

Videobasierte Erfassung wissensbasierten Verarbeitens als Teilprozess der professionellen Unterrichtswahrnehmung – Analyse eines geschlossenen und offenen Verfahrens

Madeleine M. Müller  · Bernadette Gold 

Eingegangen: 16. Juli 2021 / Überarbeitet: 7. März 2022 / Angenommen: 12. Juli 2022 / Online publiziert: 7. Oktober 2022
© Der/die Autor(en) 2022

Zusammenfassung Die professionelle Wahrnehmung als Erkennen und wissensbasiertes Analysieren relevanter Ereignisse und Situationen gilt als eine grundlegende Voraussetzung für das professionelle Handeln einer Lehrkraft. Zur Erfassung werden meist kontextualisierte Instrumente mit Videostimulus verwendet, welche sowohl geschlossene Erfassungsformate wie Ratingitems oder auch offene Erfassungsformate wie schriftliche Kommentierungen verwenden. Da die Zusammenhänge zwischen den Erfassungsformaten unter der Verwendung unterschiedlicher Videos in bisherigen Studien eher niedrig ausfallen, geht die vorliegende Querschnittsstudie ($N=309$ Lehramtsstudierende des Bachelors und Masters) der Frage nach, inwiefern sich Zusammenhänge bei Verwendung eines identischen Videostimulus zeigen, und betrachtet Hinweise auf Konstrukt- und Kriteriumsvalidität der beiden Erfassungsformate. Zusammenfassend deuten niedrige bis moderate Zusammenhänge zwischen geschlossener und offener Erfassungsart zur Messung der professionellen Wahrnehmung darauf hin, dass nicht auf die Messung desselben Konstruktes geschlossen werden kann. Weiterhin zeigte sich kein eindeutiges Zusammenhangsmuster mit dem professionellen Wissen über Klassenführung, das für die Überlegenheit einer Erfassungsart sprach. Erwartungskonform konnten zwischen beiden Erfassungsformaten Gruppenunterschiede im Sinne der Kriteriumsvalidität identifiziert werden, wobei die geschlossene Erfassung zu Teilen sensitiver war.

Madeleine M. Müller (✉)

Erziehungswissenschaftliche Fakultät, Inklusive Bildungsprozesse mit dem Schwerpunkt emotionale und soziale Entwicklung, Universität Erfurt, Nordhäuser Str. 63, 99089 Erfurt, Deutschland
E-Mail: Madeleine.mueller@uni-erfurt.de

Prof. Dr. Bernadette Gold

Fakultät für Erziehungswissenschaft, Psychologie und Bildungsforschung, Institut für Allgemeine Didaktik und Schulpädagogik, Technische Universität Dortmund, Martin-Schmeißer-Weg 4–8, 44227 Dortmund, Deutschland
E-Mail: bernadette.gold@tu-dortmund.de

Schlüsselwörter Professionelle Wahrnehmung · Klassenführung · Erfassungsmethoden/Geschlossene und offene Erfassung · Konstrukt- und Kriteriumsvalidität

Video-based measurements of knowledge-based reasoning as a process of professional vision—analysis of a closed and open task format

Abstract The professional vision as noticing and knowledge-based reasoning of relevant events and situations is a crucial prerequisite for a teacher to act professionally. It is usually measured with contextualized instruments using a video stimulus, combined with either closed formats like rating items or open formats like written comments. Since correlations between these measurement formats related to different videos have been rather low in previous studies, the present cross-sectional study ($N=309$ bachelor and master student teachers) examined whether correlations were higher when identical video stimuli are used. Furthermore, evidence for construct- and criterion validity were considered. In summary, low to moderate correlations between closed and open measurement formats for assessing professional vision indicated that different constructs are being measured. Further, there was no clear pattern between the professional knowledge of classroom management supporting one of the measurement formats as more valid. As expected, both measurement formats identified group differences in terms of criterion validity, with the closed format being more sensitive.

Keywords Professional vision · Classroom management · Closed and open measurement formats · Construct and criterion validity

1 Einleitung

Die professionelle Unterrichtswahrnehmung stellt eine entscheidende Voraussetzung für erfolgreiches Lehrer*innenhandeln dar (Barth 2017). Sie beinhaltet Prozesse selektiver Aufmerksamkeitsfokussierung (*noticing/selective attention*) und wissensbasierten Verarbeitens (*knowledge-based reasoning*) von unterrichtsrelevanten Ereignissen und Situationen (Seidel et al. 2010; Sherin und van Es 2009). Für die Erfassung der professionellen Unterrichtswahrnehmung werden meistens kontextualisierte Instrumente mit Videostimulus eingesetzt, welche den situativen Kontext des unterrichtlichen Handelns berücksichtigen (König 2015). In diesen Testinstrumenten werden sowohl geschlossene Erfassungsformate wie Ratingitems (z. B. Meschede et al. 2015; Seidel und Stürmer 2014) als auch offene Erfassungsformate verwendet, in welchen eine gezeigte Unterrichtssequenz hinsichtlich ausgewählter Schwerpunkte meistens analysiert und schriftlich kommentiert werden soll (z. B. Kersting et al. 2012; Sherin und van Es 2009). Zusammenhänge zwischen offenen und geschlossenen Formen der Erfassung der professionellen Unterrichtswahrnehmung, meist unter Verwendung unterschiedlicher Unterrichtssequenzen je Erfassungsformat, fallen eher niedrig aus (z. B. Frommelt et al. 2019; Gold et al. 2016; König et al. 2014) und stellen somit die konvergente Validität der Instrumente in Frage. Dabei

ist jedoch bislang unklar, ob die niedrigen Zusammenhänge auf die Konfundierung der jeweiligen Erfassungsformate mit unterschiedlichen Videostimuli zurückzuführen sind, sodass sich bei der Verwendung *desselben* Videostimulus höhere Korrelationen und somit Hinweise auf konvergente Validität ergeben würden. Die Frage, ob beide Erfassungsarten dasselbe Konstrukt erfassen ist jedoch für die Vergleichbarkeit von Studienergebnissen, die auf solche unterschiedlichen Instrumente zurückgreifen, zu diesem Thema überaus relevant. Beispielsweise existieren Interventionsstudien zur Förderung der professionellen Unterrichtswahrnehmung, deren Ergebnisse zur Wirksamkeit sich trotz ähnlicher Designs teilweise unterscheiden und die u. a. auf unterschiedliche Erfassungsformate für die Wirksamkeitsprüfung zurückgreifen (z. B. Frommelt et al. 2019; Hellermann et al. 2015). Auch Zusammenhänge mit performanten Maßen wie Unterrichtsqualität oder Outcomevariablen wie Schüler*innenleistung sollten differenzierter beleuchtet werden. Zudem ist nicht klar, welche der geschlossen und offen erfassten Prozesse der professionellen Unterrichtswahrnehmung prädiktiv valider sind (z. B. Gold et al. 2021a; Kersting et al. 2012).

Das Ziel der vorliegenden Querschnittstudie ist die Untersuchung ausgewählter Validitätskriterien (Konstrukt- und Kriteriumsvalidität) einer offenen (Kodierung einer schriftlichen Videoanalyse) und einer geschlossenen Erfassung (Ratingitems) der professionellen Unterrichtswahrnehmung bei einem identischen Videostimulus. Unter Kontrolle einer möglichen Konfundierung von Videoszene und Erfassungsformat wird der Frage nachgegangen, welche Zusammenhänge zwischen den beiden Erfassungen professioneller Unterrichtswahrnehmung identifiziert werden können. Dabei werden Facetten des wissensbasierten Verarbeitens als Teilprozess der professionellen Unterrichtswahrnehmung betrachtet, da Ratingitems den Prozess der selektiven Aufmerksamkeitsfokussierung bereits vorwegnehmen (Brovelli et al. 2013). Neben Zusammenhängen zwischen den beiden Erfassungsarten werden Zusammenhänge mit dem pädagogisch-psychologischen Wissen und Unterschiede zwischen Bachelor- und Masterstudierenden als Indikator für die Kriteriumsvalidität untersucht.

2 Theoretischer Hintergrund

2.1 Professionelle Unterrichtswahrnehmung

Unterricht stellt eine komplexe Situation dar, in der eine Lehrkraft die Aufmerksamkeit zielgerichtet auf lernrelevante Ereignisse lenken und eine Vielzahl an Informationen verarbeiten muss (Doyle 1986). Diese Fähigkeit der professionellen Unterrichtswahrnehmung beinhaltet die beiden Teilprozesse der selektiven Aufmerksamkeitsfokussierung (*noticing/selective attention*) sowie des wissensbasierten Verarbeitens von Unterricht (*knowledge-based reasoning*) (Seidel et al. 2010; Sherin und van Es 2009). Für letzteres lassen sich je nach Forscher*innengruppe verschiedene Subfacetten ausdifferenzieren (z. B. *Beschreiben, Interpretieren*; Gold und Holodynski 2017; *describe, evaluate, interpret*; Sherin und van Es 2009; *Beschreiben, Erklären, Vorhersagen*; Seidel et al. 2010). Neben verschiedenen angenommenen Niveaustufen des wissensbasierten Verarbeitens der zugrundeliegenden Unterrichtssituation wer-

den in manchen Ansätzen zusätzlich entweder Prognosen über weitere Verläufe des Unterrichts antizipiert (*Vorhersagen*, Seidel et al. 2010) oder Handlungsalternativen generiert (*generation*, Santagata et al. 2007; *suggestions for improvement*, Kersting et al. 2012). Beides schließen Kaiser et al. (2015) unter der situationsspezifischen Fähigkeit *decision-making* ein. Dabei folgen sie der theoretischen Modellierung des PID-Modells von Blömeke et al. (2015), in welchem situationsspezifische Fähigkeiten (*perception, interpretation, decision-making*) als Teil der Lehrkräftekompetenz gesehen werden, welche konzeptuell stark mit den dargestellten Modellierungsansätzen der professionellen Unterrichtswahrnehmung überlappen.

