

**Effekte von mündlichen versus schriftlichen Antwortformaten bei der performanznahen Messung von Deutsch-als-Zweitsprache (DaZ)-Kompetenz**

Lemmrich, Svenja; Bahls, Anika; Ehmke, Timo

*Published in:*  
Zeitschrift für pädagogische Psychologie

*DOI:*  
[10.1024/1010-0652/a000282](https://doi.org/10.1024/1010-0652/a000282)

*Publication date:*  
2022

*Document Version*  
Verlags-PDF (auch: Version of Record)

[Link to publication](#)

*Citation for published version (APA):*

Lemmrich, S., Bahls, A., & Ehmke, T. (2022). 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*, 36(1-2), 37-47.  
<https://doi.org/10.1024/1010-0652/a000282>

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# 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 performanznähtere 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<sup>1</sup> 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-

<sup>1</sup> 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 situationsspezifischen 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 erfolgende 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 relevanten 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 wissensbasiertes (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<sup>®</sup>), 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 performanznahem 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-

werten, 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.

## Methode

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 <sup>1)</sup>	Mündlich	Ja	26	44
		Nein	18	
	Schriftlich	Ja	22	44
		Nein	24	
Studienfach	Mündlich	Ja	29	39
		Nein	10	
	Schriftlich	Ja	30	40
		Nein	10	

<sup>1)</sup> 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	
Abiturdurchschnittsnote	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 vorlagen. Die mündlichen Antworten wurden als Audiodateien aufgezeichnet, die schriftlichen Antworten in Texteinga-

benfenstern 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, inwie weit 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  $r = 0.49$ .

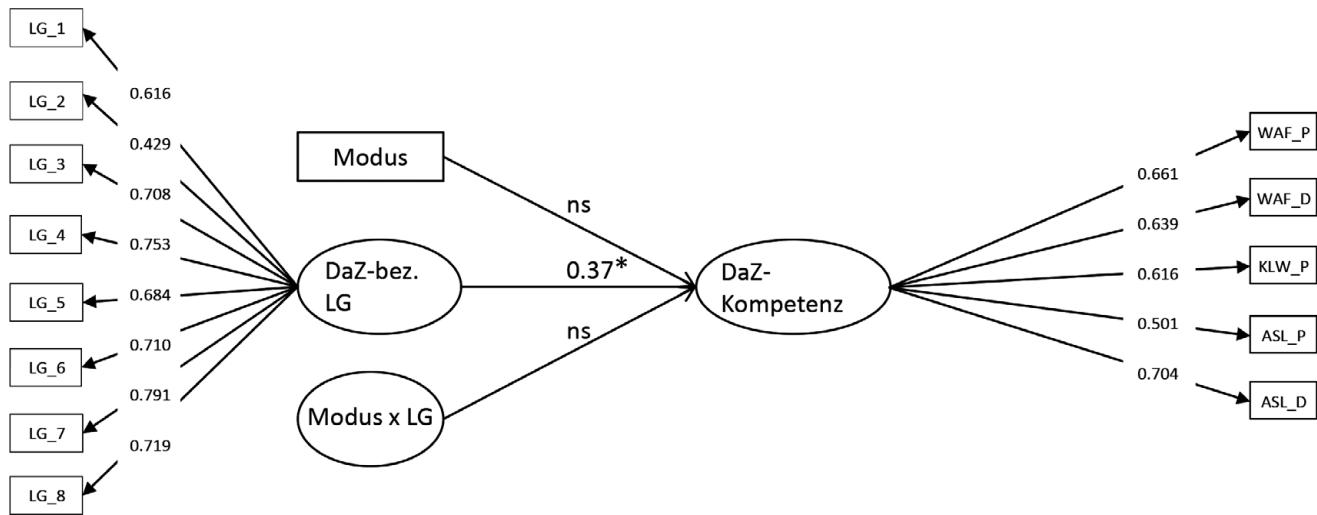
Um die zweite Forschungsfrage zu beantworten, wurden zwei latente Strukturgleichungsmodelle geschätzt, welche den Einfluss der Faktoren *Antwortmodus*, *DaZ-Lerngelegenheiten* sowie einen Interaktionseffekt (*Antwortmodus*  $\times$  *DaZ-Lerngelegenheiten*) auf die gemessene DaZ-Kompetenz (Modell 1,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

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.

