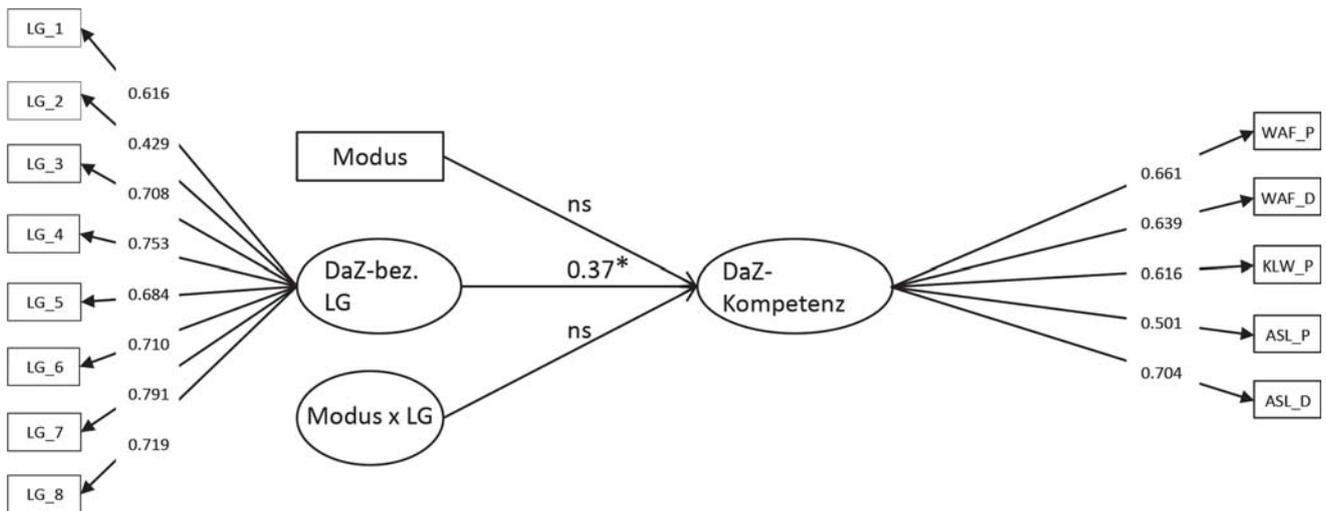


Abbildung 1. Vorhersage der DaZ-Kompetenz durch Testmodus, DaZ-bezogenen Lerngelegenheiten und deren Interaktion. Modus: 0 = mündliches Antwortformat, 1 = schriftliches Antwortformat, DaZ-bez. LG = DaZ-bezogene Lerngelegenheiten (z-standardisiert), Modus x LG = Interaktionsterm.

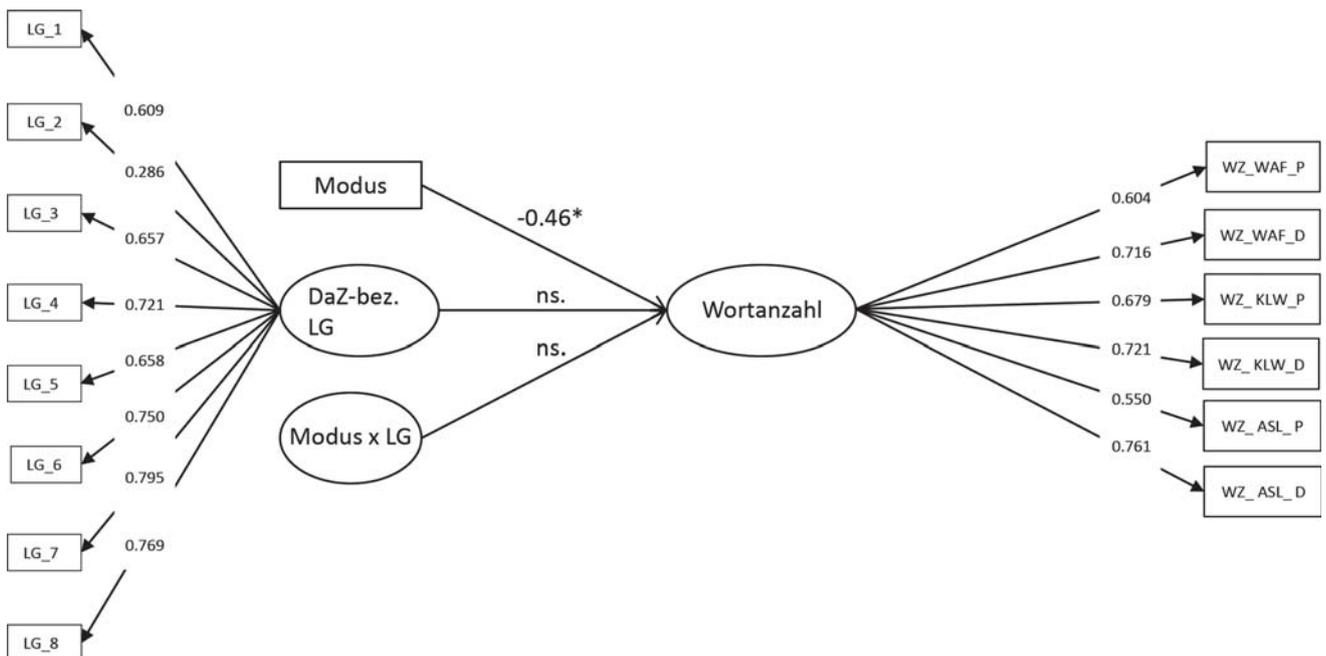


Abbildung 2. Vorhersage der Antwortlänge (gemessen an der Wortanzahl) durch Testmodus, DaZ-bezogenen Lerngelegenheiten und deren Interaktion. Modus: 0 = mündliches Antwortformat, 1 = schriftliches Antwortformat, DaZ-bez. LG = DaZ-bezogene Lerngelegenheiten (z-standardisiert), Modus x LG = Interaktionsterm

Diskussion

Ziel dieser Studie war es, zwei unterschiedliche Antwortformate (schriftlich versus mündlich) in einer computerbasierten Testumgebung gegeneinander zu testen. Es wurde zum einen untersucht, ob das variierte Antwortformat einen Effekt auf die Korrektheit und auf die Ausführlichkeit der Antworten hat. Zum anderen wurde der Einfluss von DaZ-be-

zogenen Lerngelegenheiten auf die Testergebnisse in den unterschiedlichen Modi untersucht bzw. auf Interaktionseffekte geprüft. Die Studie kommt zu folgenden Ergebnissen:

1. Die Korrektheit bleibt von dem veränderten Antwortformat unbeeinflusst. Bezüglich der Ausführlichkeit, gemessen an der Wortanzahl, zeigt sich, dass mündliche Antworten statistisch signifikant ausführlicher sind als schriftliche Antworten.

2. Testpersonen mit vielen DaZ-bezogenen Lerngelegenheiten erreichen eine höhere DaZ-Kompetenz als solche mit wenigen. Moduseffekte lassen sich unter Kontrolle der DaZ-Lerngelegenheiten nur für die Ausführlichkeit der Antworten nicht aber für die Korrektheit nachweisen.
3. Es konnten keine Interaktionseffekte (Modus \times DaZ-bezogene Lerngelegenheiten) nachgewiesen werden.

Wir hatten erwartet, dass Testaufgaben mit einem mündlichen Antwortformat eher von Probandinnen und Probanden höherer Kompetenz bewältigt werden können, da dies eine automatische, intuitive Antwort fordert. Umgekehrt sollte in dieser Stichprobe das schriftliche Antwortformat von Vorteil sein, da es sich um eine Stichprobe mit Studierenden handelt (die dennoch unterschiedlich viele Lerngelegenheiten hatten). Unterschiede zwischen den Antwortmodi zeigen sich in dieser Studie jedoch nicht. In dieser Studie bedeutet dies, dass weder das eine, noch das andere Antwortformat bestimmte Testpersonen bevorteilt oder benachteiligt. Es kann jedoch vermutet werden, dass Testaufgaben mit mündlichem Antwortformat sich als leichter erweisen als Testaufgaben mit einem schriftlichen Antwortformat, wenn Handlungskompetenz vorliegt (z.B. bei erfahrenen Lehrkräften). Die unterrichtsnahen Anforderungen sprechen schnelle, intuitive Prozesse und Handlungsrouninen an, während das schriftliche Format durch die für das Unterrichten unübliche Anforderung eine zusätzliche Schwierigkeit darstellt. Dies könnte ein Hinweis darauf sein, dass in einer Stichprobe mit erfahrenen Lehrkräften ein mündliches, spontanes Antwortformat eben genau diese identifiziert. Dies ist in zukünftigen handlungsnahen Testumgebungen mit erfahrenen Lehrkräften zu prüfen.

Hinsichtlich der Ausführlichkeit der Antworten gibt es einen eindeutigen Befund. Die Antwortlänge gemessen an der Wortanzahl ist beim mündlichen Antworten bedeutsam größer als beim schriftlichen Antwortformat. In den theoretischen Überlegungen wurde bereits gezeigt, dass Schreiben es erlaubt, Gedanken vorher zu strukturieren und mündliche Antworten im Gegensatz dazu spontaner, aber auch weniger strukturiert sind. Durch die automatisch laufende Audioaufnahme haben sich Probandinnen und Probanden vermutlich eher gedrängt gefühlt, zu antworten, auch wenn sie keine Antwort wussten, als in dem schriftlichen Format, in welchem sie dann ggf. eher gänzlich auf eine Antwort verzichtet haben.

In dieser Studie ist das Ergebnis der Kompetenzmessung bei Studierenden des Lehramts durch das Antwortformat unbeeinflusst. Es sind weitere Studien erforderlich, um zu prüfen, inwieweit sich auch in anderen Testdomänen und anderen Stichproben kein Moduseffekt für mündliche Antwortformate zeigt. Entsprechend sollte eine vergleichbare Studie auch noch einmal mit

Lehrkräften wiederholt werden, die langjährige Lehrerfahrung im Umgang mit DaZ-Schülerinnen und Schülern haben und bei denen die Kompetenz höher ist, als in dieser Studie, in der lediglich einige Studierende mehr DaZ-Lerngelegenheiten vorweisen konnten, als andere. Des Weiteren könnte in zukünftigen Studien geprüft werden, inwieweit die Beantwortung von nur einer der beiden Fragen (Wahrnehmung einer Situation und Treffen einer Handlungsentscheidung) zu einem Stimulus einen Effekt auf die Antworten zeigen. Möglicherweise ist in unserer Studie die Reaktion der Probandinnen und Probanden insbesondere auf die zweite Frage weniger spontan als intendiert, da ja bereits die erste Frage beantwortet wurde.

In zukünftigen Studien könnten auch noch einmal alternative Maße für die Ausführlichkeit der Antworten erprobt werden. Diese wurde in dieser Studie mit Hilfe der Wortanzahl ermittelt. Da es jedoch einen Unterschied zwischen Mündlichkeit und Schriftlichkeit gibt (z.B. im Hinblick auf Strukturierung und Wiederholungen), ist die Wortanzahl in den Modi möglicherweise nicht verlässlich vergleichbar, auch wenn die Aufnahmen lediglich nach Wortlaut transkribiert wurden.

Um die Messung von Kompetenzen möglichst performanznah zu realisieren, ist der Einsatz von Videos kombiniert mit einem mündlichen Antwortformat vielversprechend. Ein solches Testszenario erlaubt es, eine Situation sehr lebensnah zu präsentieren. Wir haben in dieser Studie angenommen, dass ein mündlicher Antwortmodus zu einer performanznäheren Messung beitragen kann. Es zeigen sich dabei keine Moduseffekte auf die Korrektheit der Antworten. Für die Testentwicklung kann damit die Schlussfolgerung gezogen werden, dass ein mündliches Antwortformat bei Studierenden nicht zu unfairen Testergebnissen führt, verglichen mit einem traditionellen schriftlichen Antwortformat. Beide Antwortformate realisieren in unserer Stichprobe vergleichbare Kennwerte bei den Kompetenzmessungen. Aus technischer Sicht ist ein mündliches Antwortformat mit Audiodaten jedoch mit erheblich höherer Komplexität und Aufwand in der Datenauswertung verbunden als die Verarbeitung von textbasierten Daten. Für Testnutzerinnen und -nutzer wird ein mündliches Antwortformat jedoch möglicherweise als authentischer wahrgenommen und ist mit höherer Akzeptanz verbunden als ein schriftliches. Dies müsste allerdings in nachfolgenden Studien noch geprüft werden.

Abschließend noch zwei generelle Problematiken, die in Bezug auf performanznahes Messen diskutiert werden: In den theoretischen Überlegungen wurde deutlich, dass erfahrene Lehrkräfte Situationen holistischer wahrnehmen, angehende Lehrkräfte jedoch ihre Handlungen eher rechtfertigen können (König & Lebens, 2012). In ihrer Diskussion beziehen sich König und Lebens dabei auf Neu-

weg, welcher von einer Transformation des Wissens ausgeht. Das Wissen erfahrener Lehrkräfte ist nach Neuweg so intuitiv, dass sie die Begründung ihrer situationsspezifischen Handlung nicht mehr explizit verbalisieren können (Neuweg, 2014). König und Lebens drücken es folgendermaßen aus: „Der Transfer von den Wissensinhalten hin zum situationsspezifischen Handeln wäre bereits vollzogen, wodurch sich die Antworten auf Fragebogenitems einem sprachlichen Repräsentationsmodus entziehen“ (König & Lebens, 2012, S.23). Sieland & Jordaan (2019) diskutieren in ihrem Aufsatz eine weitere Problematik: Sie fokussieren die Interaktionen, in die Lehrkräfte tagtäglich eingebunden werden, welche auf verschiedenen Ebenen viele Ressourcen kosten. Die Grundbedürfnisse von Lehrkräften würden durch die hohe Anzahl an Interaktionen stark in Anspruch genommen, wodurch Stress ausgelöst würde, was wiederum das Risiko für unprofessionelle Reaktionen erhöhen würde (Sieland & Jordaan, 2019, S.8). Dies führt laut der Autorinnen und Autoren ggf. zu einer „defizitären Performanz trotz eigentlich vorhandener Kompetenzen“ (Sieland & Jordaan, 2019, S.9). Es sind also weitere Erprobungen des mündlichen Antwortformats in einer performanznahen Messung erforderlich, um mehr über das Wissen bzw. Können von erfahrenen Lehrkräften zu erfahren, um das Verhältnis von Kompetenz und Performanz zu untersuchen und auch, um die Rolle des Antwortformats in Testungen mit Testpersonen unterschiedlicher Expertise zu überprüfen.

Literatur

- Bauer, N. & Blasius, J. (2014). Methoden der Sozialforschung. Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-531-18939-0>
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie*, 223, 3–13. <https://doi.org/10.1027/2151-2604/a000194>.
- Blömeke, S., König, J., Suhl, U., Horth, J. & Döhrmann, M. (2015). Wie situationsbezogen ist die Kompetenz von Lehrkräften? Zur Generalisierbarkeit der Ergebnisse von videobasierten Performanztests. *Zeitschrift für Pädagogik*, 61 (3), 310–327.
- Borko, H. (2016). Methodological contributions to video-based studies of classroom teaching and learning: A commentary. *ZDM Mathematics Education*, 48 (1–2), 213–218.
- Bromme, R. (2014). *Der Lehrer als Experte. Zur Psychologie des professionellen Wissens* (Standardwerke aus Psychologie und Pädagogik, Reprints, Bd. 7, [Nachdr. der Ausg.] Münster: Waxmann.
- Casale, G., Strauß, S., Hennemann, T. & König, J. (2016). Wie lässt sich Klassenführungsexpertise messen? Überprüfung eines videobasierten Erhebungsinstruments für Lehrkräfte unter Anwendung der Generalisierbarkeitstheorie. *Empirische Sonderpädagogik*, 2, 119–139.
- Ehmke, T., Hammer, S., Köker, A., Ohm, U. & Koch-Priewe, B. (2018). *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache*. Münster: Waxmann.
- Es, E. van & Sherin, M. (2002). Learning to Notice: Scaffolding New Teachers' Interpretations of Classroom Interactions. *Journal of Information Technology for Teacher Education*, 10 (4), 571–596.
- Evans, J. St. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgement and Social Cognition. *Annual Review of Psychology*, 59, 255–278.
- Gaudin, C. & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review*, 16, 41–67. <https://doi.org/10.1016/j.edurev.2015.06.001>
- Gawronski, B., & Creighton, L. A. (2013). Dual process theories. In D. E. Carlston (Ed.), *Oxford library of psychology. The Oxford handbook of social cognition* (pp. 282–312). New York, NY, US: Oxford University Press.
- Goodwin, C. (1994). Professional vision. *American anthropologist*, 96 (3), 606–633. <https://doi.org/10.1525/aa.1994.96.3.02a00100>
- Gold, B. & Holodynski, M. (2017) Using digital video to measure the elementary classroom management: Test validation and methodological challenges. *Computers & Education*, 107, 3–30. <https://doi.org/10.1016/j.compedu.2016.12.012>
- Hecker, S.-L. & Nimz, K. (2020). Expertinnenratings zur Deutschals-Zweitsprache-Kompetenz von Lehrkräften. Eine Vorstudie zur Konzipierung eines videobasierten Testinstruments. In J. Stiller, C. Laschke, T. Nesyba, U. Salaschek (Hrsg.), *Berlin-Brandenburger Beiträge zur Bildungsforschung 2020* (S. 39–64). Berlin: Peter Lang.
- Heins, J. & Zabka, T. (2019). Mentale Prozesse bei der Unterrichtsbeobachtung. *Zeitschrift für Pädagogik*, 6, 904–925.
- Hepberger, B., Moser Opitz, E., Heinze, A. & Lindmeier, A. (2019). Entwicklung und Validierung eines Tests zur Erfassung der mathematikspezifischen professionellen Kompetenzen von frühpädagogischen Fachkräften. *Psychologie in Erziehung und Unterricht*, 66, Preprint Online. München & Basel: Ernst Reinhardt Verlag. <https://doi.org/10.2378/peu2019.art24d>
- Hoth J., Kaiser G., Döhrmann M., König J., Blömeke S. (2018). A Situated Approach to Assess Teachers' Professional Competencies Using Classroom Videos. In: O. Buchbinder & S. Kuntze (Hrsg.) *Mathematics Teachers Engaging with Representations of Practice* (S. 23–45), ICME-13 Monographs. Springer, Cham
- Huxham, M., Campbell, F. & Westwood, J. (2012). Oral versus Written Assessment: A Test of Student Performance and Attitudes. *Assessment & Evaluation in Higher Education*, 37 (1), 125–136. <https://doi.org/10.1080/02602938.2010.515012>
- Kahneman, D. (2013). *Thinking, fast and slow*. New York: Farrar, Straus & Giroux.
- Kersting, N. (2008). Using video clips of mathematics classroom instruction as item prompts to measure teachers knowledge of teaching mathematics. *Educational and Psychological Measurement*, 68, 845–861.
- Klahn, J. (2016). *Analyse von DaZ-relevanten Unterrichtssituationen in Hinblick auf DaZ-Kompetenzen von Lehrkräften*. (Masterarbeit, internes Dokument).
- Koch-Priewe, B. (2018). Das DaZKom-Projekt – ein Überblick. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Hrsg.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (S. 7–37). Münster: Waxmann.
- Köker, A., Rosenbrock-Agyei, S., Ohm, U., Carlson, S. A., Ehmke, T., Hammer, S., Koch-Priewe, B., & Schulze, N. (2015). DaZKom – Ein Modell von Lehrerkompetenz im Bereich Deutsch als Zweitsprache. In B. Koch-Priewe, A. Köker, J. Seifried & E. Wuttke (Hrsg.): *Kompetenzerwerb an Hochschulen: Modellierung und Messung. Zur Professionalisierung angehender Lehrerinnen und Lehrer sowie früh-pädagogischer Fachkräfte* (S. 177–206). Bad Heilbrunn: Klinkhardt.
- König, J., & Lebens, M. (2012). Classroom Management Expertise (CME) von Lehrkräften messen: Überlegungen zur Testung mit-

