

Eva Cauet<sup>1</sup>  
 Andreas Borowski<sup>2</sup>  
 Hans E. Fischer<sup>3</sup>

<sup>1</sup>Universität Koblenz-Landau  
<sup>2</sup>Universität Potsdam  
<sup>3</sup>Universität Duisburg-Essen

## **Professionswissen und Unterrichtsqualität: Wie aussagekräftig sind unsere Ergebnisse?**

### **Zielsetzung und Ausgangspunkt**

In den vergangenen Jahren wurden zahlreiche Tests zur Erhebung des Professionswissens von Physiklehrkräften entwickelt, das als wichtige Voraussetzung für gutes und erfolgreiches Unterrichten gilt (z.B. Brovelli, Bölsterli, Rehm & Wilhelm, 2013; Kirschner, 2013; Kröger, Neumann & Petersen, 2015; Riese, 2009; Riese et al., 2015). Dabei werden die drei Professionswissensdimensionen Fachwissen (CK), fachdidaktisches Wissen (PCK) und pädagogisches Wissen (PK) allerdings unterschiedlich modelliert (vgl. z.B. Kirschner et al., 2016, S.1346). Obwohl belastbare empirische Evidenz für die Handlungsrelevanz des explizierbaren Professionswissens noch aussteht, werden derartige Tests oft mit dem Ziel eingesetzt, Aussagen über die Wirksamkeit der Lehrerausbildung zu treffen (vgl. z.B. Riese, 2009, S.11). Solange die prädiktive Validität der Tests für gutes oder erfolgreiches Unterrichten nicht nachgewiesen wird, ist die Validität solcher Aussagen allerdings nicht sichergestellt. Wenige Studien untersuchen Zusammenhänge zwischen Professionswissen, Unterrichtsqualität und Unterrichtserfolg (Ohle, 2010; Ergönenç, Neumann & Fischer, 2014; Vogelsang, 2014, Sadler, Sonnert, Coyle, Cook-Smith & Miller, 2013). Zudem ist die Interpretation der heterogenen Ergebnisse dieser Studien mit zahlreichen Unsicherheiten verbunden. Am Beispiel einer Teilstudie aus dem Projekt „Professionswissen in den Naturwissenschaften“ (ProwiN) (Borowski et al., 2010) im Fach Physik sollen die Probleme, die sich bei der Ergebnisinterpretation ergeben, dargestellt werden.

### **Kurzbeschreibung der ProwiN-Physik Teilstudie und ihrer Ergebnisse**

In der Studie von Cauet (2016) wurde untersucht, ob die von Kirschner (2013) entwickelten ProwiN CK- und PCK-Tests relevantes Wissen für gutes und erfolgreiches Unterrichten erfassen. Hierfür wurde in einer Stichprobe von 23 Gymnasiallehrkräften aus NRW (35% weiblich,  $M_{\text{Alter}}=44$  Jahre,  $SD_{\text{Alter}}=12$  Jahre) und ihren Klassen der Jahrgangsstufe 8/9 ( $N=610$  Lernende, 57% weiblich,  $M_{\text{Alter}}=14$  Jahre,  $SD_{\text{Alter}}=1$  Jahr) im Rahmen eines Prä-Post-Designs zunächst geprüft, ob die CK- bzw. PCK-Testwerte der Lehrkräfte (CK: 11 Aufgaben zum Schulwissen/vertieften Schulwissen in Mechanik, Rasch Personen Rel.=.73; PCK: 10 Aufgaben zu Schülervorstellungen, Umgang mit Experimenten und Konzepten in der Mechanik, Rasch Personen Rel.=.59) einen Prädiktor für Unterrichtserfolg darstellen. Dieser wurde über die Leistungen der Lernenden in einem Fachwissenstest (34 Multiple-Choice Aufgaben zur Mechanik, Rasch Personen Rel. Prä-/Post=.51/.61) am Ende einer mehrmonatigen Unterrichtseinheit zur Mechanik (Unterrichtszeit in Anzahl an 45-min Stunden:  $M=34$ ,  $SD=10$ ,  $Min=12$ ,  $Max=59$ ) operationalisiert. Die zwischen den Klassen liegende Varianz in den Post-Testwerten der Lernenden ( $ICC_{(1,1)}= 10\%$  bzw. 4% nach Kontrolle von Vorwissen, kognitiven Fähigkeiten, Geschlecht und Sprache) wurde lediglich durch die Unterrichtszeit, nicht aber durch das CK oder PCK der Lehrkräfte erklärt (Tabelle 1).

