

Christoph Kulgemeyer¹
 Andreas Borowski²
 Josef Riese³
 Nico Schreiber⁴
 Christoph Vogelsang⁵

¹Universität Bremen
²Universität Potsdam
³RWTH Aachen
⁴Universität Duisburg-Essen
⁵Universität Paderborn

Performanztests in der naturwissenschaftlichen Lehrerbildung

In den letzten Jahren wurden sogenannte Performanztests in der empirischen Lehr-Lern-Forschung entwickelt. Die Bedeutung solcher Testformate für die Naturwissenschaftsdidaktik wurde in einem Workshop diskutiert.

Wissen und Handeln von Lehrkräften: ein Messproblem

Die Forschung zum professionellen Wissen von Lehrkräften hat in der Naturwissenschaftsdidaktik eine lange Tradition. Im Anschluss an Shulmans (1986) Überlegungen zur Struktur von Professionswissen hat sich die Forschung auf die drei Bereiche Fachwissen, fachdidaktisches Wissen und pädagogisches Wissen konzentriert. Es zeigte sich, dass es gelingt, Tests zur Erfassung des Fachwissens und des fachdidaktischen Wissens zu entwickeln, die hohe Standards curricularer Validität erfüllen und dazu dienen können, mindestens in quasi-längsschnittlichen Designs Wissenszuwächse transparent zu machen. Studierende am Ende des Studiums schneiden in solchen Tests z.B. besser ab als Studierende der Anfangssemester (Riese & Reinhold, 2012). Curriculare Validität ist ein geeigneter Bezugspunkt um Wissensentwicklungen in der Lehrerbildung transparent zu machen. Es ist jedoch eine offene Frage, ob die Inhalte dieser Tests tatsächlich auch bestimmend für qualitativ hochwertiges Handeln im Lehrerberuf sind (Vogelsang & Reinhold, 2013). Dies ist aus zweierlei Gründen ein bedeutendes Forschungsdesiderat.

1. Zum einen ist es für evidenzgestützte Lehrerbildung essentiell, zu verstehen, ob und welche der traditionellen Inhalte z.B. der Fachdidaktikseminare für das berufliche Handeln relevantes Wissen darstellen. Eine empirische Disziplin wie die Physikdidaktik kann sich bei der Rechtfertigung ihrer Ausbildungsinhalte nicht mit tradiertem Wissen begnügen.
2. Zum anderen wird professionelles Wissen darüber konzeptualisiert, ob es dazu nützt, eine Profession auszuüben oder darüber zu kommunizieren. Kann der Nachweis dieses Zusammenhangs nicht gelingen, so wäre z.B. fachdidaktisches Wissen kein Professionswissen – es müsste nach anderen Ressourcen für qualitativvolles Handeln geforscht werden.

In den letzten Jahren ist in einigen groß angelegten Forschungsvorhaben versucht worden, den Nachweis des Einflusses von Fachwissen und fachdidaktischem Wissen auf Unterrichtshandeln zu führen. Die Ergebnisse sind jedoch uneinheitlich. Die COACTIV-Studie (Competence of Teachers, Cognitively Activating Instruction and Development of Students' Mathematical Literacy) beschreibt einen durch das fachdidaktische Wissen medierten Einfluss des Fachwissens auf kognitiv aktivierenden Unterricht (Baumert et al., 2010). ProwiN (Professionswissen in den Naturwissenschaften) konnte weder eine Korrelation von Fachwissen noch von fachdidaktischem Wissen und kognitiv anregendem Unterricht nachweisen (Cauet et al., 2015). Im Kontext von QuiP (Quality of Instruction in Physics) hingegen wurde zumindest eine geringe positive Korrelation zwischen fachdidaktischem Wissen und Lernfortschritt der Schülerinnen und Schüler beschrieben (Ergönenc, Neumann & Fischer, 2014). In all diesen Studien wurde auf die curriculare Validität der eingesetzten Tests zum Professionswissen geachtet. Die Ergebnisse dürfen