Nach dem PID-Modell basiert die professionelle Unterrichtswahrnehmung auf affektiv-motivationalen Dispositionen und dem professionellen Wissen und ist selbst Voraussetzung für beobachtbares Verhalten im Unterricht. Dementsprechend konnten Zusammenhänge zwischen professioneller Unterrichtswahrnehmung und Unterrichtsqualität (Gold et al. 2021a; König und Kramer 2016; Krauss et al. 2020; Roth et al. 2011) und sogar Schüler*innenleistungen (Kersting et al. 2012; Krauss et al. 2020) gefunden werden, die die Relevanz der Unterrichtswahrnehmung für qualitätsvollen Unterricht unterstreichen. Dabei bezieht sich die Operationalisierung der professionellen Unterrichtswahrnehmung in diesen Studien auf pädagogisch-psychologische (z. B. Seidel und Stürmer 2014) oder fachbezogene Unterrichtsaspekte (z. B. Meschede et al. 2015). In der vorliegenden Studie wird das wissensbasierte Verarbeiten als Teilprozess der professionellen Unterrichtswahrnehmung fokussiert und in Bezug auf den Aspekt *Klassenführung* betrachtet, welche sich als fachübergreifendes Qualitätsmerkmal von Unterricht wiederholt als bedeutsam für die Entwicklung von Lernenden erwiesen hat (Praetorius et al. 2016) und im Folgenden kurz erläutert wird.

Klassenführung trägt durch eine klar strukturierte Unterrichtsgestaltung zu einer Maximierung von aktiver Lernzeit bei (Seidel 2009). Als Zusammenspiel von Planungs- und Steuerungsanforderungen soll eine effektive Klassenführung das primäre Handlungsprogramm, welches sich an der Erreichung der Lernziele orientiert, präventiv durch Regeln und Prozeduren als Ordnungsrahmen des Unterrichts unterstützen, aufrecht halten und stärken (Doyle 2006; Ophardt und Thiel 2008). Gut *etablierte Regeln und Routinen* im Sinne von positiven Verhaltenserwartungen schaffen so den Handlungsrahmen einer effektiven Klassenführung (Anderson et al. 1980; Hellermann et al. 2015; Ophardt und Thiel 2008; Seidel 2009). Dennoch auftretende Unterrichtsstörungen beeinträchtigen oder unterbrechen das primäre Handlungsprogramm (Ophardt und Thiel 2013) und sollten durch präventive Maßnahmen verhindert oder durch angemessenes Eingreifen unterbunden werden (Doyle 2006; Hellermann et al. 2015; Ophardt und Thiel 2008). Voraussetzung hierfür ist das *Monitoring*, welches einen umfassenden Überblick der Lehrkraft über die Geschehnisse im Klassenraum beinhaltet, um den primären Handlungsvektor zu schützen. Die *prozessuale Strukturierung* zeichnet sich durch ein adaptives und schwungvolles Lerntempo mit reibungslosen Übergängen und die Einbeziehung und Beschäftigung möglichst vieler Schüler*innen aus (Gold et al. 2016; Hellermann et al. 2015; Kounin 2006; Ophardt und Thiel 2008) und unterstützt somit ebenfalls das primäre Handlungsprogramm. Klassenführung wird in der vorliegenden Studie somit als erfolgreiche Umsetzung der Aufrechterhaltung und Stabilisierung des primären Hand-

lungsprogramms verstanden und folgend durch die drei Komponenten Monitoring und prozessualer Strukturierung sowie der Etabliertheit von Regeln und Routinen verstanden.

Zusammenfassend setzt erfolgreiches Klassenführungshandeln somit voraus, dass eine Lehrkraft im Unterrichtsgeschehen (klassenführungs-)relevante Ereignisse wahrnehmen und interpretieren kann, um eine passende Handlungsentscheidung treffen zu können (König und Lebens 2012). In der vorliegenden Studie ist die professionelle Unterrichtswahrnehmung von Klassenführung demnach die Fähigkeit, klassenführungsrelevante Situationen in Bezug auf die oben genannten Aspekte zu erkennen und vor dem Hintergrund pädagogisch-psychologischen Wissens zu interpretieren.

2.2 Erfassung der professionellen Unterrichtswahrnehmung

Zur Erfassung der professionellen Unterrichtswahrnehmung werden vignettenbasierte Verfahren verwendet, welche eine kurze, authentische Darstellung von Unterrichtssituationen als Stimulus präsentierten (Rehm und Bölsterli 2014). Bislang existieren vor allem Vignetten zum Fortführen oder zum Bewerten von gezeigten Unterrichtssituationen, wobei authentische oder gestellte Videos, Animationen, Comics, Fotostorys oder Textvignetten bzw. Unterrichtsskripte verwendet werden (Friesen et al. 2018; Herbst und Kosko 2014). Bevorzugt wird jedoch auf videobasierte Instrumente zurückgegriffen, die der unterrichtlichen Komplexität Rechnung tragen und somit enger mit dem tatsächlichen beruflichen Handeln verknüpft sein sollen als Papier-Bleistift-Verfahren (Blömeke 2013; Brovelli et al. 2013; König 2015; Lindmeier 2013; Neuweg 2015; Rehm und Bölsterli 2014). Damit kann im Sinne der ökologischen Validität die Nähe zu realen unterrichtlichen Anforderungssituation berücksichtigt werden, welche wiederum im Sinne der Reliabilität einer gewissen Standardisierung durch die zu beobachtende Situation unterliegt (König 2015).

Bei diesen videobasierten Instrumenten schließen sich einer videografierten Unterrichtssequenz als Stimulus Aufgaben an, die je nach Antwortformat unterschiedliche Anforderungen verlangen. Bereits vorliegende Instrumente zur Erfassung professioneller Unterrichtswahrnehmung verwenden entweder offene Verfahren, wie beispielsweise Laut-Denken-Interviews oder Antworten auf offene Fragen, die anschließend kodiert werden (z. B. Kersting et al. 2012; Mischo et al. 2020; Sherin und van Es 2009), oder geschlossene Verfahren mit Ratingitems, welche mit einem Expertenrating verglichen werden (z. B. Gold und Holodynski 2017; Seidel und Stürmer 2014). Eine Kombination aus Antwortformaten wurde bislang erst selten angewandt (z. B. Frommelt et al. 2019; König et al. 2014; Schäfer und Seidel 2015; Weber et al. 2020), obwohl insbesondere die Prozesse der *selektiven Aufmerksamkeitsfokussierung* sowie das *Generieren von Handlungsalternativen* bislang durch geschlossene Verfahren nicht oder nur unzureichend abgedeckt wurden (Weber et al. 2020).

2.2.1 *Eigenschaften geschlossener und offener Erfassungsformate*

Allgemein betrachtet unterscheiden sich zwischen den verschiedenen Erfassungsformaten das kognitive Aufgabenpotenzial bzw. die erforderlichen kognitiven Prozesse bei der Bearbeitung. In einem Vergleich zwischen vorgegebenen auszuwählenden Antwortformaten (hier Multiple-Choice-Items) und Constructed-Response-Items, in welchen Antworten eigenständig durch den Befragten generiert werden müssen, zeigte Martinez (1999), dass die kognitiven Anforderungen höher bewertet werden, wenn eigenständig formuliert werden muss. Offene Erfassungsformate, welche das ergänzende oder offene Beantworten einfordern, werden deshalb als kognitiv anfordernd eingeschätzt und als schwierigkeitsgenerierendes Merkmal betrachtet (Gehrer 2017). Die Wahl des Erfassungsformats beeinflusst somit das gemessene Konstrukt durch veränderte kognitive Ansprüche (Martinez 1999; Schaper 2009). Offenen Antwortformaten wird in Bezug auf die Erfassung von Kompetenzen in der Lehrer*innenbildungsforschung eine kompetenzorientiertere Anforderungsgestaltung, welche komplexe und realitätsnahe Leistungen einfordert und so eine hohe Inhaltsvalidität erfüllt, zugesprochen (Schaper 2009). Sie erfassen eher fähigkeitsbezogene Konstrukte und sollen anspruchsvollere Aktivitäten und Fähigkeiten auf höherem Kompetenzniveau abdecken können (Schaper 2009; Seifert et al. 2009).

Geschlossene Erfassungsarten wird durch vorgegebene Antwortkategorien mangelnder Transfer zu realen Situationen und die Erfassung von wissens- und fertigungsbezogenen Konstrukten zugesprochen (Hartig und Jude 2007; Schaper 2009). Geschlossene Antwortformate profitieren von höherer Objektivität und Testökonomie. Weitere Vorteile geschlossener Erfassungen sind eine bessere Vergleichbarkeit und Auswertungsökonomie, sie beziehen jedoch ausschließlich die Antwortkategorien aus dem Bezugsrahmen der Forscher*innen ein (Martinez 1999; Schnell et al. 2018). Bei der offenen Erfassung, wie dem schriftlichen Kommentieren einer Unterrichtssituation, besteht deshalb der Hauptvorteil darin, ohne die Vorgabe von konkreten Antwortmöglichkeiten nicht bereits in eine bestimmte Richtung gelenkt zu werden, durch einen ungestützten Abruf eher an tatsächliche Wissensbestände zu gelangen und die Ratewahrscheinlichkeit zu minimieren (Schnell et al. 2018). So zeigt sich bei einer offenen Erfassung der professionellen Unterrichtswahrnehmung präziser, auf welche Unterrichtsereignisse sich Lehrkräfte in welchem Ausmaß fokussieren (Frommelt et al. 2019). Ratingitems nehmen dagegen durch vermeintliches Lenken der Aufmerksamkeit das Identifizieren von relevanten Situationen als notwendiger Prozess der selektiven Aufmerksamkeitsfokussierung vorweg und könnten somit zu einer Überschätzung der professionellen Unterrichtswahrnehmung führen (Brovelli et al. 2013).

Andererseits fällt erfahrenen Lehrkräften das Verbalisieren und Begründen des im Handeln gezeigten impliziten Wissens schwer (Neuweg 2015; Shulman 1987). Nach dem so genannten Verbalisierungsdilemma (Hecker et al. 2020) identifizieren und interpretieren Lehrkräfte in Unterrichtsvideos durchaus relevante Ereignisse automatisiert, können diese jedoch mitunter nicht explizit artikulieren, während sie bei Ratingitems in der Lage wären, eine angemessene Einschätzung zu treffen. Ein Nachteil einer offenen Erfassung der professionellen Unterrichtswahrnehmung

ist somit neben dem Voraussetzen von Artikulationsfähigkeit die Zugänglichkeit expliziten Wissens (Schnell et al. 2018).

