

Zum Verbalisierungsdilemma bei der Erfassung der situationspezifischen Fähigkeiten von Lehrkräften

Am Beispiel der Domäne Deutsch als Zweitsprache

Sarah-Larissa Hecker  · Stephanie Falkenstern · Svenja Lemmrich · Timo Ehmke

Angenommen: 4. Mai 2020 / Online publiziert: 14. Mai 2020
© Der/die Autor(en) 2020

Zusammenfassung Anforderungsnahе Testumgebungen versprechen besonders prädiKTive Hinweise auf die tatsächliche Performanz von angehenden und praktizierenden Lehrkräften. Auf Basis einer Methoden-Triangulation empirisch erhobener Daten aus einem videobasierten Kompetenztest widmet sich der Beitrag allerdings einem potentiellen Validitätsproblem bei der Erfassung der situationspezifischen Fähigkeiten von angehenden und praktizierenden Lehrkräften aller Fächer als Indikator ihrer Kompetenz im Bereich Deutsch als Zweitsprache. In der besonders performanznahen Testumgebung werden keine geschlossenen Items eingesetzt, sondern maximal offene mündliche Antworten erfragt, also zunächst eine Verbalisierung des Wahrgenommenen, gefolgt von einer verbalisierten Handlungsreaktion. Es finden sich in der Auswertung der Antworten von $N=124$ angehenden und praktizierenden Lehrkräften Hinweise auf ein Verbalisierungsdilemma: offenbar ist die geforderte Verbalisierung der Komponente der präzisen Wahrnehmung schwieriger als die der Handlungsreaktion, der Zusammenhang zwischen beidem stellt sich als komplex dar. Dies wird in einen über diese Studie hinausgehenden Kontext eingeordnet, indem Ergebnisse aus der Expertiseforschung herangezogen und Konsequenzen für videobasierte Testumgebungen diskutiert werden.

Schlüsselwörter Wahrnehmung · Lehrkräfte · Video · Kompetenztest · Verbalisierung

S.-L. Hecker (✉) · S. Falkenstern
Universität Bielefeld, Bielefeld, Deutschland
E-Mail: sarah-larissa.hecker@uni-bielefeld.de

S. Lemmrich · Prof. Dr. T. Ehmke
Leuphana Universität Lüneburg, Lüneburg, Deutschland

Indications of a verbalization dilemma in the assessment of teachers' situation-specific skills

Illustrated using the domain of German as a Second Language

Abstract Test designs using tasks that mirror actual classroom demands are understood to predict teaching performance rather closely. Based on a methodological triangulation of empirical data from a video-based competence test, however, this article deals with a potential validity issue regarding the measurement of situation-specific skills of pre- and in-service teachers of all subjects as indicators of their competence in linguistically responsive teaching. The performance-oriented test design does not use closed items, but asks for maximally open oral responses: a verbalized perception is followed by a verbalized immediate (re-)action. The responses of $n = 124$ pre- and in-service teachers point towards a dilemma concerning the demand of verbalizing perceptions: The results show that the former seems to be more difficult than verbalizing immediate reactions, while the relation between both skills emerges as complicated. The nature of the dilemma is described both theoretically and empirically and thereafter put into a broader context, all the while using results from research on teachers' expertise. Finally, consequences for measuring competences in video-based test designs are discussed.

Keywords Perception · Teachers · Video · Competence test · Verbalization

1 Einleitung

Die anforderungs- und performanznahe Erfassung von Lehrer/innen-Kompetenzen als Ergänzung zu reinen Fachwissenstests ist ein Desiderat, das sich verstärkt im letzten Jahrzehnt herausgebildet hat (bspw. Neuweg 2014). Unter anderem erhofft man sich davon eine höhere Prädiktivität in Bezug auf das tatsächliche Unterrichtsverhalten (Bromme 2014). Dabei liegt die Annahme zugrunde, dass erfahrene Lehrpersonen nicht nur auf explizites (Fach-)wissen zurückgreifen, um in ihren Klassenräumen professionell agieren zu können. Der präzisen Wahrnehmung (van Es und Sherin 2002) von Lehrkräften kommt hierbei im Expertiseparadigma eine Schlüsselrolle zu (Berliner 1992). Als situationspezifische Fähigkeit (Blömeke et al. 2015) vermittelt diese im analytischen Modell zwischen Wissen und Können, das bedeutet, die kompetente wissensbasierte Wahrnehmung einer Situation ist dort Voraussetzung einer entsprechend kompetenten Handlung (Hoth 2015). Vermehrt wird die präzise Wahrnehmung daher als performanznaher Indikator unterschiedlicher Bereiche der Lehrkräftekompetenz in häufig videobasierten Testumgebungen herangezogen (bspw. im Projekt *TEDS-FU*, vgl. Blömeke et al. 2014). Eine elementare Bedingung der performanznahen Messung situationspezifischer Fähigkeiten ist dabei, dass diese von den Testpersonen verlässlich verbalisiert werden können. Hieran geben Forschungsergebnisse allerdings Anlass zum Zweifel. So stellte bereits Shulman fest, dass Lehrkräfte außerhalb des Bereichs ihres Fachwissens Schwierigkeiten hätten, zu artikulieren was sie wüssten und weshalb (1987, S. 6). Schließlich handeln Lehrkräfte mit Berufserfahrung in realen Unterrichtssituationen intuitiv und greifen

dabei, anders als Noviz/innen, nicht mehr nur auf Regelwissen zurück, das häufig explizit verbalisierbar ist (Dreyfus und Dreyfus 1989; Neuweg 2015a; Berliner 1992).

Die Deutsch-als-Zweitsprache-(DaZ-)Kompetenz von angehenden, aber auch erfahrenen Lehrkräften kann nun ebenfalls mithilfe eines videobasierten Testinstruments anforderungsnah erfasst werden. Hinweise auf das DaZ-kompetente Handeln von Lehrkräften werden dabei unter anderem auf Basis deren Wahrnehmung sprachbildungsrelevanter Charakteristika in einer komplexen videographierten Unterrichtssituation gewonnen¹. In der Testumgebung finden sich dabei bewusst keine vorgegebenen Antwortmöglichkeiten. Stattdessen werden, um die DaZ-Kompetenz möglichst performanznah erfassen zu können, Fragetypen eingesetzt, die maximal offene mündliche Antworten verlangen. Diese sollen idealtypisch das Hineinversetzen in die gezeigte Situation erforderlich machen. Inwiefern nun die in der Videotestsituation geforderte Verbalisierung des Wahrgenommenen auch die Erfassung der situationsspezifischen Fähigkeiten als Indikatoren von DaZ-kompetenter Performanz bei erfahrenen Lehrkräften erschweren könnte, steht im Fokus dieses Beitrags.