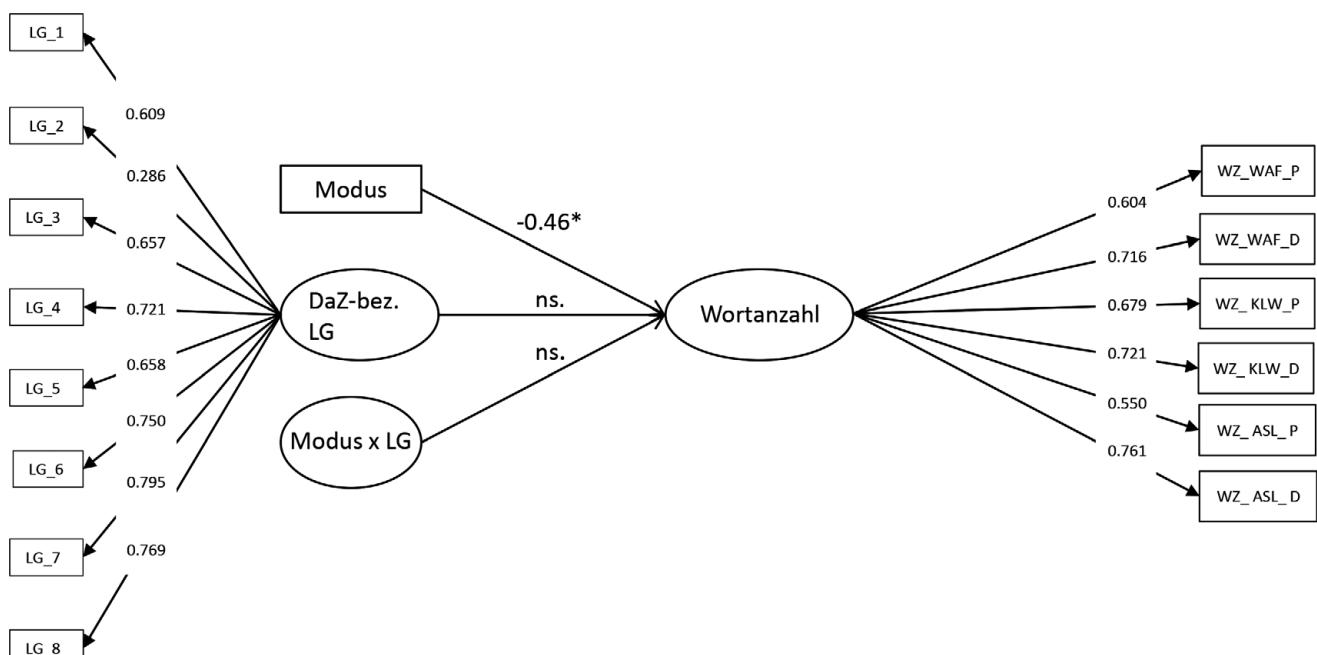
In Modell 2 zeigt sich, dass der Moduseffekt (mündliches Antwortformat = ausführlichere Antworten) auch bei Kontrolle der DaZ-Lerngelegenheiten bestehen bleibt ( $\beta = -0.46$ ,  $p < 0.05$ ). Ein positiver Zusammenhang von DaZ-bezogenen Lerngelegenheiten auf die Ausführlichkeit der Antwort zeigt sich 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 (74)</i>	<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. Bezuglich 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 × 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 bevorzugt 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 Handlungsroutinen 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 bedeutam 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 Handlungseinscheidung) 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 täglich 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 Antwortformates 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 Deutsch-als-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 fröhpädagogischen 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 Unterrichtssituativen 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: Netzwerk.
- 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 Lehrerwissen. 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 wissensbasiert 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