- hilfe von Videovignetten und erste empirische Befunde. *Lehrerbildung auf dem Prüfstand*, 5 (1), 3–28.
- König, K. & Hector, T. M. (2017). *Zur Theatralität von WhatsApp-Sprachnachrichten. Nutzungskontexte von Audio-Postings in der mobilen Messenger-Kommunikation*. Leibniz Universität Hannover: Networx.
- Lindmeier, A. (2013). Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen. In U. Riegel & K. Macha (Hrsg.), *Videobasierte Kompetenzforschung in den Fachdidaktiken*, (S. 45–62), Fachdidaktische Forschung, Band 4. Münster: Waxmann.
- Molitor-Lübbert, S. (2002). Schreiben und Denken. Kognitive Grundlagen des Schreibens. In D. Perrin, I. Böttcher, O. Kruse & A. Wrobel (Hrsg.), *Schreiben. Von intuitiven zu professionellen Schreibstrategien* (S. 33–46). Wiesbaden: Springer Fachmedien.
- Muthén, L. K., & Muthén, B. O. (2014). MPLUS (Version 7.3) [Computer software]. Los Angeles, CA: Muthén & Muthén.
- Neuweg, G. H. (2014). Das Wissen der Wissensvermittler. Problemstellungen, Befunde und Perspektiven der Forschung zum Lehrwissen. In: E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 583–614), 2. Überarb. Auflage. Münster: Waxmann.
- Paetsch, J., Darsow, A., Wagner, F. S., Hammer, S., & Ehmke, T. (2019). Prädiktoren des Kompetenzzuwachses im Bereich Deutsch als Zweitsprache bei Lehramtsstudierenden. *Unterrichtswissenschaft*, 47 (1), 51–77. <https://doi.org/10.1007/s42010-019-00038-7>
- Poole, M. E. & Field, T. W. (1976). A comparison of oral and written code elaboration. *Language and Speech*, 19 (4), 305–312. <https://doi.org/10.1177/002383097601900401>
- Pugalee, D. K. (2004). A Comparison of Verbal and Written Descriptions of Students' Problem Solving Processes. *Educational Studies in Mathematics*, 55 (1–3), 27–47. <https://doi.org/10.1023/B:EDUC.0000017666.11367.c7>
- Ryan, J. J., Townsend, J. M., & Kreiner, D. S. (2014). Comparison of Oral, Written, and Pointing Responses to WAIS-IV Digit Span. *Applied Neuropsychology: Adult*, 21 (2), 94–97.
- Rutsch, J. & Dörfler, T. (2018). Vignettentest zur Erfassung des fachdidaktischen Wissens im Leseunterricht bei angehenden Lehrkräften. *Diagnostica*, 64 (1), 2–13. <https://doi.org/10.1026/0012-1924/a000188>
- Samuels, R. (2009). The Magical Number Two, Plus or Minus: Dual Process Theory as a Theory of Cognitive Kinds. In: K. Frankish & J. Evans (Hrsg.), *In Two Minds: Dual Process and Beyond*. Oxford: Oxford University Press.
- Scheuermann, U. (2016). *Schreibdenken. Schreiben als Denk- und Lernwerkzeug nutzen und vermitteln*. Opladen & Toronto: Barbara Budrich Verlag.
- Seidel, T., Blomberg, G. & Stürmer, K. (2010). „Observer“ – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt: OBSERVE. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG- Schwerpunktprogramms und Perspektiven des Forschungsansatzes. Zeitschrift für Pädagogik, Beiheft 56*, 296–306.
- Stürmer, K., Seidel, T. & Kunina-Habenicht, O. (2015). Unterricht wissenschaftsbasiert beobachten. Unterschiede und erklärende Faktoren bei Referendaren zum Berufseinstieg. *Zeitschrift für Pädagogik*, 61 (3), 345–360.
- Sieland, B. & Jordaan, L. (2019). Kommentar: Lehrer- und Lehrerinnenbildung im Umbruch zwischen Auftrag, alten Schwächen und neuen Chancen. In: G. Uhde & B. Thies (Hrsg.), *Kompetenzentwicklung im Lehramtstudium durch professionelles Training* (S. 7–19). Braunschweig: Universität, ATUS-Tagung.
- Thorburn, M. & Collins, D. (2006). Accuracy and authenticity of oral and written assessments in high-stakes school examinations. *The Curriculum Journal*, 17 (1), 3–25. <https://doi.org/10.1080/09585170600682491>

Historie

Manuskript eingereicht: 06.03.2019

Manuskript nach Revision angenommen: 26.06.2020

Onlineveröffentlichung: 27.08.2020

Svenja Lemmrich

Leuphana Universität Lüneburg

Universitätsallee 1

21335 Lüneburg

svenja.lemmrich@leuphana.de

8. Artikel 2: Performanznahe und videobasierte Messung von DaZ-Kompetenz bei Lehrkräften: Skalierung und dimensionale Struktur des Testinstruments

Der Beitrag ist wie folgt publiziert:

Lemmrich, S., Hecker, S.-L., Klein, S., Ehmke, T., Köker, A., Koch-Priewe, B. & Ohm, U. (2019). Performanznahe und videobasierte Messung von DaZ-Kompetenz bei Lehrkräften: Skalierung und dimensionale Struktur des Testinstruments. In T. Ehmke, P. Kuhl & M. Pietsch (Hg.), *Lehrer. Bildung. Gestalten: Beiträge zur empirischen Forschung in der Lehrerbildung* (1. Aufl., S. 190–204). Beltz.

Performanznahe und videobasierte Messung von DaZ-Kompetenz bei Lehrkräften

Skalierung und dimensionale Struktur des Testinstruments

Svenja Lemmrich, Sarah-Larissa Hecker, Stephanie Klein,
Timo Ehmke, Anne Köker, Barbara Koch-Priewe und Udo Ohm

Abstract

Die vorliegende Studie verfolgt im Rahmen des DaZKom-Video-Projekts das Ziel, Deutsch-als-Zweitsprache(DaZ)-Kompetenz bei (angehenden) Lehrkräften performanznah mit einem videobasierten Testinstrument zu erfassen. Dieser Beitrag prüft die psychometrische Qualität des DaZKom-Video-Testinstruments und untersucht die Dimensionalität des Konstruktes *DaZ-Kompetenz* sowie Zusammenhänge zwischen der DaZ-Kompetenz mit soziodemographischen Angaben und DaZ-bezogenen Lerngelegenheiten. Dabei handelt es sich um eine Erprobungsstudie im Rahmen eines Testentwicklungsprozesses. Getestet wurden Lehramtsstudierende sowie Lehrkräfte unterschiedlicher Klassenstufen. Die Ergebnisse der Item- und Skalenanalyse zeigen eine befriedigende bis sehr gute psychometrische Qualität. Die Items haben eine sehr gute Passung zum Rasch-Modell und erreichen eine zufriedenstellende Trennschärfe. Die Ergebnisse zur dimensional Struktur des Konstrukts weisen darauf hin, dass es sich eher um ein eindimensionales Konstrukt handelt und nicht um ein zwei- oder dreidimensionales Konstrukt. In Bezug auf soziodemographische Angaben und DaZ-bezogene Lerngelegenheiten zeigt sich, dass die Erstsprache Deutsch sowie insbesondere Zusatzqualifikationen wie das DaZ-Zertifikat oder auch Forschungserfahrung im Bereich DaZ, statistisch signifikant mit der DaZ-Kompetenz von (angehenden) Lehrkräften zusammenhängen. Der Umfang der DaZ-bezogenen Lerngelegenheiten korreliert niedrig mit dem Testergebnis. Die Befunde dieser Erprobungsstudie werden genutzt, um das Testinstrument weiterzuentwickeln.

1 Einleitung

Sprache spielt im Fachunterricht eine zentrale Rolle: Sie ist der Schlüssel zu fachunterrichtlichen Inhalten, ohne den Schülerinnen und Schüler keinen Zugang zu diesen erhalten. Sprache im Fachunterricht hat besondere Eigenschaften (bspw. schriftsprachlich geprägt, spezifische Operatoren, vgl. dazu Schmölzer-Eibinger 2013) und ist deshalb für viele Schülerinnen und Schüler, insbesondere auch für solche mit Deutsch als Zweitsprache, ein Hindernis für ihren Lernerfolg (Köker 2018; Schmölzer-Eibinger 2013). Lehrkräfte aller Unterrichtsfächer müssen die fachsprachlichen Anforderungen an ihre Schülerinnen und Schüler kennen und Kompetenzen besitzen, sie nicht nur fachlich, sondern ebenso in sprachlicher Hinsicht dabei zu unterstützen, Zugang zu Unterrichtsinhalten zu erhalten (Thürmann/Vollmer 2013). Allerdings liegen hierzu weder Ausbildungs- noch Performanzstandards vor, wie Lehrkräfte im Bereich DaZ ausgebildet werden müssen (Koch-Priewe 2018) und es ist unklar, welche Lerngelegenheiten zur Kompetenzentwicklung der Lehrkräfte im Bereich Deutsch als Zweitsprache (DaZ) beitragen (Baumann 2017). Das *DaZKom-Video*-Projekt¹ setzt hier an mit dem Ziel, universitäre und außeruniversitäre Lerngelegenheiten im Bereich DaZ zu evaluieren und zu verbessern. Der vorliegende Beitrag greift auf Vorarbeiten aus dem Vorgängerprojekt *DaZKom*² zurück, die bereits ein DaZ-Kompetenzmodell sowie einen Paper-Pencil-Test beinhalten (Köker et al. 2015). Für eine Erfassung professioneller Kompetenzen von Lehrkräften, die nicht allein Wissen fokussiert, sondern möglichst handlungsnah operiert, muss über Papier- und Bleistift-Formate hinausgegangen werden (Aufschnaiter/Blömeke 2010). Dies setzt das Projekt mithilfe von Video-Vignetten und einem mündlichen Antwortformat um. Dadurch soll nicht nur Wissen, sondern auch Können erfasst werden. Beides sind Teil des Professionalitätsverständnisses nach Baumert et al. (2011). Der vorliegende Beitrag untersucht, inwieweit das videobasierte Testinstrument geeignet ist, das Konstrukt DaZ-Kompetenz möglichst performanznah zu erfassen, prüft verschiedene Skalierungsmodelle und untersucht als Indikatoren für die Konstruktvalidität die Zusammenhänge zwischen der gemessenen DaZ-Kompetenz mit sozio-

1 Projekt *DaZKom-Video* (2017–2019): Performanznahe Messung von Deutsch-als-Zweitsprache-Kompetenz bei (angehenden) Lehrkräften; Projektleitung: Prof. Dr. B. Koch-Priewe, Dr. A. Köker, Prof. Dr. U. Ohm, Prof. Dr. T. Ehmke, gefördert vom BMBF (Teil des Clusters KoKoHs II)

2 Projekt *DaZKom* (2015–2017): Professionelle Kompetenzen angehender LehrerInnen (Sek I) im Bereich Deutsch als Zweitsprache; Projektleitung: Prof. Dr. B. Koch-Priewe, Dr. A. Köker, Prof. Dr. U. Ohm, Prof. Dr. T. Ehmke, gefördert vom BMBF (Teil des Clusters KoKoHs I)

demographischen Merkmalen und DaZ-bezogenen Lerngelegenheiten der befragten (angehenden) Lehrkräfte.

2 Forschungslage: Deutsch als Zweitsprache – Kompetenz empirisch erfassen

In den folgenden beiden Abschnitten wird die Forschungslage dargestellt mit zweierlei Fokussierung: Zum einen wird auf die dimensionale Struktur von DaZ-Kompetenz eingegangen. Zum anderen wird die Bedeutung von Wahrnehmen und Handeln für performanznahe DaZ-Kompetenz näher beschrieben.

Theoretische Fundierung der Dimensionen von DaZ-Kompetenz

Im Rahmen des Vorgängerprojektes *DaZKom* wurde bereits ein Kompetenzmodell entwickelt, welches die inhaltliche Struktur von DaZ-Kompetenz darstellt (Köker et al. 2015). Als drei Hauptdimensionen wurden hier *Fachregister*, *Mehrsprachigkeit* und *Didaktik* identifiziert, die wiederum in Subdimensionen untergliedert sind, welche sich dann in einzelne Facetten ausdifferenzieren lassen (ebd.). In Bezug auf unterrichtliches Handeln von Lehrkräften wirken diese drei Dimensionen meist gemeinsam, wobei jedoch davon ausgegangen wird, dass je nach Situation meist eine Dimension fokussiert wird (Ohm 2018). Die Dimension *Fachregister* beschreibt das Medium Sprache, welches als Basis für jegliche Lehr-Lern-Prozesse dient, die in (schulischen) Institutionen stattfinden und dessen Umfang und Funktion sich Fachlehrkräfte bewusst sein müssen. Die Dimension *Mehrsprachigkeit* fokussiert den Lernprozess und betont besondere Anforderungen an Schülerinnen und Schüler mit DaZ, die von Lehrkräften berücksichtigt werden sollten. Die Dimension *Didaktik* ergänzt die beiden ersten Dimensionen um didaktische Kompetenzen, die notwendig sind, um einen Fachunterricht zu gestalten, der zweitsprachliche Entwicklungsprozesse berücksichtigt und fokussiert damit den Lehrprozess (Köker et al. 2015).

Wahrnehmung und Handlung als Indikatoren für Performanz

Im Folgenden wird auf die situationsspezifischen Fähigkeiten *Wahrnehmung* und *Handlung* eingegangen, welche in dieser Studie als Indikatoren für Performanz in DaZ-relevanten Unterrichtssituationen dienen (Blömeke/Gustafsson/Shavelson 2015). Die Definition von *Wahrnehmung* wird in der Forschung zu Lehrkräfteprofessionalisierung meist auf das Konzept der *professional vision* nach Goodwin (1994) bezogen (Lindmeier 2013). Goodwin beschreibt in seiner Arbeit, dass Expertinnen und Experten bedeutungsreiche Situationen für ihren Tätigkeitsbereich gezielter wahrnehmen als Novizinnen und Novizen, da ihre

Wahrnehmung dahingehend geschult ist (ebd. 1994). Van Es und Sherin (2002) greifen das Konzept auf und definieren zwei Komponenten der *professionellen Wahrnehmung*: *noticing* und *knowledge-based reasoning*, wobei ersteres eben diese gefilterte Wahrnehmung beschreibt und letzteres die Fähigkeit, auf Basis des professionellen Wissens Schlussfolgerungen aus der wahrgenommenen Situation zu ziehen (van Es/Sherin 2002). Expertinnen und Experten können demnach auf professionelles Wissen zurückgreifen, um das Wahrgenommene in Konzepte und Theorien einzuordnen und deshalb präziser darauf reagieren. Dies unterscheidet sie von Novizinnen bzw. Novizen, welche meist eine reine Beschreibung dessen liefern, was sie beobachten (Seidel/Blomberg/Stürmer 2010).

Professionelle *Handlungskompetenz* besteht aus Professionswissen und individuellen Fähigkeiten sowie motivationalen, volitionalen und sozialen Aspekten und ist abhängig von domänenspezifischen Handlungsanforderungen (Weinert 2001; Baumert/Kunter 2006). Lindmeier (2013) erfasst in ihrer Studie Handlungskompetenz als *aktionsbezogene Kompetenz* und meint damit die Fähigkeit der Lehrkräfte, ihr „professionelles Wissen spontan und unmittelbar für Situationen an[zu]wenden“ (ebd., S. 51).

Das DaZKom-Video-Testinstrument verwendet Testitems, die diese theoretische Struktur aufgreifen. Authentische Unterrichtssituationen werden durch Videostimuli präsentiert und mit gezielten Wahrnehmungsaufgaben und Handlungsaufgaben verbunden, um so die DaZ-Kompetenz der (angehenden) Lehrkräfte möglichst performanznah zu messen.

3 Forschungsfragen

Vor dem Hintergrund der Forschungslage sollen folgende Forschungsfragen in dieser Studie beantwortet werden:

1. Inwieweit kann eine ausreichende psychometrische Qualität der Items des DaZKom-Video-Testinstrumentes festgestellt werden?

Bezugnehmend auf andere Studien (u. a. Lindmeier 2013), die Performanztests nutzen, kann davon ausgegangen werden, dass eine ausreichende psychometrische Qualität erreicht werden kann. Jedoch ist das Format dieser Studie mit der Kombination aus Video-Vignetten und mündlichen Antworten sehr innovativ. Insbesondere zum mündlichen Antwortformat liegen bislang nur wenig bis keine Erfahrungswerte vor. Des Weiteren hat die Erprobungsstudie das Ziel, aus dem Itempool eine Itemauswahl zu treffen. Dabei sollen zufriedenstellende psychometrische Kennwerte für den Gesamttest erreicht werden und zugleich sollen die Items (a) alle drei Dimensionen *Fachregister*, *Mehrsprachigkeit*, *Di-*

didaktik sowie (b) die beiden Aufgabentypen *Wahrnehmung* und *Handlung* möglichst gleichverteilt abdecken.

2. Welche dimensionale Struktur weist das Testinstrument auf?

Aufgrund von Befunden des nicht-videobasierten DaZKom-Tests kann vermutet werden, dass bei einer videobasierten Messung das Konstrukt ebenfalls nicht eindimensional strukturiert ist. Orientiert an der Definition von Performanz könnte es zum einen eine zweidimensionale Struktur aufweisen (Wahrnehmungs- und Handlungsdimension). Orientiert an der Definition von DaZ-Kompetenz könnte es auch eine dreidimensionale Struktur aufweisen (Fachregister, Mehrsprachigkeit, Didaktik), siehe dazu Abschnitt 2.

3. Welche Zusammenhänge bestehen zwischen Personenmerkmalen, der individuellen DaZ-Erfahrung, der Nutzung von DaZ-bezogenen Lerngelegenheiten und der DaZ-Kompetenz von Testpersonen?

Diese Zusammenhänge zu erfassen dient einerseits der Konstruktvalidierung, andererseits sollen sie Aufschluss über die Beschaffenheit von DaZ-kompetenzförderlichen Lerngelegenheiten geben. Dabei wird angenommen, dass Probandinnen und Probanden, die bereits DaZ-bezogene Lerngelegenheiten hatten, im Kompetenztest bessere Ergebnisse erzielen. Außerdem wird davon ausgegangen, dass DaZ-erfahrene Lehrkräfte einen performanzorientierten Test besser bewältigen können als unerfahrene Studierende.

4 Methodisches Vorgehen

Das Testinstrument DaZKom-Video

Der auf Tablets dargebotene Test enthält 16 Video-Vignetten, die DaZ-relevante Unterrichtssituationen zeigen, mit jeweils zwei dazugehörigen Items. Ein Item aus dem Bereich Wahrnehmung (*Was nehmen Sie wahr?*) und eines aus dem Bereich Handlung (*Wie würden Sie anstelle der Lehrkraft in dieser Situation handeln?* oder *Sie sind Lehrkraft in dieser Situation, wie reagieren Sie wörtlich?*). Die Teilnehmerinnen und Teilnehmer sahen jeweils eine zufällige Auswahl von zehn der 20 Videos und antworteten jeweils mündlich auf die Fragen. Die Audioantworten der Probanden wurden mithilfe eines Headsets und von Tablet-PCs aufgezeichnet. Die Videos sind den Dimensionen des DaZ-Kompetenzmodells zugeordnet (*Fachregister, Mehrsprachigkeit und Didaktik*). Basierend auf Expertinnen- und Expertenratings wird davon ausgegangen, dass sich zwar häufig alle Dimensionen des Modells in jedem Video wiederfinden,

jedoch jede gezeigte Situation einen Fokus hat, welcher einer konkreten Dimension zugeordnet werden kann. Für die Ratings wurden drei Expertinnen und Experten das Modell (sowie die theoretische Fundierung des Modells), die Videos des Testinstruments und die dazugehörigen Kodierleitfäden vorgelegt, mit der Bitte, die Items den (Sub-)Dimensionen des Modells zuzuordnen. Für die Auswertung der Probandinnen- und Probandenantworten wurde mithilfe von Expertinnen- und Experteninterviews zu jedem Item ein Kodierleitfaden entwickelt, der mit Ankerbeispielen aus Erprobungsstudien unterfüttert ist (vgl. dazu auch Hecker et al., eingereicht). Alle Items werden durch einen dreistufigen Antwortcode kodiert: Code 2 (die DaZ-relevanten Aspekte wurden situationspezifisch gänzlich erfasst und in einen weiteren Kontext eingeordnet), Code 1 (die DaZ-relevanten Aspekte wurden teilweise situationspezifisch erfasst) und Code 0 (die DaZ-relevanten Aspekte wurden nicht erfasst). Für diese Teilstudie wurden Code 2 und Code 1 zusammengefasst und dementsprechend nur Code 1 (*richtig*) und 0 (*falsch*) vergeben. Die Kodiererübereinstimmung bei der Doppelkodierung für die dreistufige Kodierung lag zwischen $0.6 < PÜ < 0.8$.

