

Blömeke, Sigrid; König, Johannes; Suhl, Ute; Hoth, Jessica; Döhrmann, Martina
**Wie situationsbezogen ist die Kompetenz von Lehrkräften? Zur
Generalisierbarkeit der Ergebnisse von videobasierten Performanztests**
Zeitschrift für Pädagogik 61 (2015) 3, S. 310-327



Quellenangabe/ Reference:

Blömeke, Sigrid; König, Johannes; Suhl, Ute; Hoth, Jessica; Döhrmann, Martina: Wie situationsbezogen ist die Kompetenz von Lehrkräften? Zur Generalisierbarkeit der Ergebnisse von videobasierten Performanztests - In: Zeitschrift für Pädagogik 61 (2015) 3, S. 310-327 - URN: urn:nbn:de:0111-pedocs-153509 - DOI: 10.25656/01:15350

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-153509>

<https://doi.org/10.25656/01:15350>

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

ZEITSCHRIFT FÜR PÄDAGOGIK

Heft 3

Mai/Juni 2015

■ *Thementeil*

Kontextualisierte Erfassung von Lehrer- kompetenzen

■ *Essay*

Bildungsforschung im Kontext gesellschaftlicher Praxis.
Über (soziale) Bedingungen der Möglichkeit, Bildungs-
praxis durch Bildungsforschung zu beeinflussen

■ *Allgemeiner Teil*

Large Scale Assessments und die Bildung Erwachsener:
Erträge, Grenzen und Potenziale der Forschung

Heimann, Schulz oder Klafki? Eine quantitative
Studie zur Einschätzung der Praktikabilität allgemein-
didaktischer Planungsmodelle

Inhaltsverzeichnis

Thementeil: Kontextualisierte Erfassung von Lehrerkompetenzen

Johannes König

Kontextualisierte Erfassung von Lehrerkompetenzen.

Einführung in den Thementeil 305

Sigrid Blömeke/Johannes König/Ute Suhl/Jessica Hoth/

Martina Döhrmann

Wie situationsbezogen ist die Kompetenz von Lehrkräften?

Zur Generalisierbarkeit der Ergebnisse von videobasierten Performanztests 310

Andreas Busse/Gabriele Kaiser

Wissen und Fähigkeiten in Fachdidaktik und Pädagogik:

Zur Natur der professionellen Kompetenz von Lehrkräften 328

Kathleen Stürmer/Tina Seidel/Olga Kunina-Habenicht

Unterricht wissensbasiert beobachten – Unterschiede und

erklärende Faktoren bei Referendaren zum Berufseinstieg 345

Sarah Forster-Heinzer/Fritz Oser

Wer setzt das Maß? Eine kritische Auseinandersetzung

mit dem Advokatorischen Ansatz 361

Georg Hans Neuweg

Kontextualisierte Kompetenzmessung – Eine Bilanz

zu aktuellen Konzeptionen und forschungsmethodischen Zugängen 377

Deutscher Bildungsserver

Linktipps zum Thema „Kontextualisierte Erfassung

von Lehrerkompetenzen“ 384

Essay

Helmut Heid

Bildungsforschung im Kontext gesellschaftlicher Praxis.

Über (soziale) Bedingungen der Möglichkeit, Bildungspraxis
durch Bildungsforschung zu beeinflussen 390

Allgemeiner Teil

Josef Schrader

Large Scale Assessments und die Bildung Erwachsener:

Erträge, Grenzen und Potenziale der Forschung 410

Stephan Wernke/Jochen Werner/Klaus Zierer

Heimann, Schulz oder Klafki? Eine quantitative Studie zur Einschätzung

der Praktikabilität allgemeindidaktischer Planungsmodelle 429

Besprechungen

Heinz-Elmar Tenorth

Johannes Drerup: Paternalismus, Perfektionismus

und die Grenzen der Freiheit 452

Verónica Oelsner

Phillip Dylan Thomas Knobloch: Pädagogik in Argentinien.

Eine Untersuchung im Kontext Lateinamerikas mit Methoden

der Vergleichenden Erziehungswissenschaft 455

Dokumentation

Pädagogische Neuerscheinungen 459

Impressum U3

Table of Contents

Topic: Contextualized Assessment of Teacher Competences

Johannes König

Contextualized Assessment of Teacher Competences. An introduction 305

Sigrid Blömeke/Johannes König/Ute Suhl/Jessica Hoth/Martina Döhrmann

To What Extent Is Teacher Competence Situation-Related?

On the generalizability of the results of video-based performance tests 310

Andreas Busse/Gabriele Kaiser

Knowledge and Skills in Subject-Related Didactics and Pedagogics –

On the nature of the professional competence of teachers 328

Kathleen Stürmer/Tina Seidel/Olga Kunina-Habenicht

Knowledge-Based Classroom Observation – Differences between
trainee teachers at the start of their professional career

and explanatory factors 345

Sarah Forster-Heinzer/Fritz Oser

Who Sets the Benchmark? A critical discussion of the advocacy approach ... 361

Georg Hans Neuweg

Contextualized Competence Measurement – Taking stock

of current conceptions and research-methodological approaches 377

Deutscher Bildungsserver

Tips of links relating to the topic of “Contextualized Assessment

of Teacher Competences” 384

Contributions

Helmut Heid

Educational Research in the Context of Social Practice –

On the (social) conditions for the possibility of influencing

the educational practice through educational research 390

Josef Schrader

Large-Scale Assessments and the Education of Adults:

Results, limits, and potentials of research 410

Stephan Wernke/Jochen Werner/Klaus Zierer

Heimann, Schulz, or Klafki? – A quantitative study on the assessment of the practicability of general didactical planning models	429
Book Reviews	452
New Books	459
Impressum	U3

Wie situationsbezogen ist die Kompetenz von Lehrkräften?

Zur Generalisierbarkeit der Ergebnisse von videobasierten Performanztests

Zusammenfassung: Untersucht wird die Situationsabhängigkeit der Performanz von Lehrkräften mithilfe eines videobasierten Assessments, das im Rahmen von TEDS-FU, einer Follow-up-Studie zur *Teacher Education and Development Study: Learning to Teach Mathematics* (TEDS-M), entwickelt wurde. Dabei werden zwei Anforderungen betrachtet: pädagogische Aufgaben des Klassenmanagements und fachdidaktische Aufgaben des Unterrichtens von Mathematik. Ob und – wenn ja – wie stark die Umsetzung der dafür benötigten situationsbezogenen Fähigkeiten von den jeweiligen Unterrichtssituationen des Assessments abhängig ist, wird dabei mit denselben Videos erfasst, so dass sich Unterschiede zwischen pädagogischen und mathematikdidaktischen Fähigkeiten herausarbeiten lassen. Als konkurrierende Hypothesen werden keine, teilweise oder vollständige Situationsabhängigkeit der Itembearbeitung angenommen. Die durchgeführten Modellvergleiche zeigen, dass die intraindividuelle Variation von Lehrerperformanz in den verschiedenen Situationen im Bereich Pädagogik stärker ausgeprägt ist als im Bereich Mathematikdidaktik.

Schlagnote: Kompetenzmessung, Videotest, Mehrdimensionalität, Performanz, Lehrerforschung

Die bisherige Forschung zu Lehrerkompetenzen stützt sich überwiegend auf die Erfassung mithilfe von Papier-und-Bleistift-Tests, typischerweise mit einem hohen Anteil an Multiple-Choice-Items (z. B. Blömeke, Kaiser & Lehmann, 2010). Offene Antwortformate werden ebenfalls eingesetzt; allerdings wird vorab definiert, auf welchen Ausschnitt einer Unterrichtssituation sich ein Item bezieht. Die große Zahl an Beobachtungen (Items) und die Vorabdefinition des Gegenstands seitens der Testentwickler¹ ermöglichen, dass die professionelle Kompetenz der Lehrkräfte mit hoher Zuverlässigkeit (Reliabilität) erfasst wird.

