

Die Psychometrie des Lehrerprofessionswissens – Betrachtungen zur Validität verschiedener empirischer Zugänge

In den letzten Jahren ist eine regelrechte Industrie für die Entwicklung von Professionswissenstests für Lehrkräfte entstanden. Interessanterweise ist dies – sicher nicht zuletzt aufgrund des PISA-Schocks 2000 – vor allem ein Phänomen im deutschen Sprachraum. Mittlerweile gibt es für verschiedenste Unterrichtsfächer Tests zum Fachwissen (FW) und zum fachdidaktischen Wissen (FDW) von Lehrkräften (in Krauss et al., 2017, sind über 40 gelistet) sowie zahlreiche Testkonstruktionen zum fachübergreifenden pädagogischen Wissen (PW). Dabei geht es im Wesentlichen nicht um Individualdiagnostik, sondern um die Identifikation von Determinanten und Erwerbsverläufen (z. B. bei Studierendenpopulationen) oder die Untersuchung der prädiktiven Validität (z. B. für die Unterrichtsqualität), und zwar meist mit dem Ziel einer evidenzbasierten Optimierung der Lehrerbildung.

„First generation“: Papier-und-Bleistift-Tests

Als Pioniere derartiger Testkonstruktionen haben in den Jahren 2004/2005 drei Arbeitsgruppen im Fach Mathematik Messinstrumente zum fachbezogenen Professionswissen von (angehenden) Lehrkräften mit Fokus auf heterogene Populationen entwickelt: In der COACTIV-Studie (Kunter et al., 2011) wurden vor allem Lehrpersonen der Sekundarstufe und in MT21 (Blömeke, Kaiser & Lehmann, 2008) Lehramtsstudierende sowie Referendare der Sekundarstufe untersucht. Die Arbeitsgruppe aus Michigan zielte vornehmlich auf Primarstufenlehrkräfte ab (Ball, Hill & Bass, 2005).

Obwohl alle drei Gruppen völlig unabhängig voneinander arbeiteten, werden zwei Parallelen, eine methodische und eine inhaltliche, augenscheinlich: Erstens finden sich in den Tests aller drei Arbeitsgruppen Items, die als *hypothetische Unterrichtsszenarios* formuliert sind, in denen Lehrkräfte in typische didaktische Anforderungssituationen versetzt werden (z. B. ‚Erklären Sie Schülerinnen und Schülern ...‘). Zweitens bilden das *Erklären von Fachinhalten* und *fachbezogene Schüler(wohl)vorstellungen* einen wesentlichen Anteil der Iteminhalte, zwei zentrale Facetten, die in jüngerer Zeit in (fast) allen Testkonstruktionen zum FDW aufgegriffen werden.

Bei den meisten Untersuchungsinstrumenten der ersten Generation handelt es sich um Papier-und-Bleistift-Tests, bei denen das professionelle Wissen von Lehrkräften ‚im Labor‘, d. h. nicht im Unterrichtskontext, über ‚schriftliche Antworten‘ erhoben wird. Mittlerweile wird jedoch vor allem in der Mathematik und den Naturwissenschaften sehr rege und fruchtbar

darüber diskutiert, inwiefern mit Papier-und-Bleistift-Tests gemessenes *Wissen* mit *Können* beziehungsweise sogar mit tatsächlichem *Handeln* im Unterricht gleichgesetzt werden kann (z. B. Blömeke, Gustafsson & Shavelson, 2015; Vogelsang, 2014; Cauet et al., 2015).