2 Theoretischer Hintergrund

2.1 DaZ-Kompetenz

Lehrkräfte aller Fächer, die im Bereich *Deutsch als Zweitsprache/Sprachbildung* kompetent sind, nutzen die Tatsache, dass fachliche Inhalte stets über Sprache transportiert werden, als Ausgangspunkt für ihr Handeln im Fachunterricht sowie dessen Planung. Wissen im Bereich DaZ formt demnach die Wahrnehmung der Unterrichtsrealität: Ist etwa in einer Situation ein Schüler zu sehen, der den Kopf auf dem Tisch abgelegt hat, würde dies nicht unbedingt zuerst auf die Unmotiviertheit des Schülers zurückgeführt. Stattdessen würde als Erklärung des Verhaltens möglicherweise Resignation in Erwägung gezogen, verursacht durch fehlendes Makro-Scaffolding (Schleppegrell und O'Hallaron 2011) oder wenig sprachsensibles Material – und dementsprechend gehandelt. Als potentielle Lösung scheinbar fachlicher Probleme nehmen DaZ-kompetente Lehrkräfte also auch sprachliche Unklarheiten in den Blick (Köker 2018). Das empirisch überprüfte *DaZKom-Modell* (Köker et al. 2015) stellt die inhaltlichen Anforderungen an DaZ-kompetente Lehrkräfte detailliert dar und dient damit als Basis des videobasierten Tests, der in diesem Beitrag im Fokus steht.

2.2 Situationsspezifische Fähigkeiten als Indikatoren für Performanz

Studien zur videobasierten Kompetenzmessung von Lehrkräften beziehen sich häufig auf das *PID-Modell* von Blömeke et al. (2015), welches mit den sogenannten situationsspezifischen Fähigkeiten Indikatoren für das Handeln in realen Unterrichts-

¹ Das Projekt „DaZKom-Video: Performanznahe Messung von Deutsch-als-Zweitsprache-Kompetenz bei (angehenden) Lehrkräften“ ist BMBF-gefördert und Teil des Clusters *KoKoHs II* (FKZ: 01PK16001A). Antragsteller/innen sind Barbara Koch-Priewe, Anne Köker, Udo Ohm und Timo Ehmke.

situationen definiert: die präzise Wahrnehmung von Unterrichtssituationen („perception“ [P]), „deren zielangemessene Analyse und Interpretation [I] sowie die flexible Reaktion darauf („decision-making“ [D]) [...]“ (Blömeke et al. 2014, S. 515). Diese drei Fähigkeiten *P*, *I* und *D* nehmen dabei im analytischen Modell eine vermittelnde Rolle zwischen Disposition (kognitive sowie affektiv-motivationale Fähigkeiten) und der Performanz ein, wobei die tatsächliche Ausprägung der Disposition in der Performanz je nach Situation, Aufgabe und individueller Verfassung unterschiedlich ist. Zu beachten hierbei ist, dass es sich lediglich um eine rein analytische Trennung der Teilprozesse handelt; so geht die neuere kognitionspsychologische Theoriebildung von einer funktionalen Integration von Wissen (bzw. Kognition oder Disposition), Wahrnehmung und Handlung aus (Hommel et al. 2001). Die Wahrnehmungs- und Interpretationskomponenten aus dem *PID-Modell* finden sich ähnlich bereits bei Sherin (2001, van Es und Sherin 2002) in ihrem auf Goodwins gleichnamiges Konzept (1994) zurückgehenden Konstrukt der *Professional Vision* von Lehrkräften, das wiederum von Seidel et al. (2010) unter dem Namen *professionelle Unterrichtswahrnehmung* auf den deutschsprachigen Raum übertragen wurde. Die dort modellierten Subdimensionen *selektive Wahrnehmung* (*Noticing*) und *wissensbasiertes Schlussfolgern* (*knowledge-based reasoning* mit den Teilprozessen *Beschreiben*, *Bewerten* und *Interpretieren*) hängen dort, wie in Blömeke et al.s Modell, nicht nur voneinander, sondern auch vom Wissen ab (Sherin und van Es 2008, Seidel und Stürmer 2014, Stürmer und Seidel 2015). Genau dieser Zusammenhang ist bedeutsam für die Erfassung dieser Fähigkeiten in Testsituationen. Wie genau die Disposition (speziell das pädagogische und/oder fachspezifische Wissen von Lehrkräften und Studierenden) und Ausprägungen der situationsspezifischen Fähigkeiten bzw. der professionellen Unterrichtswahrnehmung zusammenhängen, wird derzeit häufig untersucht; die Ergebnisse sind jedoch noch uneindeutig (Hepberger et al. 2017; Hoth et al. 2018). Kaum empirisch erforscht ist bislang die Beziehung der situationsspezifischen Fähigkeiten untereinander und deren Zusammenhang mit der Performanz (vgl. aber bspw. Kersting et al. 2010).

2.3 Das Verbalisierungsdilemma bei der Erfassung situationsspezifischer Fähigkeiten

Die Voraussetzung für die isolierte Messung der situationsspezifischen Fähigkeiten in einer Testsituation als Indikatoren von Performanz ist zwangsläufig, dass diese verbalisiert werden können. Ergebnisse der Forschung zu Expert/innen lassen allerdings daran zweifeln, ob dies in jedem Fall verlässlich gelingen kann. Dies hat mit der erfahrungsbedingt unterschiedlichen Repräsentation des handlungsrelevanten Wissens zu tun, das mit den situationsspezifischen Fähigkeiten zusammenhängt. Zweifelsfälle stellen vor allem Lehrkräfte mit Berufserfahrung dar, denn unter anderem durch Erfahrung mit der Bewältigung von Situationen mit jeweils ähnlichen Anforderungen wird das explizite Wissen angereichert und umstrukturiert. So wird es teilweise in Form von Situations- und Handlungsprototypen (Wahl 2002; Ross et al. 2006) organisiert. Ist Wissen in Form von solchen Chunks repräsentiert, handelt es sich um ein implizites, nicht ohne weiteres reflexiv zugängliches Wissen. Bei Expert/innen-Lehrkräften ist die Wahrnehmung von Unterrichtssituationen

aufgrund der Wissensrepräsentation in Form von Situationsprototypen holistischer und kategorialer als bei Noviz/innen, sodass für die Handlung nicht relevante Details ausgeblendet werden können (vgl. Dreyfus und Dreyfus 1989; Bromme 2014). Gleichzeitig erleichtert die Situationsprototypenstruktur die Antizipation zu erwartender Handlungsverläufe und ihrer Wirkungen (ebd.). Dadurch sind Lehrkräften spontane und effiziente Reaktionen auf einen Stimulus möglich, ohne dass bewusste Kontrolle und Anwendung von Regeln (also Zugriff auf explizites verbalisierbares Wissen) nötig wäre, wie es bei Noviz/innen noch eher der Fall ist (Krummenacher und Müller 2017). Daher geht man mit Dreyfus und Dreyfus (1989) u. v. a. bei Expert/innen von einem intuitiven Handeln aus. Die Folge ist, dass Gründe für intuitives Handeln meist im Nachhinein nicht verlässlich verbalisiert werden können, da die Chunks, die ein solches Handeln ermöglichen, nicht reflexiv zugänglich sind (Gigerenzer und Gaissmaier 2012; Koch-Priewe 1986). Demnach ist verbalisierbares Wissen nicht unbedingt identisch mit dem erfahrungsbasierten impliziten, nicht explizierbaren Wissen, das tatsächlich handlungsleitend wirkt (Ross et al. 2006).