Kritik an diesem Herangehen bezieht sich zum einen auf die Einschränkung der unterrichtlichen Komplexität durch die analytische Kleinteiligkeit. Argumentiert wird, dass Unterricht ein holistisches Geschehen sei und die Aufspaltung komplexer Kompetenzen in ihre Einzelteile an der Realität des Klassenraums vorbeigehe. Zum anderen wird kritisiert, dass die Erhebung weitgehend kontextfrei erfolge, also eher das generelle Leistungspotenzial einer Lehrperson erfasse als seine Realisierung in unterschied-

1 Aus Platzgründen wird nachfolgend nur die männliche Geschlechtsform verwendet. Es sind jedoch stets beide Geschlechter gemeint.

lichen Unterrichtssituationen. Argumentiert wird, dass diese Realisierung von situationsspezifischen Merkmalen abhängt – beispielsweise von der Klassenkomposition oder dem Unterrichtsgegenstand bzw. dem Unterrichtsziel, aber auch von dem konkreten Geschehen im Klassenraum, insbesondere den Reaktionen der Schüler.

Diese Abhängigkeit wird von Vertretern traditioneller Assessments nicht bestritten. Ihrem Ansatz unterliegt allerdings die Annahme, dass – trotz Schwankungen in der Transformation von Kompetenz in Performanz aufgrund von situativen Bedingungen – Lehrpersonen mit höherer Kompetenz im Mittel über verschiedene Unterrichtssituationen hinweg dennoch angemessener handeln als andere Lehrpersonen. Die Frage ist, ob diese Annahme berechtigt ist oder ob die Performanz von Lehrkräften situationsabhängig so variabel ist, dass Inkonsistenzen in Relation zu ihrer Kompetenz auftreten, dass Lehrpersonen mit geringerer Kompetenz also häufiger besser handeln als andere und dass diese Häufigkeit über zufälliges Auftreten hinausgeht. Eine zweite Frage ist, ob die intraindividuelle Variation unterschiedlich ausfällt, je nachdem ob man pädagogische oder fachdidaktische Fähigkeiten betrachtet.

Diese Fragen können mithilfe von Assessments geprüft werden, die in Form von Texten, Bildern, Videos oder Realsimulationen Unterrichtsszenen darbieten und situationsbezogene Reaktionen hierauf verlangen. Für größere Stichproben weisen insbesondere Videos Potenzial auf (zuletzt u. a. Bruckmaier et al., 2013; Lindmeier, Heinze & Reiss, 2013). Sie können situationsspezifisch kontextualisieren, zu unterschiedlichen Themen längere Sequenzen an Lehrer-Schüler-Interaktionen zeigen und Lehrerkompetenz stärker holistisch abbilden, indem pädagogische und fachliche Anforderungen zugleich bewältigt werden müssen (Benjamin, 2013).

Mit videobasierten Assessments sind allerdings schwierige methodische Herausforderungen verbunden. Falls Lehrkräfte mit derselben Kompetenz in einer Situation tatsächlich signifikant unterschiedlich handeln oder falls Lehrkräfte mit höherer Kompetenz nicht systematisch angemessener handeln als andere, wäre eine Extrapolation von einer Situation auf andere nicht zulässig. Brennan (2001, S. 295) bezeichnet dies als „reliability-validity paradoxon“. Reliabilität sei als Konzept auf Stabilität angewiesen: „Reliability, broadly conceived, involves quantifying the consistencies and/or inconsistencies in examinee scores.“

In diesem Beitrag wird untersucht, wie situationsabhängig bestimmte Aspekte der Performanz von Lehrkräften tatsächlich sind. Dies erfolgt mithilfe eines videobasierten Assessments, das im Rahmen von TEDS-FU, einer Follow-up-Studie zur *Teacher Education and Development Study: Learning to Teach Mathematics* (TEDS-M), entwickelt wurde. Dabei werden zwei Anforderungen betrachtet: pädagogische Aufgaben des Klassenmanagements und fachdidaktische Aufgaben des Unterrichtens von Mathematik. Ob und wenn ja wie stark die Erfassung dieser beiden situationsbezogenen Fähigkeiten von den jeweiligen Videovignetten (Unterrichtssituationen) abhängig ist, wird dabei erstmals mit denselben Unterrichtsvideos untersucht, sodass sich Unterschiede zwischen pädagogischen und mathematikdidaktischen Fähigkeiten herausarbeiten lassen.

1. Theoretischer Rahmen und Forschungsstand

1.1 Transformation von Kompetenz in Performanz

Um Kompetenz situationsspezifisch und holistisch untersuchen zu können, muss das bisher in Deutschland dominierende analytische und stabile Verständnis erweitert werden. Blömeke, Gustafsson und Shavelson (2015) entwickelten zu diesem Zweck ein Modell, das Kompetenz als Kontinuum mit vielfachen Übergängen betrachtet (P-I-D model of competence transformation; Abb. 1). Deren empirische Erhebung ist allerdings methodisch anspruchsvoll. Da die zu bearbeitenden Aufgaben komplexer sind als in klassischen Papier-und-Bleistift-Tests, können weniger Aufgaben eingesetzt werden. Dies kann mit einem Verlust an diagnostischer Genauigkeit, Generalisierbarkeit und Reliabilität einhergehen (Brennan, 2001). In Bezug auf die Items, die zu einer Aufgabe gehören, besteht zudem die Gefahr, dass sie stärker untereinander zusammenhängen als mit den übrigen. Dieses statistische Problem stellt die empirische Manifestation von kontextgebundenen Fähigkeiten dar, deren Untersuchung am Beispiel von Mathematiklehrkräften im Mittelpunkt des vorliegenden Beitrags steht.

Entsprechende Studien liegen von Kersting, Givvin, Sotelo und Stigler (2010), Bruckmaier et al. (2013) und Dunekacke, Jenßen und Blömeke (2015) vor. Kersting untersucht videobasiert die mathematikbezogenen Analyse- und Handlungsfähigkeiten von Lehrkräften. Nach Betrachten von 10 bis 13 online präsentierten Videoclips geben diese offene Kommentare zur Lehrer-Schüler-Interaktion im dargestellten Unterricht ab, aus denen vier Teilscores zur mathematischen und lernprozessbezogenen Analysefähigkeit, zur Interpretationsqualität sowie zur Entwicklung von Handlungsoptionen gebildet werden. Zur Untersuchung von Situationsabhängigkeit stützen sich Kersting et al.

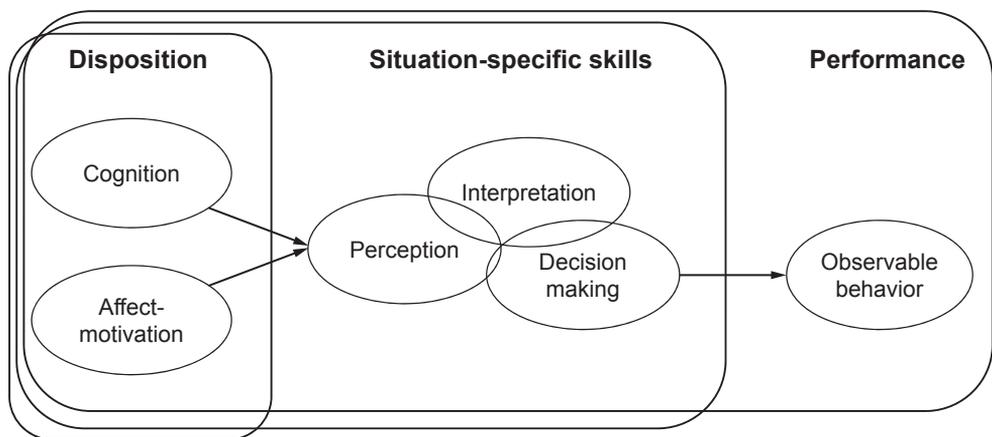


Abb. 1: Modell der Transformation von Kompetenz in Performanz, vermittelt über situationsbezogene Fähigkeiten der Wahrnehmung (P = perception), Interpretation (I = interpretation) und Entscheidungsfindung (D = decision-making)

(2010) auf die Generalizability-Theorie, die Varianz in Konstrukten möglichst differenziert erklären und Quellen identifizieren will, die Messfehler produzieren (Shavelson & Webb, 1991). Die Bearbeitung unterschiedlicher Situationen kann eine solche Quelle darstellen. Bei Kersting et al. (2010) klärt ein Modell, das diese als Fehlerquelle spezifiziert, allerdings nicht substantiell mehr Varianz auf als ein Modell ohne Videofaktor. Die Itembearbeitung erfolgt in dieser Studie also weitgehend situationsunabhängig.