jedoch nicht so interpretiert werden, dass Professionswissen, wie es in der Lehrerausbildung vermittelt wird, einen unklaren Einfluss auf Lehrerhandeln hat. Es ist vielmehr möglich, dass Einschränkungen im Studiendesign dazu führen, dass der Einfluss nicht dargestellt werden kann. In diesen Studien wurde Unterrichtsqualität durch Auswertungen von Videobeobachtungen, Ratings von Schülerinnen und Schülern oder die Analyse von Lernaufgaben realer Unterrichtsstunden gemessen. Kulgemeyer und Schecker (2017) betonen, dass jede Unterrichtsstunde Rahmenbedingungen unterworfen ist, die es teilweise unmöglich machen, dass Professionswissen zu qualitativ hochwertigem Unterricht führt. Eine Klasse ist ein komplexes soziales Gefüge: wenn z.B. Konflikte zwischen Schülerinnen und Schülern die Arbeit am Inhalt unmöglich machen, ist es für eine Lehrkraft schwierig hochqualitativen Unterricht durchzuführen, selbst wenn sie es potentiell beherrscht. Umgekehrt gibt es sicher auch Lerngruppen, mit denen auch eine Lehrkraft mit niedrigem Professionswissen in einzelnen Stunden große Fortschritte machen kann. Man kann eine ganze Reihe an solchen Rahmenbedingungen finden, die die Unterrichtsqualität stark beeinflussen. Nimmt man diese Rahmenbedingungen als zufällig an, müsste man eine große Anzahl an Stunden videographieren und auswerten, bis der evtl. kleine Effekt des Professionswissens auf Unterrichtsqualität zutage tritt. Dies ist wegen des Aufwands, den die Auswertung ganzer Stunden mit sich bringt, selbst in großen Projekten nicht möglich.

Performanztests: eine begriffliche Klärung

Performanztests stellen eine Alternative dar, die helfen könnte, den Nachweis des Einflusses von Professionswissen auf Handlungsqualität beim Unterrichten zu führen. Mit Performanztests gemeint sind dabei Verfahren, bei denen Unterrichtshandlungen unter standardisierten Rahmenbedingungen simuliert werden. Ein Beispiel dafür ist der Test von Kulgemeyer und Tomszyszyn (2015), in dem Handlungsqualität beim Erklären gemessen wird. Darin erklären angehende Lehrkräfte in einer dialogischen Situation einem Schüler unter standardisierten Bedingungen ein physikalisches Phänomen: das Thema und mögliche Darstellungsformen werden gestellt, die Zeit ist auf 10 Minuten begrenzt und der Schüler ist darauf trainiert, Fragen zu stellen, die die Erklärfähigkeit herausfordern (z.B. „Gibt es dafür auch ein Beispiel?“). Diese Situationen werden videographiert und kategorienbasiert ausgewertet. Durch die Standardisierung der Situation wird eine Vielzahl der Rahmenbedingungen konstant gehalten, die es bei der Videobeobachtung einer Unterrichtsstunde erschweren, den Effekt des Einflusses von Professionswissen auf das Unterrichtshandeln nachzuweisen. Zudem kann der Test schneller analysiert werden als ganze Stunden und dadurch mit einer höheren Anzahl an Probanden durchgeführt werden.

In Anlehnung an Miller (1990) kann man allgemein vier Testformate unterscheiden, die beim Erfassen beruflicher Fähigkeiten zum Einsatz kommen können: schriftliche Tests des deklarativen Wissens, schriftliche Kompetenztests, Performanztests und die Beobachtungen freien beruflichen Handelns. In einem schriftlichen Kompetenztest wird die Anwendung von Wissen beim Lösen beruflicher Probleme gemessen; der Begriff „Kompetenz“ bezieht sich auf das zu testende Konstrukt. Das Testformat ist ein schriftlicher Paper-and-pencil-Test. Dies ist beim Begriff „Performanztest“ grundlegend anders: hier ist mit dem Betonen der „Performanz“ die Handlungsnähe des Testverfahrens gemeint – also das Testformat, nicht das zu testende Konstrukt. Kompetenz kann verstanden werden als (latentes) Konstrukt, das man nur in der (manifesten) Performanz beobachten kann. Die Performanz bei Kompetenztests ist das beobachtbare Verhalten beim Lösen des Tests. In Performanztests ist die Performanz näher am freien beruflichen Handeln: beim Erklärtest zum Beispiel ist das Handeln beim Erklären gemeint. Die Authentizität der simulierten Handlung ist ein wichtiges Qualitätsmerkmal eines Performanztests. Ausgewertet wird in Performanztests die Handlungsqualität – ob diese ein Maß für Kompetenz ist, kann diskutiert werden.