„Next generation“: Videovignetten

Seit den letzten Jahren werden deshalb verstärkt Videovignetten als Stimuli verwendet, die eine sehr gute Approximation an kontextreiche Unterrichtssituationen bei gleichzeitiger Beibehaltung der Standardisierung bieten (z. B. Lindmeier, 2013; Bruckmaier et al., 2016; Rutsch et al., 2017). Die zahlreichen Vor- und Nachteile dieser Erhebungsmethode im Vergleich zu klassischen Papier-und-Bleistift-Verfahren werden in der Forschungsliteratur dabei umfassend und kritisch reflektiert (König & Lebens, 2012; Lindmeier, 2013; Vogelsang & Reinhold, 2013; Kaiser et al., 2015; Knievel, Lindmeier & Heinze, 2015; Neuweg, 2015; Bruckmaier et al., 2016; Rutsch et al., 2017).

Wenn auch unstrittig ist, dass es keinen singulären Königsweg gibt, ist dennoch festzuhalten, dass einige Aspekte wie das selbstständige Stoppen des Films an ‚didaktisch kritischen Stellen‘ durch den Studienteilnehmer, die Modellierung von Zeitdruck bei der Bearbeitung oder die Implementation von Animationen wie beispielweise dynamische Geometriesoftware offenbar *ausschließlich* im Videoformat zu berücksichtigen sind. Zu häufiger Kritik am Videoformat veranlasst hingegen die unnatürliche Vorselektion der Informationen aufgrund der gewählten Kameraperspektive und die hierdurch potenziell bedingte Einschränkung von Wahrnehmungskonzepten wie *professional vision* oder *noticing*. Angeführt werden auch eine mögliche Beeinflussung durch unzureichende Ton- und Bildqualität, die Tatsache, dass es sich bei den gezeigten Videovignetten letztlich um ‚fremden Unterricht‘ – meist dargestellt durch eine ‚Schauspiellehrkraft‘ – handelt, und, ob dieser *advokatorische Ansatz* überhaupt ein echtes Authentizitätserleben bei den getesteten Studienteilnehmern evozieren kann.

Validität

Spätestens seit Messick (1988) wird die Validität einer psychometrischen Messung nicht mehr als ein Einzelwert betrachtet (wie für die Reliabilität oft nur der Wert für Cronbachs Alpha), sondern als ein *integrativer und fortwährender Prozess* der empirischen und argumentativen Verteidigung, schließlich also als *Gesamtbewertung* der vorhandenen theoretischen und empirischen Evidenz angesehen.

Validitätsaspekte nach Messick (1988) und an diesen anknüpfend Leuders (2014) sind z. B:

- 1) Inhaltsvalidität (z. B. Erfassung via Augenscheinvalidität, d. h. durch Sichtung von Items durch Didaktiker oder erfahrende Lehrkräfte)
- 2) Kognitive Validität (z. B. Erfordernis ökologisch valider, d. h. auch im Unterricht hervorgerufener Kognitionen zur Bearbeitung der Items)
- 3) Strukturelle Validität (z. B. durch faktoranalytische Überprüfung der klassischen Dreiteilung in die Kernkategorien FW, FDW und PW)
- 4) Verallgemeinerbarkeit (z. B. Unabhängigkeit von der Stichprobe, aber auch von speziellen Items)
- 5) Externe Validität (z. B. konvergente, diskriminante und prädiktive Validierung durch ‚erwartungskonforme‘ Zusammenhänge und Differenzen)
- 6) Konsequentielle Validität (z. B. angemessene Nutzbarkeit der Ergebnisse in pädagogischen und bildungspolitischen Bereichen)

Interessanterweise steht in Arbeiten zu Videovignettentests oft die kognitive Validität im Vordergrund, während bei Papier-und-Bleistift-Tests häufig die externe Validität ermittelt wird. Untersuchungen zum direkten Vergleich der prädiktiven Validität beider Formate liegen bislang lediglich in Einzelfällen vor (z. B. kann durch den Papier-und-Bleistift-Test zum FDW bei COACTIV die Unterrichtsqualität der untersuchten Lehrkräfte etwas besser vorhergesagt werden als mit dem Videovignettentest zu fachdidaktischen Kompetenzen). Auch gibt es bislang kaum systematische Vergleiche zu Faktoren jenseits reiner Darbietungseffekte der Stimuli, beispielsweise ob eventuell eher das *Antwortformat* die entscheidende Rolle spielt. Denkbar ist z. B., dass das Entwickeln spontaner Handlungsoptionen eher durch offene Antwortformate, idealerweise inklusive der Instruktion, möglichst viele unterschiedliche Alternativen anzugeben, abgebildet werden kann.