Dunekacke et al. (2015) untersuchen die mathematikbezogenen Wahrnehmungs-, Analyse- und Entscheidungsfähigkeiten von Erziehern mithilfe eines aus drei Videos bestehenden Assessments. Eine potenzielle Situationsabhängigkeit wird aufgegriffen, indem sie Item-Parcels bilden (Little, Cunningham, Shahar & Widaman, 2002), mit denen sichergestellt wird, dass Messvorgänge oft weder dezidiert situationsunabhängig sind, sodass ein Testscore über alle Videos hinweg gebildet werden kann, noch dezidiert situationsabhängig sind, sodass die Items der verschiedenen Videos nichts gemeinsam haben. Entsprechend gehen auch Bruckmaier et al. (2013) in ihrem videobasierten Assessment von Mathematiklehrkräften vor. Dieses resultiert in fünf Dimensionen – Schülerorientierung, Methodenorientierung, Verständnisorientierung, fachliche Präzision und Ergreifen der didaktischen Chance –, die wiederum zu zwei übergeordneten Dimensionen pädagogisch-unterrichtsmethodischer und fachbezogener Art zusammengefasst werden.

Fächerübergreifende Erkenntnisse zur Situationsabhängigkeit insbesondere der Fähigkeit zum Klassenmanagement ergeben sich aus einer Forschung, die das Konzept der „professional vision“ von Goodwin (1994) aufgreift (Sherin, 2007). Seidel, Blomberg und Stürmer (2010) zielen mit dem videobasierten Observer-Tool auf die Erfassung von Zielklarheit, Unterstützung und Klassenklima. Lehrkräften und Studierenden werden sechs Unterrichtsszenen aus verschiedenen Fächern online präsentiert, die auf vierstufigen Likertskalen eingeschätzt werden müssen. Modellvergleiche zeigen, dass ein dreidimensionales Modell ohne Videofaktor die beste Passung an die Daten besitzt.

In der Studie von Gold et al. (im Druck) zeigt sich in Bezug auf den Sachunterricht der Grundschule dagegen eine deutliche Abhängigkeit der Itembearbeitung von den verschiedenen Videosituationen. Fünf Videoclips wurden von Lehramtsstudierenden und praktizierenden Lehrkräften mithilfe von vierstufigen Ratingskalen eingeschätzt. Im Modell, das am besten passt, wurden für jedes Item eine Ladung auf einen Generalfaktor, der die situationsübergreifende Fähigkeit zum Klassenmanagement abbildet, und eine zweite Ladung auf das betreffende Video zugelassen. In vielen Fällen luden die Items sogar höher auf den jeweiligen Video- als auf den Fähigkeitsfaktor. Ähnlich passten in einer Studie von König (eingereicht) mit Lehrkräften sowie vier Videoclips Modelle, in denen die Videos zur Erklärung von Varianz explizit berücksichtigt wurden, besser zu den Daten als Modelle ohne einen solchen Methodenfaktor. Die Situationsabhängigkeit war allerdings geringer ausgeprägt als bei Gold et al. (im Druck).

Der Forschungsstand zur Vignettenabhängigkeit ist also uneinheitlich. Die Studien von Gold et al. (im Druck) und König (eingereicht) stellen eine solche in Bezug auf die Fähigkeit zum Klassenmanagement fest, während dies für Seidel et al. (2010) eher nicht gilt. In Bezug auf mathematikbezogene Unterrichtsansforderungen greifen Dunekacke

et al. (2015) und Bruckmaier et al. (2013) auf videobezogenes Item-Parcelling zurück, was indirekt auf Situationsabhängigkeit hinweist, während Kersting et al. (2010) eine solche nicht feststellten.

1.2 Hypothesen

Wir prüfen die Situationsabhängigkeit bestimmter Aspekte der professionellen Kompetenz von Mathematiklehrkräften der Sekundarstufe I und untersuchen dabei zwei Konstrukte: die *mathematikbezogene* Wahrnehmung (*perception*), Interpretation (*interpretation*) und Handlungsgenerierung (*decision-making*), im Folgenden als M_PID bezeichnet, sowie die *pädagogische* Wahrnehmung, Interpretation und Handlungsgenerierung, P_PID genannt.

Für beide Konstrukte betrachten wir zwei konkurrierende Modelle, und zwar stellen wir der Hypothese einer gering ausgeprägten Situationspezifität eine Hypothese gegenüber, die eine starke intraindividuelle Abhängigkeit der Itembearbeitung von den einzelnen Videos annimmt. Im ersten Falle würde ein eindimensionales Modell, das nur die Fähigkeit zum Klassenmanagement bzw. zum Unterrichten von Mathematik, aber keinen Videofaktor spezifiziert, signifikant besser oder zumindest nicht schlechter passen. Im zweiten Falle würde dagegen das mehrdimensionale Modell, das die Videos und damit die Verschiedenheit der Unterrichtssituationen berücksichtigt, signifikant besser passen (siehe Abb. 2).

Neben dieser inhaltlich ausgerichteten Forschungsfrage zu Unterschieden in der Situationsabhängigkeit von pädagogischen bzw. mathematikdidaktischen Fähigkeiten stellt sich als zweite, methodisch ausgerichtete Forschungsfrage, ob die Bearbeitung der Unterrichtssituationen in allen drei oder nur in einem oder zwei Fällen von den spezifischen Videos abhängt. Variationen im Situationsbezug der Items könnten hierfür ursächlich sein.

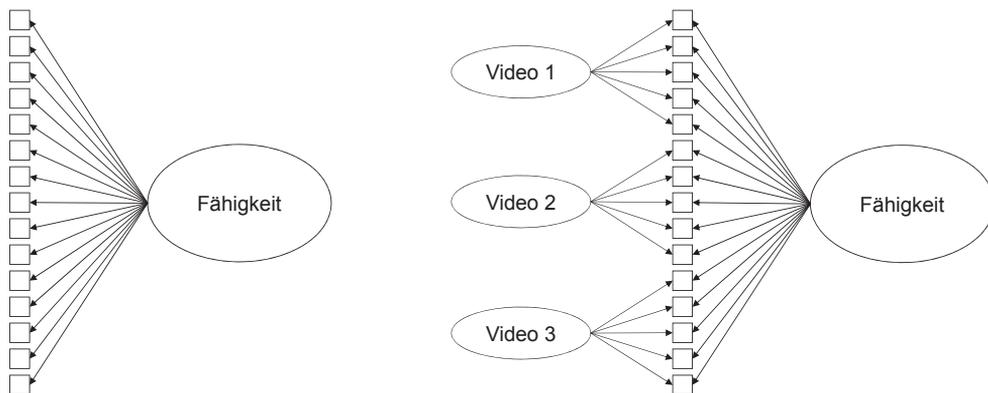


Abb. 2: Schematische Darstellung der konkurrierenden Hypothesen (links keine signifikante Situationsabhängigkeit, rechts signifikant situationsabhängige Bearbeitung der Items)

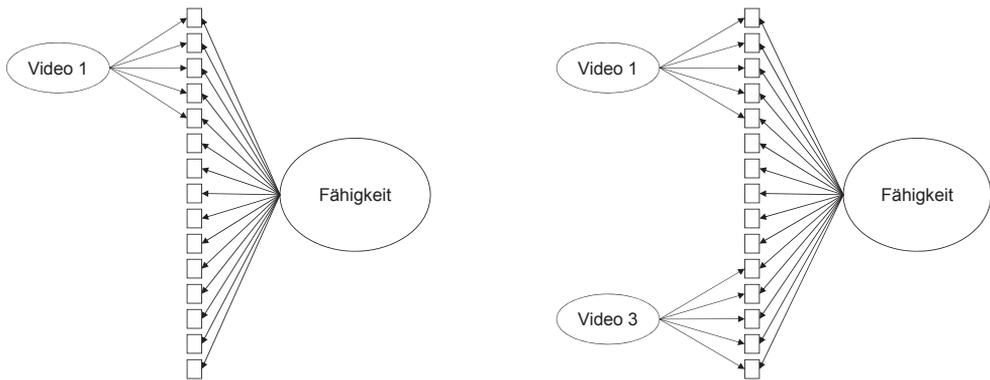


Abb. 3: Beispielhafte Darstellung der differenzierten Hypothesen (links signifikante Situationsabhängigkeit nur bei einer Videovignette, rechts bei zwei Vignetten)

Zur Prüfung werden erneut konkurrierende Hypothesen aufgestellt. Ist die Itembearbeitung nur bei einer Videovignette situationsabhängig, muss neben einer Ladung auf den Generalfaktor, der die übergreifende Fähigkeit der Lehrkräfte abbildet, eine zweite Ladung nur für jene Items zugelassen werden, die entweder an Video 1, 2 oder 3 gebunden sind (siehe das Beispiel in Abb. 3, links). Besteht bei zwei Videos Situationsabhängigkeit, werden neben dem Generalfaktor Doppelladungen für die Items aus Video 1 und 2, Video 1 und 3 bzw. Video 2 und 3 zugelassen (Abb. 3, rechts).