Dies hat Konsequenzen für Testumgebungen, die die Kompetenzen angehender, aber auch erfahrener Lehrkräfte in unterschiedlichen Domänen zu erheben suchen: Insbesondere in Antworten zur *Interpretation* (als Komponente von *PID* oder *knowledge-based reasoning*) besteht die Gefahr, explizites verbalisierbares Wissen mit der tatsächlichen Kompetenz zu verwechseln – und andersherum (Herzog 2018). Möglicherweise bedeutet dies also, dass die situationsspezifischen Fähigkeiten erfahrener Lehrkräfte, sofern sie isoliert und über den Umweg der Verbalisierung erfasst werden, keine zuverlässigen Indikatoren der Performanz sind, da in solchen Testsituationen unter Umständen lediglich eine Explizierung von eigentlich impliziten Wissensbeständen und/oder unbewusst ablaufenden Prozessen verlangt wird.

2.4 Umgang mit dem Verbalisierungsdilemma: Ein performanznaher Versuchsaufbau

Die obigen Überlegungen geben domänenübergreifend Anlass zum Zweifel, ob sich alle situationsspezifischen Fähigkeiten in einer Testumgebung für praktizierende Lehrkräfte erfassen lassen. Aufgrund der zu vermutenden Problematik von erfahrenen Lehrkräften, ihre situationsspezifischen Fähigkeiten verlässlich zu verbalisieren, werden in der vorliegenden Studie nur die Teilkomponenten in den Testaufbau zur Querschnittsdomäne DaZ einbezogen, deren performanznahe Erfassung im Einklang mit Annahmen der Expertiseforschung steht. Dies hat Auswirkungen zunächst auf die Art der Prompts, weiterhin auf die Art der über die Prompts erzielbaren Teilhandlungen. So wird im videobasierten Kompetenztest *DaZKom-Video* für angehende und praktizierende Lehrkräfte auf maximal offene Aufgabenprompts zurückgegriffen. Erstens, um die Intuitivität der Wahrnehmung und der geäußerten Handlungsoptionen der Proband/innen grundsätzlich zu ermöglichen; zweitens, damit einhergehend, um keinen Fokus vorgeben zu müssen, durch den die präzise Wahrnehmung nicht mehr von Nöten wäre (Sunder et al. 2016). Ziel der offenen Items muss es drittens sein, eine möglichst performanznahe und authentische Unterrichtshandlung (im Folgenden zur Abgrenzung *H* genannt) einzufordern. Einer solchen wird sich über die direkte mündliche Reaktion auf eine im Video gezeigte Schüler/innenäußerung

angenähert. Über dieses Adressieren wird eine verbalisierte Handlung realisiert, die ein elementarer Bestandteil auch realer Unterrichtssituationen ist und daher in der Praxis vielfach eingeübt worden sein muss. Dabei ist die Instruktion an die situationsspezifische Fähigkeit *Entscheidungen treffen* (D) zwar angelehnt, wird jedoch durch die gefragte Spontanität einer wörtlichen Reaktion noch performanznäher, da sie ein Hineinversetzen in die Situation erfordert². Eine Interpretation des Handelns wird nicht verlangt, damit die Voraussetzung gegeben ist, dass das Explizieren nicht reflexiv zugänglichen impliziten Wissens ausbleiben kann. Damit gibt dieser Prompt (*H*) vermutlich ökologisch validere Hinweise auf die Performanz in realen Unterrichtssituationen. Auf die Instruktion zu der Teilkomponente *Interpretation* des *knowledge-based reasoning* bzw. des *PID-Modells* wird damit gänzlich verzichtet. Denn da stünde das explizit verbalisierbare Wissen der Expert/innen im Fokus, sodass zwangsläufig eine Reflexionsebene angeregt würde, die die direkte Reaktion stören könnte (bspw. Baumgartner und Gruber-Mücke 2017). Implizit ist die Interpretation jedoch in der Handlungsreaktion enthalten (ähnlich Hoth 2015).

Erhalten bleibt die Teilkomponente *P* als vielversprechender Indikator von Expertise, schließlich lassen sich Noviz/innen und Expert/innen anhand von distinkten Merkmalen ihrer Wahrnehmung gut unterscheiden (z. B. Dreyfus und Dreyfus 1989). Und doch gilt, dass diese im Rahmen von Kompetenztests aller Domänen zunächst verbalisiert werden muss. Damit fordern Items zur präzisen Wahrnehmung die für Noviz/innen wie Expert/innen ungewohnte Verbalisierung von etwas in der Unterrichtsrealität unbewusst stattfindendem. Da das Verbalisieren von *P* – anders als das präzise Wahrnehmen selbst – in der Unterrichtsrealität nicht vorkommt und damit auch nicht geübt worden sein kann, ist davon auszugehen, dass die Items zur präzisen Wahrnehmung in einer performanznahen Testsituation wie dem *DaZKom-Video-Test* schwieriger sind als Items, die eine (gewohnte) direkte Reaktion (*H*) erfragen.

2.5 Erkenntnisinteresse und Fragestellungen

Unter Einbezug von Erkenntnissen aus der Expertiseforschung wurde ein Versuchsaufbau skizziert, der die Art der Wissensrepräsentation von erfahrenen Lehrkräften bei der Erfassung ihrer situationsspezifischen Fähigkeiten bestmöglich mit einzubeziehen versucht. Ein mögliches Verbalisierungsdilemma sollte demnach dabei bei der Teilkomponente *H* besser umgangen werden können, als dies bei *P* der Fall ist. Da davon ausgegangen werden kann, dass die präzise Wahrnehmung relevanter Situationscharakteristika in der Realität mit der Handlung zusammenhängt, zumindest deren „Grundlage bildet“ (Radtke 1996, S. 97; ähnlich Star und Strickland 2008) oder aber die Teilprozesse parallel und integrativ ablaufen, wie Jacobs et al. (2011) annehmen, lässt sich das Gelingen des Versuchsaufbaus überprüfen. So kann das Testen des statistischen Zusammenhangs zwischen *P* und *H* Hinweise darauf geben, ob die verbalisierte *P* dennoch verlässlich ist, also ob tatsächlich das verbalisiert wird, was auch wahrgenommen wurde und reaktionsrelevant war.

² Vgl. auch Neuweg für eine Übersicht über begründete „[...] Vorbehalte gegenüber dem Konzept der Entscheidung, das scheinbar problemlos zwischen Wissen und Handeln vermittelt.“ (2014, S. 597).

Basierend auf den oben beschriebenen theoretischen Annahmen wird anhand eigener mithilfe eines videobasierten Kompetenztests erhobener Daten zur DaZ-Kompetenz von Lehrkräften geprüft, ob sich dieser Zusammenhang zwischen der präzisen Wahrnehmung (P) und der spontanen Handlungsreaktion (H) auch in der Testsituation bestätigen lässt, *obwohl* beides über den Umweg der Verbalisierung erfasst wird. Dies geschieht zunächst quantitativ, anschließend qualitativ. Dabei werden die folgenden Hypothesen geprüft:

- I. Aufgaben, mit denen die präzise Wahrnehmung DaZ-relevanter Merkmale in einer Unterrichtssituation gemessen werden soll, werden schwieriger bewältigt als Aufgaben, die eine Handlungsoption einfordern.
- II. Werden Aufgaben zur spontanen DaZ-relevanten Handlungsreaktion korrekt gelöst, korrespondiert dies mit einer korrekten präzisen Wahrnehmung.