Vor dem Hintergrund der Erfolgsunsicherheit des Lehrerhandelns - eine Lehrkraft kann lediglich Lehrangebote zur Verfügung stellen, ob sie von den Lernenden genutzt werden, kann sie nur bedingt beeinflussen (Baumert & Kunter, 2006, S.476-477) – wurde außerdem

Prädiktoren für Post-Testwerte		Kontroll-Modell	CK-Modell	PCK-Modell
Schülerebene (N=610)		Prä-Test, KFT, Geschlecht & Sprache erklären <b>R<sup>2</sup>= (34 ± 3)%</b> der Varianz innerhalb der Klassen		
Klassenebene (N=23)				
Unterrichtszeit	$\gamma^{\text{StdYX}}$	<b>0.80 ± 0.11</b>	<b>0.81 ± 0.11</b>	<b>0.81 ± 0.11</b>
	KI <sub>95%</sub>	[0.60,1.01]	[0.61,1.00]	[0.60,1.02]
Zusätzlicher Prädiktor (CK/PCK)	$\gamma^{\text{StdYX}}$		0.07 ± 0.19	-0.15 ± 0.16
	KI <sub>95%</sub>		[-0.30,0.44]	[-0.45,0.15]
Varianzaufklärung	R <sup>2</sup>	<b>(65 ± 18)%</b>	<b>(65 ± 16)%</b>	<b>(67 ± 18)%</b>

Legende:  $\gamma^{\text{StdYX}}$  = vollstandardisierter Steigungskoeffizient; fettgedruckt= signifikant mit  $p_{1\text{-seitig}} < .001$

Tab. 1: Mehrebenenregressionen auf die Post-Testwerte der Lernenden.

untersucht, inwieweit die Qualität des Unterrichts der Lehrkräfte - beurteilt über die kognitiv aktivierende Gestaltung von zwei videographierten Unterrichtsstunden innerhalb der Unterrichtseinheit zur Mechanik - mit ihrem CK bzw. PCK zusammenhängt. Die kognitive Aktivierung wurde mit einem Ratingmanual (adaptiert nach Vogelsang, 2014) über die Bewertung von 29 Handlungsindikatoren in 7 Subskalen (Bewusstmachen des Lernstatus, Exploration des Vorwissens und der Vorstellungen, Exploration der Denkweisen, Evolutionärer Umgang mit Schülervorstellungen, Lehrkraft als Mediator, Rezeptives Lernverständnis (-), herausfordernde Lerngelegenheiten) auf einer 3-stufigen Likertskala eingeschätzt ( $\alpha_{C,1.\text{Stunde}/2.\text{Stunde}}=.91/.87$ ;  $ICC_{(2,1)unjust,1.\text{Stunde}/2.\text{Stunde}}=.64/.69$ , Subskalen:  $.21 < ICC_{(2,1)unjust} < .71$ ). Cauet (2016, S.176) konnte zeigen, dass das über beide Unterrichtsstunden gemittelte Maß für kognitive Aktivierung ein signifikanter Prädiktor für die Post-Testwerte der Lernenden ist. Es wird daher davon ausgegangen, dass das Rating ein Merkmal der Unterrichtsqualität erfasst. Theoretischen Überlegungen folgend (vgl. Cauet, 2016, S.69-73) würde man erwarten, dass die kognitiv aktivierende Gestaltung des Unterrichts stärker vom PCK als vom CK der Lehrkräfte abhängen sollte. Korrelationsanalysen zeigten jedoch lediglich signifikante Korrelationen zwischen CK und dem über beide Unterrichtsstunden gemittelten Maß für kognitive Aktivierung ( $r=.36 \pm 0.19$ ,  $p_{1\text{-seitig}}=.044$ ), aber nicht zwischen PCK und kognitiver Aktivierung ( $r=.21 \pm 0.19$ ,  $p_{1\text{-seitig}}=.165$ ). Im Rahmen von ProwiN wurde auch das PK der Lehrkräfte untersucht: Es zeigten sich positive Zusammenhänge mittlerer Effektstärke sowohl zum Unterrichtserfolg als auch zu Aspekten der Unterrichtsqualität (Klassenführung und kognitive Aktivierung) (Lenske et al., 2015; Cauet, 2016).