Kompetenz befähigt zum erfolgreichen Handeln in *verschiedenen* Situationen. In Performanztests wird nur *eine Situation* des beruflichen Handelns simuliert. In Bezug auf Lehrerbildung kann man Fachwissen und fachdidaktisches Wissen als zentrale Elemente professioneller Handlungskompetenz formulieren: sie befähigen dann zum hochqualitativen Handeln in verschiedenen beruflichen Situationen, vom Anleiten von Schülerexperimenten bis hin zum Erklären. Aus der Handlungsqualität beim Erklären allein sollte nicht auf Fachwissen und fachdidaktisches Wissen zurückgeschlossen werden.

„Objective Structured Clinical Examinations“ als Alternative für die Lehrerbildung?

Eine lange Tradition haben Performanztests in der Ausbildung von Medizinern (Harden et al., 1975). Einen hohen Verbreitungsgrad haben dabei Serien kurzer Performanztests, sogenannte „Objective Structured Clinical Examinations (OSCE)“. Simuliert werden dabei verschiedene Situationen des beruflichen Handelns, bei angehenden Ärzten z.B. Untersuchungen des Bauchraums oder das Aufnehmen einer Anamnese (Rochelson et al., 1985). Wesentlicher Bestandteil sind oftmals „standardized patients“, also Patientendarsteller, die mit einer gewissen Rollenbeschreibung reproduzierbares und standardisiertes Verhalten zeigen. Dies ist sehr ähnlich zu den standardisierten Schülern, die im Test von Kulgemeyer & Tomczyszyn (2015) den Kern darstellen. Angehende Ärzte durchlaufen in einer OSCE nacheinander mehrere Stationen mit Performanztests, jede dauert etwa zwischen 5 und 15 Minuten. Ausgewertet wird in OSCEs allerdings in der Regel kein Video, sondern es befinden sich ein oder zwei Rater direkt an einer Station. Es zeigt sich, dass mit geeigneten Scoring Sheets hohe Standards der Testgüte erreicht werden können (z.B. Walters, Osborn & Raven, 2005). Es erscheint vielversprechend, dass auch für die Lehrerbildung eine solche Serie an Performanztests einen wertvollen Beitrag dazu leisten kann, den Zusammenhang von Wissen und Handeln zu untersuchen. Wenn es gelingt, eine Serie an authentischen Handlungssituationen zu konstruieren, die Standardsituationen des Lehrerhandelns repräsentieren, kann professionelle Handlungskompetenz in verschiedenen wichtigen Situationen in der Performanz beobachtet werden. Echte Rückschlüsse auf zugrundeliegende Kompetenz werden dadurch möglich. Das Projekt ProfilLe-P+ arbeitet daran, indem über den Erklärtest hinaus Performanztests für die Planung und die Reflexion von Unterricht entwickelt werden (siehe u.a. Beitrag von Vogelsang et al. in diesem Band).

Wesentliche Nachteile sind ein immer noch hoher Aufwand in der Durchführung und Auswertung im Vergleich mit schriftlichen Testverfahren sowie die Zergliederung des Lehrerhandelns in einzelne Standardsituationen. Um dem hohen Testaufwand zu begegnen, versuchen beispielsweise Bartels und Kulgemeyer (2016) Handlungssituationen beim Erklären in einem Onlinetest zu simulieren und mit geschlossenen Aufgaben zu erheben, die ihre Distraktoren und Attraktoren aus realem Verhalten beim freien Handeln gewonnen haben. Möglicherweise geht durch die Zergliederung in Standardsituationen allerdings ein holistischer Blick auf das Lehrerhandeln verloren: Performanztests – auch als Serie – können deshalb nur Ergänzung von Videostudien sein, die komplette Unterrichtsstunde analysieren (Kulgemeyer & Schecker, 2017). Eine könnten evtl. Lehr-Lern-Labore sein. Korneck, Krüger und Szogs (2017) beschreiben zum Beispiel Microteaching-Lehrveranstaltungen, in denen in zwölfminütigen Unterrichtsminiaturen Unterricht in verschiedenen Phasen geleistet und analysiert wird. Wegen der hohen Relevanz der Frage, ob in der Ausbildung zur Physiklehrkraft vermitteltes Wissen tatsächlich bei der Bewältigung beruflicher Handlungssituationen hilft, erscheint es sinnvoll zu sein, verschiedene Wege zu gehen: von Videostudien über Unterrichtsminiaturen über Performanztests. Insbesondere wenn sich in verschiedenen Verfahren dieselben Facetten der Ausbildungsinhalte als relevant erwiesen, wäre dies ein besonders überzeugendes Argument.