3 Methodisches Vorgehen

3.1 Datenerhebung und Testinstrument

Im Rahmen der Prä-Pilotierung des Testinstruments des *DaZKom-Video*-Projekts wurden im Frühjahr 2018 20 Videovignetten samt je zwei Items (P und H) an $N=124^3$ Proband/innen erprobt ($n=84$ Studierende des Lehramts oder Linguistik und $n=40$ berufserfahrene Lehrkräfte). Die Ergebnisse dieser Gesamtstichprobe wurden zur deskriptiv-quantitativen Beantwortung der Hypothesenprüfung herangezogen und anschließend nach Fällen untersucht, in denen eine laut Testergebnis auf Item-Ebene nicht kompetente P der Nennung einer korrekten Handlungsoption H voranging. Diese wurden für eine qualitative Auswertung der Antworten ausgewählt.

Die standardisierte Datenerhebung fand an Universitäten, Weiterbildungseinrichtungen und Schulen in Nordrhein-Westfalen, Niedersachsen und Berlin statt. Zunächst gab eine der jeweils zwei geschulten Testleiterinnen eine kurze leitfadensbasierte Einführung, in der u. a. auf den Ablauf und Hintergrund des Tests, insbesondere die DaZ-Relevanz aller Items, eingegangen wurde. Die Teilnehmer/innen bearbeiteten eine von zwei Testversionen anschließend selbständig mithilfe von Headsets auf Tablets. Eine Beispielvignette inklusive -antwort wurde zu Beginn gezeigt, es folgten zehn Videovignetten mit jeweils zwei offenen Items⁴. Zusätzlich wurde ein Multiple-Choice-Fragebogen zum beruflichen Hintergrund der Proband/innen, ihrer Unterrichtserfahrung in Jahren, Erfahrung mit DaZ-Lerner/innen sowie zu ihren Lerngelegenheiten im Bereich DaZ eingesetzt (vgl. Lemmrich et al. 2019).

³ Die Prä-Pilotierungsstudie umfasste zudem $n=13$ Forschende aus dem Bereich Didaktik, die jedoch nicht Gegenstand der Auswertungen dieses Aufsatzes sind.

⁴ Zwei der Videovignetten wurden aus unterschiedlichen Gründen (Missverstehen durch Testpersonen, testtheoretische Erwägungen) direkt im Anschluss an die Erhebungsphase aus dem Instrument entnommen und sind daher nicht Teil der nachfolgenden Auswertung.

Die Videovignetten zeigen bis zu dreiminütige authentische Situationen aus dem Fachunterricht der Sek. I (Mathematik, Deutsch, Naturwissenschaften), deren inhaltliche Passung zu den Subdimensionen des DaZ-Kompetenzmodells (Köker et al. 2015) und DaZ-Relevanz in einem Rating mit $N=3$ universitären DaZ-Expert/innen bestätigt wurde. Um die Unmittelbarkeit von Lehrer/innenhandeln in realen Situationen bestmöglich zu simulieren und damit die ökologische Validität wie oben dargestellt zu erhöhen, wurden maximal offene Fragestellungen gewählt. Einer Fragestellung zur präzisen Wahrnehmung P („Was nehmen Sie wahr?“) folgte eine zur simulierten Handlung H („Sie sind die Lehrkraft in dieser Situation, wie reagieren Sie wörtlich?“ oder „Wie würden Sie in dieser Situation handeln?“), die jeweils mündlich beantwortet werden mussten. Die Vignetten endeten zu diesem Zweck meist mit einer Schüler/innenäußerung, auf die direkt („wörtlich“) reagiert werden sollte.

3.2 Punktevergabe mithilfe eines Kodierleitfadens: Was gilt als „richtig“ Antwort?

Die zu Anfang des Tests gezeigte Beispielvignette samt -antwort sollte die Fokussierung der Teilnehmer/innen auf DaZ-relevante Situationscharakteristika trotz der Offenheit der Prompts sicherstellen. Dabei ist grundsätzlich die Verbindung DaZ-relevanter Merkmale mit solchen, die bspw. der Klassenführung zuzuordnen sind, nicht auszuschließen. Eine mit „richtig“ kodierte Antwort setzt jedoch bspw. voraus, dass scheinbar fachlich begründete Probleme als Merkmale wahrgenommen werden, die mit Aspekten der Sprachbildung in Verbindung stehen bzw. mit Methoden des sprachsensiblen Fachunterrichts möglicherweise gelöst werden können. Daher wurden ausschließlich Videovignetten ausgewählt, in denen Situationsmerkmale auftraten, die mit einer „DaZ-Brille“ salient waren. Diese Salienz wurde mithilfe der Generierung eines datenbasierten Kodierleitfadens indirekt getestet, indem Expert/innen dieselben Prompts vorgelegt wurden wie später den Proband/innen. Dies geschah im Rahmen von Expert/innenratings ($N=12$, davon $n=6$ Universitätsangehörige aus dem Bereich DaZ sowie Lehrkräfte ($n=6$) mit jeweils mehr als 30-jähriger Berufserfahrung und nebenberuflicher Tätigkeit als DaZ-Weiterbildner/in). Zusätzliche Informationen erhielten die Expert/innen nicht. Ihre Antworten auf die Prompts wurden qualitativ-inhaltsanalytisch ausgewertet und durch die Projektleitung final den folgenden Punktzahlen zugeordnet: die situationsspezifischen DaZ-relevanten Aspekte wurden *nicht* (0 Pkt., *falsch*), *teilweise* (1 Pkt.) oder *gänzlich* (2 Pkt., *richtig*) erfasst. Aus den Expert/innen-Antworten entstand datenbasiert der Kodierleitfaden. Dieser wurde damit bewusst nicht aufgrund von theoretischen Forschungsannahmen konstruiert, sondern bildet tatsächliche Wahrnehmungen und Handlungsoptionen von Expert/innen aus Wissenschaft *und* Praxis ab und nutzt ausschließlich diese als Grundlage der Norm (Hecker und Nimz 2020). Die finale Zuordnung zu den drei Codes erfolgte durch das Projektteam. Die Anzahl der möglichen Aspekte bei den jeweiligen Punktzahlen variiert je nach Item, abhängig von der Zahl der durch die Expert/innen genannten Optionen. Inhaltlich korrespondieren die Aspekte zwischen P und H häufig, sind aber jeweils datengeleitet entstanden und wurden nicht nachträglich aufeinander abgestimmt.