Je nach Erfassungsformat können sich also Unterschiede hinsichtlich testökonomischer und psychometrischer Güte sowie der kognitiven Anforderungskomplexität zeigen, obwohl mit beiden Formaten dieselbe Fähigkeit erfasst werden soll. In Bezug auf die professionelle Unterrichtswahrnehmung von Klassenführung stellt sich deshalb die Frage, inwiefern mit einer offenen und einer geschlossenen Erfassung dasselbe Konstrukt gemessen wird.

2.2.2 *Stand der Forschung zu Zusammenhängen zwischen geschlossenen und offenen Formaten zur Erfassung der professionellen Unterrichtswahrnehmung*

Die unterschiedlichen Eigenschaften der beiden Erfassungsformate spiegeln sich auch in Studien wider, welche nur geringe oder gar Nullkorrelationen zwischen geschlossenen und offen erhobener professioneller Unterrichtswahrnehmung finden konnten. Die Interpretation dieser Befunde wird jedoch dadurch erschwert, dass die beiden Erfassungsformate häufig mit der Erhebung der Teilprozesse von professioneller Unterrichtswahrnehmung konfundiert sind und sich teilweise auf unterschiedliche Videostimuli beziehen. Gold et al. (2016) berichteten beispielsweise bei dem Einsatz unterschiedlicher Videostimuli, dass die offenen mittels einer schriftlichen Analyse erfassten erkannten Klassenführungsevents (noticing) und ihre Interpretation (knowledge-based reasoning) kaum mit einer geschlossenen Erfassung der Interpretation von Klassenführung (knowledge-based reasoning) korrelierten. König et al. (2014) verwendeten ein videobasiertes Instrument mit Fokus auf pädagogischen Aspekten, das aus drei Videosequenzen bestand und zu jedem Video sowohl Multiple-Choice-Items als auch offene Fragen stellte. Sie fanden lediglich einen niedrigen Zusammenhang zwischen der mittels Multiple-Choice-Items erhobenen Skala noticing und der durch offene Fragen erfassten Skala knowledge-based reasoning. Schäfer und Seidel (2015) untersuchten zur Validierung eines offenen Verfahrens zur Erfassung der qualitativen Ebenen Beschreiben, Erklären und Vorhersagen des knowledge-based reasoning den Zusammenhang mit dem geschlossenen Instrument „Observer“ (Seidel und Stürmer 2014). Je höher die Punktzahl im Observer war, desto höher war auch die qualitative Ebene im knowledge-based reasoning; jedoch ebenfalls mit niedrigen Korrelationen von $0,07 < r < 0,22$. Frommelt et al. (2019) folgten diesem Ansatz bei der Erfassung der professionellen Unterrichtswahrnehmung von Zielklarheit, Lernbegleitung und Lernatmosphäre. Als offenes Format verwendeten sie verschriftlichte Kommentaren zu videografierten Unterrichtssequenzen. Zusätzlich wurde als geschlossenes Verfahren das Instrument „Observer“ genutzt. Es zeigten sich signifikante, niedrige bis moderate Korrelationen mit $0,27 < r < 0,35$ zwischen dem offen erfassten noticing/identification von relevanten Events und den geschlossen erfassten Facetten des knowledge-based reasoning. Zwischen dem Begründen in der offenen Erfassung (reasons) und dem knowledge-based reasoning ergaben sich jedoch auch hier nur niedrige, nicht signifikante Zusammenhänge. Im Gegensatz zu Gold et al. (2016) verwendeten Schäfer und Seidel (2015) sowie Frommelt et al. (2019) jeweils eine identische Videosequenz für die Zusammenhänge der offenen und geschlossenen Erfassung, was die jeweiligen

Testwerte messmethodisch vergleichbarer macht und die höheren Zusammenhänge erklären könnte. König et al. (2014) verwendeten zwar für jede Videoszene gleichermaßen ein offenes und geschlossenes Verfahren, diese korrelierten jedoch ebenfalls nur niedrig. Insgesamt könnten die niedrigen Zusammenhänge somit auf die Verwendung unterschiedlicher Videos, aber auch auf die unterschiedlichen Antwortformate rückführbar zu sein. Ob unter Kontrolle des Antwortformats *und* des Videostimulus' tatsächlich niedrige Korrelationen zwischen den beiden Facetten der professionellen Unterrichtswahrnehmung vorliegen und, ob über ihren Zusammenhang im Sinne der konvergenten Validität hinaus, beide Erfassungsarten weitere Validitätskriterien erfüllen, bleibt weiter offen. Dafür werden im folgenden Kapitel gängige Ansätze zur Validierung von Instrumenten zur Erfassung der professionellen Unterrichtswahrnehmung dargestellt.

2.2.3 Validierungsansätze zur Erfassung der professionellen Unterrichtswahrnehmung

Bestehende Instrumente zur Erfassung der professionellen Unterrichtswahrnehmung wurden hauptsächlich auf Inhalts-, Konstrukt- und Kriteriumsvalidität geprüft. Die Inhaltsvalidität wird meist betrachtet, indem Expert*innen die Repräsentativität, Authentizität oder Verständlichkeit des Instruments einschätzen (z. B. Gold und Holodynski 2017; Plöger und Scholl 2014; Seidel et al. 2010). Hinsichtlich der *Konstruktvalidität* wird neben der Untersuchung der angenommenen Dimensionalität (Gold und Holodynski 2017; König 2015; Meschede et al. 2015; Seidel und Stürmer 2014) bei einigen Instrumenten auch die diskriminante Validität betrachtet. Beispielsweise zeigen einige Studien, dass das professionelle Wissen und die professionelle Unterrichtswahrnehmung eng miteinander verknüpft sind, sich empirisch jedoch gut trennen lassen (z. B. Gold und Holodynski 2017; König et al. 2014; Steffensky et al. 2015). Zur Überprüfung der *konkurrenten Kriteriumsvalidität* werden meist Unterschiede zwischen Gruppen unterschiedlicher Expertise untersucht, wobei erfahrenere Studierendengruppen, Lehramtsanwärter*innen sowie erfahrene Lehrkräfte erwartungsgemäß höhere Testwerte aufweisen (z. B. Gold und Holodynski 2017; Meschede et al. 2015; Plöger und Scholl 2014; Zucker 2019).

3 Ziel, Fragestellungen und Hypothesen der Studie

Ziel der vorliegenden Studie war die Betrachtung der Konstrukt- und Kriteriumsvalidität einer offenen (schriftliche Analyse) und einer geschlossenen Erfassung (Ratingitems) der professionellen Unterrichtswahrnehmung von Klassenführung mit *demselben* Videostimulus, um eine mögliche Konfundierung von Videoszene und Aufgabenformat zu kontrollieren. Da Ratingitems die Aufmerksamkeit bereits auf konkrete Aspekte oder Situationen lenken, wurde der Teilprozess der selektiven Aufmerksamkeitsfokussierung in der vorliegenden Untersuchung nicht berücksichtigt.

Zusammenhänge zwischen den beiden Erfassungsarten und ihre jeweiligen Zusammenhänge mit dem pädagogisch-psychologischen Wissen über Klassenführung sollten Aufschlüsse über die *Konstruktvalidität* geben. Darüber hinaus wurde die

Kriteriumsvalidität über die Sensitivität für Expertiseunterschiede zwischen Bachelor- und Masterstudierenden betrachtet.

Fragestellung 1a – Hinweise auf konvergente Konstruktvalidität: Hängt die geschlossene Erfassung des wissensbasierten Verarbeitens mittels Ratingitems mit ihrer offenen Erfassung zusammen?

Bisherige Studien zeigen keine oder lediglich niedrige Korrelationen zwischen verschiedenen Erfassungsformaten, wenn jeweils unterschiedliche Videos als Stimuli genutzt wurden. Dementsprechend wurde bei Verwendung desselben Videostimulus zumindest eine moderate Korrelation zwischen dem geschlossenen und offenen Erfassungsformat angenommen.

Um weitere Hinweise für die konvergente Validität der beiden Erfassungsarten zu erhalten, wurde in Anlehnung an den Multitrait-Multimethod-Ansatz faktoranalytisch überprüft, ob ein Modell mit spezifizierten Methodenfaktoren zu einer besseren Modellpassung führt als ein Modell ohne Berücksichtigung des Erfassungsformats.

Fragestellung 1b – Hinweise auf diskriminante Konstruktvalidität: Gibt es Unterschiede in der Korrelation zwischen dem geschlossen und dem offen erfassten wissensbasierten Verarbeiten mit dem professionellen Wissen über Klassenführung?

Fragestellung 2 – Hinweise auf Kriteriumsvalidität: Gibt es Unterschiede in der Sensitivität für Expertiseunterschiede (zwischen Bachelor- und Masterstudierenden) zwischen dem geschlossen und dem offen erfassten wissensbasierten Verarbeiten?

Masterstudierende verfügen über höhere Ausprägungen im pädagogisch-psychologischen Wissen (König et al. 2008; Tachtsoglou und König 2017), welches als Voraussetzung für die professionelle Unterrichtswahrnehmung von Klassenführung gesehen wird. Masterstudierende sollten somit aufgrund von mehr Lerngelegenheiten (höheren Semesterzahl, Absolvieren des Praxissemesters) höhere Ausprägungen in beiden Erfassungsarten erreichen. Möglicherweise zeigen sich vor allem bei der offenen Erfassung größere Unterschiede zugunsten der Masterstudierenden, da die Ratewahrscheinlichkeit bei geschlossenen Formaten höher ist und sich somit Expertiseunterschiede eher beim freien Verschriftlichen der Videoanalyse zeigen könnten.