2. Methodisches Vorgehen

2.1 Stichprobe

2008 wurden in Deutschland im Rahmen von TEDS-M angehende Sekundarstufen-I-Lehrkräfte für das Fach Mathematik im letzten Jahr ihrer Ausbildung getestet (Blömeke et al., 2010). Vier Jahre später wurden von dieser Stichprobe 171 Lehrkräfte, die sich meist im dritten Berufsjahr befanden, im Rahmen der Online-Folgestudie TEDS-FU ein weiteres Mal untersucht. Gymnasiallehrkräfte (56%) sowie Haupt- und Realschullehrkräfte (44%) nahmen teil. Sie waren im Mittel 32 Jahre alt und zu 59% weiblich. 74% geben an, einen Leistungskurs in Mathematik besucht zu haben; die mittlere Abiturnote lag bei 2.1 (SD = 0.6). Die Gesamtnote im ersten Staatsexamen lag im Mittel bei 1.9 (SD = 0.5), die des zweiten bei 2.1 (SD = 0.7).

Auch wenn die TEDS-FU-Stichprobe in diesen Hintergrundmerkmalen der repräsentativen Ausgangsstichprobe von TEDS-M ähnlich ist, handelt es sich um eine positiv verzerrte Gelegenheitsstichprobe. Die Leistungsmerkmale in Bezug auf das Professionswissen zeigen, dass vorwiegend Lehrpersonen aus dem oberen TEDS-M-Leistungsspektrum an TEDS-FU teilgenommen haben.

2.2 Videobasiertes Testinstrument

Auswahl der Unterrichtssituationen, technische Umsetzung und Validierung

Die mathematikbezogene bzw. pädagogische Wahrnehmung, Interpretation und Entscheidung (M_PID bzw. P_PID) wurden mithilfe von drei Videovignetten erfasst. Der Expertiseforschung zufolge nehmen Expertenlehrer Unterricht aufgrund besserer Informationsverarbeitung – z. B. durch Verbinden von Einzelhandlungen zu Mustern (*chunking*) – schneller und genauer wahr, sie antizipieren mögliche Probleme und sie können flexibler reagieren.

Die 2,5- bis 4-minütigen Videovignetten zeigen Mathematikunterricht der Klassen 8 bis 10 aus allen Schulformen der Sekundarstufe I, wobei die Unterrichtsszenen gestellt waren („*scripted*“ *lessons*). In den drei Videos wird ein Eindruck von der gesamten Unterrichtsstunde gegeben, indem Teilsequenzen aus den Eingangs-, Erarbeitungs- und Auswertungsphasen gezeigt werden. Vor dem Ansehen erhielten die TEDS-FU-Teilnehmenden Kontextinformationen zur Zusammensetzung der Klasse, den Vorkenntnissen der Schüler sowie zum Inhalt der letzten Stunden. Es wurden zudem der mathematische Gegenstand der Vignette und die Lösungen der Schüleraufgaben erklärt, um zu vermeiden, dass sich die Versuchspersonen vorrangig damit befassen. Der Fokus ihrer Aufmerksamkeit sollte auf fachdidaktischen und pädagogischen Aspekten der Unterrichtsführung liegen. Jedes Video konnte nur einmal angesehen werden.

Die ausgewählten Unterrichtssituationen gehören inhaltlich zu den Leitideen *Raum und Form*, *Messen* sowie *Funktionaler Zusammenhang*. Für diese wurden anknüpfend an das TEDS-M-Framework und unter Hinzuziehung von Indikatoren, die auf der Basis der Experten-Novizen-Forschung nachweislich gut zwischen diesen beiden Gruppen diskriminieren, über Analysen der Bildungsstandards und Lehrpläne für den Mathematikunterricht der Sekundarstufe I sogenannte *critical incidents* identifiziert, denen eine hohe Bedeutung für den Erwerb mathematischer Kompetenz zugesprochen wurde. In Bezug auf die Fähigkeit des Klassenmanagements wurde anhand der Literatur zur Unterrichtsqualität vergleichbar vorgegangen.

Die Analyseergebnisse wurden durch systematische Diskussionen mit in der Lehrplanarbeit und Aufgabenentwicklung erfahrenen Lehrkräften validiert, wobei wichtige Kriterien eine angemessene Zentralität sowie Häufigkeit des Auftretens der Situationen im Mathematikunterricht der Sekundarstufe I waren. Eine Validierung der videografischen Umsetzung erfolgte durch Reviews eines 15-köpfigen mathematikdidaktischen sowie eines 15-köpfigen pädagogischen Expertenpanels. Weitere Validierungen erfolgten über Pilotstudien mit Studierenden, Referendaren und erfahrenen Lehrkräften.

Erfassung von Wahrnehmung, Interpretation und Entscheidungsfindung

Die Tests zu den Videoclips wurden im 2×3-Design konstruiert. Für jedes Video wurden spezifische Items entwickelt, die pädagogische bzw. mathematikdidaktische Anforderungen sowie Fähigkeiten der Wahrnehmung, Interpretation bzw. Entscheidung erfassen.

	Mathematikbezogene Items			Pädagogische Items			Gesamt		
	Rating (P)	Offen (ID)	Gesamt	Rating (P)	Offen (ID)	Gesamt	Rating (P)	Offen (ID)	Gesamt
Froschkönig	6	5	11	10	5	15	16	10	26
Körper	5	8	13	7	7	14	12	15	27
Schachtel	5	5	10	5	6	11	10	11	21
Gesamt	16	18	34	22	18	40	38	36	74

Tab. 1: Ratingskalen zur Erfassung der Wahrnehmungsfähigkeit (P) und offene Aufgaben zur Erfassung der Interpretations- und Entscheidungsfähigkeit (ID)

Zur Erfassung der Wahrnehmungsgenauigkeit wurden adaptierte 4-stufige Ratingskalen eingesetzt, eine 26-köpfige Expertengruppe entschied in einem standardisierten zweistufigen Prozess, welche Antworten als korrekt akzeptiert werden konnten (Kaiser, Benthien, Döhrmann, König & Blömeke, 2013). Über offene Aufgaben wurde die Fähigkeit der Lehrkräfte erfasst, Mathematikunterricht interpretieren und Entscheidungen über Handlungsoptionen treffen zu können. Für deren Kodierung wurden Kodierhandbücher mit Kriterien für richtige und falsche Antworten sowie Ankerbeispielen und Grenzfällen erstellt. Doppelkodierungen von 20% der Fälle ergaben eine gute Inter-Koder-Reliabilität ($\kappa = .86$).

Aus dem mathematikdidaktischen Videotest flossen 34 von ursprünglich 39 Items in die Skalierungen ein. Diese setzten sich etwa je zur Hälfte aus Ratingskalen und offenen Aufgaben zusammen, die wiederum relativ gleichmäßig über die drei Videovignetten verteilt waren. Aus dem pädagogischen Videotest flossen entsprechend 40 von 45 Items in die Skalierungen ein. Jeweils fünf Items wurden wegen schlechter Fitwerte ausgeschlossen. Eine genaue Übersicht gibt Tab. 1.