Soziodemographische Angaben und DaZ-bezogene Lerngelegenheiten

Neben der DaZ-Kompetenz wurden Personenmerkmale wie Geschlecht und Studien- bzw. Unterrichtsfächer (6 Items) erfasst. Zudem wurden DaZ-relevante Lerngelegenheiten durch eine Skala mit 8 Items ($\alpha = 0.83$) (Ehmke/Lemmerich 2018) und zwei Einzelitems erfasst. Für ersteres wurde eine fünfstufige Likert-Skala verwendet: (1) *gar nicht*, (2) *in einer Sitzung*, (3) *in mehreren Sitzungen*, (4) *in einer ganzen Lehrveranstaltung*, (5) *in mehreren Lehrveranstaltungen*. Dabei wurden die Items wie folgt abgefragt: „In Veranstaltungen in Ihrem Lehramtsstudium sowie in Fort- und Weiterbildungen haben Sie bereits für das eigene Studienfach typische Texte hinsichtlich ihrer sprachlichen Besonderheiten für Schülerinnen und Schüler mit Deutsch als Zweitsprache analysiert (z. B. mathematische Textaufgaben)?“

Stichprobe und Untersuchungsablauf

Für die Durchführung der Studie wurden $N = 137$ Personen befragt, davon sind 83.2 % weiblich und 16.8 % männlich bzw. 77.4 % Lehramtsstudierende deutscher Universitäten und 22.6 % Lehrkräfte. Es waren fast alle Studien- bzw. Unterrichtsfächer (Mathematik, Deutsch, Englisch, Physik, Biologie, Chemie, Geographie, Sachunterricht, Geschichte, Musik, Kunst, Religion, Politik, Sport) vertreten, wobei 61.9 % das (Studien-)Fach Deutsch angaben. 13.4 % gaben an, Zusatzqualifikationen im Bereich DaZ zu besitzen. 88.3 % gaben an, Deutsch als Erstsprache zu haben, wobei 8.8 % eine andere Erstsprache haben, darunter vor allem Russisch, Polnisch und Kroatisch (2.9 % haben keine Angabe gemacht).

Für die technische Umsetzung des Testinstrumentes auf Tablet-PCs wurde das Programm *LimeSurvey* verwendet, welches die Umsetzung des Tests mithilfe einer App ermöglicht. Die Teilnehmenden sahen auf den Tablets zunächst eine Einführung in das Testinstrument. Diese besteht zum einen aus einer technischen Einführung, zum anderen ist eine Beispiel-Vignette mit dazugehörigen Beispielitems und -antworten enthalten. Nach der Einführung bearbeiten die Probandinnen und Probanden zunächst den Videotest, bevor sie dann, ebenfalls auf den Tablets, die Fragen zu soziodemographischen Daten und den Lerngelegenheiten beantworteten. Insgesamt dauerte die Erhebung 60 Minuten.

Statistisches Vorgehen

Die Dateneingabe und Auswertung der Häufigkeiten (Codes des DaZKom-Video-Tests, soziodemographische Daten und Lerngelegenheiten) erfolgten in SPSS 25 (IBM 2018). Die Item- und Skalenanalysen wurden auf Basis des Rasch-Modells (Rost 2004) mithilfe des Programms *ConQuest* (Adams/Wu/Wilson 2015) durchgeführt. Die DaZ-Kompetenz der Probandinnen und Probanden wurde als Personenmesswert mithilfe von Personenfähigkeitsschätzern, Weighted Likelihood Estimates (WLEs), bestimmt. *Personenfähigkeitsschätzer* impliziert, dass die Beantwortung eines Items einerseits davon abhängt, welche Fähigkeiten (in dieser Studie DaZ-Kompetenz) Probandinnen und Probanden mitbringen und andererseits auch berücksichtigt, wie hoch die Itemschwierigkeit ist (Wilson 2005). Es werden also hiermit nicht die tatsächlichen Antworten abgebildet, sondern die Wahrscheinlichkeit, die sich aus der Personenfähigkeit und Itemschwierigkeit ergibt, eine richtige Antwort zu erzielen (ebd.). Um die Dimensionalität des Konstruktes zu untersuchen, wurden mit *ConQuest* ein-, zwei- und dreidimensionale Modelle geschätzt.

5 Ergebnisse

Psychometrische Qualität

Das erste Kriterium für die psychometrische Qualität des entwickelten Instruments ist die Passung zum Rasch-Modell. In Tabelle 1 sind hierzu die Ergebnisse der IRT-Skalierung dargestellt. Es zeigt sich, dass die Items im Mittel eine sehr gute Passung zum Rasch-Modell aufweisen ($1.0 < MNSQ^3 < 1.25$; OECD 2005). Einige Items haben eine weniger gute Passung ($Min = 0.85$; $Max = 1.14$), jedoch befindet sich zumindest der *Max*-Wert noch in der Spanne des zufriede-

3 *Weighted Mean Square Fit Statistics* (MNSQ) zeigt an, ob es bei einem Item eine gute Übereinstimmung zwischen der beobachteten und der erwartete Lösungswahrscheinlichkeit gibt.

denstellenden Item-Fits. Das zweite Kriterium ist die Trennschärfe der Items, da diese der Indikator dafür ist, wie verlässlich das Instrument zwischen Personen mit hoher und Personen mit niedriger Merkmalsausprägung unterscheiden kann. Items mit einer Trennschärfe von unter 0.30 wurden aus testtheoretischen Gründen aus den Berechnungen ausgeschlossen, mit der Ausnahme von vier Items, die eine Trennschärfe von 0.21 bis 0.25 ($Min = 0.21$) aufweisen. Diese wurden in den Berechnungen belassen, da es sich bei zweien davon jeweils um eines der Partneritems der Dimension *Wahrnehmung* zur Dimension *Handlung* handelt und das jeweils andere Item der Dimension *Handlung* eine zufriedenstellende Trennschärfe aufweist. Ein weiteres Itempaar wurde aus inhaltlichen Gründen im Test belassen, obwohl beide Items eine Trennschärfe von kleiner 0.3 aufweisen. Insgesamt wurden acht Items (also vier Vignetten) aus der finalen Skalierung ausgeschlossen. Die Trennschärfe der Items ist danach zufriedenstellend ($M = 0.44$; $SD = 0.12$). Die mittlere Itemschwierigkeit ($M = 1.56$; $SD = 0.88$) und die Personenfähigkeit ($M = 0.08$; $SD = 1.10$) überlappen allerdings zu wenig. Dies legt nahe, dass die Items zu schwierig sind bzw. die Personenfähigkeit zu gering. Die EAP-Reliabilität beträgt eindimensional skaliert $\alpha = 0.62$.

Tabelle 1: Item- und Skalenkennwerte

	M	SD	Min	Max
Weighted fit MNSQ	1.01	0.07	0.85	1.14
Trennschärfe	0.44	0.12	0.21	0.74
Itemschwierigkeit	1.56	0.88	-0.71	3.05
WLE	0.08	1.10	-2.73	2.77

Dimensionalitätsprüfung

Die Dimensionalität des eingesetzten Testinstrumentes wurde mithilfe von ein-, zwei- und dreidimensionalen IRT-Modellen geprüft. Die Fit-Indices des Modellvergleichs (vgl. Abb. 1) sind in Tabelle 2 dargestellt. Die Devianz ist die Abweichung vom Idealwert und sollte dementsprechend möglichst niedrig sein. Sie ist, genau wie das AIC (Akaike's Information Criterion), „ein Maß für Anpassungsgüte des geschätzten Modells an die vorliegenden empirischen Daten (Stichprobe) unter Berücksichtigung des Modells“ (Moosbrugger 2007, S. 390). Aus dem AIC sind die Informationskriterien BIC und CAIC hervorgegangen, wobei BIC viele unnötige Modellparameter (Missachtung des Sparsamkeitsprinzips) bestraft und CAIC im Gegensatz zu AIC den Stichprobenumfang mit berücksichtigt. All diese Maße sollten möglichst niedrig sein (ebd.). Die Informationskriterien für das eindimensionale Modell sind für die Devianz, den BIC und CAIC niedriger als für das zweidimensionale Modell (Skalierung nach den

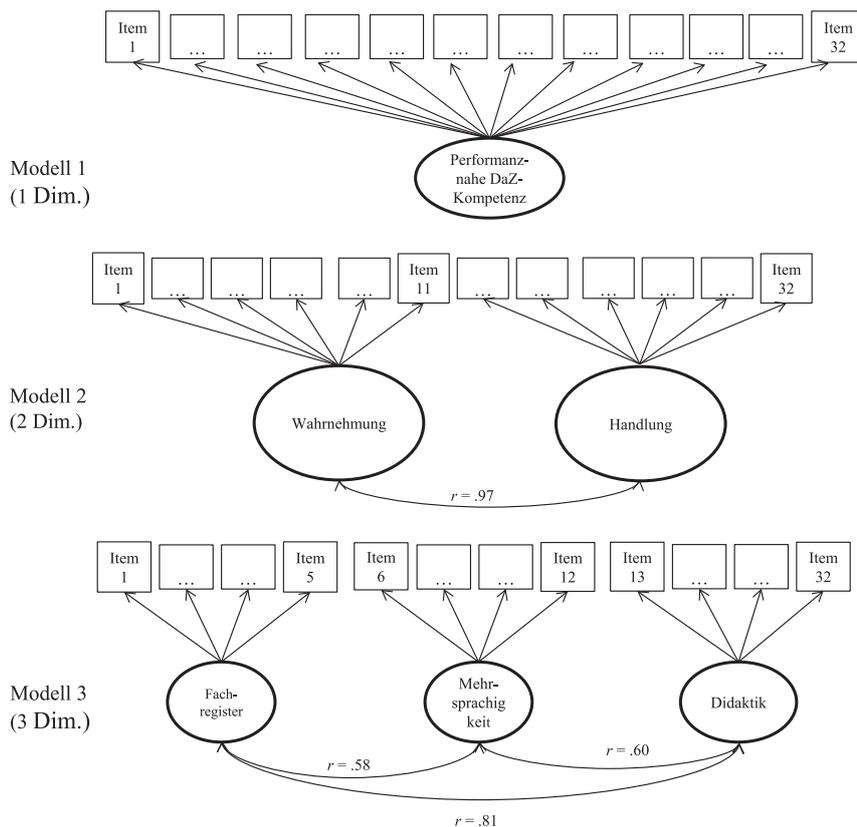
Dimensionen *Wahrnehmung* und *Handlung*) und für das dreidimensionale Modell (Skalierung nach den Dimensionen *Fachregister*, *Mehrsprachigkeit* und *Didaktik*). Dieses Ergebnis zeigt, dass es sich – entgegen der aufgestellten Hypothese – offenbar eher um ein eindimensionales Konstrukt handelt.

Tabelle 2: Dimensionalitätsvergleich

	N	Parameter	Devianz	AIC	BIC	CAIC
1-dimensional	137	31	1757	1819	1910	1942
2-dimensional	137	33	1789	1855	1951	1984
3-dimensional	137	36	1778	1850	1955	1991

In Abbildung 1 zum Modellvergleich wird in Modell 2 deutlich, dass die zwei Dimensionen *Wahrnehmung* und *Handlung* stark korrelieren ($r = 0.97$), was darauf hinweist, dass das Konstrukt der *performanznahen DaZ-Kompetenz* hier keine zwei trennbaren Dimensionen aufweist.

Abbildung 1: Modellvergleich



In Modell 3 wird deutlich, dass Dimension 2 mit Dimension 1 ($r = 0.58$) und mit Dimension 3 ($r = 0.60$) eine mittlere Korrelation aufweist, Dimension 1 und 3 ($r = 0.81$) dagegen hoch korrelieren. Die Abgrenzung nach Dimensionen ist für die Dimensionen *Fachregister* (1) und *Didaktik* (3) offenbar schwierig, wohingegen die Dimension *Mehrsprachigkeit* (2) sich eher als separates Teilkonstrukt von den beiden anderen abgrenzt.

Zusammenhänge von DaZ-Kompetenz mit weiteren Variablen wie soziodemographischen Merkmalen und Lerngelegenheiten

In Tabelle 3 sind die Korrelationen ausgewählter soziodemographischer Merkmale und DaZ-Lerngelegenheiten mit dem Personenfähigkeitsschätzer (WLE) für die DaZ-Kompetenz (eindimensionale Skalierung) dargestellt. Weibliche und männliche Testpersonen unterscheiden sich nicht in ihrer DaZ-Kompetenz. Testpersonen, die Deutsch als Erstsprache haben, erreichen im Test höhere Werte als Personen, die eine andere Sprache als Erstsprache sprechen. Deutsch als Studien- bzw. Unterrichtsfach steht in keinem signifikanten Zusammenhang mit der DaZ-Kompetenz. Testpersonen mit Zusatzqualifikationen im Bereich DaZ (DaZ-Zertifikat) hingegen erreichen im Test erwartungsgemäß statistisch hochsignifikant höhere Testwerte als Personen ohne DaZ-Zertifikat. Gleiches gilt für Personen, die Forschung im Bereich DaZ betreiben. Lehrerfahrung allein zeigt dagegen keinen statistisch signifikanten Zusammenhang mit der Fähigkeit, die Items des Tests zu bearbeiten, ebenso wenig wie die Summe der DaZ-bezogenen Lerngelegenheiten. Letzteres ist nicht erwartungskonform, weshalb die Rolle der Lerngelegenheiten im Anschluss auf Einzelitemebene detaillierter untersucht wurde. Von acht Items der DaZ-Lerngelegenheiten-Skala (LG) kann eines identifiziert werden (LG_8), welches statistisch signifikant mit dem WLE korreliert: Die Probandinnen und Probanden, die angeben, in Lehrveranstaltungen bereits sprachensible Unterrichtsstunden entworfen zu haben, erzielen im Test höhere Werte als Personen, die dies noch nicht gelernt haben ($r = 0.17^*$).

Tabelle 3: Korrelationen der Ergebnisse aus dem DaZKom-Video-Test mit Personenmerkmalen

	WLE
Geschlecht (m = 0; w = 1)	0.01
Erstsprache (Deutsch = 1; andere = 0)	0.18*
Fach Deutsch (nein = 0, ja = 1)	0.03
DaZ-Zertifikat (nein = 0, ja = 1)	0.22**
Lehrerfahrung DaZ (nein = 0; ja = 1)	0.13
Forschung DaZ (nein = 0; ja = 1)	0.24**
Skala DaZ-Lerngelegenheiten (DaZ-LG)	0.14

* Die Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant

** Die Korrelation ist auf dem Niveau von .01 (2-seitig) signifikant

6 Zusammenfassung, Diskussion und Ausblick

Ziel der vorliegenden Studie war es, die psychometrische Qualität und die dimensionale Struktur des Testinstruments zur performanznahen Messung von DaZ-Kompetenz zu prüfen und Zusammenhänge zwischen der gemessenen DaZ-Kompetenz mit soziodemographischen Merkmalen und DaZ-Lerngelegenheiten zu untersuchen. Die zentralen Erkenntnisse der Untersuchung sind folgende:

1. Das untersuchte performanznahe Testinstrument zeigt – nach Ausschluss von 8 Items – eine gute Passung zum Rasch-Modell und die Items weisen eine zufriedenstellende Trennschärfe auf. Die Skalenreliabilität ist ausreichend. Noch nicht optimal ist die Passung von Itemschwierigkeit zu den Fähigkeiten der getesteten Probandinnen und Probanden. Diese Testitems sind etwas zu schwer für die aktuelle Stichprobe.
2. Die angenommene Mehrdimensionalität konnte nicht bestätigt werden. Das eindimensionale Modell wies die beste Passung auf.
3. Bezüglich der Zusammenhänge der Testergebnisse mit soziodemographischen Merkmalen und Lerngelegenheiten konnte festgestellt werden, dass Deutsch als Erstsprache als Merkmal sowie insbesondere Lerngelegenheiten in Form von Zusatzqualifikationen und Forschung im Bereich DaZ einen statistisch signifikanten Zusammenhang mit dem Testergebnis aufweisen. Die Summe der Skala der DaZ-Lerngelegenheiten zeigt wider Erwarten keinen statistisch signifikanten Zusammenhang mit dem Testergebnis.

Die Auswertung des Testinstruments auf Grundlage der Item-Response-Theorie nach ersten Erprobungen bietet den Vorteil, dass das Konstrukt theoretisch

definiert wird und Modelle zunächst geprüft werden, um dann aus den psychometrischen Kennwerten und dem Modellvergleich Rückschlüsse für die weitere Testentwicklung zu ziehen. Im Hinblick auf die psychometrischen Kennwerte kann festgehalten werden, dass die Items aktuell überwiegend noch zu schwer für die Testpersonen sind. Dies wird in Zukunft insofern bearbeitet, als dass die Video-Stimuli mit weiteren Kontextinformationen versehen werden, um eine noch realere Einbettung der Situation zu erreichen und damit die Situationsspezifität zu erhöhen. Gleichzeitig wird die Stichprobe der erfahrenen Lehrkräfte in nachfolgenden Studien vergrößert, sodass mehr Probandinnen und Probanden der Stichprobe bereits einen höheren Grad an Expertise erreicht haben, die benötigt wird, um performanznahe Testsituationen bewältigen zu können (vgl. Kapitel 2).

Im Hinblick auf die Dimensionalität des Konstruktes können unterschiedliche Perspektiven herangezogen werden, um die Ergebnisse zu reflektieren. Die Korrelation zwischen den zwei Dimensionen *Wahrnehmung* und *Handlung* in der zweidimensionalen Skalierung sprechen dafür, dass diese beiden Dimensionen nicht getrennt voneinander für das Konstrukt festgelegt werden können. Hinweise dazu finden sich auch in anderen Arbeiten zur professionellen Wahrnehmung oder zur Handlungskompetenz bei Lehrkräften. Niesen (2018) fasst in ihrer Studie zusammen: „Zunehmend häufen sich Hinweise für die Annahme, dass ein Zusammenhang zwischen professioneller Wahrnehmungs- und Handlungskompetenz von Lehrkräften besteht, dergestalt, dass sich erstere positiv auf letztere auswirkt.“ (ebd., S. 122). Die kognitionspsychologische Forschung geht davon aus, dass beide Dimensionen in ihrer kognitiven Repräsentation überlappen (Hommel et al. 2001). Die dreidimensionale Struktur des Konstruktes DaZ-Kompetenz bestätigt sich für die Kompetenzstruktur im Bereich des Wissens (Ehmke/Hammer 2018). Die vorliegende Studie gibt jedoch Hinweise darauf, dass im Bereich der performanznahen Messung keine klare Trennung der Dimensionen mehr möglich ist, sondern dass die drei Dimensionen *Fachregister*, *Mehrsprachigkeit* und *Didaktik* einheitlicher auftreten. Einschränkend muss allerdings darauf hingewiesen werden: nicht alle Probandinnen und Probanden haben auch alle Items bearbeitet. Es sind in der Summe also nur wenige Items pro Dimension, die von einer Testperson bearbeitet wurden. In zukünftigen Studien soll die dreidimensionale Struktur des Konstrukts daher noch einmal, unter Berücksichtigung einer größeren Stichprobenzahl, untersucht werden. In der DaZKom-Normierungsstudie sollen rund 300 Probandinnen und Probanden befragt werden, von denen mindestens die Hälfte Lehrkräfte sind. Alle Probandinnen und Probanden werden dann jeweils auch alle Items bearbeiten. Außerdem sollen in zukünftigen Analysen Testlet-Effekte kontrolliert werden, da der Test zu einem Stimulus immer zwei Items vorgibt und dadurch ggf. Kontexteffekte entstehen.

Dass sich Dimension 2 im Modellvergleich klarer von Dimension 1 und 3 abgrenzen lässt, könnte auf die inhaltliche Fokussierung der Items zurückzuführen sein. Items dieser Dimension fokussieren entweder die Subdimension *Zweitspracherwerb* oder *Migration*. Besonders in letzterer spielen häufig Differenzkonstruktionen im Unterricht eine zentrale Rolle, welche für Lehrkräfte offenbar eine besondere Herausforderung darstellen. Der Fortbildungsbedarf im reflexiven Umgang mit ethno-kulturellen Differenzordnungen im Unterricht wurde bereits in einer anderen Studie festgestellt (Klein/Hecker/Lemmerich, eingereicht). Ähnliche Korrelationen der drei Dimensionen zeigen sich im Übrigen auch im Modellvergleich der Skalierung des DaZKom-Testinstruments (Ehmke/Hammer 2018).