Einstweilen ist somit festzuhalten, dass bei Testkonstruktionen jeweils genau überlegt werden sollte, ob durch das Videografieren einer Situation tatsächlich substantiell andere Antworten zu erwarten sind als bei deren schriftlicher Schilderung. Zu erwägen ist künftig gewiss auch eine theoriebasierte Kombination von Papier-und-Bleistift-Items und Videoformaten.

Literatur

- Ball, D. L., Hill, H. H. & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator*, 29 (3), 14–46.
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. J. (2015). Beyond Dichotomies. Competence Viewed as a Continuum. *Zeitschrift für Psychologie*, 223 (1), 3–13.
- Blömeke, S., Kaiser, G. & Lehmann, R. (Hrsg.) (2008). *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten*

- deutscher Mathematikstudierender und -referendare. Erste Ergebnisse zur Wirksamkeit der Lehrerausbildung.* Münster: Waxmann.
- Bruckmaier, G., Krauss, S., Blum, W. & Leiss, D. (2016). Measuring mathematics teachers' professional competence by using video clips (COACTIV video). *ZDM Mathematics Education*, 48 (1+2), 111–124.
- Cauet, E., Liepertz, S., Borowski, A. & Fischer, H. E. (2015). Does it matter what we measure? Domain-specific professional knowledge of physics. *Revue suisse des sciences de l'éducation*, 37 (3), 462–479.
- 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.
- Knievel, I., Lindmeier, A. & Heinze, A. (2015). Beyond knowledge: Measuring primary teachers' subject-specific competences in and for teaching mathematics with items based on video vignettes. *International Journal of Science and Mathematics Education*, 13 (2), 309–329.
- König, J. & Lebens, M. (2012). Classroom Management Expertise (CME) von Lehrkräften messen: Überlegungen zur Testung mithilfe von Videovignetten und erste empirische Befunde. *Lehrerbildung auf dem Prüfstand*, 5 (1), 3–29.
- Krauss, S., Lindl, A., Schilcher, A., Fricke, M., Göhring, A., Hofmann, B., Kirchhoff, P. & Mulder, R. H. (Hrsg.) (2017). *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik* (S. 381–438). Münster: Waxmann.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.) (2011). *Professionelle Kompetenz von Lehrkräften, Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Leuders, T. (2014). Modellierungen mathematischer Kompetenzen – Kriterien für eine Validitätsprüfung aus fachdidaktischer Sicht. *Journal für Mathematik-Didaktik*, 35 (1), 7–48.
- 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.
- Messick, S. (1988). The once and future issues of validity. Assessing the meaning and consequences of measurement. In H. Wainer & H. Braun (Hrsg.), *Test validity* (S. 3–45). Hillsdale: Lawrence Erlbaum.
- Neuweg, G. H. (2015). Kontextualisierte Kompetenzmessung: Eine Bilanz zu aktuellen Konzeptionen und forschungsmethodischen Zugängen. *Zeitschrift für Pädagogik*, 61 (3), 377–383.
- Rutsch, J., Rehm, M., Vogel, M., Seidenfuß, M. & Dörfler, T. (Hrsg.), *Effektive Kompetenzdiagnose in der Lehrerbildung*. Wiesbaden: Springer.
- Vogelsang, C. & Reinhold, P. (2013). Zur Handlungsvalidität von Tests zum professionellen Wissen von Lehrkräften. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 103–128.
- Vogelsang, C. (2014). *Validierung eines Instruments zur Erfassung der professionellen Handlungskompetenz von (angehenden) Physiklehrkräften*. Berlin: Logos.