Inhalte der Videovignetten

In der ersten Videovignette „Schachtel“ wird Mathematikunterricht in einer leistungsheterogenen neunten Gymnasialklasse gezeigt. Die Schüler führen zu der in Abb. 4 gezeigten Aufgabe eine Partnerarbeit durch. Die Art und Weise, in der die Jugendlichen das Problem bearbeiten, unterscheidet sich aus mathematikdidaktischer und pädagogischer Sicht. Im Anschluss an die Partnerarbeit findet ein Plenumsgespräch zum Austausch der Ergebnisse statt.

Ein Beispiel für die Erfassung der mathematikbezogenen Wahrnehmungsfähigkeit zeigt Abb. 5. Als korrekt wurde „trifft voll und ganz zu“ kodiert. Zur Erfassung der mathematikbezogenen Interpretationsfähigkeit lautet die Aufgabe in Bezug auf die drei Schülerpaare (Abb. 6): „Beschreiben Sie kontrastierend die wesentlichen Aspekte der Herangehensweisen aus mathematikdidaktischer Sicht (in Stichworten). Nennen Sie dabei – falls möglich – auch die dazugehörigen Fachbegriffe.“ Eine richtige Antwort musste den Arbeitsprozess orientiert an den drei Repräsentationsmodi nach Bruner

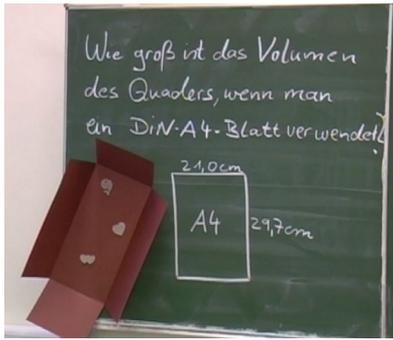


Abb. 4: Screenshot aus der Video-
vignette „Schachtel“

ALLE Aussagen beziehen sich auf die im Video gezeigten Unterrichtsausschnitte.
Markieren Sie zu den folgenden Aussagen jeweils den Grad Ihrer Zustimmung.

Bitte treffen Sie pro Zeile EINE Auswahl!

	trifft voll und ganz zu	trifft eher zu	trifft eher nicht zu	trifft überhaupt nicht zu
Die Aufgabe enthält einige Merkmale einer offenen Aufgabenstellung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abb. 5: Item zur Erfassung der mathematikunterrichtsbezogenen Wahrnehmungsfähigkeit (M_PID)



1



2



3

Im Video wurden drei Paare in ihrem Arbeitsprozess genauer betrachtet. Diese Arbeitsprozesse sollen im Folgenden aus zwei Perspektiven – einer (a) MATHEMATIKDIDAKTISCHEN und einer (b) PÄDAGOGISCHEN – betrachtet werden:

(a) MATHEMATIKDIDAKTISCHE PERSPEKTIVE
In jeder der drei gezeigten Herangehensweisen wird die Aufgabe MATHEMATISCH auf eine GANZ EIGENE ART DARGESTELLT UND BEARBEITET.
Beschreiben Sie kontrastierend die WESENTLICHEN Aspekte der Herangehensweisen aus mathematikdidaktischer Sicht (in Stichworten).
Nennen Sie dabei – falls möglich – auch die dazugehörigen Fachbegriffe.

(b) PÄDAGOGISCHE PERSPEKTIVE
Beschreiben Sie kontrastierend für jedes der drei Paare das WESENTLICHE DER ART UND WEISE in der die beiden jeweiligen Jugendlichen ihre ZUSAMMENARBEIT gestaltet haben (in Stichworten).

Abb. 6: Item zur Erfassung der mathematikunterrichtsbezogenen Interpretationsfähigkeit (M_PID)



In den Bildungsstandards der Kultusministerkonferenz werden ALLGEMEINE MATHEMATISCHE KOMPETENZEN formuliert. Nehmen Sie an, Sie müssten die Stunde an der Stelle, an der das Video endet, INHALTLICH SINNVOLL ANKNÜPFEND WEITERFÜHREN:
Formulieren Sie in einem Satz wörtlich jeweils einen Arbeitsauftrag, den Sie der Klasse geben würden, ...

... wenn Sie die allgemeine mathematische Kompetenz MATHEMATISCHE DARSTELLUNGEN VERWENDEN betonen wollen.

..., wenn Sie die allgemeine mathematische Kompetenz MATHEMATISCH ARGUMENTIEREN betonen wollen.

Abb. 7: Item zur Erfassung der mathematikunterrichtsbezogenen Entscheidungsfähigkeit (M_PID)

Der Lehrer strebt an, die vorhandene Unterrichtszeit effektiv für Lernaktivitäten zu nutzen. Nennen Sie drei organisatorische Mittel, die er zu diesem Zweck verwendet.

Abb. 8: Item zur Erfassung der pädagogischen Entscheidungsfähigkeit (P_PID)

(1966) ikonisch, enaktiv und symbolisch analysieren, wobei auch andere Fachbegriffe als korrekt akzeptiert wurden.

Auch die Fähigkeit, zielangemessen eine Entscheidung zum weiteren Handeln zu treffen, wurde mit offenen Aufgaben erfasst (Abb. 7). Damit das Fortführen der Situationen nicht in einer Beliebigkeit der Antworten und damit ihrer Unvergleichbarkeit resultiert, wurde der Raum möglicher Antworten eingegrenzt, indem beispielsweise eine Anknüpfung an das Video in Form eines Arbeitsauftrags zur Förderung einer spezifischen mathematischen Kompetenz gefordert wurde.

Die zweite Videovignette „Körper“ behandelt aus mathematischer Perspektive inhaltlich die Oberflächen- und Volumenberechnung bei einem Kirchturm, während die dritte Vignette „Froschkönig“ die Bestimmung der Masse einer Goldkugel thematisiert. Bei beiden Vignetten geht es unter dem Gesichtspunkt mathematischer Prozesse um das Modellieren. Dabei steht bei der Körper-Vignette die inhaltliche Gestaltung von geeigneten Mathematikaufgaben durch eine Lehrkraft im Vordergrund, bei der Froschkönig-Vignette sind es Schülerantworten. Aus pädagogischer Perspektive werden jeweils Fragen des Klassenmanagements behandelt. Abb. 8 zeigt hierzu ein Beispiel.

2.3 Datenanalysen

Die Hypothesenprüfung erfolgt, indem mithilfe der Software *ConQuest* die Anpassung von ein- und mehrdimensionalen Modellen an die Daten verglichen wird (Wu, Adams & Wilson, 1997). Ein eindimensionales Modell mit einem Generalfaktor, der über alle Videos hinweg ohne Situationsabhängigkeit die zugrunde liegende mathematische bzw. pädagogische Fähigkeit abbildet, wird mehrdimensionalen Modellen gegenübergestellt, die unterschiedliche Varianten an situationsabhängiger Itembearbeitung annehmen (Abb. 2 und 3). Bei diesem Vorgehen wird die Varianz der Messergebnisse jeder Vignette zerlegt in eine Varianz, die durch einen Generalfaktor aufgeklärt werden kann (d. h. eine allen Vignetten gemeinsame Varianz), sowie in Varianzanteile, die jeweils durch die spezifischen Vignetten aufgeklärt werden können (plus Messfehlervarianz). Dies entspricht dem hierarchischen Denkmodell, das der Kognitionspsychologie insgesamt zugrunde liegt, die oft einen Generalfaktor allgemeiner Fähigkeit postuliert, darüber hinaus aber auch subtest-spezifische Faktoren, die in der vorliegenden Studie durch die drei Vignetten mit ihren spezifischen Fragen repräsentiert sind.

Zur Beurteilung der Modellanpassung an die Daten werden die Fitstatistik in Relation zur zu schätzenden Zahl der Parameter sowie die Reliabilitäten der Faktoren betrachtet. Wenn die Verbesserung der Anpassung durch die Berücksichtigung der Situation größer als null ist, deutet dies darauf hin, dass einige Personen die Items eines Videos trotz gleicher Ausprägung ihrer Kompetenz besser lösen können als andere.