Im Hinblick auf Zusammenhänge der Testergebnisse mit weiteren Merkmalen sticht insbesondere heraus, dass die Skala für die DaZ-bezogenen Lerngelegenheiten nicht statistisch signifikant mit dem Testergebnis korreliert. Gleichzeitig gibt es zwischen dem Erwerb von Zusatzqualifikationen und dem Testergebnis sehr wohl einen statistisch signifikanten Zusammenhang. Dies ist ein Indikator dafür, dass DaZ-spezifische Lerngelegenheiten offenbar zum Erfolg führen können und ggf. die genutzte Skala zur Erfassung der Lerngelegenheiten a) nicht umfassend genug ist und b) die Skala in Bezug auf handlungsnaher Kompetenzerfassung irrelevante Lerngelegenheiten abfragt. Die Entwicklung der Fragenbogenskalen ist Teil eines Testentwicklungsprozesses. In künftigen Studien wird demnach der Fokus auf die DaZ-bezogenen Lerngelegenheiten verstärkt und eine umfassende Überarbeitung und Erweiterung der Skala vorgenommen, die auch Lerngelegenheiten in Fort- und Weiterbildungen erfasst.

Diese Studie hatte das Ziel, das Testinstrument zu erproben und psychometrische Kennwerte zu gewinnen. Dabei wurden jedoch keine deskriptiven Ergebnisse der erhobenen Daten untersucht. Dies wird in zukünftigen Datenerhebungen mit einem Testinstrument von guter psychometrischer Qualität das Ziel sein, um schließlich Veranstaltungskonzeptionen und lernförderliche DaZ-bezogene Lerngelegenheiten zu identifizieren, mit dem langfristigen Ziel, angehende Lehrkräfte optimal auf den Unterricht vorzubereiten.

Literatur

- Adams, Raymond J./Wu, Margaret L./Wilson, Mark R. (2015): ACER ConQuest: Generalised Item Response Modelling Software [Computer software]. Version 4. Camberwell, Victoria: Australian Council for Educational Research.
- Aufschnaiter, Claudia v./Blömeke, Sigrid (2010): Professionelle Kompetenz von (angehenden) Lehrkräften erfassen – Desiderata. In: Zeitschrift für Didaktik der Naturwissenschaften 16, S. 361–367.

- Baumann, Barbara (2017): Sprachförderung und Deutsch als Zweitsprache in der Lehrerbildung. In: Becker-Mrotzek, Michael/Rosenberg, Peter/Schroeder, Christoph/Witte, Annika (Hrsg.): Deutsch als Zweitsprache in der Lehrerbildung. Münster: Waxmann, S. 9–26.
- Baumert, Jürgen/Kunter, Mareike/Blum, Werner/Klusmann, Uta/Krauss, Stefan/Neubrand, Michael (2011): Professionelle Kompetenz von Lehrkräften, kognitiv aktivierender Unterricht und die mathematische Kompetenz von Schülerinnen und Schülern (COACTIV) – Ein Forschungsprogramm. In: Kunter, Mareike/Baumert, Jürgen/Blum, Werner/Klusmann, Uta/Krauss, Stefan/Neubrand, Michael (Hrsg.): Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV. Münster: Waxmann, S. 7–26.
- Baumert, Jürgen/Kunter, Mareike (2006): Stichwort: Professionelle Kompetenz von Lehrkräften. In: Zeitschrift für Erziehungswissenschaft 9, H. 4, S. 469–520.
- Blömeke, Sigrid/Gustafsson, Jan-Eric/Shavelson, Richard J. (2015): Beyond dichotomies: Competence viewed as a continuum. In: Zeitschrift für Psychologie 223, H. 1, S. 3–13.
- Ehmke, Timo/Hammer, Svenja (2018): Skalierung und dimensionale Struktur des DaZKom-Testinstruments. In: Ehmke, Timo/Hammer, Svenja/Köker, Anne/Ohm, Udo/Koch-Priewe, Barbara (Hrsg.): Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache. Münster: Waxmann, S. 129–148.
- Ehmke, Timo/Lemmerich, Svenja (2018): Bedeutung von Lerngelegenheiten für den Erwerb von DaZ-Kompetenz. In: Ehmke, Timo/Hammer, Svenja/Köker, Anne/Ohm, Udo/Koch-Priewe, Barbara (Hrsg.): Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache. Münster: Waxmann, S. 201–219.
- Goodwin, Charles (1994): Professional vision. In: American Anthropologist 96, H. 3, S. 606–633.
- Hecker, Sarah-Larissa/Klein, Stephanie/Lemmerich, Svenja/Ehmke, Timo (eingereicht): Ist DaZ-kompetentes Wahrnehmen von Unterrichtssituationen notwendig, um DaZ-kompetente Handlungsoptionen zu nennen?
- Hommel, Bernhard/Müsseler, Jochen/Aschersleben, Gisa/Prinz, Wolfgang (2001): The theory of event coding (TEC): A framework for perception and action planning. In: Behavioral and Brain Sciences 24, H. 5, S. 849–878.
- Klein, Stephanie/Hecker, Sarah-Larissa/Lemmerich, Svenja (eingereicht): Lehrkräfteprofessionalisierung im Kontext von Heterogenität. DaZ-kompetenter Umgang mit Differenzordnungen im Unterricht.
- Koch-Priewe, Barbara (2018): Das DaZKom-Projekt – ein Überblick. In: Ehmke, Timo/Hammer, Svenja/Köker, Anne/Ohm, Udo/Koch-Priewe, Barbara (Hrsg.): Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache. Münster: Waxmann, S. 7–37.
- Köker, Anne (2018): Zur Relevanz der Ausbildung von Kompetenzen von Lehrkräften im Bereich DaZ. In: Ehmke, Timo/Hammer, Svenja/Köker, Anne/Ohm, Udo/Koch-Priewe, Barbara (Hrsg.): Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache. Münster: Waxmann, S. 57–71.
- Köker, Anne/Rosenbrock-Agyei, Sonja/Ohm, Udo/Carlson, Sonja A./Ehmke, Timo/Hammer, Svenja/Koch-Priewe, Barbara/Schulze, Nina (2015): DaZKom – Ein Modell von Lehrerkompetenz im Bereich Deutsch als Zweitsprache. In: Koch-Priewe, Barbara/Köker, Anne/Seifried, Jürgen/Wuttke, Eveline (Hrsg.): Kompetenzerwerb an Hochschulen: Modellierung und Messung. Zur Professionalisierung angehender Lehrerinnen und Lehrer sowie früh-pädagogischer Fachkräfte. Bad Heilbrunn: Klinkhardt, S. 177–206.

- Lindmeier, Anke/Heinze, Aiso/Reiss, Kristina (2012): Eine Machbarkeitsstudie zur Operationalisierung aktionsbezogener Kompetenz von Mathematiklehrkräften mit videobasierten Maßen. In: *Journal für Mathematik-Didaktik* 34, H. 1, S. 99–119.
- Lindmeier, Anke (2013): Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen. In: Riegel, Ulrich/Macha, Klaas (Hrsg.): *Videobasierte Kompetenzforschung in den Fachdidaktiken. Fachdidaktische Forschungen. Band 4.* Münster: Waxmann, S. 45–62.
- Moosburger, Helfried/Kelava, Augustin (2007): *Testtheorie und Fragebogenkonstruktion.* Heidelberg: Springer Medizin Verlag.
- Niesen, Heike (2018): Förderung mehrsprachigkeitssensibler professioneller Handlungskompetenz angehender Englischlehrkräfte. In: *Zeitschrift für Interkulturellen Fremdsprachenunterricht: Didaktik und Methodik im Bereich Deutsch als Fremdsprache* 23, H. 1, S. 121–134.
- OECD (2005): *PISA 2003. Technical report.* Paris: OECD.
- Ohm, Udo (2018): Das Modell von DaZ-Kompetenz bei angehenden Lehrkräften. In: Ehmke, Timo/Hammer, Svenja/Köker, Anne/Ohm, Udo/Koch-Priewe, Barbara (Hrsg.): *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache.* Münster: Waxmann, S. 73–91.
- Rost, Jürgen (2004): *Testtheorie Testkonstruktion.* 2. Auflage. Bern: Huber.
- Schmölzer-Eibinger, Sabine (2013): Sprache als Medium des Lernens im Fach. In: Becker-Mrotzek, Michael/Schramm, Karen/Thürmann, Eike/Vollmer, Helmut J. (Hrsg.): *Sprache im Fach. Sprachlichkeit und fachliches Lernen.* Münster: Waxmann, S. 25–40.
- Seidel, Tina/Blomberg, Geraldine/Stürmer, Kathleen (2010): „Observer“ – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE. In: Bayrhuber, Marianne/Leuders, Timo/Bruder, Regina/Wirtz, Markus (Hrsg.): *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes.* Weinheim und Basel: Beltz, S. 296–306.
- Thürmann, Eike/Vollmer, Helmut J. (2013): Schulsprache und Sprachsensibler Fachunterricht: Eine Checkliste mit Erläuterungen. In: Röhner, Charlotte/Hövelbrinks, Britta (Hrsg.): *Fachbezogene Sprachförderung in Deutsch als Zweitsprache. Theoretische Konzepte und empirische Befunde zum Erwerb bildungssprachlicher Kompetenzen.* Weinheim und Basel: Beltz Juventa, S. 121–233.
- van Es, Elizabeth A./Sherin, Miriam G. (2002): Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. In: *Journal of Technology and Teacher Education* 10, H. 4, S. 571–596.
- Weinert, Franz E. (2001): Concept of Competence: A Conceptual Clarification. In: Rychen, Dominique S./Salganik, Laura H. (Hrsg.): *Defining and Selecting Key Competencies.* Göttingen: Hogrefe, S. 45–66.
- Wilson, Mark (2005): *Constructing Measures. An Item Response Modeling Approach.* New York und London: Psychology Press, Taylor & Francis Group.

9. Artikel 3: Linguistically Responsive Teaching in Multilingual Classrooms. Development of a Performance-Oriented Test to Assess Teachers' Competence

Der Beitrag ist wie folgt publiziert:

Lemmrich, S., Hecker, S.-L., Klein, S., Ehmke, T., Koch-Priewe, B., Köker, A. & Ohm, U. (2020). Linguistically Responsive Teaching in Multilingual Classrooms. Development of a Performance-Oriented Test to Assess Teachers' Competence. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper & C. Lautenbach (Hg.), *Student Learning in German Higher Education* (1. Aufl., S. 125–140). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27886-1_7



2.6 Linguistically Responsive Teaching in Multilingual Classrooms

Development of a Performance-Oriented Test to Assess Teachers' Competence

Lemmrich, S., Hecker, S.-L., Klein, S., Ehmke, T., Koch-Priewe, B., Köker, A., and Ohm, U.

Abstract

The *DaZKom-Video* project aims at performance-oriented measurement of German-as-a second-language (GSL)-competence of pre- and in-service teachers using video-based stimuli and oral responses. The present study is part of this project and focuses on the psychometric quality of the *DaZKom-Video* test instrument and evaluates the dimensionality of the construct *GSL competence (Deutsch-als-Zweitsprache-Kompetenz)*. Additionally, we explore correlations between GSL competence and sociodemographic data as well as learning opportunities. The sample consists of pre- and in-service teachers from different school forms and universities across Germany. The items and scale analyses reveal good psychometric quality: The item-fit to the Rasch-model as well as the discrimination is satisfactory. The dimensional analyses show best model fit for the one-dimensional and three-dimensional models. Test persons with GSL experience such as in-service teachers (in contrast to pre-service teachers), test persons with teaching experience, those with an additional qualification in the field of GSL and test persons with many learning opportunities are statistically more likely to respond to items correctly. These results indicate a reliable and valid test instrument for performance-oriented measurement of GSL competence.

Keywords

Video-based testing, teacher education, linguistically responsive teaching, teachers' competence

1 Introduction

Language is crucial in teaching and learning: Every subject-specific classroom uses language to represent content. Language has specific characteristics in content classrooms and every subject adds its own subject-specific register (e.g. written language or specific operators, Schmölzer-Eibinger 2013). For some pupils, especially but not only for those with German as a second language (GSL), it therefore can be challenging to apprehend content successfully. Teachers of all subjects must be aware that specific registers come with their own demands and may pose challenges for their pupils. Accordingly, teachers need competencies to support their pupils in gaining access to the content by teaching linguistically responsive (Thürmann and Vollmer 2013, Cummins 2001). Kimanen et al. (2019) implicate that teachers' competence in linguistically and culturally responsive teaching could increase teachers' self-esteem and therefore have an impact on teachers' motivation and commitment to teaching. How teachers must be trained concerning teaching multilingual learners is still unclear, however, and performance standards of teachers' competencies in this field are missing. Furthermore, understanding which learning opportunities may contribute to the development of teachers' competencies in teaching in linguistically diverse classrooms is still a desideratum (Baumann 2017).

The *DaZKom-Video*¹ project aims at evaluating and improving learning opportunities for pre- and in-service teachers in teaching multilingual learners and developed a performance-oriented test instrument. The theoretical and assessment framework includes the GSL competence model and the paper-pencil-test of the previous project (*DaZKom*²). To measure teachers' professional competencies and not solely capture teachers' knowledge, but operate as closely to performance as possible, it is necessary to use performance-oriented test formats (Aufschnaiter

1 The project *DaZKom-Video: Performanznahe Messung von Deutsch-als-Zweit-sprache-Kompetenz bei (angehenden) Lehrkräften (2017–2020)*.

2 The previous project *DaZKom (2012–2015)* was funded by the German Federal Ministry for Education and Research (BMBF) (01PK11010A, 01PK11010B).

and Blömeke 2010). Consequently, the test instrument uses video vignettes of teaching situations and asks for orally given responses that are then recorded.

This article examines the extent to which the developed performance-oriented test instrument is suitable for capturing the construct of GSL competence. It evaluates different scaling models concerning the dimensionality of the GSL construct. As indicators for convergent and discriminant validity, the correlations of the measured GSL competence and socio-demographic characteristics as well as GSL-related learning opportunities of test persons are examined in this study.

Concerning the terminology, this paper uses ‘teaching pupils with German as a second language (GSL)’ synonymously to ‘teaching multilingual learners’ and ‘teaching in linguistically diverse classrooms’. GSL competence therefore refers to the competency to teach multilingual learners/in linguistically diverse classrooms.

2 Theoretical Background

2.1 Structure of GSL Competence

A structural model of teachers’ GSL-competence served as the theoretical framework and basis for the performance-oriented test development. This model was devised in a previous project on the basis of document analysis, followed by an expert rating (Ehmke and Hammer 2018; Köker et al. 2015). Due to the premise that GSL competency is relevant for teaching all subjects in the mainstream classroom, the model consists of different partial competencies which are important for subject content integrated language facilitation and dealing adequately with multilingualism. The model illustrates the question of how subject content can be taught in such a way that the linguistic needs of all pupils, especially those with German as their second language, are taken into account (Carlson et al. 2018). In addition, the model depicts what teachers need to know to be able to act linguistically responsive. This knowledge refers to linguistic aspects, for instance, with regard to the specificity of specialized texts, aspects of second language acquisition and multilingualism as well as the didactic aspects of lesson planning and implementation. Since GSL competence is considered a generic competence in all subjects, it always refers to subject content.

The structural model discerns three dimensions: *subject-specific register*, *multilingualism*, and *didactics* (Ohm 2018). All dimensions are differentiated into subdimensions and further characterized by content-related facets.

The first dimension, *subject-specific register*, concerns language as a medium for classroom interaction and understanding technical concepts as well as a

learning object. It is assumed that all subjects have their specific registers that are essential for the students' knowledge construction and must therefore be taught explicitly in connection with subject contents (Schlepppegrell 2004).

The second dimension, *Multilingualism*, includes the two subdimensions *Second Language Acquisition* and *Migration* and focuses on the learning process (Ohm 2018). To support the (GSL) students according to their individual language learning levels, teachers require knowledge about the development of students' linguistic competence, for instance, regarding milestones of second language acquisition. Second language learning resp. the increasingly differentiated use of linguistic registers occur in the context of multilingualism. At the same time, multilingualism refers not only to external but also to internal multilingualism, such as the use of dialects. Therefore, the second dimension considers both linguistic diversity in school and dealing with heterogeneity.

The third dimension, *Didactics*, includes the subdimensions *formative assessment* in classrooms and *language facilitation* (Ohm 2018) and complements the first two dimensions. Based on knowledge about subject-specific registers and the specific characteristics of second language development, teachers plan and conduct their lessons. Consequently, the teaching process is at the focus of this dimension.

2.2 Indicators of Performance: Perception and Decision-Making

The situation-specific skills, *perception* and *decision-making* (Blömeke et al. 2015), serve as indicators of performance in GSL-relevant teaching situations. In research on teacher professionalization, perception often is defined by referring to the concept of *professional vision* by Goodwin (1994). He explains experts perceiving meaningful situations in their domain of expertise more selectively than novices do, as experts' perception has been trained to that effect in years of practice. Van Es and Sherin (2002) define two components of teachers' *professional vision*: *noticing* and *knowledge-based reasoning*. While the first term describes the filtered selective perception, the second refers to the ability to draw conclusions based on the situations perceived (Es and Sherin 2002). Therefore, experts are able to resort to their professional knowledge to classify the perceived into concepts and theories and react more precisely. In contrast, novices often deliver only a mere description of what they observed (Seidel et al. 2010).

Professional competence consists of professional knowledge and individual abilities, as well as motivational, volitional and social aspects and is dependent on domain-specific requests of action (Weinert 2001; Baumert and Kunter 2006).

Lindmeier (2013) captures the ability of *decision making* as *action-oriented* competency and thereby refers to the ability of teachers to use their professional knowledge spontaneously and immediately in situations (ibid.; see also Kuhn et al. in this volume).

The *DaZKom-Video* test instrument uses test items which pick up this theoretical structure. Video-stimuli present authentic teaching situations with specific *perception* and *decision-making* tasks to measure the German-as-a-Second-Language competence of the (pre-service) teachers as closely to performance as possible. The video-vignettes were carefully chosen after reviewing many teaching situations (video and audio recordings) and were validated for their authenticity by experts of universities and teacher experts in schools.

3 Research Questions

Based on the theoretical framework, three research questions will be answered in this paper:

1. How satisfactory is the *DaZKom-Video* test instrument's psychometrical quality?

Based on the evaluation of data of a first pilot study (Lemmrich et al. 2019), it can be assumed that the data will have a good item fit to the Rasch-model. Additionally, the discrimination in the first pilot study was satisfactory, even though the items were too difficult for the respective sample then.

2. Which dimensional structure can be determined for the test instrument?

Considering the results of the paper-pencil test of the previous project (*DaZKom*) it can be assumed that the construct *GSL competence* is structured multidimensionally (Hammer et al. 2015), also in a performance-oriented test. However, taking the results of the first pilot study into account (Lemmrich et al. 2019), it is possible that the theoretically determined dimensions do not appear clearly separated. As the two-dimensional analysis (*perception (1)* and *decision-making (2)*) showed strong correlations between these two dimensions in the first pilot study (ibid.) we expect a similar result in the second pilot study. Referring to *GSL competence*, three dimensions can be determined based on the theoretical assumptions (*subject-specific register (1)*, *multilingualism (2)*, *didactics (3)*, see *DaZKom-Model* in Köker et al. 2015, Ohm 2018; Section 2), which either appear as one-dimensional

or can be determined separately. In the first pilot study, results showed a medium correlation between dimension 1 and 3 and a low correlation between both of these and dimension 2. As the present pilot study was carried out with a revised test instrument and a different sample including more in-service teachers than the first pilot study, the results are expected to clarify the dimensionality of the construct.

3. Which correlations can be identified between the GSL competence of the test persons, the sociodemographic characteristics and learning opportunities?

The correlations examined give insight into the convergent and discriminant validity. We assume that test persons who had GSL-relevant learning opportunities are more likely to respond correctly. Additionally, experienced teachers are expected to pass the performance-oriented test more successfully than unexperienced teachers or pre-service teachers.

4 Methods

4.1 Test Instrument

The project team collected data in a standardized way at four universities and six schools in four German federal states (Berlin, Lower Saxony, North-Rhine Westphalia, Hamburg). The conductors first introduced the test closely following a manual, covering both the test procedure as well as the test's background. In their introductions, the conductors invariably pointed out the items' GSL focus. Afterwards, the test persons took two of four sub tests. Each subtest contained four video stimuli. Headsets and tablets were provided to enable the participants to take the performance-oriented test independently. After watching an example item including a sample solution, eight video-vignettes were shown, each followed by two items on the situation-specific skills *perception* and *decision-making*. The video-vignettes used as prompts take 30 seconds to three minutes, each dealing with authentic, GSL-relevant teaching situations that were taken from different subject lessons at secondary school classes. Subjects are Mathematics, German, Ethics and Science. All situations were matched both to the dimensions and subdimensions of the DaZKom-Model (Köker et al. 2015). Their adequacy, GSL relevance and typicality for the chosen dimension were assessed and approved in expert ratings with $N = 3$ university experts in the field.