### **Diskussion der Ergebnisse und ihrer Aussagekraft**

Auf eine inhaltliche Diskussion der Ergebnisse wird an dieser Stelle verzichtet und auf Cauet (2016, S.206-211) verwiesen. Vielmehr soll hier die Aussagekraft der Ergebnisse diskutiert werden. Als Hauptprobleme für eine valide Ergebnisinterpretation ergeben sich folgende Punkte:

#### *1. Untersuchung einer kleinen Gelegenheitsstichprobe*

Aufgrund der niedrigen Teststärke könnten Zusammenhänge in der vorliegenden Stichprobe „übersehen“ werden. Die Wahrscheinlichkeit Zusammenhänge der Größenordnung der gefundenen Zusammenhänge zwischen PK und Unterrichtserfolg/-qualität nachzuweisen,

liegt bei lediglich  $1-\beta=49-58\%$ . Ein Vergleich mit Daten aus ProwiN I zeigt zudem, dass die Stichprobe der Lehrkräfte eine Positivauswahl bzgl. des Fachwissens darstellt ( $t_{\text{Mann-Whitney}}=-.18$ ,  $p_{1\text{-seitig}}=.039$ ). Da in Extremgruppen im Vergleich zur Grundgesamtheit Zusammenhänge zwischen Merkmalen unterschiedlich stark ausgeprägt sein können, kann es daher zu einer Unterschätzung des Zusammenhangs zwischen CK und Unterrichtserfolg kommen (Bortz & Döring, 2006, S. 509).

### 2. Beschränkung auf die Untersuchung korrelativer Zusammenhänge

Das Design der ProwiN-Studie ermöglicht keine Untersuchung kausal bedingter Zusammenhänge, der Einfluss von Störvariablen kann daher nicht ausgeschlossen werden. Zum einen kann nicht sichergestellt werden, dass beobachtete Zusammenhänge nicht lediglich Scheinzusammenhänge darstellen, die auf die Existenz konfundierender Variablen zurückzuführen sind. Zum anderen könnten möglicherweise vorhandene Effekte des fachspezifischen Professionswissen auf Unterrichtserfolg erst unter Kontrolle weiterer Variablen, wie z.B. Klassenführung, sichtbar werden.

### 3. Einsatz unpräziser Messinstrumente

Ein weiteres Problem stellen die teils erheblichen Messungenauigkeiten der Testinstrumente dar (siehe Reliabilität des PCK-Tests und des Schülerfachwissenstests). In Korrelationsanalysen führen diese meist zu einer Unterschätzung von Zusammenhängen, während Messungenauigkeiten in Prädiktorvariablen in Mehrebenenanalysen sowohl zu einer Über- als auch zu einer Unterschätzung von Zusammenhängen führen können (vgl. z. B. Kromrey et al., 2006; Woodhouse et al., 1996).

Die angesprochenen Probleme stellen keine Einzelfälle dar, sondern erschweren auch in den anderen bisher durchgeführten Studien zum Zusammenhang zwischen dem mit schriftlichen Tests erhobenen Professionswissen von Physiklehrkräften, Unterrichtsqualität und/oder Unterrichtserfolg eine valide Interpretation der Ergebnisse. So standen in der Studie von Vogelsang (2014) zur prädiktiven Validität des Paderborner Professionswissenstests oder in der Studie „Professionswissen von Lehrkräften, naturwissenschaftlicher Unterricht und Zielerreichung im Übergang von der Primar- zur Sekundarstufe“ (PLUS) zum Teil ähnlich kleine Stichproben zur Verfügung ( $N=22$  bzw.  $N=30$ ) (Ohle, 2010). Die PLUS-Studie konnte zudem zeigen, dass Zusammenhänge zwischen PCK und Schülerleistungen erst unter Kontrolle der Klassenführung nachweisbar waren (Lange, 2010, S.168). Daten zur Klassenführung werden allerdings in anderen Studien meist nicht erhoben oder können (wie im Fall von ProwiN) aufgrund der Stichprobengröße nicht statistisch kontrolliert werden. Geringe Reliabilitäten von Schülerprätests könnten auch in der Studie „Quality of Instruction in Physics“ (QuiP) ( $Rel.=.51$ , Geller, 2015, S.96) oder in einer Large-Scale Untersuchung zu Zusammenhängen zwischen Professionswissen und Schülerleistungen von Sadler et al. (2013, S.1031) ( $Rel.=.53$ ) einen Einfluss auf die Ergebnisse haben.

### Fazit

Die Ergebnisse bisheriger empirischer Studien in der Physik sind nicht belastbar genug, um eindeutige Aussagen über Handlungsrelevanz des gemessenen Wissens zu treffen. Für einen kumulativen Erkenntnisgewinn bezüglich der Bedeutsamkeit des Professionswissens müssten - um die Vergleichbarkeit verschiedener Studien zu erleichtern - die Professionswissensdimensionen einheitlicher modelliert werden, es müsste mehr Arbeit in „Hilfsinstrumente“ wie z.B. Schülertests investiert werden – schließlich ist eine Argumentationskette lediglich so stark wie ihr schwächstes Glied – und die Ergebnisse bisheriger Studien müssten in größeren Stichproben überprüft und repliziert werden.