4 Methode

4.1 Stichprobe

Insgesamt lagen Daten von 309 Bachelor- und Masterstudierenden aus zwei deutschen Universitäten zur Reanalyse (aus den Studien von Gold et al. (2021b) und Gippert et al. (2022, under review)) vor, weshalb a priori keine identische Gruppengröße geplant und umgesetzt werden konnte (siehe Tab. 1). Die Masterstudierenden wurden nach Absolvieren ihres Praxissemesters befragt. Durch curricular verankerte vereinzelte Sitzungen im Studium zum Thema Klassenführung hatten diese Studierenden mehr Lerngelegenheiten als die Bachelorstudierenden, welche zu Beginn ihres Studiums befragt wurden. Erwartungsgemäß zeigte ein *t*-Test in Bezug auf das selbsteingeschätzte Wissen über Klassenführung (Skala von 1 bis 5), dass die Masterstudierenden ihr Wissen über die drei Klassenführungsfacetten ($M=3,74$, $SD=0,54$) signifikant höher einschätzten als die Bachelorstudierenden ($M=2,96$,

Tab. 1 Demographische Daten der Studierenden nach Gruppen

Gruppe	Anzahl	Geschlecht (Anteile in %)		Alter (in Jahren)	
		Männlich	Weiblich	<i>M</i>	<i>SD</i>
Bachelor	266	12,5	87,5	23,22	4,56
Master	43	4,7	95,3	24,47	1,59
Gesamt	309	11,3	88,7	23,42	4,26

Anmerkung. *M* Mittelwert, *SD* Standardabweichung

$SD = 0,67$) ($\Delta M = 0,78$, 95 % KI [0,57, 0,99]), $t(306) = 7,30$, $p < 0,001$, $d = 1,19$). Die Datenerhebung fand entweder in Seminaren an der Universität statt oder konnte auf freiwilliger Basis individuell zu Hause durchgeführt werden. Alle Studierenden erhielten den Zugang zum Instrument über eine Online-Plattform.

Eine Teilstichprobe ($N = 92$, 91,3 % weiblich, $M_{\text{Alter}} = 23$ Jahre, $SD_{\text{Alter}} = 2,34$) bearbeitete neben den videobasierten Instrumenten ebenfalls einen Test zur Erfassung des Klassenführungswissens. 51 Studierende davon befanden sich im Bachelorstudium und 43 im Masterstudiengang.

4.2 Instrumente

4.2.1 Pädagogisch-psychologisches Wissen über Klassenführung

Zur Erfassung des pädagogisch-psychologischen Professionswissens wurde ein Paper-Pencil-Test von Lense et al. (2015) eingesetzt, der deklarative und konditional-prozedurale Komponenten aller drei Klassenführungsfacetten Monitoring und prozeduraler Strukturierung sowie der Etabliertheit von Regeln und Routinen abdeckte. Für die vorliegende Studie wurden nur die Items zum Inhalt Klassenführung ausgewählt. Der Teil des deklarativen Wissens bestand aus komplexen Multiple-Choice-Items mit einem Itemstamm und jeweils vier bis sechs Antwortoptionen. Das konditional-prozedurale Wissen wurde mittels Textvignetten zu verschiedenen problematischen Unterrichtssituationen und verschiedenen angebotenen Handlungsoptionen erfasst. Diese sollten von den Teilnehmenden anhand von Schulnoten auf ihre Angemessenheit hin eingeschätzt werden. Die Ergebnisse des Wissenstests wurden getrennt nach Wissensart (deklarativ und konditional-prozedural) als prozentuale Anteile des Gesamtscores ausgegeben.

4.2.2 Wissensbasiertes Verarbeiten

Zur Erfassung des wissensbasierten Verarbeitens als Teilprozess der professionellen Unterrichtswahrnehmung von Klassenführung sollte eine Unterrichtsszene in Form einer schriftlichen Analyse (offene Erfassung) und mittels Ratingitems (geschlossene Erfassung) hinsichtlich der gezeigten Klassenführung eingeschätzt werden. Die verwendete Unterrichtsszene sowie die Ratingitems wurden aus dem validierten Instrument von Gold und Holodynski (2017) entnommen. Der ausgewählte, circa 2-minütige Videoclip stammt aus einem naturwissenschaftlichen Sachunterricht

zum Thema Schwimmen und Sinken und zeigte einen Phasenübergang von einer Gruppenarbeit zu einer Schülerpräsentation. Acht Expert*innen bestätigten im Sinne der Inhaltsvalidität, dass das gewählte Szenario für die Beobachtung von Klassenführung geeignet ist. Die Expert*innen setzten sich aus Wissenschaftler*innen mit einem Forschungsschwerpunkt auf Klassenführung sowie promovierten Lehrer*innenfortbildner*innen zusammen, welche zum Thema Klassenführung Fortbildungen anbieten.

Bei der Bearbeitung der Befragung erhielten die Studierenden, neben einer kurzen Erläuterung des Ablaufs, eine kurze Definition der drei Klassenführungsfacetten Monitoring, prozessuale Strukturierung und Etabliertheit von Regeln und Routinen sowie kurze Kontextinformationen zur Jahrgangsstufe und Stundenthematik.

Offene Erfassung Nach der ersten Beobachtung der Videoszene erfolgte die Aufforderung zu einer schriftlichen Analyse: „Welche klassenführungsspezifischen Maßnahmen ergreift die Lehrerin in diesem Ausschnitt und wie wirken diese Maßnahmen auf das Schüler*innenverhalten? Bitte begründen Sie.“

Die schriftlichen Analysen wurden durch zwei Rater*innen mit der Analysesoftware MAXQDA 18 in Bezug auf den Teilprozess des wissensbasierten Verarbeitens kodiert. Das Kodiermanual basierte auf einem Masterrating, das sieben Klassenführungsevents als zentral identifizierte, die erkannt werden sollten. Diese sieben Events wurden dann, falls erkannt, in zwei qualitativ unterschiedliche Analyseniveaus kodiert und bepunktet: *Beschreibung* und *Interpretation* (in Anlehnung an Blömeke et al. 2015; Gold et al. 2016; Sabers et al. 1991; siehe Tab. 2). Für ein beschriebenes, zum Masterrating passendes, Event wurde ein Punkt vergeben. Für jede zum Masterrating passende Interpretation des Events wurde ebenfalls ein Punkt vergeben, sodass bei beiden Analyseniveaus jeweils maximal sieben Punkte erlangt werden konnten.

Von 309 vorliegenden Datensätzen wurden nach drei Kodiertrainings 25 % der Dokumente von zwei Rater*innen bepunktet, wobei eine gemittelte substantielle Interraterreliabilität von Cohens Kappa $\kappa = 0,67$ erreicht wurde.

Geschlossene Erfassung Nach erneuter Präsentation der Unterrichtsszene wurden 29 Ratingitems randomisiert präsentiert. Diese sollten auf einer vierstufigen Likertskala (1 = *trifft zu*/4 = *trifft nicht zu*) in Bezug auf die in der Szene gezeigte Klassenführung eingeschätzt werden. Die Ratingitems wurden ebenfalls in die zwei Analyseniveaus *Beschreibung* und *Interpretation* aufgeteilt ($N_{\text{Bes}} = 13$, $N_{\text{Int}} = 16$, siehe Tab. 3). Wenn ein Item lediglich die Unterrichtssituation beschreibt, dabei nur auf das beobachtbare Verhalten der Lehrkraft abzielt und nur unbegründet und sekundär

Tab. 2 Beispielhafte Zuordnung der Analyseniveaus zum relevanten Klassenführungsevent „Schlagen des Gongs ist ineffektiv – Fehlende Etabliertheit von Routinen“

Analyseniveaus	Beispiel
Beschreibung	„Die Lehrerin lässt eine Glocke ertönen. Die Schüler arbeiten größtenteils weiter.“
(begründete)	„Die Schüler scheinen zwar zu wissen, was es bedeutet, aber nur ein Teil reagiert und folgt weiter dem Geschehen, der Rest ignoriert es und geht seinen eigenen Aktivitäten weiter nach. Das Klangsignal verfehlt seine Wirkung.“
Interpretation	

Tab. 3 Beispielitems geordnet nach Klassenführungsfacetten und Analyseniveau

	Beschreibung	Interpretation
Monitoring	Die Lehrerin hat nur die vortragenden Schüler*innen im Blick	Die Lehrerin leitet die vortragenden Schüler*innen und den Rest der Klasse gleichzeitig an
Prozessuale Strukturierung	Die Lehrerin benutzt keine effektiven Hilfsmittel, die den Ablauf des Unterrichts strukturieren	Die Lehrerin versäumt es, vor der neuen Unterrichtsaktivität die Aufmerksamkeit der Schüler*innen zu sammeln
Etabliertheit von Regeln und Routinen	Die Lehrerin etabliert in der gezeigten Szene Rituale im Unterricht	Da die Lehrerin nicht konsequent auf die Einhaltung ihrer aufgestellten Regel achtet, sind einige Schüler*innen nicht aufmerksam

auf das Schüler*innenverhalten eingeht, fällt es in die erste Niveaustufe *Beschreibung*. Die Proband*innen müssen somit einschätzen, ob ein bestimmtes Ereignis vorkommt oder nicht. Erfordert die Beantwortung des Ratingitems eine Integration der Lehrkräfteperspektive mit der Auswirkung auf die Schüler*innen und deren Verhalten oder deren Lernprozess, wird es weiter zur Niveaustufe *Interpretation* gezählt („multiple analytic points to form a coherent argument“; Kersting et al. 2010, S. 174).

Zur Auswertung wurden die Antworten der Studierenden mit einem Masterrating verglichen, welches auf Basis von 16 Expert*innen erstellt wurde (Gold und Holodynski 2017). Die Expert*innen setzten sich hier ebenfalls aus Wissenschaftler*innen und Lehrer*innenfortbildner*innen mit Expertise im Bereich Klassenführung zusammen.

Zur besseren Vergleichbarkeit mit der offenen Erfassung wurden die vier Antwortmöglichkeiten auf *Zustimmung* und *Ablehnung* des Ratingitems reduziert. Diese wurden dann in „1 = *übereinstimmend mit Expertenrating*“ und „0 = *nicht übereinstimmend mit Expertenrating*“ quantifiziert.