To best simulate the immediacy and spontaneity of teacher behavior in real situations and thereby achieve a higher ecological validity, only open-response

items on the test persons' perception (*What do you perceive?*) and on their simulated (re-)action (a. *You are the teacher in this situation, how do you react word for word?*, or b. *How would you act in this situation if you were the teacher?*) were used, each asking for an oral response using the headsets' built-in microphones. To enable these kinds of open responses, the video vignettes usually closed with a student's statement the test persons could react to. At last, the test persons took an additional questionnaire both on their learning opportunities and sociodemographic data (Section 4.3).

4.2 Coding Manual

The coding manual, required to measure difference in quality in the test persons' responses reliably, was developed in a complex multi-step process. It included expert ratings ($N = 6$) with $n = 3$ researchers from the field (PhD and above) and $n = 3$ experienced teachers (different subjects, 30+ years of experience in schools while also working as instructors offering further training in GSL). The experts were taken through the same procedure as the regular test persons: they first watched the video vignettes before giving oral responses to the items on their situation-specific skills, independently and without a project member being present. Their responses were audio-taped and afterwards used for a qualitative content analysis employed to identify codes (Nimz, Hecker & Köker 2018). These codes were then ascribed to the scores 0, 1 and 2 in the project team. Based on this allocation, the actual item specific coding manuals were designed. Several rounds of optimization followed. The manuals were tested using data from the first pilot study conducted in Germany ($N = 137$; $n = 40$ in-service teachers, $n = 83$ pre-service teachers, $n = 11$ scientists). All responses were double coded by two independent raters. During the rating, doubtful cases were identified and discussed in the project team, more anchors added and necessary adjustments to the manuals made. Finally, the team achieved a satisfactory interrater agreement (Cohen's Kappa $\kappa = .76$). The final manuals include closed descriptions of the values in addition to many anchors. In the process of scoring the responses, only the best aspects in every response count, the rest is not taken into account. A maximum of two points can be reached in each item. For the analysis performed in this study, code 2 and code 1 were merged into code 1.

4.3 Additional Questionnaires: Sociodemographic Data and Learning Opportunities

After giving their oral responses, the test persons filled in additional questionnaires on sociodemographic characteristics and learning opportunities. The first questionnaire presented questions on gender, subjects of studies, teaching experience, etc. (6 items). The learning opportunities questionnaire contains two scales: One scale of 16 items on GSL-relevant topics that the test persons might have discussed in their studies/teacher training ($\alpha = 0.91$; Ehmke and Lemmrich 2018). Furthermore, it contains a scale of eight items ($\alpha = 0.83$, *ibid.*) on GSL-relevant actions the test persons might have taken before in their learning contexts (e.g. studies, further teacher training, etc.). Both scales use a five-point Likert scale ((1) *never* (2) *in one session*, (3) *in several sessions*, (4) *in a module*, (5) *in several modules*). Additionally, there were three single items that aimed for answers on GSL-specific teaching experience, additional qualifications and experience in research in the field of teaching multilingual learners.

4.4 Sample

The sample size of this study was $N = 184$ test persons, 79% female and 21% male; 51.2% were pre-service teachers (teacher students) from German universities and 39.6% were in-service teachers from different areas of Germany and different types of schools; 8.2% were “other”, such as scientists, teacher educators and students of other disciplines. The test persons studied/taught different kinds of subjects (Math, German, English, Physics, Biology, Chemistry, Geography, Social Studies/Science, History, Music, Arts, Religion, Politics, Sports); 25.6% had additional qualifications in the field of GSL/teaching multilingual learners (e.g. GSL-certificate); 88.6% had German as their first language, while 11.4% of the test persons had a different first language, which mostly were Turkish (3.8%) and Russian (1.6%).

4.5 Statistical Procedure

The collected data (coding of the *DaZKom-Video* test, sociodemographic data and learning opportunities) were analyzed with SPSS 25 (IBM 2018), including frequencies and information on the sample. The psychometric analysis of items and scales was carried out in *ConQuest* (Adams et al. 2015) on the basis of the

Rasch-model (Rost 2004). The GSL competence of the test persons was determined with the Weighted Likelihood Estimates (WLEs). WLEs take two factors into account: Firstly, the correct reply to an item depending on the respondents' abilities (in this case GSL competence); secondly, the item difficulty (Wilson 2005). Hence, the WLE does not map the actual responses of the test persons, but the probability of a correct answer based on the test persons' skills and the item difficulty (ibid.). To examine the dimensionality of the construct GSL-competence, a one, a two and a three-dimensional model were estimated with *ConQuest*.

5 Results

5.1 Psychometric Quality

Table 1 shows the results of the IRT-Scaling. The first criterion for psychometric quality is the item fit based on the Rasch-model, which is very good ($1.0 < MNSQ^3 < 1.25$; OECD 2005). A few items show a lower item fit ($Min = 0.88$; $Max = 1.10$), but at least the maximum item fit is located within the satisfactory range. The second criterion is discrimination, which indicates how reliable the instrument distinguishes between test persons with high and low ability estimates. Three items that showed a discrimination of under 0.3 were excluded from all analysis. Thereafter, the average discrimination of all items was satisfactory ($M = 0.43$; $SD = 0.09$). The overlap of the average item difficulty ($M = 0.88$; $SD = 0.78$) and the respondents' abilities ($M = 0.00$; $SD = 1.08$) is not fully satisfactory yet. Therefore, the items either are too difficult or the respondents' abilities not sufficient enough. The EAP-reliability of the one-dimensional scaling amounts $\alpha = 0.62$.

3 *Weighted Meansquare Fit Statistics* (MNSQ): indicates if the expected and the observed likelihood to respond correctly relate.

Table 1 Item and scale indices

	M	SD	Min	Max
Weighted fit MNSQ	1.00	0.05	0.88	1.10
Discrimination	0.43	0.09	0.30	0.58
Item difficulty	0.88	0.78	-0.53	2.79
WLE	0.00	1.08	-2.87	3.28

5.2 Verification of Dimensions

The dimensionality of the test instrument was tested with one-, two- and three-dimensional IRT models. Table 2 shows the fit-indices of the comparison of the three models: the deviance, the AIC, BIC and the CAIC. All four of them are criteria that give information about the goodness of fit of the estimated model. They take into account the empirical data as well as the model (Moosbrugger and Kelaiva 2007, p. 390). All these criteria should be as low as possible (*ibid.*). The BIC and the CAIC are at the lowest level for the one-dimensional model. The deviance and the AIC are at lowest for the three-dimensional model. The construct *GSL competence* therefore is likely to be a one-dimensional or three-dimensional construct rather than a two-dimensional construct. The reliability of the three-dimensional model was not satisfactory due to a very low number of items per dimension. The following analyses therefore are operated with data of the one-dimensional model.

Table 2 Comparison of dimensions

	N	Parameter	Deviance	AIC	BIC	CAIC
1-dimensional	176	30	2606	2666	2761	2791
2-dimensional	176	32	2607	2671	2773	2805
3-dimensional	176	35	2585	2655	2766	2801

Looking at the two-dimensionally modeled construct, the two assumed dimensions *perception* and *decision-making* (Section 2.2) highly correlate ($r = 0.97$, Figure 1). It therefore can be assumed that the construct *GSL competence* is not separable into these two dimensions. The three-dimensional modeled construct (Section 2.1) shows moderate correlations: dimension 1 (*subject specific register*) and dimension 2 (*multilingualism*) correlate by $r = 0.46$, dimension 2 and 3 (*didactics*) by $r = 0.51$ and dimension 1 and 3 by $r = 0.55$. This result indicates that these three theoretical dimensions could be separated within the construct *GSL competence*.

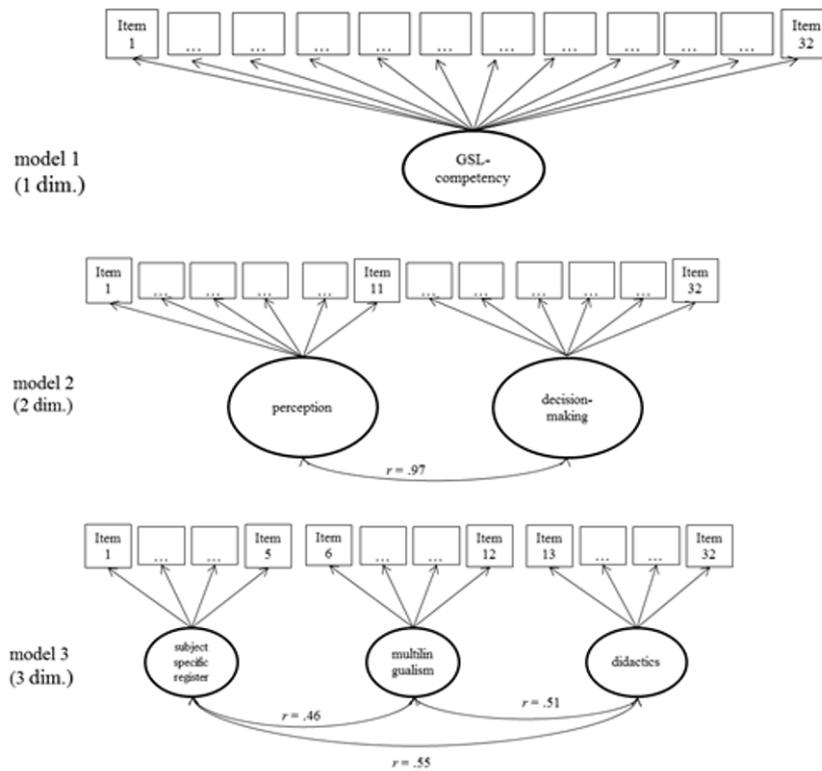


Figure 1 Comparison of models

5.3 Correlations between GSL Competence, Personal Characteristics and Learning Opportunities

Table 3 shows correlations between sociodemographic characteristics and GSL-relevant learning opportunities with the weighted likelihood estimates (WLE). There is no statistically significant difference between the female and male test persons probability to respond correctly. The results show a tendency that test persons with German as their first language are more likely to respond correctly, even though the effect is statistically non-significant. Teachers' responses are more often correct than those of students' (pre-service teachers). This effect is confirmed by the statistically significant correlation of teaching experience with the

WLE. Also, test persons who have a GSL-certificate, which is a specific qualification for teaching multilingual learners, are more likely to respond correctly than those without. Lastly, learning opportunities (sum score of both scales) correlate with the WLE: The more GSL-relevant learning opportunities test persons had, the more likely they were to respond correctly.

Table 3 Correlations between the results of the DaZKom-Video-test and sociodemographic data

	WLE
Gender (male = 0; female = 1)	0.14
Status (teacher student = 0; teacher = 1)	0.04
GSL – certificate ¹⁾ (no = 0, yes = 1)	0.20**
Teaching experience GSL ²⁾ (no = 0; yes = 1)	0.16*
Research GSL ³⁾ (no = 0; yes = 1)	0.09
Scale GSL-learning opportunities ⁴⁾	0.18*

* The correlations are statistically significant at a level of .05 (two-sided).

** The correlations are statistically significant at a level of .01 (two-sided).

- 1) additional certificate that proves the attendance of a specific course on teaching students with German as a second language
- 2) experience in teaching students with German as a second language
- 3) research experience in the field of German as a second language/teaching multilingual learners
- 4) learning opportunities in the field of German as a second language/teaching multilingual learners

6 Discussion and Outlook

This study has been carried out to evaluate the psychometric quality and the dimensionality of the performance-oriented test instrument. Additionally, correlations of GSL competence with sociodemographic data and learning opportunities were examined. The main results were the following:

1. The test instrument shows a good fit to the Rasch-model. Three items were excluded from the test due to a low discrimination value (< 0.3). The discrimination was satisfactory. At the same time, the test did not fit perfectly to the ability of the test persons, which indicates that the items were too difficult.
2. The dimensional analyses showed the best model fits for the one-dimensional and the three-dimensional models. The three-dimensional structure represents

the theoretical structure of the GSL assessment framework. The two dimensions of perception and decision making could not be separated statistically from each other.

3. The correlations between the WLE of the test persons concerning GSL competence and the sociodemographic data/learning opportunities indicate a higher likelihood of responding correctly for test persons who are experienced teachers. Teaching experience as well as more learning opportunities correlated statistically significant with the WLE.

The psychometric analysis showed that some items were still too difficult for the test persons. However, compared to the first pilot study (Lemmrich et al. 2019), the item fit increased. After the first evaluation, the test instrument was revised: Stimuli that did not work were excluded and more context information was added to increase the specificity of the situations. Additionally, the current sample consists of half students (pre-service teachers) and half teachers, whereas the first pilot study was carried out with a majority of students, who are less experienced than teachers and therefore not as likely to perform successfully in a performance-oriented test instrument (Section 2).

Concerning the dimensionality of the construct, there is no evidence for a two-dimensional construct. *Perception* and *decision-making* cannot be determined separately for the construct. This result is in line with results of a first pilot study (Lemmrich et al. 2019) and is compliant with theories of cognition psychology as well as other studies on professional vision and performance (Niesen 2018; Hommel et al. 2019). Another possible explanation for the results is the fact that the items corresponding to either *perception* or *decision-making* both demand answers that are based on one and the same video vignette. The results indicate a one-dimensional construct in a performance-oriented environment. However, in the three-dimensional modeling (three dimensions of GSL competence: *subject specific register*, *multilingualism* and *didactics*; Section 2), the three dimensions show only medium correlations, which indicates that a differentiated analysis might be reasonable. Further analysis and evaluation will examine the dimensionality in detail with the final test instrument and a larger sample with a higher number of teachers. The standardization study will examine a sample of $N = 300$ ($N_{\text{teachers}} = 150$) with a revised test instrument (12 final video vignettes in one test version).

Looking at the correlations of GSL-competence with sociodemographic data and learning opportunities, the results comply with the theoretical background. Experience and learning opportunities correlate with the WLEs of the test persons: test persons with teaching experience as well as with additional qualifica-

tions in the field of GSL or learning opportunities are more likely to respond correctly to the items.

The instrument contains 16 vignettes that show very short teaching examples. It therefore offers only a selection of possible contexts and cannot portray the broad variety and complexity of everyday teaching situations. This study aimed at evaluating the psychometric quality and dimensionality of the test instrument and did not analyze descriptive values. Future studies (with an equally high psychometric quality) will take this into account and identify GSL-relevant learning opportunities to then comprehensively train teachers in this field and prepare them optimally to teach in linguistically diverse classrooms.

Funding

DaZKom-Video was funded by the KoKoHs-program of the German Federal Ministry for Education and Research (01PK16001A, 01PK16001B).

References

- Adams, R. J., Wu, M. L., & Wilson, M. R. (2015). ACER ConQuest: Generalised Item Response Modelling Software [Computer software]. Version 4. Camberwell, Victoria: Australian Council for Educational Research.
- Aufschnaiter, C. v., & Blömeke, S. (2010). Professionelle Kompetenz von (angehenden) Lehrkräften erfassen – Desiderata. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, pp. 361–367.
- Baumann, B. (2017). Sprachförderung und Deutsch als Zweitsprache in der Lehrbildung. In M. Becker-Mrotzek, P. Rosenberg, C. Schroeder & A. Witte (Eds.), *Deutsch als Zweitsprache in der Lehrbildung* (pp. 9–26). Münster: Waxmann.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), (pp. 469–520).
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie*, 223 (1), pp. 3–13.
- Carlson, S. A., Köker, A., Rosenbrock-Agyei, S., Ohm, U., Koch-Priewe, B., Hammer, S., Fischer, N., & Ehmke, T. (2018). DaZKom – a Structure Model of Pre-service Teachers' Competency for Teaching German as a Second Language in the Mainstream Classroom. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Eds.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (pp. 261–283). Münster and New York: Waxmann.
- Cummins, J. (2001). *Negotiating Identities: Education for Empowerment in a Diverse Society*. Second Edition. Los Angeles: California Association for Bilingual Education.

- Ehmke, T., & Hammer, S. (2018). Skalierung und dimensionale Struktur des DaZ-Kom-Testinstruments. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Eds.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (pp. 129–148). Münster: Waxmann.
- Ehmke, T., & Lemmrich, S. (2018). Bedeutung von Lerngelegenheiten für den Erwerb von DaZ-Kompetenz. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Eds.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (pp. 201–219). Münster: Waxmann
- Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96 (3), pp. 606–633.
- Hammer S., Carlson, S. A., Ehmke, T., Koch-Priewe, B., Köker, A., Ohm, U. Rosenbrock, S., & Schulze, N. (2015). Kompetenz von Lehramtsstudierenden in Deutsch als Zweitsprache. Validierung des GSL-Testinstruments. *Zeitschrift für Pädagogik*, Beiheft 61, pp. 32–54.
- Hommel, B., Müsseler, J., Aschersleben, G., Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. In: *Behavioral and Brain Sciences* 24, H. 5, S. 849–878.
- Kimmanen, A., Alisaari, J., & Kallioniemi, A. (2019). In-service and pre-service teachers' orientations to linguistic, cultural and worldview diversity. In: *Journal of Teacher Education and Educators*, 8 (1), 35–54.
- Köker, A., Rosenbrock-Agyei, S., Ohm, U., Carlson, S. A., Ehmke, T., Hammer, S., Koch-Priewe, B., & Schulze, N. (2015). DaZKom – Ein Modell von Lehrerkompetenz im Bereich Deutsch als Zweitsprache. In B. Koch-Priewe, A. Köker, J. Seifried & E. Wuttke (Eds.): *Kompetenzerwerb an Hochschulen: Modellierung und Messung. Zur Professionalisierung angehender Lehrerinnen und Lehrer sowie früh-pädagogischer Fachkräfte* (pp. 177–206). Bad Heilbrunn: Klinkhardt.
- Lemmrich, S., Bahls, A., & Ehmke, T. (under review). Effekte von mündlichen versus schriftlichen Antwortformaten bei der performanznahen Messung von Deutsch-als-Zweitsprache (DaZ)-Kompetenz – eine experimentelle Studie mit angehenden Lehrkräften.
- Lemmrich, S., Hecker, S.-L., Klein, S., Ehmke, T. Köker, A., Koch-Priewe, B., & Ohm, U. (2019). Performanznahe und videobasierte Messung von DaZ-Kompetenz bei Lehrkräften. Skalierung und dimensionale Struktur des Testinstruments. In T. Ehmke, P. Kuhl & M. Pietsch (Eds.), *Lehrer. Bildung. Gestalten. Beiträge zur empirischen Forschung in der Lehrerbildung* (pp. 188–202). Weinheim und Basel: Beltz Juventa.
- Lindmeier, A. (2013). Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen. In U. Riegel & K. Macha (Eds.), *Videobasierte Kompetenzforschung in den Fachdidaktiken. Fachdidaktische Forschungen, Vol. 4*, pp. 45–62. Münster: Waxmann.
- Moosbrugger, H., & Kelava, A. (2007). Testtheorie und Fragebogenkonstruktion. Heidelberg: Springer Medizin Verlag.
- Nimz, K., Hecker, S. – L., & Köker, A. (2018). Videobasierte Messung von DaZ-Kompetenz bei Lehrkräften. In C. Caruso, J. Hofmann, A. Rohde & K. Schick (Eds.), *Sprache im Unterricht. Ansätze, Konzept, Methoden* (pp. 439–452). Trier: WVT.
- Niesen, H. (2018). Förderung mehrsprachigkeitssensibler professioneller Handlungskompetenz angehender Englischlehrkräfte. In: *Zeitschrift für Interkulturellen Fremdsprachenunterricht: Didaktik und Methodik im Bereich Deutsch als Fremdsprache* 23, H. 1, S. 121–134.
- OECD (2005): PISA 2003. Technical report. Paris: OECD.

- Ohm, U. (2018). Das Modell von DaZ-Kompetenz bei angehenden Lehrkräften. In T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Eds.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (pp. 73–91). Münster: Waxmann.
- Rost, J. (2004). Testtheorie Testkonstruktion (second edition). Bern: Huber.
- Schmölzer-Eibinger, S. (2013). Sprache als Medium des Lernens im Fach. In M. Becker-Mrotzek, K. Schramm, E. Thürmann & H. J. Vollmer (Eds.), *Sprache im Fach. Sprachlichkeit und fachliches Lernen* (pp. 25–40). Münster: Waxmann.
- Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer“ – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE. In M. Bayrhuber, T. Leuders, R. Bruder & M. Wirtz (Eds.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (pp. 296–306). Weinheim und Basel: Beltz.
- Thürmann, E., & Vollmer, H. J. (2013). Schulsprache und Sprachsensibler Fachunterricht: Eine Checkliste mit Erläuterungen. In C. Röhner & B. Hövelbrinks (Eds.), *Fachbezogene Sprachförderung in Deutsch als Zweitsprache. Theoretische Konzepte und empirische Befunde zum Erwerb bildungssprachlicher Kompetenzen* (pp. 121–233). Weinheim: Beltz Juventa.
- van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, 10 (4), pp. 571–596.
- Schleppegrell, M. J. (2004). The language of school: A functional linguistics perspective. Mahwah: Lawrence Erlbaum Associates.
- Weinert, F. E. (2001). Concept of Competence: A Conceptual Clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and Selecting Key Competencies* (pp. 45–66). Göttingen: Hogrefe.
- Wilson, M. (2005). Constructing Measures. An Item Response Modeling Approach. New York and London: Psychology Press, Taylor & Francis Group.