Tab. 1 Übersicht über die Mittelwerte in *P* und *H*

Video-Nr	<i>N</i> (<i>P</i>)	MW (<i>P</i>)	SD (<i>P</i>)	<i>N</i> (<i>H</i>)	MW (<i>H</i>)	SD (<i>H</i>)
1	60	0,05	0,22	54	0,31	0,47
2	58	0,03	0,18	57	0,05	0,22
3	59	0,17	0,38	57	0,46	0,68
4	55	0,29	0,63	54	0,63	0,49
5	61	0,23	0,42	56	0,14	0,52
6	53	0,09	0,30	53	0,30	0,46
7	53	0,53	0,58	54	0,44	0,66
8	59	0,17	0,46	55	0,31	0,61
9	53	0,15	0,46	52	0,27	0,45
10	54	0,26	0,48	52	0,52	0,78
11	61	0,08	0,28	58	0,21	0,45
12	57	0,12	0,33	53	0,21	0,45
13	59	0,39	0,70	57	0,33	0,61
14	59	0,07	0,25	55	0,24	0,61
15	58	0,07	0,26	50	0,14	0,35
16	58	0,26	0,52	54	0,41	0,71
17	57	0,14	0,40	56	0,16	0,53
18	57	0,09	0,29	56	0,30	0,54

Der Leitfaden wurde anschließend mithilfe der Antworten aus der Prä-Pilotierung mehrmals im Sinne einer Präzisierung überarbeitet und durch weitere Ankerbeispiele ergänzt, wobei alle Antworten von jeweils zwei geschulten unabhängigen Raterinnen doppelt kodiert wurden. Änderungen an dem Leitfaden resultierten in einer erneuten Kodierung der betroffenen Antworten und einer erneuten Besprechung des Leitfadens im Team. Schließlich wurde eine gute Interrater-Übereinstimmung erzielt (PÜ \bar{O} = 88,84, Cohens Kappa κ = 0,76). Für jedes der 18×2 Items liegen Antworten von bis zu 61 Testpersonen⁵ vor (vgl. Tab. 1). Pro Item konnten in der Prä-Pilotierung höchstens zwei Punkte erreicht werden, wobei jeweils die beste genannte Antwort zählte. Wurden etwa zwei Aspekte genannt, die jeweils der Rubrik *teilweise richtig* (=1 Punkt) zuzuordnen sind, wurden diese Punkte nicht addiert, da die Qualität der Antwort schließlich bei *teilweise richtig* blieb. Wurde zusätzlich ein falscher Aspekt genannt, wurde dennoch ein Punkt vergeben.

Ausgewertet werden die Ergebnisse des Gesamttests mithilfe der Item-Response-Theory, insbesondere einem Partial-Credit-Modell. Hierbei werden die Abstände zwischen den Antwortkategorien für jedes Item separat geschätzt (ausführliche Informationen zu den psychometrischen Kennwerten des Instruments finden sich in Lemmrich et al. (2019)). Dabei stellte sich heraus, dass der *DaZKom-Video-Test* in der Prä-Pilotierung noch zu schwer ist (ebd.). Die nachfolgenden Ergebnisse müs-

⁵ In der Prä-Pilotierungsphase wurden sowohl Items als auch Instrument getestet. Schwierigkeiten der Proband/innen bei der Handhabung der Tablets und technische Probleme beim Abspielen einzelner Videos führten zu einer itemabhängigen Quote fehlender Antworten. Da hier aber lediglich Analysen auf Einzelitem-Ebene durchgeführt werden, ist dies unproblematisch.

sen unter diesem Vorbehalt betrachtet werden. Zudem werden in diesem Aufsatz zugunsten der Anschaulichkeit lediglich Mittelwerte auf Einzel-Item-Basis miteinander verglichen. Diese können nicht im Hinblick auf die Kompetenz der Testpersonen interpretiert werden.

Weiterhin sind die Items untereinander unterschiedlich schwierig (vgl. Tab. 1). Auf mögliche inhaltliche Gründe hierfür, wie z. B. eine mehr oder weniger saliente DaZ-Spezifik in den zugehörigen Videos, kann in diesem Rahmen nicht eingegangen werden. Auf mögliche methodische Gründe hingegen weisen die im Folgenden präsentierten Ergebnisse zu den Fragestellungen dieses Beitrags hin.

4 Ergebnisse

4.1 Werden Aufgaben zur DaZ-relevanten P schwieriger bewältigt als solche zu H ?

In der quantitativen Auswertung wurde auf Item-Ebene ein deutliches Ungleichgewicht in Bezug auf die Schwierigkeit der Aufgabenstellungen zu Wahrnehmung und Handlungsreaktion ersichtlich (vgl. Tab. 1). So zeigten die Mittelwerte der Aufgabenschwierigkeiten, dass die Fragen zur Handlungsreaktion (H) den Proband/

Tab. 2 Vergleich der Mittelwerte von und Korrelationen zwischen P und H

Video-Nr	N	$H > P$	$p(H > P)$	r	$p(r)$
1	53	✓**	0,001	-0,17	0,228
2	55	✓	0,659	-0,05	0,735
3	55	✓**	0,002	-0,03	0,858
4	54	✓**	0,001	0,24	0,081
5	56	–	0,228	0,34**	0,010
6	51	✓**	0,006	0,22	0,119
7	52	–	0,389	0,18	0,192
8	53	✓	0,164	0,25	0,075
9	51	✓	0,182	0,08	0,591
10	50	✓	0,063	0,08	0,582
11	54	✓	0,070	0,03	0,833
12	53	✓*	0,033	0,57**	<0,001
13	57	–	0,822	0,68**	<0,001
14	55	✓*	0,028	0,47**	<0,001
15	50	✓	0,371	-0,12	0,411
16	54	✓	0,073	0,57**	<0,001
17	56	✓	0,811	0,32*	0,017
18	56	✓**	0,002	0,41**	0,002

N Anzahl vorliegender Antworten, MW Mittelwert, SD Standardabweichung

r : Korrelation nach Pearson zwischen P und H . */** : Der Mittelwertsunterschied oder die Korrelation ist auf dem Niveau von 0,05(*) bzw. 0,01(**) signifikant

$H > P$: der beim jeweiligen Item in der H -Antwort erzielte Mittelwert ist größer (✓) als der in der P -Antwort

innen leichter fielen als die Fragen zur präzisen Wahrnehmung (*P*): In einer Frage zur Handlungsreaktion wurde im Mittel ein Wert von $MW=0,30$ ($SD=0,53$) erreicht, in einer Frage zu *P* dagegen im Mittel nur $MW=0,18$ ($SD=0,40$); der Unterschied ist statistisch signifikant. Nach Bildung der Summenscores aller in den Items zu *P* bzw. *H* erzielten Punkte wurde insgesamt im Mittel fast ein ganzer Punkt mehr in den Handlungsreaktionen erreicht ($MW(P)=1,46$ ($SD=1,48$), $MW(H)=2,39$ ($SD=1,23$)). Auch dieser Unterschied ist signifikant. Auf individueller Paar-Item-Ebene ist der erzielte Wert für die Frage nach der Handlungsreaktion in fünfzehn von achtzehn Vignetten höher, in sechs davon signifikant (vgl. Tab. 2).

Zudem ist auffällig, dass die Kombination von zwei richtigen oder zwei teilweise richtig gelösten Aufgaben bei *P* und auch bei *H* weitaus seltener vorkam (46 Mal) als die Kombination von falsch gelösten Aufgaben bei *P* mit einer teilweise richtigen oder richtigen Antwort bei *H* (183 Mal). In vielen Fällen scheint also kein Zusammenhang zwischen beidem zu bestehen.