Nach Abschluss eines Beschreibungs-Items aufgrund schlechter Itemkennwerte lag die interne Konsistenz des geschlossenen Teils des Instruments für den Gesamttest (Cronbachs $\alpha = 0,81$) und auch für die Subskalen Beschreibung ($\alpha = 0,70$) und Interpretation ($\alpha = 0,68$) in einem guten bis akzeptablen Bereich.

4.3 Statistische Analyse

Um Hinweise auf die Konstruktvalidität der beiden Erfassungsarten des wissensbasierten Verarbeitens zu erhalten (Fragestellung 1a), wurde in Anlehnung an Multi-trait-Multimethod-Analysen (MMTM) eine faktorenanalytische und eine korrelative Herangehensweise kombiniert. Da die Codes der offenen Erfassung keinen konkreten Items zugeordnet werden konnten, waren keine MMTM-Analysen möglich. Dafür wurde der methodenbedingte Einfluss bei der Erfassung des wissensbasierten Verarbeitens anhand eines Bi-Faktor-Modells (Reise 2012) untersucht. Im ersten Schritt wurden zur Reduktion des zu schätzenden Modells die 28 Ratingitems zu neun Parcels mit jeweils drei Items (mit Ausnahme des letzten Parcels mit vier Items) nach dem domain-representative approach (Kishton und Widaman 1994) zusammengefasst, die jeweils ein Item pro Klassenführungsfacette enthielten. Vier

Parcels enthielten die Beschreibungsitems und fünf Parcels die Interpretationsitems. Danach wurde geprüft, ob ein Modell, welches die zwei Erfassungsarten in Form von Methodenfaktoren berücksichtigt, einen besseren Modellfit aufweist als ein Modell ohne Methodenfaktoren.

Somit wurde zunächst ein zweidimensionales Modell mit den Inhaltsfaktoren *Beschreibung* und *Interpretation* spezifiziert. In einem zweiten Modell wurden zwei Methodenfaktoren ergänzt, die die Erfassungsmethoden repräsentierten (*geschlossene* bzw. *offene* Erfassung). In diesem Modell lud jede Variable also auf einen Faktor der beiden Teilfacetten des wissensbasierten Verarbeitens und auf einen Faktor der entsprechenden Erfassungsart. Korrelationen zwischen den Methodenfaktoren und den Inhaltsfaktoren wurden nicht zugelassen (Geiser et al. 2010). Diese Berechnungen wurden mit der Statistiksoftware *Mplus 8* (Muthén und Muthén 2017) durchgeführt.

Zusätzlich wurden die Zusammenhänge zwischen der offenen und geschlossenen Erfassung (Fragestellung 1a) sowie dem pädagogisch-psychologischen Wissen (Fragestellung 1b) anhand von Pearson-Korrelationen berechnet sowie mittels der Prüfgröße z nach Pearson und Filon (1898) statistisch verglichen. Für die Mittelwertvergleiche zwischen den Studierendengruppen (Fragestellung 2) wurden bei einem festgelegten Signifikanzniveau von 0,05 mittels Mann-Whitney- U -Tests berechnet, da die Gruppen nicht normalverteilt (Shapiro-Wilk-Test: $p < 0,001$) waren. Alle Berechnungen wurden mit der Statistiksoftware SPSS (Version 25) durchgeführt.

5 Ergebnisse

5.1 Ergebnisse zur Konstruktvalidität: Faktoranalytische Modellprüfung und Zusammenhänge zwischen den Erfassungsarten des wissensbasierten Verarbeitens sowie mit dem pädagogisch-psychologischen Wissen (Fragestellung 1)

Die Fitstatistiken der beiden Strukturgleichungsmodelle (mit und ohne Methodenfaktoren) zeigten eine bessere Passung des Bi-Faktor-Modells, $\chi^2(198) = 320,41$, $p < 0,001$; RMSEA = 0,045 (C.I. 90 % = 0,036; 0,054), CFI = 0,93, TLI = 0,91, im Vergleich zu dem zweidimensionalen Modell ohne Methodenfaktoren, $\chi^2(222) = 608,23$, $p < 0,001$, RMSEA = 0,075 (C.I. 90 % = 0,068; 0,082), CFI = 0,77, TLI = 0,74. Trotzdem lagen bei dem Modell mit Methodenfaktoren die deskriptiven Gütemaße des CFI und TLI annähernd unter dem angestrebten Cut-Off-Werten von 0,95 (Hu und Bentler 1999). Die beiden Inhaltsfaktoren Beschreibung und Interpretation korrelierten moderat mit 0,40 ($p < 0,001$), während die beiden Erfassungsarten hoch mit 0,60 ($p < 0,001$) zusammenhingen. Insgesamt sprachen die Ergebnisse des Modellvergleichs für eine geringe konvergente Validität.

Diese wurde auch in den Ergebnissen der bivariaten Korrelationen deutlich: Es zeigten sich statistisch signifikante Zusammenhänge zwischen den zwei qualitativen Stufen Beschreibung (BES) und Interpretation (INT) (siehe Tab. 4). Diese fielen innerhalb eines Erfassungsformats jedoch höher aus als innerhalb der Facetten des

Tab. 4 Korrelationen zwischen offener und geschlossener Erfassung des wissensbasierten Verarbeitens sowie dem pädagogischen Wissen

Kategorie	Geschlossene Erfassung (N=309)		Offene Erfassung (N=309)		Pädagogisches Wissen (N=92)	
	Beschreibung	Interpretation	Beschreibung	Interpretation	DE	KO
GE	BES	0,67***	0,13*	0,36***	0,20	0,12
	INT		0,13*	0,32***	0,25*	0,16
OE	BES			0,45***	0,11	0,19
	INT				0,06	0,26*
PW	DE					0,05
	KO					

Anmerkung. *** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$, GE Geschlossene Erfassung, OE Offene Erfassung, BES Beschreibung, INT Interpretation, PW Pädagogisches Wissen, DE deklarativ, KO konditional-prozedural

Tab. 5 Statistischer Vergleich der Korrelationen mittels Prüfgröße z

Annahme	Prüfgröße z	Ergebnisse
$BES_{GE} - BES_{OE} > BES_{GE} - INT_{GE}$	8,752***	$BES_{GE} - BES_{OE} < BES_{GE} - INT_{GE}$
$INT_{GE} - INT_{OE} > BES_{GE} - INT_{GE}$	6,655***	$INT_{GE} - INT_{OE} < BES_{GE} - INT_{GE}$
$BES_{GE} - BES_{OE} > BES_{OE} - INT_{OE}$	5,416***	$BES_{GE} - BES_{OE} < BES_{OE} - INT_{OE}$
$INT_{GE} - INT_{OE} > BES_{OE} - INT_{OE}$	1,961*	$INT_{GE} - INT_{OE} < BES_{OE} - INT_{OE}$

Anmerkung. *** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$, GE Geschlossene Erfassung, OE Offene Erfassung, BES Beschreibung, INT Interpretation

wissensbasierten Verarbeitens zwischen den Erfassungsformaten. Dies verdeutlicht auch der statistische Vergleich der Korrelationskoeffizienten (Pearson und Filon 1898): Die Korrelationen zwischen den beiden Analyseniveaus innerhalb eines Erfassungsformates waren signifikant höher als die Korrelationen zwischen denselben Analyseniveaus aus unterschiedlichen Erfassungsformaten (siehe Tab. 5).

Hinsichtlich der Zusammenhänge der beiden Erfassungsarten mit dem pädagogisch-psychologischen Wissen korrelierten sowohl bei der geschlossenen als auch bei der offenen Erfassung des wissensbasierten Verarbeitens lediglich das Analyseniveau Interpretation signifikant mit dem Wissen. So zeigte sich sowohl ein schwacher, signifikanter Zusammenhang zwischen dem deklarativen Wissen und den Interpretationsitems als auch zwischen dem konditional-prozeduralen Wissen und der offen erfassten Interpretation (siehe Tab. 4).

5.2 Ergebnisse zur Kriteriumsvalidität: Expertiseunterschiede im wissensbasierten Verarbeiten (Fragestellung 2)

In Bezug auf die Sensitivität für Expertiseunterschiede beim wissensbasierten Verarbeiten von erkannten klassenführungsrelevanten Events zeigten sich erwartungskonforme Unterschiede zugunsten der Masterstudierenden (siehe Tab. 6). Nach einer Poweranalyse lag die Power der Mittelwertvergleiche bei drei der vier Tests bei mindestens 90%. Bei der geschlossenen Erfassung wurden diese mit einem mittleren Effekt auf der Ebene der Beschreibung signifikant, $U = 3387,00$, $p < 0,001$, sowie auch

Tab. 6 Mittlere Testleistung aufgeteilt nach Studierendenkohorten

	Studiengang	Offene Erfassung			Geschlossene Erfassung		
		<i>M</i>	<i>SD</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>d</i>
Beschreibung	Bachelor	3,18	1,35	0,30*	7,89	2,57	0,65***
	Master	3,58	1,31		9,53	2,33	
Interpretation	Bachelor	1,36	1,22	0,59**	11,79	2,75	0,56***
	Master	2,12	1,68		13,28	2,09	

Anmerkung. *M* Mittelwert, *SD* Standardabweichung, *d*=Effektstärke Cohens *d*, *** $p < .0001$, ** $p < 0,01$, * $p < 0,05$

mit kleiner Effektstärke auf der Ebene der Interpretation, $U = 3782,00$, $p < 0,001$. Bei der offenen Erfassung unterschieden sich die Master- und Bachelorstudierenden signifikant sowohl auf Ebene der Beschreibung, $U = 4665,50$, $p = 0,046$, als auch auf Ebene der Interpretation mit einem mittleren Effekt, $U = 4280,50$, $p = 0,006$.