3. Ergebnisse

Tabelle 2 dokumentiert die Anpassung der getesteten Modelle für die mathematikdidaktische Wahrnehmung, Interpretation und Entscheidungsfindung an die Daten sowie die Reliabilitäten und Varianzen der jeweiligen Faktoren. Zunächst wurde ein Modell geschätzt, das keine Situationsabhängigkeit annimmt, sondern nur eine videoübergreifende Fähigkeit (1). Anschließend folgten Modelltestungen, bei denen für jeweils eines der Videos eine Abhängigkeit der Itembearbeitung angenommen wird (2, 3 und 4), bevor Modelle getestet wurden, die für jeweils zwei Videos Abhängigkeit annehmen (5, 6 und 7). Als letztes wurde ein Modell spezifiziert, das für alle drei Videovignetten annimmt, dass ihre Bearbeitung von den dargestellten Situationen abhängt, die zu einem Video gehörenden Items also nicht lokal stochastisch unabhängig sind (8).

In mathematikdidaktischer Hinsicht zeigt Modell 2, das eine situationsübergreifende Fähigkeit annimmt (M_PID) und für die Froschkönig-Vignette (F) eine Situationsabhängigkeit modelliert, die beste Anpassung. Ähnlich gute Werte erreichen die Modelle 5, 6 und 8, denen gemeinsam ist, dass jeweils die Bearbeitung der Froschkönig-Items als situationsabhängig spezifiziert ist. Diese Hinweise auf eine Situationsabhängigkeit der Vignette spiegeln sich auch in den Reliabilitäten und Varianzen wider (Spalten 4 ff. in Tab. 2). Während für die Items, die zur Körper- (Spalten 6 und 10) bzw. Schachtel-Vignette (7 und 11) gehören, kaum systematische Kovariationen ihrer Items und infol-

1	2	3	4	5	6	7	8	9	10	11
Modell	Anpassung	# Par.	Rel. M_PID	Rel. F	Rel. K	Rel. S	Var. M_PID	Var. F	Var. K	Var. S
1 M_PID	6 343,1	38	0.62				0.33			
2 M_PID + F	6 338,8	39	0.58	0.21			0.31	0.18		
3 M_PID + K	6 343,5	39	0.61		0.04		0.33		0.03	
4 M_PID + S	6 343,2	39	0.61			0.05	0.33			0.04
5 M_PID + F + K	6 339,3	40	0.57	0.20	0.04		0.31	0.18	0.03	
6 M_PID + F + S	6 339,0	40	0.57	0.20		0.05	0.31	0.18		0.04
7 M_PID + K + S	6 343,5	40	0.61		0.03	0.05	0.33		0.03	0.05
8 M_PID + F + K + S	6 339,5	41	0.56	0.20	0.03	0.05	0.31	0.18	0.02	0.04

F: Frosch-Vignette, K: Körper-Vignette, S: Schachtel-Vignette. Anpassung: *final deviance* in ConQuest, entspricht $-2 \cdot \text{Log-Likelihood}$, # Par: Anzahl der freien Parameter, Rel.: Reliabilität, Var.: Varianz.

Tab. 2: Modellanpassungen, Reliabilitäten und Varianzen für M_PID

gedessen niedrige Reliabilitäten festzustellen sind, weisen die Items der Froschkönig-Vignette einen sichtbaren gemeinsamen Varianzanteil auf (Spalte 9). Allerdings ist auch dieser niedrig, wie an der geringen Reliabilität deutlich wird (5).

Ob die Unterschiede in den Modellanpassungen tatsächlich signifikant sind, ist in Tab. 3 dokumentiert, wo die einzelnen Modelle miteinander verglichen werden. In der Mehrheit der Modellvergleiche existiert keine signifikante Situationsabhängigkeit der mathematikdidaktischen Itembearbeitung von den Videos. Lediglich vier der Modelle, die die Froschkönig-Vignette enthalten, zeichnen sich durch eine systematisch unterschiedliche Bearbeitung ihrer Items aus.

Tabelle 4 dokumentiert in ähnlicher Form die Anpassung der getesteten Modelle für die pädagogische Wahrnehmung, Interpretation und Entscheidungsfindung an die Daten sowie die Reliabilitäten und Varianzen der jeweiligen Faktoren. Die beste Anpassung zeigt hier Modell 8, das eine situationsübergreifende Fähigkeit annimmt (P_PID) und für alle drei Vignetten (F, K und S) eine Situationsabhängigkeit modelliert. Ähnlich gute Werte erreichen die Modelle 2, 5 und 6, denen wie im Falle der Mathematikdidaktik gemeinsam ist, dass jeweils die Bearbeitung der Froschkönig-Items als situationsabhängig spezifiziert ist. Die Hinweise auf Situationsabhängigkeit spiegeln sich auch in den Reliabilitäten und Varianzen wider (Spalten 4 ff. in Tab. 4). Alle Faktoren weisen systematische Kovariationen ihrer Items auf, die bei der Froschkönig-Vignette am stärksten ausfallen. Allerdings muss man festhalten, dass dieser gemeinsame Varianzanteil ausweislich der niedrigen Reliabilitäten vergleichsweise gering ist.

Die Anpassung an die Daten wird bei den meisten Modellen signifikant besser, wenn für die Froschkönig- und/oder die Körper-Vignette eine Situationsabhängigkeit modelliert wird (Tab. 5). Die Schachtel-Vignette kann dagegen von einer solchen situations-spezifischen Modellierung in der Regel ausgenommen und auf die Varianzerklärung durch den Generalfaktor P_PID begrenzt werden.

Vergleich der Modelle		Diff. Anpassung	Diff. # Par.	Signifikanz	Situationsabhängigkeit
M_PID	M_PID + F	4.3	1	*	F
M_PID	M_PID + K	-0.4	1	ns	Nein
M_PID	M_PID + S	-0.1	1	ns	Nein
M_PID + F	M_PID + F + K	-0.5	1	ns	Nein
M_PID + F	M_PID + F + S	-0.2	1	ns	Nein
M_PID + K	M_PID + F + K	4.2	1	*	F
M_PID + K	M_PID + K + S	-0.1	1	ns	Nein
M_PID + S	M_PID + F + S	4.1	1	*	F
M_PID + S	M_PID + K + S	-0.4	1	ns	Nein
M_PID + F + K	M_PID + F + K + S	-0.2	1	ns	Nein
M_PID + F + S	M_PID + F + K + S	-0.5	1	ns	Nein
M_PID + K + S	M_PID + F + K + S	4.1	1	*	F
M_PID + F	M_PID + F + K + S	-0.7	2	ns	Nein
M_PID + K	M_PID + F + K + S	4	2	ns	Nein
M_PID + S	M_PID + F + K + S	3.7	2	ns	Nein
M_PID	M_PID + F + K + S	3.6	3	ns	Nein

F: Froschkönig-Vignette, K: Körper-Vignette, S: Schachtel-Vignette. Diff. Anpassung: 2*Differenz der Log-Likelihood, Diff. # par: Differenz der Anzahl freier Parameter, Signifikanz: Chi-Quadrat-Test
* p < 0.05, ns = nicht signifikant.

Tab. 3: Ergebnisse der Signifikanztests auf Situationsabhängigkeit von M_PID

1	2	3	4	5	6	7	8	9	10	11
Modell	Anpassung	# Par.	Rel. P_PID	Rel. F	Rel. K	Rel. S	Var. P_PID	Var. F	Var. K	Var. S
1 P_PID	7 078,6	44	0.65				0.34			
2 P_PID + F	7 066,3	45	0.59	0.31			0.30	0.31		
3 P_PID + K	7 074,1	45	0.63		0.20		0.34		0.15	
4 P_PID + S	7 075,9	45	0.63			0.16	0.34			0.13
5 P_PID + F + K	7 063,1	46	0.56	0.29	0.17		0.29	0.30	0.13	
6 P_PID + F + S	7 064,0	46	0.56	0.31		0.15	0.29	0.31		0.12
7 P_PID + K + S	7 072,4	46	0.61		0.17	0.14	0.34		0.14	0.11
8 P_PID + F + K + S	7 061,8	47	0.54	0.31	0.15	0.13	0.28	0.30	0.11	0.10

F: Froschkönig-Vignette, K: Körper-Vignette, S: Schachtel-Vignette. Anpassung: final deviance in ConQuest, entspricht -2*Log-Likelihood, # Par: Anzahl der freien Parameter, Rel.: Reliabilität, Var.: Varianz.