Die Korrelationen nach Pearson zwischen *P* und *H* spiegeln dies nur zum Teil. Während sie zu elf der Videovignetten nicht signifikant sind, finden sich zu den übrigen sieben dagegen schwache bis zufriedenstellende signifikante Korrelationen (zwischen $r=0,32$, $p=0,017$, und $r=0,68$, $p<0,001$, vgl. Tab. 2). Zwischen diesen Partner-Items besteht also bei der Antwort zu *P* und der zur Handlungsreaktion statistisch ein Zusammenhang: Wer das Richtige erkannt und verbalisiert hat, nennt auch eine richtige Handlungsoption. Bei allen Items gibt es jedoch Fälle, in denen die Proband/innen falsch wahrnehmen, aber richtig oder teilweise richtig reagieren (*H*). Die erste Hypothese kann demnach vorerst angenommen werden: In vielen Fällen scheint das Nennen einer Handlungsoption leichter als präzises Wahrnehmen zu sein. Außerdem zeigen sich häufig keine Zusammenhänge zwischen dem Lösungserfolg in einem und dem Abschneiden im Partner-Item (jeweils *P* oder *H*).

Da in der quantitativen Auswertung kein inhaltliches Matching der bei *P* und *H* genannten Aspekte durchgeführt wurde, werden im Folgenden die Fälle qualitativ untersucht, die den Kontrast einer falschen *P*-Antwort und einer richtigen Handlungsoption abbildeten. Bei elf Testpersonen konnte auf diese Weise insgesamt 14 Mal ein Antwortmuster identifiziert werden, bei dem die Proband/innen die Höchstpunktzahl in der *H*-Antwort erreichten, bei Aufgaben zur präzisen Wahrnehmung aber scheiterten⁶.

4.2 Inwiefern korrespondieren die Antworten zu *P* und *H*?

Bei der qualitativen Auswertung der erhobenen Daten konnten drei Antworten identifiziert werden, bei denen in der *H*-Antwort bestimmte Situationsmerkmale ergänzt wurden, die in der *P*-Antwort zuvor nicht verbalisiert worden waren, aber offenbar als relevant für die Handlungsoption empfunden wurden. Inhaltlich zeigte sich damit in der Antwort auf die *H*-Frage ein neuer Aspekt der Wahrnehmung, der

⁶ Der umgekehrte Fall (zwei Punkte in *P*, gefolgt von keinem Punkt in *H*), trat fünf Mal auf. Statt ein Validitätsproblem zu offenbaren, weist diese Kombination darauf hin, dass Proband/innen (noch) nicht adäquat DaZ-relevant und situationsspezifisch auf ein Problem reagieren können, obwohl sie es wahrgenommen haben.

vorab nicht genannt worden war. In drei weiteren Fällen fehlte diese nachträgliche Verbalisierung der für die Handlungsoption relevanten Situationsmerkmale, sodass die kompetente *H*-Antwort gar nicht mit einer entsprechenden Wahrnehmung in Verbindung zu stehen schien. In den *P*-Antworten wurden zwar bestimmte Situationsmerkmale genannt, für die Handlungsoption schienen diese aber keine Rolle gespielt zu haben. Beispielsweise verbalisierte eine Probandin, dass sie die Äußerungen der Schüler/innengruppe im Video als Provokation verstehe, nahm in ihrer *H*-Antwort ihre in *P* geäußerte Bemerkung aber nicht mehr auf. Stattdessen nannte sie eine angemessene Handlungsoption, die zeigte, dass ihr die DaZ-Relevanz der Situation dennoch aufgefallen sein musste, ohne dass sie dies zuvor verbalisiert hatte. Auch bei einer weiteren Lehrkraft ließ sich erst in der Handlung rekonstruieren, dass offenbar etwas wahrgenommen worden war, das aber nicht verbalisiert wurde: Obwohl die Lehrkraft in *P* auf einer rein beschreibenden und nicht DaZ-relevanten Ebene verblieb (der Schüler sei genervt und wenig motiviert), erkannte sie das Potential der Situation für eine sprachensible Lösung offenbar doch – obwohl nicht verbalisiert – implizit, wie in ihrer *H*-Antwort deutlich wurde: *„Ich würde dem Schüler eine Checkliste geben, in der die Kriterien für eine gute Einleitung drinstehen, und würde ihn fragen, ob er seine Einleitung mit diesen Kriterien, also mit dieser Checkliste, nochmal überarbeiten möchte.“*

Dies spricht gegen die zweite Hypothese. Es ist also möglich, dass Handlungsrelevantes wahrgenommen wird, dies jedoch nicht oder nicht vollständig verbalisiert wird. Die Bestätigung der ersten Hypothese kann demnach vorsichtig ergänzt werden: Möglicherweise ist das präzise Wahrnehmen nicht *schwieriger* als die Handlungsreaktion, sondern tatsächlich nur *nicht verbalisierbar*.

5 Diskussion

In der Kognitionspsychologie wird davon ausgegangen, dass Wahrnehmungs- und Handlungsprozesse zusammenhängen und häufig unbewusst ablaufen (vgl. Abschn. 2.3). In performanznahen Testsituationen, die diese Prozesse analytisch getrennt simulieren, werden beide Komponenten⁷ direkt erfragt und über den Umweg der Verbalisierung erfasst, um Hinweise auf die tatsächliche Performanz zu gewinnen. Geprüft wurde in dieser Studie am Beispiel der Domäne *DaZ*, ob es Anzeichen dafür gibt, dass die notwendige Verbalisierung die Erfassung der situationsspezifischen Fähigkeit *präzise Wahrnehmung (P)* erschwert. Dies gelang, indem der Zusammenhang mit den ökologisch valideren Antworten (vgl. dazu insbesondere 2.4) zur spontanen Handlungsreaktion (*H*) von Lehrkräften im *DaZKom-Video-Test* quantitativ und qualitativ geprüft wurde. Dabei finden sich in der quantitativen Analyse tatsächlich erste Indizien für ein Verbalisierungsdilemma: Items zu *P* sind offenbar schwieriger als die zu *H*.

Das Verhältnis zwischen den beiden Indikatoren für Performanz stellt sich in der Testsituation zudem nicht eindeutig dar: Zwar zeigt sich häufig ein Zusammenhang

⁷ Wie bereits in Abschn. 2.3 dargestellt wurde, ist die häufig erhobene dritte situationsspezifische Fähigkeit *Interpretation* vermutlich ebenfalls vom Verbalisierungsdilemma betroffen.

zwischen dem, was in der *P*-Antwort und dem, was in der *H*-Antwort verbalisiert wird. Dieser auf den ersten Blick widersprüchliche statistische Zusammenhang zwischen *P* und *H* kann aber mit der Tatsache in Verbindung gebracht werden, dass die Items insgesamt zu schwer sind. Viele Testpersonen erreichten daher in den Partner-Items *P* und *H* dieselbe Punktzahl, nämlich 0. In Fällen allerdings, in denen zumindest in einem der Partner-Items Punkte erzielt werden konnten, scheint die präzise Wahrnehmung der Unterrichtssituationen nur zum Teil mit den genannten Handlungsoptionen zusammenzuhängen. Die eingangs angestellte Vermutung, dass möglicherweise die Verbalisierung, nicht aber das tatsächliche Wahrnehmungsvermögen bei den Lehrkräften unzureichend sein könnte, verstärkte sich durch die qualitative Auswertung.