6 Diskussion

In der vorliegenden Studie wurden Hinweise auf die Konstrukt- und Kriteriumsvalidität einer geschlossenen und einer offenen Erfassung des wissensbasierten Verarbeitens als Teil der professionellen Unterrichtswahrnehmung von Klassenführung untersucht. Diese wurden mit *demselben* Videostimulus offen mittels Kodierung einer schriftlichen Analyse und geschlossen mittels Ratingitems gemessen. Da bisherige Studien mit unterschiedlichen Videostimuli lediglich niedrige oder Nullkorrelationen zwischen den Erfassungsarten zeigten, sollte die vorliegende Studie Aufschluss darüber geben, ob die Erfassungsarten höher miteinander korrelieren, wenn sie sich auf denselben Videostimulus beziehen. Weiterhin wurden für Validitätshinweise Zusammenhänge mit Professionswissen und Unterschiede zwischen Bachelor- und Masterstudierenden betrachtet.

6.1 Fragestellung 1 – Konstruktvalidität

Bei den Zusammenhängen zwischen der geschlossenen und offenen Erfassung zeigten sich die höchsten Korrelationen zwischen den Analyseniveaus des wissensbasierten Verarbeitens *innerhalb* einer Erfassungsart. Ein Vergleich der Korrelationen zwischen den Analyseniveaus der beiden Erfassungsarten lieferte weitere Hinweise für eine methodenbedingte Abhängigkeit. Podsakoff et al. (2003) benennen jenen systematischen Einfluss durch die gemeinsame Methode *Common Method Bias*, der beispielsweise durch Multitrait-Multimethod-Analysen kontrolliert werden kann (Campbell und Fiske 1959). Dafür müssten jedoch zur Vergleichbarkeit die einzelnen Klassenführungsmerkmale innerhalb der geschlossenen und offenen Erfassung zwischen den Erfassungsarten parallelisiert werden. Eine Annäherung an MMTM-Ansätze war der Vergleich eines zweidimensionalen Modells mit einem Bi-Faktor-Modell, das über die Spezifikation von Methodenfaktoren die Erfassungsart berücksichtigt. Den oben genannten Korrelationen entsprechend, bildete das Bi-Faktor-

Modell mit Methodenfaktoren die Daten besser ab als ein zweidimensionales Modell.

Die hohen Zusammenhänge der beiden Teilprozesse Beschreiben und (begründetes) Interpretieren innerhalb einer Erfassungsart bestärken aber auch Ergebnisse aus Validierungsstudien von geschlossenen Instrumenten, die verschiedene Facetten des wissensbasierten Verarbeitens zu trennen versuchten (Gold und Holodynski 2017; Meschede et al. 2015; Seidel und Stürmer 2014). Hierbei korrelierten die angenommenen Dimensionen der professionellen Unterrichtswahrnehmung ebenfalls sehr hoch miteinander.

Beim Analyseniveau Beschreibung zeigten sich zwischen den beiden Erfassungsformaten niedrigere Zusammenhänge als bei dem höheren Analyseniveau Interpretation. Dies steht entgegen der Ergebnisse von Frommelt et al. (2019), die zwischen höheren Niveaus nur sehr geringe Korrelationen finden konnten. Allerdings fielen die Zusammenhänge der geschlossen erfassten Variablen mit den offenen Beschreibungen insgesamt niedrig aus. Eine Ursache könnte in der teils inkohärenten Passung zwischen Aufgabenstellung und Kodierung liegen: Die Aufgabenstellung für die offene Erfassung forderte explizit eine Begründung für die wahrgenommenen klassenführungsrelevanten Ereignisse. Kodiert wurde jedoch sowohl die Beschreibung als auch die begründete Interpretation.

Hinsichtlich der diskriminanten Validität zeigten sich signifikante Zusammenhänge mit dem pädagogischen-psychologischen Wissen über Klassenführung, jedoch ausschließlich auf den Analyseniveaus der Interpretation. Dabei korrelierte das anwendungsbezogene konditional-prozedurale Wissen lediglich mit der offen erfassten Interpretation und das deklarative Wissen nur mit der mittels Ratingitems erfassten Interpretation. Möglicherweise ist für die Einschätzung einer vorgegebenen Situationsinterpretation im Rahmen von Ratingitems deklaratives Wissen ausreichend, während für das eigenständige Generieren einer Interpretation (und somit das Einbeziehen der Komplexität und Kontextabhängigkeit von Unterricht) möglicherweise gerade anwendungsbezogenes Wissen notwendig ist. Dies ebenfalls bestärkend zeigte sich das anwendungsbezogene Wissen im Vergleich zum deklarativen Wissen bedeutsamer für das Lehrkräftehandeln (Lenske et al. 2016).

Zusammenfassend korrelierten die geschlossene und die offene Erfassungsart zur Messung der professionellen Unterrichtswahrnehmung nicht hoch genug, um auf die Messung desselben Konstrukts schließen zu können. Die Betrachtung der Validierung über das konstruktnahe professionelle Wissen zu Klassenführung zeigte kein eindeutiges Muster, das für die Überlegenheit einer Erfassungsart sprach. Diese Ergebnisse spiegeln somit ebenfalls bestehende Ergebnisse wider: Bei geschlossenen Instrumenten haben sich bereits in mehreren Studien erwartungsgemäße Zusammenhänge mit dem Professionswissen gezeigt (Gold und Holodynski 2017; König et al. 2014; Meschede et al. 2017), die jedoch zwischen niedrigen (z. B. $r=0,13$ bei König et al. 2014) und hohen Korrelationen (z. B. $r=0,56$ bei Meschede et al. 2017) streuen. Ein ähnlich großer Range zwischen niedrigen (z. B. $r=0,12$ bei Bruns et al. 2020) und hohen Korrelationen (z. B. $r=0,62$ bei Kersting et al. 2010) zeigt sich bei Instrumenten offenen Formaten (Blömeke et al. 2016; Kersting et al. 2012; König et al. 2014; Krauss et al. 2020) (oder offenen, videobasierten Wissenstests, die in der konkreten Operationalisierung jedoch ähnlich sind). Erklärungen für die variiere-

renden Zusammenhänge könnten sich u. a. in unterschiedlich praxiserfahrenen Proband*innen begründen. Bei Bruns et al. (2020) wurden Studierende und bei König et al. (2014) am Berufsbeginn stehende Lehrkräfte (ca. 4 Jahre nach Berufseinstieg) im Rahmen einer Follow-up-Untersuchung befragt. Bei Meschede et al. (2017) wurden sowohl Masterstudierende als auch erfahrene Lehrkräfte (im Mittel ca. 10 Jahre Berufserfahrung) erfasst, wobei der Zusammenhang zwischen der professionellen Wahrnehmung und dem professionellen Wissen bei den Lehrkräften wie bei einer ähnlich erfahrenen Stichprobe in der Studie von Kersting et al. (2012) vergleichsweise hoch war. Dies lässt vermuten, dass praktische Lerngelegenheiten bedeutsam für die Höhe der Zusammenhänge sind und unterstützt Befunde aus der Expertiseforschung. Durch die spezifische Wissensstruktur in kognitive Schemata, wird davon ausgegangen, dass durch eine stärkere Vernetzung und Integration von Wissen mit Erfahrungen komplexe Unterrichtssituationen von Expert*innenlehrkräften gezielter wahrgenommen werden (Berliner 2001, de Jong und Ferguson-Hessler 1996). Deshalb bedarf es einer weiteren Prüfung mit verschiedenen Stichproben oder beispielsweise Längsschnittstudien, um die Vermutungen sowohl mit deklarativen als auch prozeduralen Wissensbausteinen zu prüfen und zu generalisieren.

6.2 Fragestellung 2 – Kriteriumsvalidität

Um Aspekte der Kriteriumsvalidität zu betrachten, wurde die Sensitivität der beiden Erfassungsarten für Expertiseunterschiede zwischen Bachelor- und Masterstudierenden der vorliegenden Studierendenkohorten untersucht. Erwartungskonforme Unterschiede im wissensbasierten Verarbeiten zeigten sich in beiden Analyseniveaus und in beiden Erfassungsarten mit kleinen bis mittleren Effektstärken. Trotz der etwas geringeren Sensitivität der offen erfassten Beschreibungen konnten im Sinne der Kriteriumsvalidität beide Erfassungsarten erwartungskonforme Gruppenunterschiede zwischen den Studierenden mit niedriger und höherer Semesteranzahl erfassen (Barth 2017; Meschede et al. 2015; Zucker 2019).

6.3 Limitationen und Ausblick

Eine Einschränkung der Untersuchung ist die Verwendung *eines* Videoclips, um den Teststimulus identisch zu halten. Dementsprechend kann eine Video- oder Situationsabhängigkeit, auch bezüglich der getroffenen Fach- und Schulauswahl, nicht ausgeschlossen werden und die Generalisierbarkeit auf weitere Videoclips aus unterschiedlichen Schularten, Fächer oder Situationen müsste geprüft werden. Einschränkung ist außerdem die ungleiche Stichprobengröße ($n_{\text{Bachelor}} = 266$, $n_{\text{Master}} = 43$) zu nennen, welche der Reanalyse bereits vorliegender Daten im Sinne eines „intensive use of scientific data“ geschuldet ist (Machado 2015, S. 209). Für einen ersten Schritt zur Betrachtung möglicher Hinweise auf Konstrukt- und Kriteriumsvalidität stellt dieses Vorgehen eine ressourcenschonende Möglichkeit dar. Um trotzdem die Stabilität der Ergebnisse zu prüfen, wäre, neben einem für die Kriteriumsvalidität bedeutsamen Vergleich zu Lehramtsanwärter*innen oder bereits praktizierenden Lehrkräften, ebenfalls eine Replikation mit weiteren verhältnismäßigen Stichpro-

bengrößen anzustreben. In diesem Zusammenhang könnten ebenfalls Standortunterschiede ausgeschlossen werden.