10. Artikel 4: Performance-Oriented Measurement of Teachers' Competence in Linguistically Responsive Teaching: Connection of LRT Competence with Relevant Learning Opportunities, Personality Factors and Beliefs

Lemmrich, S. & Ehmke, T. (eingereicht, Pre-Print). Performance-Oriented Measurement of Teachers' Competence in Linguistically Responsive Teaching: Connection of LRT Competence with Relevant Learning Opportunities, Personality Factors and Beliefs. *Language Awareness*.

Der Beitrag wird hier als Pre-Print veröffentlicht.

Performance-Oriented Measurement of Teachers' Competence in Linguistically Responsive Teaching: Connection of LRT Competence with Relevant Learning Opportunities, Personality Factors and Beliefs

0. Abstract

This paper presents a quantitative study on teachers' competence in linguistically responsive teaching (LRT). To assess performance-oriented competence rather than professional knowledge, the test instrument uses video vignettes and corresponding items based on situation-specific skills (*perception* and *decision making*). The participants are required to respond orally. The research questions focus (a) on performance-oriented competence and (b) on the connection between (preservice and inservice) teachers' LRT competence and individual characteristics such as subjects of studies/teaching, teaching experience in LRT, research experience in this field, LRT-relevant learning opportunities and beliefs concerning multilingualism in school and teaching. To validate the oral response format, we used an additional questionnaire on personality factors (*big five*). The findings illustrate a higher test score for participants who are experienced in teaching and those who are studying/teaching English. Also, participants with positive beliefs are more likely to perform better in the test. Personality factors do not show a connection to LRT competence. Based on the results, an oral response format does not benefit or disadvantage certain personalities and therefore is likely to be a fair test format. We conclude that, besides linguistic knowledge and language awareness, positive beliefs seem to play a fundamental role in teacher education concerning LRT.

1. Introduction

The increasing number of students learning English as their second language (Lucas & Villegas, 2013) and, in the context of this study, students learning German as their second language in Germany (Wernicke et al., 2021), leads to a culturally and linguistically diverse student population in schools. (Lucas & Villegas, 2013). As teachers are responsible for enabling all students to access content, differentiation becomes increasingly important (Cho et al., 2020). The idea of decreasing intolerance towards linguistic diversity in society through education was already established 40 years ago with the concept of language awareness. That also involves teaching teachers beforehand and providing appropriate material for them to use in the classroom (Hawkins, 1999; Svalberg, 2016). Since then, much research has been done and led to concrete suggestions of how language awareness should be included into teaching (Svalberg, 2016). However, this brings with it the need to prepare teachers so that they can take the

responsibility and teach linguistically and culturally responsively (Cho et al., 2020). Focusing on the role of language in content classrooms, on teachers' competence to embrace language support in their teaching and on relevant learning opportunities that teachers need to be prepared to teach linguistically responsive, is the starting point of our research. Additionally, and in contrast to many other studies that use a sample of preservice teachers/ students due to accessibility (Svalberg, 2016) , our research examines a sample of preservice and inservice teachers and a small number of teacher educators and researchers.

The research was conducted within the project *DaZKom-Video*. The goal of the project is to enable a performance-oriented measurement of teachers' competence to teach linguistically responsively (We hereafter use the terms *Linguistically Responsive Teaching (LRT)* and *LRT-competence*). The aim is to capture performance, rather than to assess declarative knowledge. Therefore, we demand spontaneous decisions for action in the assessment. For this purpose, video vignettes were developed that show LRT-relevant teaching situations to serve as stimuli in the test instrument. Based on our assumption that oral responses increase spontaneity, authenticity and therefore enable a measurement close to real performance, the response format in the LRT-test requires oral responses from the participants (Lemmrich, Bahls, & Ehmke, 2020).

Focusing on the validation of the test instrument, the overall research question is: Which connections exist between LRT competence, individual and academic characteristics, LRT-relevant learning opportunities, personality factors (*big five*, Schupp & Gerlitz, 2014) and beliefs concerning multilingualism in school and teaching (Fischer, 2020)? Within the framework of this study, we also aim at conclusions about the test fairness regarding the oral response mode: It is unclear whether certain personality traits affect the ability in this innovative setting, which would in turn falsify the test results.

Firstly, this paper provides an overview of the underlying theoretical background and the current state of research. The focus here lies on the construct of LRT competence and on performance-oriented measurement of competence. Secondly, our validation approach is presented here. The theoretical framework is followed by the research questions, method and results as well as the discussion.

2. Theoretical Framework

The theoretical framework introduces our understanding of *LRT competence* and the definition of *competence* and *performance* we refer to.

2.1. LRT competence

As inclusion of all English language learners into general education classrooms increases, teachers of all subjects need to be prepared to teach linguistically and culturally responsively (Lucas & Villegas, 2013). The same applies to German language learners in German general education classrooms (Berkel-Otto et al., 2021). In this study, we follow the idea that linguistically and culturally responsive teaching applies to all language learners in the mainstream classroom (Lucas et al., 2008). We refer to second language learners, to different language registers, regional dialects and other varieties or accents (*inclusive understanding of multilingualism*, Schroedler, 2021) and hereafter use the term multilingual learners (MLLs) for the linguistically and culturally diverse student population.

Linguistically responsive teaching (Lucas & Villegas, 2013) is a concept that defines types of pedagogical knowledge and skills teachers need to teach multilingual learners in the mainstream classroom, e.g. *linguistic knowledge, principles of second language learning and understanding of academic language*. The concept also includes fundamental orientations, linguistically responsive teachers should follow and thus set a strong focus on valuing linguistic diversity and having positive beliefs towards linguistic diversity. To be prepared, teachers not only need to acquire linguistic knowledge, they also need to reflect on their beliefs and be able to diagnose the language of the classroom (Lucas & Villegas, 2013). Besides linguistically responsive teaching, some studies explicitly focus on culturally responsive teaching (Acquah et al., 2016; Gay, 2010; Villegas & Lucas, 2002): To create a meaningful learning situation to all students in a classroom, teachers should address cultural knowledge and experiences of students. By doing so, students are enabled to engage, not only contentwise but also socially and emotionally (Acquah et al., 2016). In this study, we focus on linguistically responsive teaching. Nevertheless, we are very aware that both, linguistically and culturally responsive teaching are linked (Lucas & Villegas, 2011). As the conception of linguistically responsive teaching has its' roots in the framework of culturally responsive teaching (Villegas & Lucas, 2002), this aspect is still represented although not explicitly addressed.

Culturally and linguistically competent teachers improve their students' academic success (Gay, 2010; Ladson-Billings, 2009; Lucas & Villegas, 2013). In contrast, teachers can disadvantage their students by not being sufficiently prepared to teaching content and academic language simultaneously (Villegas, 2018). The lack of preparation can as well lead to insecurities, because teachers' are often unable to teach in a confident and linguistically responsive way (Villegas, 2018). Therefore, they do not take the responsibility and consequently miss out on supporting their students (Villegas, 2018). Also, unprepared teachers

often have a deficit view on MLLs, which in turn has a (negative) impact on their students learning success (Villegas, 2018). In Germany, a deficit view on MLLs is often associated with the concept of the *monolingual habitus* of the German school system (Gogolin, 2008).

In the German research context of this study, the term *German as a second language (GSL) competence* is prominent. The term originates from studies on German as a second or foreign language (Berkel-Otto et al., 2021). Meanwhile, we know that language awareness and linguistically responsive teaching are substantial for all students, especially in regard to academic language at school (Köker, 2018; Lucas & Villegas, 2013; Schleppegrell, 2009). GSL competence refers to teachers' competence in teaching their subject content in a way that also takes linguistic aspects into account, as all students should have access to learning content, regardless of their language skills (Carlson & Präg, 2018). The construct of GSL competence is defined in a structure model and includes three main dimensions: *subject-specific language*, *multilingualism* and *didactics* (Carlson et al., 2018; Ohm, 2018). All three dimensions subdivide into different subdimensions and facets (Carlson et al., 2018; Ohm, 2018).

Although GSL competence refers to a model that defines different dimensions and facets of teachers' competence in this field (Carlson et al., 2018; Ohm, 2018) and LRT is a concrete set of knowledge and skills teachers need to be prepared (Lucas & Villegas, 2011), both have a high level of conceptual overlap, which is why we decided to use the term *LRT competence* in this study. A central agreement of both concepts is the role of subject-specific and academic language: Both should be addressed in content teaching as "language is a key element for content learning as content is acquired and expressed through language" (Berkel-Otto et al., 2021, p. 87). Academic language is more abstract and technical and its' meaning often lies in the language itself (Lucas et al., 2008), while conversational language in contrast mostly occurs in situations where people talk about own experiences and is supported by gestures and facial expressions which help to understand the meaning (Lucas et al., 2008). Also, the purposes of academic language use differ from those of conversational language. Purposes can be discussing different point of views on an academic rather than a personal topic, or making hypothesis about certain topics. This can be extremely challenging for pupils, also if they are fluent in everyday conversation, because different linguistic or stylistic devices are needed (Lucas et al., 2008).

The goal of this study is to measure teachers' competence in LRT in a performance-oriented format and to evaluate LRT-relevant learning opportunities to assess, which of these make competent linguistically responsive teachers, as teacher education in the field of LRT is still highly heterogenous in Germany (Berkel-Otto et al., 2021). For this purpose, we developed a

questionnaire on various LRT-relevant topics (table 4, chapter 5) along the theoretical background on LRT, that might have been addressed in participants' teacher trainings (Lucas & Villegas, 2013; Ohm, 2018) (see 4.1. Instrument).

2.2. Measuring Teachers' Professional Competences

In this study we follow the concept *competence as a continuum* (Blömeke, Gustafsson, & Shavelson, 2015, Figure 1). Accordingly, professional knowledge is part of teachers' individual dispositions, besides affect motivation and beliefs. The situation-specific skills such as *perception* or *decision making* are mediating between the dispositions and the performance (Blömeke, Gustafsson, & Shavelson, 2015): Teachers dispositions influence their situation-specific skills (*perception, interpretation, and decision making*) and affect their performance.

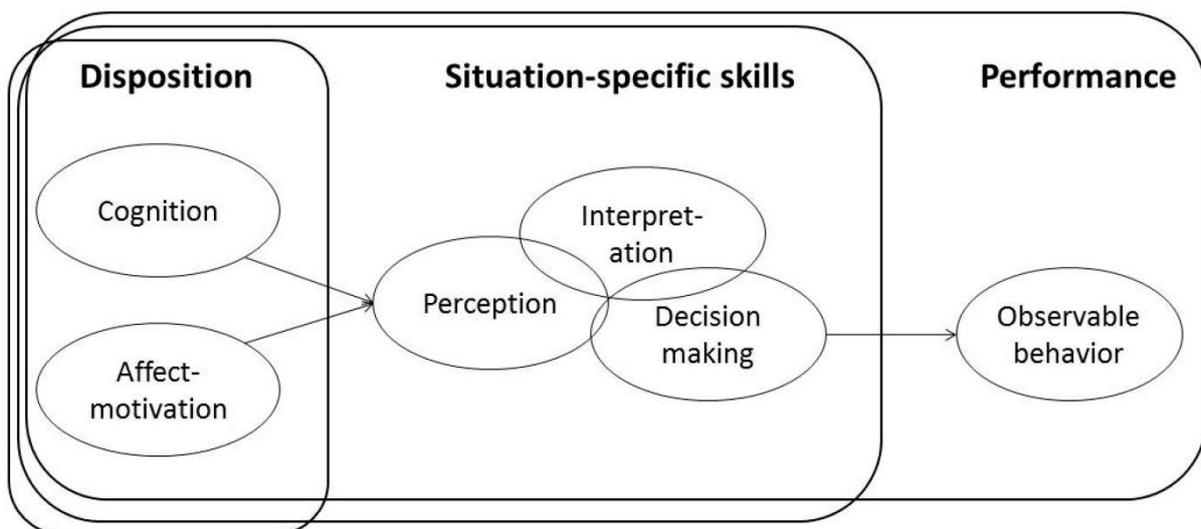


Figure 1. Competence as a Continuum (Blömeke, Gustafsson, & Shavelson, 2015, S.9).

The performance (observable behavior) then allows conclusions about the level of competence (Blömeke, Gustafsson, & Shavelson, 2015). The idea of competence being a continuum applies bi-directionally: Also teaching experience can serve as a starting point for new knowledge, which is processed by situation-specific skills (deliberate reflection) and then transformed into new beliefs and knowledge (Kramer et al., 2020; Meschede et al., 2017; Santagata & Yeh, 2016).

Not only educational research but also research on cognition and on expertise have focused on perception (in teacher education referred to as *professional vision* (Goodwin, 1994)) as a key link between dispositions (e.g. knowledge) and performance (Star & Strickland, 2008). Only the perception of relevant teaching situations enables teachers to perform in the classroom (van

Es & Sherin, 2002). Novice teachers, in contrast to experienced teachers, often only describe what they observe, whereas experts identify the relevant event in a complex teaching situation with different events and perspectives and relate the observed with concepts and theories they know. As a result they decide more quickly and accurately on actions (Blömeke, König, et al., 2015; Santagata & Yeh, 2016; van Es & Sherin, 2002).

To assess LRT-relevant performance rather than dispositions, we used video vignettes in our assessment that showed LRT-relevant teaching situations as they illustrate the complexity of teaching more appropriately (Blömeke, König, et al., 2015; Casale et al., 2016). Each vignette was related to two items based on the situation-specific skills *perception* and *decision making*. The test requested oral responses as we assume that an oral response format captures an actual performance more realistically than written responses as they represent a natural action. Oral responses create time pressure, the participants are deprived of time to structure their thoughts and to ask for learned knowledge (Lemmrich, Bahls, & Ehmke, 2020).

2.3.Validation Approach

Validation is a process that aims at checking the plausibility, consistency and completeness of the assumptions underlying the test instrument and the interpretations based on test results (Kane, 2013). Kane identifies three main conclusions for the interpretation of test results: Evaluation, generalization, and extrapolation. Extrapolation implies an interpretation of test results related to the future performance of subjects in "real life" (Kane, 2013). It is therefore assumed that the performance of participants in the LRT-test of this study can be used as an indicator for the actual performance as a teacher in content classrooms. This study will examine discriminant validity (as part of construct validity) to ensure that measurements of one characteristic of this test instrument do not allow conclusions to be drawn about another independent characteristic (Michalos, 2014). Since oral responses are used in this test instrument for the purpose of performance-related measurement, a distinction in mean test scores should be ensured between competent participants and those who are not. In this study personality factors are included as covariates to test if certain types of personalities have advantages in responding orally. If so, this would cause distortion in the measurement of competence. An existing test instrument is used for this purpose (Schupp & Gerlitz, 2014), which focuses on the *big five* personality factors (Rammstedt & Danner, 2017): *extraversion* (outgoing, assertive, active), *agreeableness* (kind, gentle, trusting), *conscientiousness* (thorough, responsible, hardworking), *neuroticism* (anxious, fearful, moody) and *openness* (imaginative, adventurous, creative) (Goldberg, 1990; Rammstedt & Danner, 2017). The

recording of the dimensions has proven to be a reliable predictor in numerous studies of different objectives and research fields and has been used in educational research studies such as the German National Education Panel (NEPS) (Rammstedt & Danner, 2017). With the *big five* a broad spectrum of human experience and behavior can be described with few dimensions (Rammstedt & Danner, 2017).

3. Research Questions

The present study examines the relationship between the results of the *LRT*-test and various other characteristics of the participants and validation features. Accordingly, we focus on the following three research questions:

- 1.) Is there a correlation between the *LRT* competence, and individual and academic characteristics, and the use of *LRT*-relevant learning opportunities?

Participants who have German as a second language rather than their first are expected to perform worse on average in the *LRT*-test. The studies PISA 2012 (OECD, 2014) and PIAAC 2013 (Programme for the International Assessment of Adult Competencies, (OECD, 2013)) confirm this assumption as they show that respondents (in this case students) with German as their first language have higher language skills on average. Referring to results of these two studies, female participants are expected to achieve higher results in the *LRT*-test than male. One reason can be that girls learn words earlier than boys and start speaking at a younger age (Chipere, 2014). They are also more likely to write and read before they start school while boys tend to develop more weaknesses in reading and/or writing than girls (Chipere, 2014; Halpern, 2013; Stanat et al., 2018). And the described differences between girls and boys continue or reoccur beyond infant age in terms of e.g. punctuation or comprehension (Chipere, 2014; Maccoby & Jacklin, 1978). Female respondents are also more likely to have positive beliefs about linguistic and cultural heterogeneity in school and teaching (Fischer & Ehmke, 2019). Moreover, positive beliefs concerning multilingualism in school and teaching are associated with high levels of *LRT* competence (Hammer et al., 2016).

Participants who study German as a subject (preservice teachers) or teach German (inservice teachers) are expected to achieve higher test scores in the *LRT*-test than other participants, because they are offered more *LRT*-relevant learning opportunities and experience in the field of language education/teaching multilingual learners (e.g. linguistic skills). *LRT*-relevant learning opportunities of participants, such as in specific seminars on *LRT*, should lead to higher test scores as well. The *DaZKom*-paper-pencil-test showed similar results (Ehmke &

Lemmrich, 2018; Hammer et al., 2015): Participants with LRT-relevant learning opportunities achieved a higher test score in the paper-pencil-Test (Ehmke & Lemmrich, 2018). We also assume that specific teaching experience in the field of LRT should lead to higher LRT competence.

2.) Which connections can be identified between the LRT-test results and the participants personality factors?

The format of the LRT test requires an oral response of the participants with a video vignette as a stimulus, showing a LRT-relevant teaching situation they have watched before. Therefore, the participants need so answer spontaneously. This also creates time pressure and the participants are deprived of time to structure their thoughts and to think of what they know. Besides the level of LRT competence, personality traits like a higher degree of extraversion or openness might affect the response behavior. One could expect that participants who are more open or extroverted find it easier to answer orally and to be audiotaped and therefore perform better in the test, compared to introverted participants.

3.) Which relationship can be identified between beliefs of participants concerning multilingualism in school and teaching and the LRT-test results?

Beliefs are fundamental in professional development (Blömeke, Gustafsson, & Shavelson, 2015) and beliefs on multilingualism in the classroom influence future teachers' actions (Fischer & Lahmann, 2020; Pajares, 1992). Other than GSL competence, the concept of LRT includes orientations (which are influenced by beliefs (Lucas & Villegas, 2011; Richardson, 1996)): linguistically responsive teachers understand that language, culture and identity are intertwined, they value linguistic diversity, and they develop awareness of language (Lucas & Villegas, 2011, 2013) Therefore, we decided to assess beliefs on multilingualism in school and teaching, which will also help to validate the instrument. We expect that positive *beliefs concerning multilingualism in school and teaching* (Fischer & Ehmke, 2020) and teaching are linked to high levels of LRT competence. Other studies have already come to similar conclusions: Students who showed high levels of LRT competence in the paper-pencil-test also had positive beliefs concerning multilingualism in school (Hammer et al., 2016). This is also supported by the fact that other studies have already found a high mediation effect between LRT-relevant learning opportunities and beliefs. In turn, the number of LRT-relevant learning opportunities is related to the measured LRT competence (Fischer & Ehmke, 2019). Also, studies on beliefs have shown, that “(preservice) teachers may only pay attention to those

classroom events that correspond with their existing beliefs about teaching, subject matter and students' learning” (Meschede et al., 2017, p. 161). Beliefs are a fundamental factor in the acquisition and processing of knowledge (Pajares, 1992; Ricart Brede, 2019). These assumptions suggest that participants with positive beliefs concerning multilingualism in school and teaching are more likely to also have more knowledge, and perceive as well as perform better in the test.

4. Method

The following chapter introduces the LRT- test instrument, the additional questionnaire instruments as well as the procedure and gives information on the sample.

4.1. Instruments

LRT- test instrument

The performance-oriented measurement of LRT competence was conducted with a video-based test instrument, called LRT-test. The test instrument contains 12 video vignettes with two corresponding items each:

- a) What do you perceive?
- b) If you were a teacher in this situation, how would you act?