Momentan bleibt damit zweifelhaft, ob die Antworten von angehenden und erfahrenen Lehrkräften auf die Frage „Was nehmen Sie wahr?“ tatsächlich geeignet sind, um Rückschlüsse auf deren Kompetenz ziehen zu können. Die oben formulierten theoretischen Annahmen konnten also durch empirische Hinweise darauf ergänzt werden, dass zumindest bei der Zielgruppe der Lehrkräfte die Messwerte für die präzise Wahrnehmung in der indirekten Erfassung über die Verbalisierung in der Domäne DaZ nicht valide sind.

Nun werden die meisten Lehrkräfte im Berufsleben auch nicht damit konfrontiert, sich bewusst zu machen und zu verbalisieren, was sie wahrnehmen. Darin konnten die Proband/innen demnach ohnehin auch keine Expertise entwickeln. Während also begründet davon ausgegangen werden kann, dass die Komponente *H* aufgrund von Erfahrung im Umgang mit Schüler/innen-Äußerungen leichter zu beantworten ist, kann zunächst nicht ausgeschlossen werden, dass allein fehlende Routine im Verbalisieren des eigentlich unbewusst ablaufenden Prozesses der Wahrnehmung für die Schwierigkeiten der befragten Lehrpersonen verantwortlich ist, diese aber nach entsprechenden Trainings sehr wohl verbalisiert werden könnte. Auch nach einem externen Training, etwa im Rahmen von Lehrveranstaltungen oder Weiterbildungen, ist aber fraglich, ob der sich bewusstgemachte Prozess der Wahrnehmung dann mehr mit dem unbewusst ablaufenden intuitiven Wahrnehmen übereinstimmt, der tatsächlich mit einer Handlungsreaktion zusammenhängt. Polanyi zumindest geht davon aus, dass eine solche Bewusstmachung den Charakter einer „destruktiven Analyse“ hat: sobald über einen Teilprozess einer Handlung nachgedacht werde, verändere sich diese (1974, S. 50–52). Danach wäre die Komponente *P* auch nach Trainings nicht valider erfassbar.

Daher stellen sich mehrere Herausforderungen an aktuelle und zukünftige Forschungsdesigns. Zwar wurde erst in dieser Studie, also sobald die situationsspezifischen Fähigkeiten mithilfe von maximal offenen Items erfasst wurden, überhaupt deutlich, dass die nötige Verbalisierung der präzisen Wahrnehmung bei Lehrkräften ein Validitätsproblem darstellen könnte. Auf geschlossene Items zurückzugreifen, die den Fokus stärker vorgeben, stellt jedoch keine Lösung hierfür dar. Denn ein tatsächliches präzises Wahrnehmen wird in solchen Designs schlicht unnötig, und auch die Intuitivität der Wahrnehmung beim *realen* Handeln erfahrener Lehrkräfte kann in einer geschlosseneren Testsituation nicht besser abgebildet werden. Die hier berichteten Überlegungen und Ergebnisse sind dabei besonders relevant für das performanznahe Erheben von Kompetenzen *erfahrener* Lehrkräfte. Fachwissen von

Studierenden dagegen, noch nicht durch Erfahrungen angereichert und somit umstrukturiert, lässt sich vermutlich auch mit weniger offenen Fragestellungen valide erheben.

Einschränkend in Bezug auf die Generalisierbarkeit der Ergebnisse dieser Erprobungsstudie muss gelten, dass vergleichsweise geringe Fallzahlen betrachtet wurden. Auch waren die Items insgesamt noch zu schwer für die Proband/innen. Die Ergebnisse können deshalb vor allem als Hinweise interpretiert werden, die die oben ausgeführten theoretischen Annahmen stützen. Zudem kann auch bei einer performanznahen Methodik weder sicher davon ausgegangen werden, dass die im Test geäußerten *H*-Antworten, also direktes spontanes Adressieren, *tatsächlich* ökologisch valider sind, noch, dass diese auch im eigenen Unterricht so umgesetzt werden würden (vgl. Bischoff et al. 2005). Dazu kommt, dass der Fokus der Vignetten allein durch deren Auswahl, Anfang und Ende bereits vorgegeben ist – damit gilt dies zu einem gewissen Grad auch für das präzise Wahrnehmen. Des Weiteren wurden in dieser Studie nur Ergebnisse aus der Domäne *DaZ* empirisch betrachtet. Das Verbalisierungsdilemma allerdings sollte laut theoretischer Herleitung dennoch nicht auf diesen Bereich beschränkt, sondern ein domänenübergreifendes sein. Dies müsste anhand von ähnlich erhobenen Antworten aus einer anderen Domäne überprüft werden.

Weitere Untersuchungen sind zudem notwendig, um alternative, möglicherweise validere Erfassungsmethoden der präzisen Wahrnehmung bei Lehrkräften zu identifizieren. So könnten Eye-Tracking-Verfahren hilfreich sein, die bereits eingesetzt werden, um kognitive Entscheidungs- oder Lernprozesse besser zu verstehen (vgl. Beach und McConnel 2018). Denkbar wäre auch, die Frage nach der präzisen Wahrnehmung einer Situation vorerst ganz außen vor zu lassen und sich ausschließlich auf die Frage nach einer als direkte Reaktion simulierten Handlung zu konzentrieren (ähnlich Sunder et al. 2016). In diesem Fall könnte man die Wahrnehmung auf Basis der geäußerten Handlungsreaktion rekonstruieren und nicht explizit erfragen (Levin et al. 2009). Forschungsergebnisse weisen bereits darauf hin, dass informierte Handlungsalternativen ein signifikanter Prädiktor besonders hoher Schüler/innenleistungen sind (Kersting et al. 2010); demnach genügen Handlungsreaktionen möglicherweise als Indikatoren von Performanz.

Empfehlenswert erscheint es also, in Testumgebungen für Lehrkräfte vornehmlich anforderungsnahen Aufgaben (wie solche zur spontanen Handlungsreaktion in Bezug auf eine videographierte Unterrichtssituation) zu stellen, um in Kompetenztests nicht „expertise in verbalising“ zu erfassen, sondern – performanznah – „expertise in doing“ (Neuweg 2015b, S. 90).

Funding Open Access funding provided by Projekt DEAL.