Im Workshop wurde deutlich, dass die Teilnehmenden Performanztests generell als Format begrüßen, aber die simulierten Situationen möglicherweise eher als Lerngelegenheiten als nutzen wollen. Ein möglicher nächster Schritt wäre das Spezifizieren von „Standardsituationen des Physikunterrichts“, die als Performanztests nachgestellt werden.

Literatur

- Bartels, H. & Kulgemeyer, C. (2016). Entwicklung eines computerbasierten Testinstruments für Erklärfähigkeit. *PhyDid B - Didaktik der Physik - Beiträge zur DPG Frühjahrstagung*.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M. & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 133–180.
- Cauet, E., Liepertz, S., Kirschner, S., Borowski, A., & Fischer, H. E. (2015). Does it Matter What We Measure? Domain-specific Professional Knowledge of Physics Teachers. *Revue suisse des sciences de l'éducation*, 37(3), 462–479.
- Ergönenc, J., Neumann, K. & Fischer, H. (2014). The Impact of Pedagogical Content Knowledge on Cognitive Activation and Student Learning. In H. E. Fischer, P. Labudde, K. Neumann & J. Viiri (Hrsg.), *Quality of Instruction in Physics. Comparing Finland, Germany and Switzerland* (pp. 145–160). Münster: Waxmann.
- Harden, R., Stevenson, M., Wilson, W. & G. W. (1975). Assessment of Clinical Competence using Objective Structured Examination. *British Medical Journal*, 1, 447–451.
- Korneck, F., Krüger, M. & Szogs, M. (2017). Professionswissen, Lehrerüberzeugungen und Unterrichtsqualität angehender Physiklehrkräfte unterschiedlicher Schulformen. In H. Fischler & E. Sumfleth (Hrsg.), *Professionelle Kompetenz von Lehrkräften der Chemie und Physik* (pp. 77-96). Berlin: Logos.
- Kulgemeyer, C. & Tomczyszyn, E. (2015). Physik erklären. Messung der Erklärens-fähigkeit angehender Physiklehrkräfte in einer simulierten Unterrichtssituation. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 111–126.
- Kulgemeyer, C. & Schecker, H. (2017). Handlungsqualität beim Erklären von Physik: Performanztests in der Lehrerbildung. In H. Fischler & E. Sumfleth (Hrsg.), *Professionelle Kompetenz von Lehrkräften der Chemie und Physik* (pp. 97-112). Berlin: Logos.
- Miller, G. (1990). The Assessment of Clinical Skills / Competence / Performance. *Journal of the Association of American Medical Colleges*, 65(9), 63–67.
- Riese, J. & Reinhold, P. (2012). Die professionelle Kompetenzn angehender Physiklehrkräfte in verschiedenen Ausbildungsformen – Empirische Hinweise für eine Verbesserung des Lehramtsstudiums. *Zeitschrift für Erziehungswissenschaft*, 15., S. 111-143.
- Rochelson, B., Baker, D., Mann, W., Monheit, A. & Stone, M. (1985). Use of male and female professional patient teams in teaching physical examination of the genitalia. *The Journal of Reproductive Medicine*, 30(11), 864–866.
- Shulman, L. (1986): Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15(2), 4-14.
- Vogelsang, C. & Reinhold, P. (2013). Zur Handlungsvalidität von Tests zum professionellen Wissen von Lehrkräften. *Zeitschrift für Didaktik der Naturwissenschaften* 19, 129-157.
- Walters, K., Osborn, D. & Raven, P. (2005). The development, validity and reliability of a multimodality objective structured clinical examination in psychiatry. *Medical Education*, 39, 292–298.