The test was carried out in a computer-based setting on tablet computers (in the App *Limesurvey*) with headsets. The duration of the test was approximately 60 minutes. In total, the LRT-test contains 24 items and has a reliability of $\alpha=0.76$. The responses of the test subjects were given orally, were recorded and then coded according to a coding manual (0=false, 1=unspecific elements of LRT, 2=specific LRT-aspects, 3=situation-specific LRT-relevant). The coding manual provides specific indications for every vignette. The video vignettes were carefully selected during the project. The vignettes show 2-3 minutes long LRT-related teaching situations e.g. showing the teacher who verbally explains a mathematical topic or students in the classroom discussing about to form groups based on their migration background. The test items in connection with the video vignettes were tested in two successive pilot studies (Lemmrigh et al., 2019; Lemmrigh, Hecker, et al., 2020). They were also tested in an expert rating for authenticity, relevance and fit to the theoretical framework of LRT. Fitting here refers to the fit and assignment of the individual vignettes to the structural model for LRT competence. The items were also divided into three subscales that correspond to the dimensions of the model: *subject specific register* (15 items), *multilingualism* (4 items) and *didactics* (5 items)

(Lemmrich, Hecker, et al., 2020). However, in this study we combined all items in a one-dimensional measurement scale. The corresponding coding manual for the video test was developed with the help of expert interviews and qualitative content analysis and was finalized in elaborate revision loops (Hecker & Nimz, 2020). Therefore, experts completed the LRT-test under the same conditions as the later participants. The responses were evaluated qualitatively and content-analytically and formed the basis of the coding manual. For each score, a detailed description including anchor examples from prior pilot studies can be found in the manual (Hecker, Falkenstern, Lemmrich, & Ehmke, 2020; Lemmrich, Hecker, et al., 2020). On average, a satisfactory interrater agreement is achieved (Cohen's Kappa $\kappa=0.76$) (Hecker, Falkenstern, Lemmrich, & Ehmke, 2020)

Individual and academic characteristics and LRT-relevant learning opportunities

The questions on individual and academic characteristics cover information on gender, first language, subjects of studies and/or teaching and additional qualifications. The scale of LRT-relevant learning opportunities (22 items, $\alpha=0.90$) contains LRT-relevant topics (e.g. *linguistics*, *scaffolding*; table 4), which might have already been addressed in participants' courses/training, and uses a five-point-likert-scale as response format (not at all (1) - in several courses (5)). Descriptive statistics are provided in Table 1 and 2.

Personality traits (big five)

In order to record personality traits of the participants, this study uses an established *Big Five*-test instrument (Schupp & Gerlitz, 2014). There are different tests, all of which are designed to capture the dimensions of personality. The difference usually lies in the number of items. In general, many questions on a personality trait lead to a more precise depiction of the personality structure. It has been shown that a test version with 5 items already achieves appropriate levels, but 10 or 15 items lead to a psychometrically better result. In this study the Big Five Inventory-SOEP (Schupp & Gerlitz, 2014) questionnaire (15 items) is used. It contains three items for each of the five characteristics, arranged on a seven-level Likert scale (does not apply at all (1) - applies fully (7)). The five subscales have the following reliabilities: Cautiousness $\alpha=0.66$, Extraversion $\alpha=0.79$, Compatibility $\alpha=0.53$, Openness $\alpha=0.64$ and Neuroticism $\alpha=0.71$. Means and standard deviations of the sum scores for the sub-scales are provided in Table 2.

Beliefs regarding Multilingualism in School and Teaching

In order to measure the participants' beliefs concerning multilingualism in school and teaching, two scales were used (Fischer & Ehmke, 2020). The instrument was developed based on literature (Fischer & Ehmke, 2019), was extensively validated (Fischer, 2020) and has a good psychometric quality (Fischer & Ehmke, 2019). The response format is presented as a four-level Likert scale (do not agree at all (1) - agree completely (4)). The scales are: (1) Multilingualism in content classroom (7 items, $\alpha=0.83$) and (2) Responsibility for language support in content classroom (9 items, $\alpha=0.72$). Descriptive statistics are provided in Table 2.

4.2. Procedure

Data collection

The data collection took place between April and July 2019. The use of the test instrument was always supervised by project staff. In addition, a test leader instruction ensured a standardized test procedure. Since the video items were answered orally, the participants were placed in different parts of the room (one test run always contains a maximum of 30 participants), so that there was sufficient distance between all participants. Additionally, shielding headsets with built-in microphones were used. All participants worked on all test parts. First the video-based LRT-test was answered. The video items were presented in a randomly changing order to avoid sequence effects. All other question blocks then followed the same order: individual and academic characteristics, LRT-relevant learning opportunities, personality traits (Big Five) and beliefs concerning multilingualism in school and teaching. The questionnaires in addition to the video test were presented in a multiple-choice format and were technically implemented in such a way that answers had to be given in order to move on to the next question.

Statistical procedure

SPSS 26 (IBM 2019) and *ConQuest* (Adams et al., 2015) were used for item and scale analysis in this study. SPSS was used for the evaluation of the frequencies and correlations of the various scales: results of the *DaZKom*-Video Test, individual and academic data, LRT-relevant learning opportunities, personality traits and beliefs. The data was scaled according to the Rasch model (Rost, 2004) and further analysis, such a latent regression analysis, was carried out in *ConQuest*. Persons' measurement values were determined using the persons' ability estimator WLE (Weighted Likelihood Estimates), which was calculated from the data of the *LRT*- Test. The answering of an item depends on the one hand on the abilities (in this study, LRT competence)

of the participants and on the other hand on the level of item difficulty (Wilson, 2004). The person's ability estimators take both into account. WLE therefore is not about the actual response scores, but about the probability of achieving a correct response based on the person's ability and item difficulty (Wilson, 2004).

Sample

The data collection took place at 16 different locations across Germany. We collected data in every region: east, west, north and south as well as in urban and rural areas to ensure that there are no location related effects on the findings. Data was collected in preservice teacher training at universities, in further teacher training for inservice teachers at universities, in inhouse teacher training for inservice teachers at schools (public elementary and secondary schools) and teachers or group of teachers that agreed to participate in the test independently of teacher training. The sample (N = 295) comprises 57% preservice teachers, 39% inservice teachers and 4% teacher educators and researchers in the field. Apart from testing LRT competence, we aimed at receiving a great variety of different responses to implement the variety of responses into our standard setting procedure, as this study was part of a test development process. Table 1 shows the distribution of other characteristics within the sample, such as gender, first language, subjects of teaching or teaching experience. Table 2 gives an overview of the descriptive results of the questionnaire instruments that were used with the LRT-test (LRT-relevant learning opportunities, personality factors, beliefs). On average, participants in this sample have a medium amount of learning opportunities. The personality factor *compatibility* is more prominent amongst the participants than the other factors, whereas *neuroticism* is the least represented factor within the sample. The participants had rather positive beliefs on multilingualism in school and teaching, especially towards taking responsibility for language support. The average LRT competence of the participants is rather low.

		N	%
Gender	female	240	81.4
	male	55	18.6
First language	German	264	89.5
	other	31	10.5
Subject of studies/teaching: German	yes	177	60.0
	no	118	40.0
Subject of studies/teaching: English	yes	42	14.2
	no	253	85.8
Teaching experience regarding LRT	yes	194	65.8
	no	101	34.2
Research experience regarding LRT	yes	16	5.4
	no	279	94.6

Table 1. Sample.

	Min	Max	M	SD
LRT-relevant learning opportunities (min=22, max=110)*	22	110	56.74	17.38
Personality Factor Cautiousness (min=3, max=21)	8	21	16.27	2.88
Personality Factor Extraversion (min=3, max=21)	6	21	16.26	3.39
Personality Factor Compatibility (min=3, max=21)	4	21	17.40	2.67
Personality Factor Openness (min=3, max=21)	6	21	15.90	3.31
Personality Factor Neuroticism (min=3, max=21)	3	21	12.29	3.79
Beliefs on multilingualism in content classroom (min=7, max=28)	9	28	20.70	4.20
Beliefs on responsibility for language support (min=9, max=36)	16	36	29.70	3.65
LRT competence (min=0, max=72)	1	48	21.39	10.26

* lowest/highest possible scores

Table 2. Descriptive statistics.

Missing Data

From the sample of $N = 302$, $n = 7$ participants cancelled the test without filling in any response. These cases were excluded from all analyses. We therefore conducted analyses on $n = 295$. In the analyses of the LRT-test results, all cases that did not contain evaluable responses, such as audio files with only breathing or fragmentary expressions, were coded as wrong responses.

Results

The following section presents the results of the data analysis. Table 3 shows the manifest correlations of the LRT score (WLE) and the participants' variables such as individual and academic characteristics, personality traits and beliefs concerning multilingualism in school and teaching.

	Manifest correlations	Latent regression models								
		M1		M2		M3		M4		
		r	B	SE				B	SE	
	Intercept		-1.422	0.147	-1.099	0.053	-1.097	0.050	-1.315	0.149
RQ 1: Individual and academic characteristics and LRT-relevant learning opportunities	Gender (0=male, 1=female)	0.164	0.439	0.129					0.341	0.132
	First language (0=German, 0=other language)	-0.102	-0.450	0.163					-0.384	0.162
	Subject of studies German (0=no, 1=yes)	0.080	-0.114	0.112					-0.166	0.110
	Subject of studies EFL (0=no, 1=yes)	-0.044	-0.239	0.146					-0.285	0.142
	Teaching experience regarding LRT (0=no, 1=yes)	0.122	0.106	0.106					0.128	0.106
	Research experience regarding LRT (0=no, 1=yes)	0.251	0.913	0.224					0.731	0.224
	LRT learning opportunities	0.044	0.043	0.055					0.001	0.055
RQ 2: Personality traits (big five)	Big five scale conscientiousness	-0.004			0.047	0.055			0.003	0.050
	Big five scale extraversion	-0.030			0.002	0.055			0.047	0.051
	Big five scale agreeableness	0.014			-0.015	0.055			-0.034	0.050
	Big five scale openness	-0.014			-0.004	0.056			-0.078	0.053
	Big five scale neuroticism	0.049			0.049	0.054			0.059	0.050
RQ 3: Beliefs on multilingualism in school and teaching	Beliefs - multilingualism in content classroom	0.210					0.153	0.054	0.157	0.054
	Beliefs - responsibility for language support	0.250					0.180	0.054	0.128	0.054
	R²		0.248		0.076		0.166		0.260	

Table 3. Manifest correlations and latent regressions of the LRT scores in relation to person's characteristics.

Note: Statistically significant values are in bold face.

Relationship between LRT competence and individual and academic characteristics

The bivariate correlations between the LRT competence of the participants and individual and academic characteristics are given in column 3 of table 3. The results show that female respondents achieve statistically significantly higher results than male respondents ($r = .641$). Furthermore, a higher LRT competence is statistically significantly correlated with a higher LRT teaching experience ($r = .122$) and a higher LRT research experience ($r = .251$). No significant associations with LRT proficiency were found for the other variables of individual and academic background. Especially for the indicator of LRT relevant learning opportunities in university and non-university teacher training, this non-significant relationship was unexpected. For this reason, we conducted an additional analysis which is presented at the end of this paragraph. With this analysis, we wanted to find out to what extent learning opportunities in the different topics are predictive with LRT competence.

The results of the latent regression analysis include all variables at once and report the specific predictions of each variable under control of the other variables (model M1, column 4, table 3). In line with the bivariate correlations, the variables gender and LRT research experience are statistically significant predictors for the LRT competence. However, under control of all other predictors the LRT teaching experience has no statistically significant effect. Then again, we could find a significant prediction effect for the language background of the participants. (Preservice) teachers with a non-German language background receive lower LRT scores than persons whose first language is German. The regression analysis indicates that almost 25 percent of the variance in LRT scores can be explained by the participants' individual and academic background ($R^2 = 24.8\%$).

LRT competence and personality factors

Research question 2 asks for the relationship between the LRT competence and personality factors. The results reveal that the five scale scores of the main personality factors (openness, neuroticism, conscientiousness, agreeableness and extraversion) show no statistically significant correlations with the LRT competence and no significant predictors in model 2 of the latent regression analysis (model M2, column 6, table 3). These results disprove our assumptions: We assumed participants who are more open and extroverted, might respond more easily in this oral response format. The regression analysis indicates that only a low portion of variance could be explained by the personality factors ($R^2 = 7.6\%$).

LRT competence and beliefs

Regarding the beliefs, two scales were evaluated in the analysis: *multilingualism in content classroom* and *responsibility for language support in content classroom*. Positive beliefs towards multilingualism in subject teaching ($r = 0.21^{**}$) as well as positive beliefs concerning the responsibility for language support ($r = 0.25^{**}$) correlate highly with the measured LRT competence. The results of the latent regression model (Model M3, column 8, Table 3) also indicate a significant prediction of LRT competence by both indicators of beliefs concerning multilingualism in school and teaching. The explained variance is about $R^2=16.6\%$ and therefore higher as the personality factors and somehow lower than the individual and academic background.

Finally, in the latent regression model 4 (Table 3, column 10) all predictors of the three blocks are included in the analysis simultaneously. The pattern of the results is relatively consistent. The results show no effects for personality factors, meaning that LRT competence is independent of differences in big-5 test scores of the teacher's personality. The most predictive factors can be found concerning the individual and academic background and the (preservice) teachers' LRT related beliefs. The full model explains 26 percent of the variance in LRT competence ($R^2 = 26,0 \%$).

LRT-relevant learning opportunities, single Items

Because the scale of LRT-relevant learning opportunities did not show any significant effect, we conducted further analysis on individual topics surveyed (table 4). The results show that respondents who had a higher amount of learning experience with the topic of "support of the language learning process through scaffolding" in courses performed significantly better ($r=0.12^*$) in the LRT-test. For all other indicators, for example "(sub) areas of linguistics (e.g.: syntax, semantics, morphology)", the regressions coefficients are not statistically significant.

Item	Topic	Correlation (r) with LRT-test results
TH01	(sub-) areas of linguistics (e.g.: syntax. semantics. morphology)	0.105
TH02	German grammar	0.005
TH03	differences between oral and written language	0.025
TH04	linguistic requirements of different forms of presentation	-0.012
TH05	acquisition of academic language	0.030
TH17	written language acquisition	0.000
TH06	phenomena of second language acquisition	0.064
TH18	milestones in language development	0.044
TH19	differentiation of language acquisition contexts	0.046
TH09	migration and multilingualism	0.045
TH10	language diversity in the school	0.038
TH11	dealing with heterogeneity	0.017
TH12	language and identity	-0.015
TH13	diagnostics of language levels	0.093
TH14	language support and language training	0.025
TH15	support of the language learning process through scaffolding	0.124
TH16	language systems of migrant languages (e.g. Turkish. Russian)	-0.042
TH20	feedback procedures for the correction of errors	0.028
TH21	educational success and language competence	0.091
TH07	language in the respective content classroom	-0.017
TH08	supporting the language learning process through pupil interaction	-0.004
TH22	communication within the family and outside the school	0.017

Table 4. Bivariate correlations between LRT competence and LRT-relevant learning opportunities.

Note: Statistically significant values are in bold face.

5. Discussion

In this study, we aimed at a performance-oriented measurement of LRT competence of preservice and inservice teachers and at evaluating possible predictors for LRT competence: (1) individual and academic background and LRT-learning opportunities, (2) personality factors, (3) participants beliefs towards multilingualism in school and teaching.

In summary, the analysis yields the following results:

1. There is a connection between LRT-test results and individual characteristics, such as teaching experience and research experience. Both show statistically significant correlations. Female participants and those who study/teach English achieve higher test

scores. Surprisingly, the scale of LRT-relevant learning opportunities in contrast does not show any significant results.

2. In regard to the response format, the analysis of the personality factors does not show statistically significant results.
3. Participants with positive beliefs concerning multilingualism in school and teaching perform significantly better in the LRT-test than participants with less positive beliefs.

Regarding research questions 1, which asks for the connection between LRT-test results and individual and academic characteristics, some results are in line with our theory and assumptions (*gender, German as a first language and teaching experience*). However, some results are surprising or contrary to our expectations: German as a *subject of studies/teaching* does not affect the test results, whereas English (EFL) does. We assumed participants who study or teach German, reach higher scores in LRT-test due to their linguistic education and their knowledge about linguistic didactics. Compared to other subjects, participants who study/teach English (EFL-teachers) have experiences (personally or in teaching contexts), which confirm the fundamental role of positive beliefs concerning multilingualism in school and teaching. In the official guidelines in Germany for teaching English as a foreign language (EFL) in e.g. primary school, published by the *Standing Conference of the Ministers of Education and Cultural Affairs*, it is stated that communicative and intercultural competences in other languages are fundamental for the participation in a cosmopolitan and mobile society. Furthermore, The guidelines explicitly acknowledge that EFL teaching provides opportunities for all students to develop language awareness (Niedersächsisches Kultusministerium, 2018). Hence, future EFL-teachers are likely to be better prepared for teaching intercultural competence and raising language awareness than others and are therefore more aware of linguistic diverse classrooms.

Participants who are experienced in research in this field also reach higher test scores than those who are not. Although researchers may not have any teaching routine, they engage with the subject matter more intensively than any other group: They look at actual teaching (e.g. by collecting data through tests or observation), and apply and set up theories. This fits the theoretical assumptions concerning competence being a bi-directionally continuum between dispositions (LRT-relevant knowledge and beliefs) via professional vision and decision making to performance, and then back via processes of reflection to transformed knowledge and beliefs (Blömeke, König, et al., 2015; Kramer et al., 2020; Santagata & Yeh, 2016).

These results indicate what several other studies suggested as well: Not only personal teaching experience, but also observing teaching and theory-based reflection contribute to teachers' professionalization (Seidel & Stürmer, 2014; Weber et al., 2020).

Surprisingly, no connection was found between the scale of LRT-relevant learning opportunities and the LRT-test scores. Looking at the competence model described in the theoretical framework, LRT-relevant learning opportunities promote professional knowledge. But, due to weaknesses in the situation-specific skills, the participants may still not be able to perform in the LRT-test: Professional (LRT-relevant) knowledge and (LRT-relevant) teaching performance develop, as teachers go through various stages of education. However, achievement in different dimensions of competence does not happen equally (Blömeke & Kaiser, 2017). In additional analyses with single items of the scale with LRT-relevant learning opportunities, only one topic showed a significant correlation with LRT competence: *support of the language learning process through scaffolding*. Scaffolding is a concrete strategy for organizing the learning process in a lesson and combines many facets of LRT competence such as diagnostics, language support or difference between conversational and academic language (Gibbons, 2002). So, participants who know about scaffolding, may have an overall and practical idea of what LRT involves.

Looking at the results for research question 2, no statistically significant relations between LRT-test scores and personality factors were found. We considered shy participants to perform poorer in the test, as extroverted, more open participants, which is not confirmed in this study. For reasons of test fairness, the response format in this study makes no difference, certain personality traits (e.g. openness or extraversion) are not preferred or disadvantaged. So, the measurement of LRT competence is independent of personality factors, which is one indicator for an innovative and promising test format.

The results for research question 3 confirm our theoretically based expectations. From this we derive that beliefs are a fundamental element in the education of linguistically responsive teachers. Studies on beliefs confirm that beliefs are a filter for what we learn and how we perceive our environment (Ricart Brede, 2019; Wischmeier, 2012). People win beliefs by experiences, which then become firm opinions they are convinced of. Therefore, they are action-leading in the everyday life of the teachers (Reusser et al., 2011). In the research area of teachers' competence and professional development, beliefs are seen as part of the dispositions in the competence-continuum. They are part of bi-directional processes described above and in chapter two. Research on beliefs assumes a direct connection between beliefs and action/performance as well (Ricart Brede, 2019). Beliefs have an impact on teachers' actions

in the classroom (in the competence-model on perception and decision making and then on the performance), but reflection on action also affects beliefs (Ricart Brede, 2019).

We consider the presented findings as an indication for a valid performance-oriented measurement of LRT competence. However, the findings also show that the acquisition of performance-oriented LRT competence requires specific learning opportunities.