### Literatur

- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Borowski, A., Neuhaus, B. J., Tepner, O., Wirth, J., Fischer, H. E., Leutner, D., ... Sumfleth, E. (2010). Professionswissen von Lehrkräften in den Naturwissenschaften (ProwiN) - Kurzdarstellung des BMBF-Projekts. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 341–349.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer.
- 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(4), 306–329.
- Cauet, E. (2016). Testen wir relevantes Wissen? Zusammenhang zwischen dem Professionswissen von Physiklehrkräften und gutem und erfolgreichem Unterrichten. *Studien zum Physik und Chemielernen*. Berlin: Logos.
- Ergönenç, J., Neumann, K. & Fischer, H. E. (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: Quality of Instruction in Physics* (S. 13–30). Münster u.a.: Waxmann.
- Geller, C. (2015). Lernprozessorientierte Sequenzierung des Physikunterrichts im Zusammenhang mit Fachwissenserwerb: Eine Videostudie in Finnland, Deutschland und der Schweiz. *Studien zum Physik- und Chemielernen*. Berlin: Logos.
- Kirschner, S. (2013). Modellierung und Analyse des Professionswissens von Physiklehrkräften. *Studien zum Physik- und Chemielernen*. Berlin: Logos.
- Kirschner, S., Borowski, A., Fischer, H. E., Gess-Newsome, J. & von Aufschnaiter, C. (2016). Developing and Evaluating a Paper-and-Pencil Test to Assess Components of Physics Teachers' Pedagogical Content Knowledge. *International Journal of Science Education*, 38(8), 1343-1372.
- Kröger, J., Neumann, K. & Petersen, S. (2015). Struktur und Entwicklung des Professionswissens angehender Physiklehrkräfte. In S. Bernholt (Hrsg.), *Heterogenität und Diversität - Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht: Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Bremen 2014* (S. 106–108). Kiel: IPN.
- Kromrey, J. D., Coraggio, J. T., Phan, H. T., Romano, J. L., Hess, M. R., Lee, R. S., ... Luther, S. L. (2006). The Impact of Measurement Error in Predictor Variables in Multilevel Models: An Empirical Investigation of Statistical Bias and Sampling Error. Paper presented at the annual meeting of the Florida Educational Research Association, 2006, Jacksonville. Zugriff unter <http://www.coedu.usf.edu/main/departments/me/documents/theimpactofmeasurementerrorinpredictorvariablesinhierarchicallinearmodelsfera2006.pdf>
- Lange, K. (2010). Zusammenhänge zwischen naturwissenschaftsbezogenem fachspezifisch-pädagogischem Wissen von Grundschullehrkräften und Fortschritten im Verständnis naturwissenschaftlicher Konzepte bei Grundschülerinnen und -schülern. Münster: *Didaktik des Sachunterrichts*. Zugriff unter <http://nbnresolving.de/urn:nbn:de:hbz:6-75459654103>
- 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. Zugriff unter <http://dx.doi.org/10.1007/s11618-015-0659-x>
- Ohle, A. (2010). Primary school teachers' content knowledge in physics and its impact on teaching and students' achievement. *Studien zum Physik- und Chemielernen*. Berlin: Logos.
- Riese, J. (2009). Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften. *Studien zum Physik- und Chemielernen*. Berlin: Logos.
- Riese, J., Kulgemeyer, C., Borowski, A., Fischer, H., Gramzow, Y., Reinhold, P., ... Zander, S. (2015). Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik. In S. Blömeke & O. Zlatkin-Troitschanskaia (Hrsg.), *Kompetenzen von Studierenden* (S. 55–79). *Zeitschrift für Pädagogik*. Beiheft. 61. Weinheim u.a.: Beltz.
- Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N. & Miller, J. L. (2013). The Influence of Teachers' Knowledge on Student Learning in Middle School Physical Science Classrooms. *American Educational Research Journal*, 50(5), 1020–1049.
- Vogelsang, C. (2014). Validierung eines Instruments zur Erfassung der professionellen Handlungskompetenz von (angehenden) Physiklehrkräften: Zusammenhangsanalysen zwischen Lehrerkompetenz und Lehrerperformanz. *Studien zum Physik- und Chemielernen*. Berlin: Logos.
- Woodhouse, G., Yang, M., Goldstein, H. & Rasbash, J. (1996). Adjusting for measurement error in multilevel analysis. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 201–212.