Tab. 4: Modellanpassungen, Reliabilitäten und Varianzen für P_PID

Vergleich der Modelle		Diff. Anpassung	Diff. # Par.	Signifikanz	Situationsabhängigkeit
P_PID	1 P_PID + F	12.3	1	***	F
P_PID	1 P_PID + K	4.5	1	*	K
P_PID	1 P_PID + S	2.6	1	ns	Nein
P_PID + F	1 P_PID + F + K	3.2	1	ns	Nein
P_PID + F	1 P_PID + F + S	2.3	1	ns	Nein
P_PID + K	1 P_PID + F + K	10.9	1	***	F
P_PID + K	1 P_PID + K + S	1.7	1	ns	Nein
P_PID + S	1 P_PID + F + S	11.9	1	***	F
P_PID + S	1 P_PID + K + S	3.5	1	ns	Nein
P_PID + F + K	1 P_PID + F + K + S	1.4	1	ns	Nein
P_PID + F + S	1 P_PID + F + K + S	2.3	1	ns	Nein
P_PID + K + S	1 P_PID + F + K + S	10.6	1	***	F
P_PID + F	1 P_PID + F + K + S	4.5	2	ns	Nein
P_PID + K	1 P_PID + F + K + S	12.3	2	***	F und S
P_PID + S	1 P_PID + F + K + S	14.2	2	***	F und K
P_PID	1 P_PID + F + K + S	16.8	3	***	F und K und S

F: Froschkönig-Vignette, K: Körper-Vignette, S: Schachtel-Vignette. Diff. Anpassung: 2*Differenz der Log-Likelihood, Diff. # par: Differenz der Anzahl freier Parameter, Signifikanz: Chi-Quadrat-Test.

*** $p < .001$, * $p < 0.05$, ns = nicht signifikant.

Tab. 5: Ergebnisse der Signifikanztests auf Situationsabhängigkeit von P_PID

4. Zusammenfassung und Diskussion

Mithilfe von drei Videovignetten wurde geprüft, wie situationsabhängig die mathematikdidaktischen und pädagogischen Fähigkeiten zur Wahrnehmung, Interpretation und Entscheidungsfindung bei Mathematiklehrkräften sind. Als konkurrierende Hypothesen wurde keine, teilweise oder vollständige Situationsabhängigkeit der Itembearbeitung angenommen. Die durchgeführten Modellvergleiche haben gezeigt, dass die intraindividuelle Variation von Lehrerperformanz im Bereich Pädagogik stärker ausgeprägt war als im Bereich Mathematikdidaktik. Besonders von Situationsabhängigkeit betroffen war in beiden Fällen die Froschkönig-Vignette, im Bereich der Pädagogik zum Teil zusätzlich die Körper-Vignette.

Die Froschkönig-Vignette unterscheidet sich von den übrigen Vignetten durch ihre Fokussierung auf die Wahrnehmung und Interpretation von Schülerantworten, während die anderen beiden Vignetten mathematikdidaktische Leitideen und Unterrichtsmethoden behandeln. Insofern deuten sich hier ggf. unterschiedliche Stärken und Schwächen unabhängig vom mittleren M_PID-Fähigkeitsniveau an. Insbesondere müssen zur

Froschkönig-Vignette *mehrere* Schüler mit ihren Antworten erinnert und unterschieden werden, während es in den anderen beiden Vignetten eher Einzelfälle sind.

Allerdings ist der gemeinsame Varianzanteil der zu einem Video gehörenden Items vergleichsweise niedrig. Zieht man zudem die Vielzahl an Modelltestungen in Betracht, bei denen ein fälschlicherweise als signifikant markiertes Ergebnis nicht ausgeschlossen werden kann, erscheint es gerechtfertigt, insgesamt von einer eher geringen Situationsabhängigkeit des Lehrerhandelns in unserer Studie auszugehen, und im Sinne des Sparsamkeitsgebotes die eindimensionale Modellierung des M_PID-Faktors in weiterführenden Arbeiten zu verwenden. Hierfür spricht auch, dass die situationsübergreifende Fähigkeit zur mathematikdidaktischen Wahrnehmung, Interpretation und Entscheidungsfindung für alle Items signifikant Varianz aufklärt, was für die Items der videobezogenen Faktoren in den übrigen Modellen nicht immer der Fall ist.

Die vergleichende Analyse von mathematikbezogenen und pädagogischen Anforderungen im Rahmen eines *gemeinsamen* Testdesigns und mit denselben Videos als Stimulus bot einen wichtigen Einblick, den die eingangs zitierten Studien aufgrund ihrer jeweils gewählten Fokussierung (fachbezogen Kersting et al., 2010; fächerübergreifend Gold et al., im Druck; Seidel et al., 2010; König, eingereicht) bislang nicht leisteten. Nur so konnte herausgearbeitet werden, dass die Situationsabhängigkeit der Transformation von Kompetenz in Performanz im Bereich der Pädagogik-Skala unseres Videovignettestes höher ist.

Eine Erklärung für dieses Ergebnis kann inhaltlicher oder methodischer Art sein. Inhaltlich ist die größere Variabilität von pädagogischen Anforderungen im Vergleich zu mathematikdidaktischen im Unterricht zu betonen. Möglicherweise ist das pädagogische Lehrerwissen auch weniger eng vernetzt als das Wissen über mathematikbezogene Inhalte. Lehrkräfte unterrichten immer mehrere Unterrichtsfächer, in denen sie unterschiedliche pädagogische Erfahrungen machen, und werden in der Lehrerbildung selten stringent auf diese vorbereitet. Die mathematikbezogene Ausbildung zeichnet sich dagegen durch eine stärker kanonisierte Struktur aus. Festzuhalten ist jedoch, dass die Kovarianz der videospezifischen Items selbst in Bezug auf P_PID begrenzt ist. Je nach Fragestellung erscheint es daher gerechtfertigt, mit den komplexeren Modellen oder aber im Sinne des Sparsamkeitsgebotes auch in diesem Falle mit der eindimensionalen Modellierung weiterzuarbeiten.

5. Methodische Begrenzungen und Schlussfolgerungen

Wir haben drei kurze Videoclips eingesetzt, womit die Unterrichtskomplexität reduziert ist. Auch wenn wir auf die Zentralität und Häufigkeit der gezeigten Szenen im Mathematikunterricht geachtet haben und diese in Expertenreviews bestätigt wurden, schließt dies größere Variabilität in Bezug auf andere Unterrichtssituationen nicht aus. Zudem hat an TEDS-FU eine positiv verzerrte Gelegenheitsstichprobe teilgenommen, womit offenbleiben muss, wie generalisierbar die Ergebnisse auf das gesamte Leistungsspektrum von Lehrkräften sind.

So stellt sich die Frage, inwieweit unsere Ergebnisse zur Situationsabhängigkeit von pädagogischen bzw. mathematikdidaktischen Fähigkeiten auf das in dieser Untersuchung verwendete Testinstrument begrenzt sind. Dieses fokussiert die Wahrnehmung, Interpretation und Entscheidung von Lehrkräften und kann nicht als Beleg für Situations(un)abhängigkeit von Lehrerhandeln generell verwendet werden, da so mögliche Unterschiede im Hinblick auf andere Teilaspekte des Lehrerhandelns verwischt würden. Denkbar ist zum Beispiel, dass eine Verwendung anderer Dimensionen – zum Beispiel Zielklarheit, Unterstützung und Klassenklima nach Seidel et al. (2010) – zu anderen Ergebnissen führt. So könnte auch die bisher vorliegende Widersprüchlichkeit der Ergebnisse zur Situationsabhängigkeit von Lehrerhandeln erklärt werden.