Practical Implications

Based on the agreement in this field of research about the need of linguistically responsive teachers, this study leads to implications of how teacher training should take this into account. Firstly, it can be assumed that videos are a chance for future teachers to develop professional knowledge, professional vision and therefore as well competent performance (Gaudin & Chaliès, 2015; Seidel & Stürmer, 2014). Videos to observe (LRT-) relevant situations and to reflect on them will help in the professionalization process. In particular, a strengthening professional vision seems to be fundamental, due to its' central role for competent decision making and therefore performance in the classroom (Blömeke, König, et al., 2015). Secondly, the study provides support for implementing a learning environment with opportunities for reflection, also regarding (preservice) teachers' beliefs. Although the connection of beliefs and competence is still unclear (Ricart Brede, 2019), different studies indicate, beliefs have an impact on teacher's teaching as well as on their education (Fischer & Lahmann, 2020; Pettit, 2014). The test results allow similar conclusions. Besides educating (future) teachers' in linguistically responsive teaching, a greater focus on need to be placed on teachers' beliefs and on how to develop and shape them, for instance by evaluating beliefs (Fischer & Lahmann, 2020) and by creating enough space for reflection on experiences (Ricart Brede, 2019). According to Sieland and Jordaan (2019), professionalization only succeeds with the ability of self-reflection, as this is necessary to raise awareness to blind spots. Thirdly, referring to individual characteristics, such as subjects of studies/teaching, we understand that not only linguistic- / language-knowledge are elementary in teacher education. Raising language awareness and intercultural competencies amongst (future) teachers seem to be equally important (Gage, 2020; Hu & Gao, 2020).

Limitations and Scientific Implications

We see scientific implications in three aspects. First, this study assumes that LRT competence is a continuum between dispositions and performance, with situation-specific skills in between.

We operated the performance-oriented measurement by evaluating situation-specific skills (*perception* and *decision making*). However, it is still unclear how an actual performance in the classroom is related to situation-specific skills. In research on expertise, knowledge is assumed to transform from a conscious, declarative disposition into an unconscious, more intuitive state. Therefore, it is unclear if the actual performance of teacher experts is congruent with what they verbally express on how they would perform (Hecker, Falkenstern, & Lemmrich, 2020; Herzog, 2018).

Second, further work is required on LRT-relevant learning opportunities. We measured a scale of possible topics/activities that participants may have come across during their teacher training. But what these LRT-relevant learning opportunities look like (structure, didactics, methods, etc.) remains unclear. We suggest that research should aim at evaluating LRT-relevant learning opportunities of preservice teachers (pre-/post) at different teacher educating universities across Germany to identify LRT-relevant learning opportunities. Third, from a test theoretical / technical point of view, further research needs to be conducted on measuring performance / performance-oriented measurement. In future investigations, a different response format (written, multiple-choice, etc.) will be tested as the oral format is time consuming and more complicated in coding responses of huge samples.

In spite of its limitations, the study suggests that this innovative response format is independent of participants' personality factors and therefore promising for performance-oriented testing. The study certainly adds to our understanding of factors fostering LRT competence amongst (future) teachers: Besides linguistic knowledge, concrete strategies such as scaffolding and particularly teachers' beliefs and reflecting on them play a key role in the professional development of linguistically responsive teachers.

References

- Acquah, E. O., Commins, N. L., & Niemi, T. (2016). Preparing Teachers for Linguistic and Cultural Diversity: Experiences of Finnish Teacher Trainees. In B. Koch-Priewe & M. Krüger-Potratz (Eds.), *Die deutsche Schule Beiheft: Vol. 13. Qualifizierung für sprachliche Bildung: Programme und Projekte zur Professionalisierung von Lehrkräften und pädagogischen Fachkräften* (pp. 111–129). Waxmann.
- Adams, R. J., Wu, M. L., & Wilson. (2015). *ACER ConQuest: Generalised item response modelling software [computer software]*. Version.
- Berkel-Otto, L., Hansen, A., Hammer, S., Lemmrich, S., Schroedler, T., & Uribe, Á. (2021). Multilingualism in Teacher Education in Germany: Differences in Approaching Linguistic Diversity in Three Federal States. In M. Wernicke, S. Hammer, A. Hansen, & T. Schroedler (Eds.), *Preparing Teachers to Work with Multilingual Learners* (pp. 82–103). Multilingual Matters.
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond Dichotomies. *Zeitschrift Für Psychologie*, 223(1), 3–13. <https://doi.org/10.1027/2151-2604/a000194>
- Blömeke, S., & Kaiser, G. (2017). Understanding the development of teachers' professional competencies as personally, situationally and socially determined. In D. J. Clandinin & J. Husu (Eds.), *The SAGE handbook of research on teacher education* (pp. 783–812). SAGE Publications Ltd.
- Blömeke, S., König, J., Suhl, U., Hoth, J., & Döhrmann, M. (2015). Wie situationsbezogen ist die Kompetenz von Lehrkräften? Zur Generalisierbarkeit der Ergebnisse von videobasierten Performanztests. *Zeitschrift Für Pädagogik*, 61(3), 310–327.
- Carlson, S. A., Köker, A., Rosenbrock-Agyei, S., Ohm, U., Koch-Priewe, B., Hammer, S., Fischer, N., & Ehmke, T. (2018). DaZKom - a Structure Model of Pre-service Teachers' Competency for Teaching German as a second Language in the Mainstream Classroom. In T. Ehmke, S. Hammer, A. Köker, U. Ohm, & B. Koch-Priewe (Eds.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (pp. 261–279). Waxmann.
- Carlson, S. A., & Präg, D. (2018). Der Prozess der Aufgabenentwicklung im DaZKom-Projekt: von der Rahmenkonzeption bis zur Pilotierung des Testinstruments. In T. Ehmke, S. Hammer, A. Köker, U. Ohm, & B. Koch-Priewe (Eds.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (pp. 93–108). Waxmann.
- Casale, G., Strauß, S., Hennemann, T., & König, J. (2016). Wie lässt sich Klassenführungsexpertise messen? Überprüfung eines videobasierten Erhebungsinstruments für Lehrkräfte unter Anwendung der Generalisierbarkeitstheorie. *Empirische Sonderpädagogik*(2), 119–139.
- Chipere, N. (2014). Sex differences in phonological awareness and reading ability. *Language Awareness*, 23(3), 275–289. <https://doi.org/10.1080/09658416.2013.774007>
- Cho, S., Lee, H.-J., & Herner-Patnode, L. (2020). Factors Influencing Preservice Teachers' Self-Efficacy in Addressing Cultural and Linguistic Needs of Diverse Learners. *The Teacher Educator*, 55(4), 411–429. <https://doi.org/10.1080/08878730.2020.1805835>
- Ehmke, T., & Lemmrich, S. (2018). Bedeutung von Lerngelegenheiten für den Erwerb von DaZ-Kompetenz. In T. Ehmke, S. Hammer, A. Köker, U. Ohm, & B. Koch-Priewe (Eds.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (pp. 201–220). Waxmann.

- Fischer, N. (2020). *Überzeugungen angehender Lehrkräfte zu sprachlich-kultureller Heterogenität in Schule und Unterricht: theoretische Struktur, empirische Operationalisierung und Untersuchung der Veränderbarkeit*. Leuphana Universität Lüneburg, Lüneburg.
- Fischer, N., & Ehmke, T. (2019). Empirische Erfassung eines „messy constructs“: Überzeugungen angehender Lehrkräfte zu sprachlich-kultureller Heterogenität in Schule und Unterricht. *Zeitschrift für Erziehungswissenschaft*, 22(2), 411–433. <https://doi.org/10.1007/s11618-018-0859-2>
- Fischer, N., & Ehmke, T. (2020). Theoretische Abbildung und empirische Erfassung der Überzeugungen angehender Lehrkräfte zu sprachlich-kultureller Heterogenität in Schule und Unterricht. *DIPF | Leibniz Institute for Research and Information in Education*. Advance online publication. <https://doi.org/10.7477/395:246:1>
- Fischer, N., & Lahmann, C. (2020). Pre-service teachers' beliefs about multilingualism in school: an evaluation of a course concept for introducing linguistically responsive teaching. *Language Awareness*, 29(2), 114–133. <https://doi.org/10.1080/09658416.2020.1737706>
- Gage, O. (2020). Urgently needed: a praxis of language awareness among pre-service primary teachers. *Language Awareness*, 1–16. <https://doi.org/10.1080/09658416.2020.1786102>
- Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review*, 16, 41–67. <https://doi.org/10.1016/j.edurev.2015.06.001>
- Gay, G. (2010). *Culturally responsive teaching: Theory, research, and practice* (2. ed.). *Multicultural education series*. Teachers College.
- Gibbons, P. (2002). *Scaffolding language, scaffolding learning*. Portsmouth, NH: Heinemann.
- Gogolin, I. (2008). *Der monolinguale Habitus der multilingualen Schule*. Zugl.: Hamburg, Univ., Habil.-Schr., 1991 (2., unveränd. Aufl.). *Internationale Hochschulschriften: Bd. 101*. Waxmann.
- Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, 59(6), 1216–1229. <https://doi.org/10.1037/0022-3514.59.6.1216>
- Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96(3), 606–633.
- Halpern, D. F. (2013). *Sex Differences in Cognitive Abilities*. Psychology Press. <https://doi.org/10.4324/9780203816530>
- Hammer, S., Carlson, S. A., Ehmke, T., Koch-Priewe, B., Köker, A., Ohm, U., Rosenbrock, S., & Schulze, N. (2015). Kompetenz von Lehramtsstudierenden in Deutsch als Zweitsprache. Validierung des GSL-Testinstruments. In S. Blömeke & O. Zlatkin-Troitschanskaia (Eds.), *Zeitschrift für Pädagogik: Vol. 61. Kompetenzen von Studierenden*. Beltz Juventa.
- Hammer, S., Fischer, N., & Koch-Priewe, B. (2016). Überzeugungen von Lehramtsstudierenden zu Mehrsprachigkeit in der Schule. *Qualifizierung für sprachliche Bildung*, 147–171.
- Hawkins, E. W. (1999). Foreign Language Study and Language Awareness. *Language Awareness*, 8(3-4), 124–142. <https://doi.org/10.1080/09658419908667124>
- Hecker, S.-L., Falkenstern, S., & Lemmrich, S. (2020). Zwischen DaZ-Kompetenz und Performanz. Advance online publication. <https://doi.org/10.4119/hlz-3374> (565-584 Seiten / Herausforderung Lehrer*innenbildung - Zeitschrift zur Konzeption, Gestaltung und Diskussion, Bd. 3 Nr. 1 (2020): Herausforderung Lehrer_innenbildung-Ausgabe 3 / Herausforderung Lehrer*innenbildung - Zeitschrift zur Konzeption, Gestaltung und Diskussion, Bd. 3 Nr. 1 (2020): Herausforderung Lehrer_innenbildung - Ausgabe 3).

- Hecker, S.-L., Falkenstern, S., Lemmrich, S., & Ehmke, T. (2020). Zum Verbalisierungsdilemma bei der Erfassung der situationsspezifischen Fähigkeiten von Lehrkräften. *Zeitschrift für Bildungsforschung*, 10(2), 175–190. <https://doi.org/10.1007/s35834-020-00268-1>
- Hecker, S.-L., & Nimz, K. (2020). Expertinnenratings zur Deutsch-als-Zweitsprache-Kompetenz von Lehrkräften. Eine Vorstudie zur Konzipierung eines videobasierten Testinstruments. J. Stiller, C. Laschke, T. Nesyba, U. Salaschek (Hg.): *Berlin-Brandenburger Beiträge zur Bildungsforschung*.
- Herzog, W. (2018). Die ältere Schwester der Theorie: Eine Neubetrachtung des Theorie-Praxis-Problems. *Zeitschrift für Pädagogik*, 64(6), 812–830.
- Hu, J., & Gao, X. (2020). Understanding subject teachers' language-related pedagogical practices in content and language integrated learning classrooms. *Language Awareness*, 1–20. <https://doi.org/10.1080/09658416.2020.1768265>
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Köker, A. (2018). Zur Relevanz (bildungs-) sprachlicher Förderung in Schule. In T. Ehmke, S. Hammer, A. Köker, U. Ohm, & B. Koch-Priewe (Eds.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache*. Waxmann.
- Kramer, M., Förtsch, C., Stürmer, J., Förtsch, S., Seidel, T., & Neuhaus, B. J. (2020). Measuring biology teachers' professional vision: Development and validation of a video-based assessment tool. *Cogent Education*, 7(1). <https://doi.org/10.1080/2331186X.2020.1823155>
- Ladson-Billings, G. (2009). *The dreamkeepers: Successful teachers of African American children*. John Wiley & Sons.
- Lemmrich, S., Bahls, A., & Ehmke, T. (2020). Effekte von mündlichen versus schriftlichen Antwortformaten bei der performanznahen Messung von Deutsch-als-Zweitsprache (DaZ)-Kompetenz – eine experimentelle Studie mit angehenden Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 1–11. <https://doi.org/10.1024/1010-0652/a000282>
- Lemmrich, S., Hecker, S.-L., Klein, S., Ehmke, T., Koch-Priewe, B., Köker, A., & Ohm, U. (2020). Linguistically Responsive Teaching in Multilingual Classrooms. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper, & C. Lautenbach (Eds.), *Student Learning in German Higher Education* (1st ed., Vol. 2020, pp. 125–140). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27886-1_7
- Lemmrich, S., Hecker, S.-L., Klein, S., Ehmke, T., Köker, A., Koch-Priewe, B., & Ohm, U. (2019). Performanznahe und videobasierte Messung von DaZ-Kompetenz bei Lehrkräften: Skalierung und dimensionale Struktur des Testinstruments. In T. Ehmke, P. Kuhl, & M. Pietsch (Eds.), *Lehrer. Bildung. Gestalten: Beiträge zur empirischen Forschung in der Lehrerbildung* (1st ed., pp. 190–204). Beltz.
- Lucas, T., & Villegas, A. M. (2011). A Framework for Preparing Linguistically Responsive Teachers. In T. Lucas (Ed.), *Teacher preparation for linguistically diverse classrooms: A resource for teacher educators* (pp. 55–72). Routledge.
- Lucas, T., & Villegas, A. M. (2013). Preparing Linguistically Responsive Teachers: Laying the Foundation in Preservice Teacher Education. *Theory into Practice*, 52(2), 98–109. <https://doi.org/10.1080/00405841.2013.770327>

- Lucas, T., Villegas, A. M., & Freedson-Gonzalez, M. (2008). Linguistically Responsive Teacher Education. *Journal of Teacher Education*, 59(4), 361–373.
<https://doi.org/10.1177/0022487108322110>
- Maccoby, E. E., & Jacklin, C. N. (1978). *The psychology of sex differences*. Stanford University Press.
- Meschede, N., Fiebranz, A., Möller, K., & Steffensky, M. (2017). Teachers' professional vision, pedagogical content knowledge and beliefs: On its relation and differences between pre-service and in-service teachers. *Teaching and Teacher Education*, 66, 158–170.
- Michalos, A. C. (2014). *Encyclopedia of Quality of Life and Well-Being Research*.
- Niedersächsisches Kultusministerium. (2018). *Kerncurriculum für die Grundschule, Schuljahrgänge 3-4, Englisch*.
https://www.cuvo.nibis.de/cuvo.php?key_lev0_0=Dokumentenart&svalue_lev0_0=Kerncurriculum&key_lev0_1=Schulbereich&svalue_lev0_1=Primarbereich&key_lev0_2=Fach&svalue_lev0_2=Englisch&docid=1153&p=detail_view
- OECD. (2013). *Oecd Skills Outlook 2013: First Results from the Survey of Adult Skills*. OECD Publishing.
- OECD. (2014). *What students know and can do: Student performance in mathematics, reading and science* (Rev. ed., Febr. 2014). *PISA 2012 results: / Organisation for Economic Co-operation and Development, Programme for International Student Assessment; Vol. I*. OECD.
- Ohm, U. (2018). Das Modell von DaZ-Kompetenz bei angehenden Lehrkräften. In T. Ehmke, S. Hammer, A. Köker, U. Ohm, & B. Koch-Priewe (Eds.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (pp. 73–91). Waxmann.
- Pajares, M. F. (1992). Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct. *Review of Educational Research*, 62(3), 307–332. <https://doi.org/10.3102/00346543062003307>
- Pettit, S. (2014). Middle School Mathematics Teachers' Beliefs About ELLs in Mainstream Classrooms.
- Rammstedt, B., & Danner, D. (2017). Die Facettenstruktur des Big Five Inventory (BFI). *Diagnostica*, 63(1), 70–84. <https://doi.org/10.1026/0012-1924/a000161>
- Reusser, K., Pauli, C., & Elmer, A. (2011). Berufsbezogene Überzeugungen von Lehrerinnen und Lehrern. *Handbuch der Forschung zum Lehrerberuf*, 2.
- Ricart Brede, J. (2019). Einstellungen – beliefs – Überzeugungen – Orientierungen: Zum theoretischen Konstrukt des Projektes „Einstellungen angehender LehrerInnen zu Deutsch als Zweitsprache in Ausbildung und Unterricht“. In D. Maak & J. Ricart Brede (Eds.), *Mehrsprachigkeit: Band 46. Wissen, Können, Wollen - sollen?! (angehende) LehrerInnen und äußere Mehrsprachigkeit* (pp. 29–38). Waxmann.
- Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. *Handbook of Research on Teacher Education*, 2(102-119), 273–290.
- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion* (2., vollst. überarb. und erw. Aufl.). *Psychologie Lehrbuch*. Huber.
- Santagata, R., & Yeh, C. (2016). The role of perception, interpretation, and decision making in the development of beginning teachers' competence. *ZDM*, 48(1-2), 153–165.
<https://doi.org/10.1007/s11858-015-0737-9>
- Schleppegrell, M. J. (2009). *Language in academic subject areas and classroom instruction: what is academic language and how can we teach it?* Riverside County Office Of Education (RCOE).

- https://www.rcoe.us/educational-services/files/2012/08/What_is_Academic_Language_Schleppegrell.pdf
- Schroedler, T. (2021). What is Multilingualism? Towards an Inclusive Understanding. In M. Wernicke, S. Hammer, A. Hansen, & T. Schroedler (Eds.), *Preparing Teachers to Work with Multilingual Learners* (pp. 17–37). Multilingual Matters.
- Schupp, J., & Gerlitz, J.-Y. (Eds.) (2014). *Big Five Inventory-SOEP (BFI-S)*. ZIS - GESIS Leibniz Institute for the Social Sciences.
- Seidel, T., & Stürmer, K. (2014). Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. *American Educational Research Journal*, 51(4), 739–771.
<https://doi.org/10.3102/0002831214531321>
- Sieland, B., & Jordaan, L. (2019). Kommentar: Lehrerinnen-und Lehrerbildung im Umbruch zwischen Auftrag, alten Schwächen und neuen Chancen. *Kompetenzentwicklung im Lehramtsstudium durch professionelles Training*, 7–19.
- Stanat, P., Bergann, S., & Taraszow, T. (2018). Geschlechtsbezogene Disparitäten im deutschen Bildungswesen. In R. Tippelt & B. Schmidt-Hertha (Eds.), *Handbuch Bildungsforschung* (pp. 1321–1338). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-19981-8_59
- Star, J. R., & Strickland, S. K. (2008). Learning to observe: Using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education*, 11(2), 107–125.
- Svalberg, A. M.-L. (2016). Language Awareness research: where we are now. *Language Awareness*, 25(1-2), 4–16. <https://doi.org/10.1080/09658416.2015.1122027>
- van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, 10(4), 571–596.
- Villegas, A. M. (2018). Introduction to "Preparation and Development of Mainstream Teachers for Today's Linguistically Diverse Classrooms". *The Educational Forum*, 82(2), 131–137.
<https://doi.org/10.1080/00131725.2018.1420848>
- Villegas, A. M., & Lucas, T. (2002). Preparing Culturally Responsive Teachers. *Journal of Teacher Education*, 53(1), 20–32. <https://doi.org/10.1177/0022487102053001003>
- Weber, K. E., Prilop, C. N., Viehoff, S., Gold, B., & Kleinknecht, M. (2020). Fördert eine videobasierte Intervention im Praktikum die professionelle Wahrnehmung von Klassenführung? – Eine quantitativ-inhaltsanalytische Messung von Subprozessen professioneller Wahrnehmung. *Zeitschrift für Erziehungswissenschaft*, 23(2), 343–365. <https://doi.org/10.1007/s11618-020-00939-9>
- Wernicke, M., Hammer, S., Hansen, A., & Schroedler, T. (Eds.). (2021). *Preparing Teachers to Work with Multilingual Learners*. Multilingual Matters.
- Wilson, M. (2004). *Constructing Measures*. Routledge. <https://doi.org/10.4324/9781410611697>
- Wischmeier, I. (2012). „Teachers' Beliefs“: Überzeugungen von (Grundschul-) Lehrkräften über Schüler und Schülerinnen mit Migrationshintergrund–Theoretische Konzeption und empirische Überprüfung. In W. Wiater & D. Manschke (Eds.), *Verstehen und Kultur: Mentale Modelle und kulturelle Prägungen* (pp. 167–189). Springer; Springer VS.