Kritisch zu beleuchten ist an den reanalysierten Daten außerdem die gewählte, nicht randomisierte Reihenfolge des Testverfahrens, in welchem dem ersten Präsentieren des Videos die offene Erfassung folgte und nach erneutem Anschauen des Videos die Ratingitems. Demzufolge könnten auch Erinnerungs-, Wiederholungs- oder Ermüdungseffekte durch zweimalige Betrachtung der Unterrichtsszene die Bearbeitung des geschlossenen Teils beeinflusst haben. Hellermann et al. (2015) bemerkten in diesem Zusammenhang, dass durch den Videostimulus und eine durch die Items hervorgerufene angenommene Aufmerksamkeitslenkung auf klassenführungsrelevante Aspekte bereits eine minimale Intervention im Sinne einer Förderung der professionellen Unterrichtswahrnehmung stattfinden könne. Um jenen Priming-Effekt durch die Ratingitems auszuschließen, wurde auf eine Randomisierung bewusst verzichtet. Um jedoch auch Reihenfolgeeffekte von der offenen Erfassung auf die geschlossene Erfassung auszuschließen und die interne Validität zu erhöhen, wäre hingegen ein experimentelles Design in Zukunft bedeutsam, welches die Studierenden randomisiert auf die zwei Reihenfolgebedingungen verteilt.

Die professionelle Unterrichtswahrnehmung von Klassenführung lässt sich wie eingangs vorgestellt unterschiedlich konzeptualisieren. Unsere Studie fokussierte lediglich eine ausgewählte Auffassung von Klassenführung und deckt somit unter Umständen nicht alle relevanten Aspekte ab, wie beispielsweise das Management von Lehrkräfte-Schüler*innen-Beziehung (Piwovar 2013). Außerdem wird im PID-Modell von Blömeke et al. (2015) und auch in breiter gefassten Operationalisierungen der professionellen Unterrichtswahrnehmung (Seidel et al. 2010) zunehmend auch der Prozess des Antizipierens von Handlungsverläufen oder des Alternativenformulierens betrachtet. Da Expert*innen insbesondere elaboriertere Fähigkeiten auf den kognitiv höheren Anforderungen des wissensbasierten Verarbeitens als Noviz*innen zeigen (z. B. Berliner 2001; Seidel und Prenzel 2007) sowie durch gefundene signifikante Zusammenhänge zwischen dem Formulieren von Alternativen und Schüler*innenleistung (Kersting et al. 2012), sollte in zukünftiger Forschung eine Ergänzung geschlossener Instrumente durch offene Erfassungsformate hervor gehoben werden (Weber et al. 2020). So kann eine differenzierte Erfassung der professionellen Unterrichtswahrnehmung mit den drei Facetten (PID-Modell, Blömeke et al. 2015) gewährleistet werden. Wegen fehlender Referenz in der geschlossenen Erfassung wurde in der vorliegenden Studie auf die Auswertung der in den Originaldaten von Gold et al. (2021b) enthaltenen offen erfassten Handlungsalternativen verzichtet. Eine geschlossene Variante, in welcher ähnlich wie bei einem *Situational Judgement Test* (Gold und Holodyski 2015) explizite Handlungsalternativen eingeschätzt werden könnten, wäre prinzipiell vorstellbar. Jedoch würde ein solches Vorgehen wiederum kein eigenständiges Generieren von Handlungsvorschlägen erfordern und somit ein Rückschluss auf vorhandenes, handlungsleitendes Wissen eingeschränkt werden. Wegen einer vergleichsweise schwierigeren Erfassung der professionellen Unterrichtswahrnehmung durch den Umweg über das Verbalisieren der erkannten und interpretierten relevanten Unterrichtsereignisse, plädieren Hecker et al. (2020) für die Betrachtung der Handlungsalternativen. Darin würde sich indirekt durch eine formulierte Reaktion zeigen, welche Unterrichtsereignisse fokussiert

wurden und nicht nur, wie gut die Fähigkeit zum Verbalisieren der präzisen Wahrnehmung ist (Hecker et al. 2020). Als eine weitere Möglichkeit der methodischen Überwindung, um an die unbewussten Intentionen und Handlungspläne zu gelangen, benennt Nückles (2020) beispielsweise den Einsatz von Eye-tracking.

Trotz der aufgeführten Limitationen liefert die Studie Hinweise auf eine methodenabhängige Erfassung professioneller Unterrichtswahrnehmung von Klassenführung. Allerdings geht durch die Studie nicht hervor, welches Erfassungsformat auch prädiktiv valider wäre. Frommelt et al. (2019) resümieren kritisch, dass erst durch die benötigte Flexibilität beim adaptiven Abrufen von notwendigem Wissen zu einer (neuen) Unterrichtssituation, wie sie bei einer offenen Kommentierung zu einem Videostimulus benötigt wird, Rückschlüsse auf das tatsächliche Handeln zulässig sind. Um diesem Anliegen gerecht zu werden, sollte in zukünftigen Studien mittels umfassender Mediationsmodelle versucht werden, neben professionellen Wissenskomponenten auch tatsächlich performante Maße durch beobachtete Unterrichtsqualität oder Outcomevariablen wie Schüler*innenleistung einzubeziehen. Eine offene Erfassung der professionellen Unterrichtswahrnehmung sollte nach der Argumentation von Frommelt et al. (2019) höhere Zusammenhänge zeigen als mittels Ratingitems. Dies kann ebenfalls weitere Aufklärung über die Bedeutsamkeit der professionellen Unterrichtswahrnehmung von Klassenführung für Unterrichtsqualität liefern.

Abschließend wäre es erstrebenswert, offene und geschlossene Erfassungen *bewusst* kombiniert einzusetzen, um die Vorteile beider Erfassungsformate zu verknüpfen und jeweilige Stärken zu kombinieren. So könnten die Lernvoraussetzungen von Studierenden differenzierter analysiert und darauf aufbauend die Gestaltung von adaptiven Lehr-Lern-Gelegenheiten gezielter gefördert werden.

Förderung Das Manuskript ist nicht im Rahmen einer Drittmittelförderung entstanden.

Funding Open Access funding enabled and organized by Projekt DEAL.

Interessenkonflikt M.M. Müller und B. Gold geben an, dass kein Interessenkonflikt besteht.

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

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

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

Literatur

Anderson, L. M., Evertson, C. M., & Emmer, E. T. (1980). Dimensions in classroom management derived from recent research. *Journal of Curriculum Studies*, 12, 343–356.

- Barth, V.L. (2017). *Professionelle Wahrnehmung von Störungen im Unterricht*. Wiesbaden: Springer VS.
- Berliner, D.C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, 35(5), 463–482.
- Blömeke, S. (2013). Moving to a higher state of confusion. Der Beitrag der Videoforschung zur Kompetenzforschung. In U. Riegel & K. Macha (Hrsg.), *Videobasierte Kompetenzforschung in den Fachdidaktiken* (S. 25–44). Münster: Waxmann.
- Blömeke, S., Gustafsson, J.E., & Shavelson, R. (2015). Beyond dichotomies: Viewing competence as a continuum. *Zeitschrift für Psychologie*, 223, 3–13.
- Blömeke, S., Busse, A., Kaiser, G., König, J., & Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. *Teaching and Teacher Education*, 56, 35–46.
- Brovelli, D., Bölsterli, K., Rehm, M., & Wilhelm, M. (2013). Erfassen professioneller Kompetenzen für den naturwissenschaftlichen Unterricht: Ein Vignettest mit authentisch komplexen Unterrichtssituationen und offenem Antwortformat. *Unterrichtswissenschaft*, 41, 306–329.
- Bruns, J., Strahl, C., & Gasteiger, H. (2020). Situative Beobachtung und Wahrnehmung angehender frühpädagogischer Fachpersonen im Bereich Mathematik – Entwicklung und Validierung eines Testinstruments. *Unterrichtswissenschaft*, 49(3), 345–371.
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Doyle, W. (1986). Classroom organization and management. In M.C. Wittrock (Hrsg.), *Handbook of research on teaching* (S. 392–431). New York: Macmillan.
- Doyle, W. (2006). Ecological management and classroom management. In C.M. Evertson & C.S. Weinstein (Hrsg.), *Handbook of classroom management. Research, practice, and contemporary issues* (S. 97–126). Mahwah: Erlbaum.
- Friesen, M., Kuntze, S., & Vogel, M. (2018). Videos, Texte oder Comics? Die Rolle des Vignettenformats bei der Erhebung fachdidaktischer Analysekompetenz zum Umgang mit Darstellungen im Mathematikunterricht. In J. Rutsch, M. Rehm, M. Vogel, M. Seidenfuß & T. Dörfler (Hrsg.), *Effektive Kompetenzdiagnose in der Lehrerbildung* (S. 153–177). Wiesbaden: Springer.
- Frommelt, M., Hugener, I., & Krammer, K. (2019). Fostering teaching-related analytical skills through case-based learning with classroom videos in initial teacher education. *Journal for Educational Research Online*, 11, 37–60.
- Gehrer, K. (2017). *Der Einfluss von Aufgaben- und Textmerkmalen auf Itemschwierigkeit. Eine experimentelle Studie zur Einschränkung der wiederholten Textsicht bei der Bearbeitung von Lesekompetenztestaufgaben* (NEPS working paper, Bd. 67). Bamberg: Leibniz-Institut für Bildungsverläufe.
- Geiser, C., Eid, M., Nussbeck, F.W., Lischetzke, T., & Cole, D.A. (2010). Multitrait-Multimethod-Analyse. Multitrait-Multimethod Analysis. In H. Hollig & B. Schmitz (Hrsg.), *Handbuch Statistik, Methoden und Evaluation* (S. 679–685). Göttingen: Hogrefe.
- Gippert, C., Gold, B., & Holodynski, M. (2022). *Does pre-service teachers' professional knowledge moderate the relationship between their professional vision and general cognitive abilities?* (under review, manuscript submitted)
- Gold, B., & Holodynski, M. (2015). Development and construct validation of a situational judgment test of strategic knowledge of classroom management in elementary schools. *Educational Assessment*, 20(3), 226–248.
- Gold, B., & Holodynski, M. (2017). Using digital video to measure the professional vision of elementary classroom management: Test validation and methodological challenges. *Computers & Education*, 107, 13–30.
- Gold, B., Hellermann, C., & Holodynski, M. (2016). Professionelle Wahrnehmung von Klassenführung – Vergleich von zwei videobasierten Erfassungsmethoden. In K. Schwippert & D. Prinz (Hrsg.), *Der Forschung-Der Lehre-Der Bildung: Aktuelle Entwicklungen der Empirischen Bildungsforschung* (S. 103–118). Münster: Waxmann.
- Gold, B., Junker, R., Wissemann, M., Klassen, C., & Holodynski, M. (2021a). Are good observers good classroom managers? The relationship between teachers' professional vision and their students' ratings on classroom management. *International Journal of Educational Research*, 109, 101811.
- Gold, B., Pfirrmann, C., & Holodynski, M. (2021b). Promoting professional vision of classroom management through different analytic perspectives in video-based learning environments. *Journal of Teacher Education*, 72(4), 431–447.
- Hartig, J., & Jude, N. (2007). Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik* (S. 17–36). Berlin: BMBF.