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

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betref-

fende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

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

Literatur

- Baumgartner, P., & Gruber-Mücke, T. (2017). Externalisierung und Kommunikation von Tacit Knowledge: Mustersprachen im bildungstechnologischen Design. In A. Kraus, J. Budde, M. Hietzge & C. Wulf (Hrsg.), *Handbuch Schweigendes Wissen: Erziehung, Bildung, Sozialisation und Lernen* (S. 57–68). Weinheim: Beltz Juventa.
- Beach, P., & McConnel, J. (2018). Eye tracking methodology for studying teacher learning: a review of the research. *International Journal of Research & Method in Education*. <https://doi.org/10.1080/1743727x.2018.1496415>.
- Berliner, D.C. (1992). The nature of expertise in teaching. In F.K. Oser, A. Dick & J.-L. Patry (Hrsg.), *Effective and resposable teaching: the new synthesis* (S. 277–248). San Francisco: Jossey-Bass.
- Bischoff, S., Brühwiler, C., & Baer, M. (2005). Videotest zur Erfassung „adaptiver Lehrkompetenz“. *Beiträge zur Lehrerbildung*, 23(3), 382–397.
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R.J. (2015). Beyond dichotomies: competence viewed as a continuum. *Zeitschrift für Psychologie*, 223, 3–13.
- Blömeke, S., König, J., Busse, A., Suhl, U., Benthien, J., Döhrmann, M., & Kaiser, G. (2014). Von der Lehrerausbildung in den Beruf – Fachbezogenes Wissen als Voraussetzung für Wahrnehmung, Interpretation und Handeln im Unterricht. *Zeitschrift für Erziehungswissenschaft*, 17(3), 509–542.
- Bromme, R. (2014). *Der Lehrer als Experte: Zur Psychologie des professionellen Wissens*. Münster: Waxmann. Zugl.: Bielefeld, Univ., Habil.-Schr., 1992 u.d.T.: Bromme, R.: *Der Lehrer als Experte*
- Dreyfus, H.L., & Dreyfus, S.E. (1989). *Mind over machine: the power of human intuition and expertise in the era of the computer*. Oxford: Blackwell.
- van Es, E. A., & Sherin, G. M. (2002). Learning to notice: scaffolding new teachers Interpretations of classroom interactions. *Journal of Technology and Teacher Education*, 10, 571–596.
- Gigerenzer, G., & Gaissmaier, W. (2012). *Intuition und Führung: Wie gute Entscheidungen entstehen*. Gütersloh: Bertelsmann Stiftung.
- Goodwin, C. (1994). Professional Vision. *American Anthropologist*, 96 (3), 606–633.
- 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 (Hg.): *Berlin-Brandenburger Beiträge zur Bildungsforschung 2020* (S. 39–64). Peter Lang: Berlin.
- Hepberger, B., Lindmeier, A. M., Moser Opitz, E., & Heinze, A. (2017). „Zähl’ nochmal genauer!“ – Handlungsnahe mathematikbezogene Kompetenzen von pädagogischen Fachkräften erheben. In S. Schuler, C. Streit & G. Wittmann (Hrsg.), *Perspektiven mathematischer Bildung im Übergang vom Kindergarten zur Grundschule* (S. 239–253). Wiesbaden: Springer.
- Herzog, W. (2018). Die ältere Schwester der Theorie: Eine Neubetrachtung des Theorie-Praxis-Problems. *Zeitschrift für Pädagogik*, 64, 812–830.
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): a framework for perception and action planning. *Behavioral and Brain Sciences*, 24, 849–878.
- Hoth, J. (2015). *Situationsbezogene Diagnosekompetenz von Mathematiklehrkräften: Eine Vertiefungsstudie zur TEDS-Follow-Up-Studie*. Dissertation. Wiesbaden: Springer Spektrum.
- Hoth, J., Kaiser, G., Dohrmann, M., & König, J. (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). Cham: Springer.
- Jacobs, V.R., Lamb, L.L., Philipp, R. A., & Schappelle, B.P. (2011). Deciding how to respond on the basis of children’s understandings. In M.G. Sherin, V.R. Jacobs & R. A. Philipp (Hrsg.), *Mathematics teacher noticing: Seeing through teachers’ eyes* (S. 97–116). New York: Routledge.
- Kersting, N.B., Givvin, K.B., Sotelo, F.L., & Stigler, J.W. (2010). Teachers’ analyses of classroom video predict student learning of mathematics: further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education*, 61(1–2), 172–181.

- Koch-Priewe, B. (1986). *Subjektive didaktische Theorien von Lehrern: Tätigkeitstheorie, bildungstheoret. Didaktik u. alltäg. Handeln im Unterricht*. Frankfurt a.M.: Haag & Herchen. Zugl.: Marburg, Univ., Diss.:1985.
- Köker, A. (2018). Zur Relevanz (bildungs-)sprachlicher Förderung in Schule und Fachunterricht. In I. T. Ehmke, S. Hammer, A. Köker, U. Ohm & B. Koch-Priewe (Hrsg.), *Professionelle Kompetenzen angehender Lehrkräfte im Bereich Deutsch als Zweitsprache* (S. 39–57). 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 (Hrsg.), *Kompetenzerwerb an Hochschulen: Modellierung und Messung: Zur Professionalisierung angehender Lehrerinnen und Lehrer sowie frühpädagogischer Fachkräfte* (S. 177–205). Bad Heilbrunn: Klinkhardt.
- Krummenacher, J., & Müller, H. (2017). Aufmerksamkeit. In J. Müsseler & M. Rieger (Hrsg.), *Allgemeine Psychologie* (S. 103–151). Berlin: Springer.
- 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* (S. 190–204). Weinheim: Beltz Juventa.
- Levin, D. M., Hammer, D., & Coffey, J. E. (2009). Novice teachers' attention to student thinking. *Journal of Teacher Education*, 60(2), 142–154.
- Neueweg, G. H. (2014). Das Wissen der Wissensvermittler: Problemstellungen, Befunde und Perspektiven der Forschung zum Lehrerverwissen. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 583–614). Münster: Waxmann.
- Neueweg, G. H. (2015a). Kontextualisierte Kompetenzmessung: Eine Bilanz zu aktuellen Konzeptionen und forschungsmethodischen Zugängen. *Zeitschrift für Pädagogik*, 61(3), 377–383.
- Neueweg, G. H. (2015b). Tacit knowing and implicit learning. In *Das Schweigen der Könner: Gesammelte Schriften zum impliziten Wissen* (S. 81–98). Münster: Waxmann.
- Polanyi, M. (1974). *Personal knowledge: towards a post-critical philosophy*. Chicago: UCP.
- Radtko, F.-O. (1996). *Wissen und Können: Die Rolle der Erziehungswissenschaft in der Erziehung*. Wiesbaden: VS.
- Ross, K. G., Shafer, J. L., & Klein, G. (2006). Professional judgments and “naturalistic decision making”. In K. A. Ericsson (Hrsg.), *The Cambridge handbook of expertise and expert performance* (S. 403–419). Cambridge: CUP.
- Schleppegrell, M. J., & O'Hallaron, C. L. (2011). Teaching academic language in L2 secondary settings. *Annual Review of Applied Linguistics*, 31, 3–18.
- Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal*, 51(4), 739–771.
- Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer“: Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung: Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (S. 296–306). Weinheim: Beltz.
- Sherin, M. G. (2001). Developing a professional vision of classroom events: Teaching elementary school mathematics. In *Beyond classical pedagogy: Teaching elementary school mathematics* (S. 75–93). Hillsdale, N.J.: Erlbaum.
- Sherin, G. M., & van Es, E. A. (2008). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, 60(1), 20–37.
- Shulman, L. S. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review*, 57, 1–23.
- 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.
- Stürmer, K., & Seidel, T. (2015). Assessing professional vision in teacher candidates. *Zeitschrift für Psychologie*, 223, 54–63.
- Sunder, C., Todorova, M., & Möller, K. (2016). Kann die professionelle Unterrichtswahrnehmung von Sachunterrichtsstudierenden trainiert werden? – Konzeption und Erprobung einer Intervention mit Videos aus dem naturwissenschaftlichen Grundschulunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 22, 1–12.
- Wahl, D. (2002). Mit Training vom trägen Wissen zum kompetenten Handeln? *Zeitschrift für Pädagogik*, 48, 227–241.