In Bezug auf die von uns festgestellte Situationsabhängigkeit ist zu bedenken, dass mit dem hier gewählten Ansatz der Modellierung auf Videoebene nur Abhängigkeiten aufgedeckt werden können, die alle Items einer Unterrichtssituation betreffen. Wenn die Modellvergleichstests ergeben, dass die Varianz eines Vignettenfaktors nicht signifikant von null verschieden ist, können dennoch lokale stochastische Abhängigkeiten zwischen einzelnen Items bestehen. Ein stützender Beleg hierfür zeigt sich bei einem ergänzenden Vergleich der Anpassung von Modellen mit frei geschätzten und fixierten Faktorladungen. Durch die Restriktion tritt eine signifikante Verschlechterung des Modellfits auf, d. h. die verschiedenen Items passen unterschiedlich gut zu den verschiedenen Videos. Um Abhängigkeiten zwischen einigen wenigen Items einer Vignette aufdecken zu können, bieten sich verschiedene Ansätze an, z. B. die Q3-Statistik nach Yen (1993), bei der es sich um Partialkorrelationen von Items handelt, die lokale stochastische Abhängigkeiten einzelner Items anzeigen. Hierfür werden allerdings deutlich größere Stichproben benötigt.

6. Perspektiven für Forschung und Praxis

Zukünftige Forschung sollte direkt an diesen Einschränkungen ansetzen. Replikationen unserer Ergebnisse mit anderen Videos (längere Unterrichtssequenzen, größere Anzahl) und einer vergleichbaren Gruppe an Mathematiklehrkräften, vor allem aber auch mit anderen Gruppen (anderes mathematikbezogenes Leistungsspektrum, andere Schulstufe) bzw. Unterrichtsfächern (Deutsch, Englisch etc.) wären wünschenswert. Erst dann lässt sich genauer bestimmen, wie variabel oder stabil das Handeln von Lehrkräften unter unterschiedlichen Bedingungen ist. Dies gilt auch für den Vergleich der Situationsabhängigkeit von pädagogischen und mathematikdidaktischen Anforderungen.

Eine besonders interessante weiterführende Forschungsfrage ist herauszufinden, worauf die in unserer und anderen Studien identifizierten Schwankungen zurückzuführen sind, was also die Ursachen dieser intraindividuellen Variation der Kompetenz(messung) sind. Denkbar sind hier unsystematische Ursachen wie Tagesform, aber auch systematische Ursachen wie sehr spezielle Stärken und Schwächen. Eine erneut weiterführende Frage wäre dann, ob bei starker Situationsabhängigkeit diese explizit modelliert werden sollte. Sie würde in diesem Falle als „construct-irrelevant task specificity“ (Gold et al.,

im Druck) quasi herausgerechnet, um die vermeintlich „wahre“ Kompetenz herauszuarbeiten. Eine Gegenposition wäre, hiervon abzusehen, weil die Variabilität zur Umsetzung von Kompetenz in Performanz natürlicherweise dazu gehört.

Unter Ausbildungs- und Auswahlgesichtspunkten ist die deutlich gewordene weitgehende Stabilität des Lehrerhandelns über verschiedene Situationen hinweg eine wichtige Erkenntnis. Sollte sie in Folgestudien weiter gestützt werden können, bedeutet sie eine große Verlässlichkeit im Hinblick auf Qualitätsmerkmale von Lehrkräften.

Literatur

- Benjamin, R. (2013). The principles and logic of competency testing in higher education (S. 127–136). In S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn & J. Fege (Hrsg.), *Modeling and Measuring Competencies in Higher Education: Tasks and Challenges* (S. 127–136). Boston: Sense.
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. (2015). Beyond dichotomies: Viewing competence as a continuum. *Zeitschrift für Psychologie*, 223, 3–13.
- Blömeke, S., Kaiser, G., & Lehmann, R. (Hrsg.) (2010). *TEDS-M 2008 – Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich*. Münster: Waxmann.
- Brennan, R. L. (2001). *Generalizability Theory*. New York: Springer.
- Bruckmaier, G., Krauss, S., Leiss, D., Blum, W., Neubrand, M., & Brunner, M. (2013). COACTIV-Video: Eine unterrichtsnahe Erfassung fachdidaktischen Wissens mittels Videovignetten. In *Beiträge zum Mathematikunterricht 2013* (S. 212–215). Münster: WTM.
- Bruner, J. S. (1966). *Toward a Theory of Instruction*. Cambridge/London: Harvard University Press.
- Dunekacke, S., Jenßen, L., & Blömeke, S. (2015). Mathematikdidaktische Kompetenz von Erzieherinnen und Erziehern: Validierung des KomMa-Leistungstests durch die videogestützte Erhebung von Performanz. *Zeitschrift für Pädagogik*, 61. Beiheft, 80–99.
- Gold, B., et al. (im Druck). *Measuring the Professional Vision of Classroom Management: Test Validation and Methodological Challenges*.
- Goodwin, C. (1994). Professional Vision. *American Anthropologist*, 96, 606–633.
- Kaiser, G., Benthien, J., Döhrmann, M., König, J., & Blömeke, S. (2013). Expert Ratings as an Instrument for Validating Results of Video-Based Testing. In A. Lindmeier & A. Heinze (Hrsg.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Bd. 1, S. 83). Kiel: PME.
- Kersting, N. B., Givvin, K., Sotelo, F., & Stigler, J. W. (2010). Teacher's analysis of classroom video predicts student learning of mathematics: Further exploration of a novel measure of teacher knowledge. *Journal of Teacher Education*, 61, 172–181.
- König, J. (eingereicht). *Measuring Classroom Management Expertise (CME) of Teachers: A Video-Based Assessment Approach and Statistical Results*.
- Lindmeier, A., Heinze, A., & Reiss, K. (2013). Eine Machbarkeitsstudie zur Operationalisierung aktionsbezogener Kompetenz von Mathematiklehrkräften mit videobasierten Maßen. *Journal für Mathematik-Didaktik*, 34, 99–119.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9, 151–173.
- Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer“ – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. *Zeitschrift für Pädagogik*, 56. Beiheft, 296–306.

- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park: Sage.
- Sherin, M. G. (2007). The development of teachers' professional vision in video clubs. In R. Goldman, R. Pea, B. Barron & S. J. Derry (Hrsg.), *Video research in the learning sciences* (S. 383–395). Mahwah: Lawrence Erlbaum.
- Wu, M. L., Adams, R. J., & Wilson, M. (1997). *ConQuest: Generalised item response modelling software*. Hawthorn: ACER.
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal for Educational Measurement*, 30, 187–213.

Abstract: Using a video-based assessment developed in the follow-up study of the *Teacher Education and Development Study: Learning to Teach Mathematics* (TEDS-M), we examine whether and to what extent teacher performance depends on differences in classroom situations. Two teaching challenges are focused upon: general pedagogical tasks related to classroom management and subject-specific tasks related to teaching mathematics. The assessment of situation-specific skills necessary to master these two tasks is done with video-vignettes. Since each time the same video-vignettes are used, differences between general pedagogical and subject-specific skills can be examined. We test competing hypotheses of having no, partial, or complete item-dependency on classroom situations. Findings reveal that intra-individual variation of teacher performance is more pronounced in the area of general pedagogy than in the area of mathematics pedagogy.

Keywords: Competence Assessment, Video Test, Multidimensionality, Performance, Teacher Research

Anschrift der Autorinnen/des Autors

Prof. Dr. Sigrid Blömeke, Centre for Educational Measurement at the University of Oslo (CEMO), Postboks 1072/Blindern, 0316 Oslo, Norwegen
E-Mail: sigribl@cemo.uio.no

Prof. Dr. Johannes König, Universität zu Köln, Institut für Allgemeine Didaktik und Schulforschung, Gronewaldstraße 2, 50931 Köln, Deutschland
E-Mail: johannes.koenig@uni-koeln.de

Dr. Ute Suhl, Humboldt-Universität zu Berlin, Systematische Didaktik und Unterrichtsforschung, Ziegelstraße 13c, Unter den Linden 6, 10099 Berlin, Deutschland
E-Mail: ute.suhl@staff.hu-berlin.de

Jessica Hoth, Universität Vechta, Institut für Didaktik der Mathematik und des Sachunterrichts (IfD), Postfach 15 53, 49364 Vechta, Deutschland
E-Mail: jessica.benthien@uni-vechta.de

Prof. Dr. rer. nat. Martina Döhrmann, Universität Vechta, Institut für Didaktik der Mathematik und des Sachunterrichts (IfD), Postfach 15 53, 49364 Vechta, Deutschland
E-Mail: martina.doehrmann@uni-vechta.de