- Hecker, S.-L., Falkenstern, S., Lemmrich, S., & Ehmke, T. (2020). Zum Verbalisierungsdilemma bei der Erfassung der situationspezifischen Fähigkeiten von Lehrkräften. Am Beispiel der Domäne Deutsch als Zweitsprache. *Zeitschrift für Bildungsforschung*, *10*, 175–190.
- Hellermann, C., Gold, B., & Holodynski, M. (2015). Förderung von Klassenführungsfähigkeiten im Lehramtsstudium. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *47*(2), 97–109.
- Herbst, P., & Kosko, K. W. (2014). Using representations of practice to elicit mathematics teachers' tacit knowledge of practice: a comparison of responses to animations and videos. *Journal of Mathematics Teacher Education*, *17*, 515–537.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.
- de Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist*, *31*(2), 105–113.
- Kaiser, G., Busse, A., Hoth, J., König, J., & Blömeke, S. (2015). About the complexities of video-based assessments: Theoretical and methodological approaches to overcoming shortcomings of research on teachers' competence. *International Journal of Science and Mathematics Education*, *13*(2), 369–387.
- 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*, 172–181.
- Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, *49*, 568–589.
- Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement*, *54*, 757–765.
- König, J. (2015). Kontextualisierte Erfassung von Lehrerkompetenzen. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, *61*, 305–309.
- König, J., & Kramer, C. (2016). Teacher professional knowledge and classroom management: on the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME). *ZDM*, *48*, 1–13.
- König, J., & Lebens, M. (2012). Classroom Management Expertise (CME) von Lehrkräften messen: Überlegungen zur Testung mithilfe von Videovignetten und erste empirische Befunde. *Lehrerbildung auf dem Prüfstand*, *5*, 3–28.
- König, J., Peek, R., & Blömeke, S. (2008). Zum Erwerb von pädagogischem Wissen in der universitären Ausbildung: Unterscheiden sich Studierende verschiedener Lehrämter und Kohorten? *Lehrerbildung auf dem Prüfstand*, *1*, 664–682.
- König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education*, *38*, 76–88.
- Kounin, J. S. (2006). *Techniken der Klassenführung*. Münster: Waxmann.
- Krauss, S., Bruckmaier, G., Lindl, A., Hilbert, S., Binder, K., & Steib, N. (2020). Competence as a continuum in the COACTIV study: the "cascade model". *ZDM Mathematics Education*, *52*, 311–327.
- Lenske, G., Thillmann, H., Wirth, J., Dicke, T., & Leutner, D. (2015). Pädagogisch-psychologisches Professionswissen von Lehrkräften: Evaluation des ProwiN-Tests. *Zeitschrift für Erziehungswissenschaft*, *18*(2), 225–245.
- Lenske, G., Wagner, W., Wirth, J., Thillmann, H., Cauet, E., & Leutner, D. (2016). Die Bedeutung des pädagogisch-psychologischen Wissens für die Qualität der Klassenführung und den Lernzuwachs der Schüler/innen im Physikunterricht. *Zeitschrift für Erziehungswissenschaft*, *19*(1), 211–233.
- Lindmeier, A. (2013). Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen. In U. Riegel & K. Macha (Hrsg.), *Videobasierte Kompetenzforschung in den Fachdidaktiken* (S. 45–62). Münster: Waxmann.
- Machado, J. (2015). Open data and open science. In S. Albagli, M. L. Maciel & A. H. Abdo (Hrsg.), *Open science, open issues* (S. 189–214). Brasília, Rio de Janeiro: IBICT, Unirio.
- Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*, *34*, 207–218.
- Meschede, N., Steffensky, M., Wolters, M., & Möller, K. (2015). Professionelle Wahrnehmung der Lernunterstützung im naturwissenschaftlichen Grundschulunterricht. Theoretische Beschreibung und empirische Erfassung. *Unterrichtswissenschaft*, *43*, 317–335.

- Meschede, N., Fiebranz, N., Möller, K., & Steffensky, M. (2017). Teachers' professional vision, pedagogical content knowledge and beliefs: On its relation and differences between pre-service and in-service teachers. *Teaching and Teacher Education*, *66*, 158–170.
- Mischo, C., Wolstein, K., & Peters, S. (2020). Professionelle Wahrnehmung und Fachkraft-Kind-Interaktion von KiTa-Fachkräften. *Zeitschrift für Pädagogische Psychologie*. <https://doi.org/10.1024/1010-0652/a000285>.
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8. Aufl.). Los Angeles: Muthén & Muthén.
- Neuweg, G. H. (2015). Kontextualisierte Kompetenzmessung. Eine Bilanz zu aktuellen Konzeptionen und forschungsmethodischen Zugängen. *Zeitschrift für Pädagogik*, *61*, 377–383.
- Nückles, M. (2020). Investigating visual perception in teaching and learning with advanced eye-tracking methodologies: Rewards and challenges of an innovative research paradigm. *Educational Psychology Review*, *33*, 149–167.
- Ophardt, D., & Thiel, F. (2008). Klassenmanagement als Basisdimension der Unterrichtsqualität. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (S. 259–284). Wiesbaden: Verlag für Sozialwissenschaften.
- Ophardt, D., & Thiel, F. (2013). *Klassenmanagement: Ein Handbuch für Studium und Praxis*. Stuttgart: Kohlhammer.
- Pearson, K., & Filon, L. N. G. (1898). Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *191*, 229–311.
- Piwovar, V. (2013). Multidimensionale Erfassung von Kompetenzen im Klassenmanagement: Konstruktion und Validierung eines Beobachter- und eines Schülerfragebogens für die Sekundarstufe 1. *Zeitschrift für Pädagogische Psychologie*, *27*(4), 215–228.
- Plöger, W., & Scholl, D. (2014). Analysekompetenz von Lehrpersonenc – Modellierung und Messung. *Zeitschrift für Erziehungswissenschaft*, *17*(1), 85–112.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879–903.
- Praetorius, A. K., Vieluf, S., Saß, S., Bernholt, A., & Klieme, E. (2016). The same in German as in English? Investigating the subject-specificity of teaching quality. *Zeitschrift für Erziehungswissenschaft*, *19*(2), 191–209.
- Rehm, M., & Bölsterli, K. (2014). Entwicklung von Unterrichtsvignetten. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 213–225). Berlin, Heidelberg: Springer Spektrum.
- Reise, S. P. (2012). Invited paper: The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696.
- Roth, K. J., Garnier, H., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. Z. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching*, *48*, 117–148.
- Sabers, D. S., Cushing, K. S., & Berliner, D. C. (1991). Differences among teachers in a task characterized by simultaneity, multidimensional, and immediacy. *American Educational Research Journal*, *28*, 63–88.
- Santagata, R., Zannoni, C., & Stigler, J. W. (2007). The role of lesson analysis in pre-service teacher education: An empirical investigation of teacher learning from a virtual video-based field experience. *Journal of Mathematics Teacher Education*, *10*, 123–140.
- Schäfer, S., & Seidel, T. (2015). Noticing and reasoning of teaching and learning components by pre-service teachers. *Journals for Educational Research Online*, *7*, 34–58.
- Schaper, N. (2009). Aufgabenfelder und Perspektiven bei der Kompetenzmodellierung und -messung in der Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, *2*, 166–199.
- Schnell, R., Hill, P., & Esser, E. (2018). *Methoden der empirischen Sozialforschung*. De Gruyter: Oldenbourg.
- Seidel, T. (2009). Klassenführung. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 135–148). Berlin: Springer.
- Seidel, T., & Prenzel, M. (2007). Wie Lehrpersonen Unterricht wahrnehmen und einschätzen – Erfassung pädagogisch-psychologischer Kompetenzen mit Videosequenzen. *Zeitschrift für Erziehungswissenschaft*, *8*(1), 201–216.
- Seidel, T., & Stürmer, K. (2014). Modeling the structure of professional vision in preservice teachers. *American Educational Research Journal*, *51*, 739–771.

- 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* (S. 296–306). Weinheim: Beltz.
- Seifert, A., Hilligus, A., & Schaper, N. (2009). Entwicklung und psychometrische Überprüfung eines Messinstruments zur Erfassung pädagogischer Kompetenzen in der universitären Lehrerbildung. *Lehrerbildung auf dem Prüfstand*, 2, 82–103.
- Sherin, M. G., & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education*, 60, 20–37.
- Shulman, L. S. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review*, 57, 1–22.
- Steffensky, M., Gold, B., Holodynski, M., & Möller, K. (2015). Professional vision of classroom management and learning support in science classrooms — Does professional vision differ across general and content-specific classroom interactions? *International Journal of Science and Mathematics Education*, 13(2), 351–368.
- Tachtsoglou, S., & König, J. (2017). Der Einfluss universitärer Lerngelegenheiten auf das pädagogische Wissen von Lehramtsstudierenden. *Zeitschrift für Bildungsforschung*, 7, 291–310.
- Weber, K. E., Prilop, C. N., Viehoff, S., Gold, B., & Kleinknecht, M. (2020). Fördert eine videobasierte Intervention im Praktikum die professionelle Wahrnehmung von Klassenführung? Eine quantitativ-inhaltsanalytische Messung von Subprozessen professioneller Wahrnehmung. *Zeitschrift für Erziehungswissenschaft*, 23(2), 343–365.
- Zucker, V. (2019). *Erkennen und Beschreiben von formativem Assessment im naturwissenschaftlichen Grundschulunterricht. Einwicklung eines Instruments zur Erfassung von Teilfähigkeiten der professionellen Wahrnehmung von Lehramtsstudierenden*. Berlin: Logos.