

Melanie Jordans¹
Jannis Zeller¹
Rike Große-Heilmann¹
Josef Riese¹

¹RWTH Aachen University

Weiterentwicklung eines physikdidaktischen Tests zum Online-Assessment

Ausgangslage und Ziele des Projekts

In den letzten Jahren wurde das Professionswissen von angehenden Lehrkräften insbesondere im Fach Physik durch verschiedene Projekte der Lehrerbildungsforschung intensiv untersucht (z.B. ProwiN: Tepner et al., 2012; KiL: Kröger, Neumann & Petersen, 2013; ProfiLe-P: Riese et al., 2015). Es zeigte sich, dass der Erwerb handlungsnaher Fähigkeiten (z.B. Erklären physikalischer Sachverhalte: Kulgemeyer et al., 2020; Reflektieren von Physikunterricht: Kulgemeyer, Kempin & Weißbach, 2021; Planen von Physikunterricht: Riese et al., eingereicht) angehender Physiklehrkräfte in Praxisphasen durch den Stand ihres Professionswissens zu Beginn einer Praxisphase tendenziell beeinflusst wird. Dabei kann das fachdidaktische Wissen (FDW) in Anlehnung an Shulman (1986) neben dem Fachwissen und dem pädagogischen Wissen in Strukturmodellen zur professionellen Handlungskompetenz als einer der drei zentralen Bereiche des Professionswissens betrachtet werden (z.B. Baumert & Kunter, 2006). Vor diesem Hintergrund können eine Diagnostik sowie Rückmeldung zum Stand des fachdidaktischen Wissens an Lehrende in Praxisphasen (z.B. für die Adaption der Lehre an eine Lerngruppe) oder an Studierende (z.B. zu Reflexionszwecken) wichtig sein.

Im Rahmen dieses Teilprojekts von ProfiLe-P-Transfer wurde ein erprobter fachdidaktischer Leistungstest zu einem vollständig geschlossenen, online-gestützten Format weiterentwickelt, um damit (teil-)automatisierte Rückmeldungen zum Stand des FDW an Dozierende und Studierende bereitstellen zu können. Langfristig ist ein nachhaltiger Transfer des entwickelten Testinstruments in die Lehrpraxis der Ausbildung von Physiklehrkräften geplant.

Weiterentwicklung des Testinstruments

Ausgangspunkt ist der Leistungstest zum fachdidaktischen Wissen im Bereich Mechanik, der von Gramzow (2015) entwickelt und in den ersten beiden Projektphasen von ProfiLe-P/ P+ (vgl. Vogelsang et al., 2019) über viele Jahre eingesetzt wurde. Die Aufgabenkonstruktion erfolgte auf Grundlage eines Itementwicklungsmodells, welches das FDW über die drei Dimensionen *kognitive Anforderung*, *Facette* und *Inhaltsbereich* beschreibt (vgl. Gramzow, 2015). Jede der 26 Aufgaben (davon 22 offen) bzw. jedes der 43 Einzelitems des Testinstruments lässt sich somit genau einer der drei kognitiven Anforderungen (Reproduzieren, Anwenden, Analysieren), einer der vier Facetten (Instruktionsstrategien, Schülervorstellungen, Experimente, fachdidaktische Konzepte) und dem Inhaltsbereich Mechanik im dreidimensionalen Modell zuordnen (vgl. Riese, Gramzow & Reinhold, 2017). Diese Version (im Folgenden als offenes Testinstrument bezeichnet) liegt als paper-pencil-Test vor und erfordert eine manuelle Bepunktung der offenen Aufgaben mittels eines Kodiermanuals durch die Entwicklergruppe, da zur Beantwortung offener Aufgaben frei formulierte Stichworte oder Kurztexte von den Teilnehmenden erwartet werden.

Um das langfristige Ziel eines nachhaltigen Transfers des Testinstruments in die Lehrpraxis erreichen zu können, wird eine Implementation als Online-Test mit (teil-)automatisierter, zeitökonomischer Auswertung angestrebt. Dazu wurden in einem ersten Schritt die offenen

Aufgaben unter Nutzung und Kategorisierung der in bisherigen Erhebungen erhaltenen frei formulierten Antworten sowie unter Berücksichtigung des Kodiermanuals und aktueller physikdidaktischer Literatur zu Multiple-Choice-Aufgaben (Multi-Select) modifiziert. Es ergaben sich im Gesamten 157 Einzelitems, welche in 31 Aufgabeteile zusammengefasst sind. Die Aufgabenstämme und Fragestellungen konnten bis auf wenige sprachliche Anpassungen aus dem offenen Testinstrument übernommen werden. Das ursprünglich offene Antwortformat wurde in ein Matrix-Design überführt, welches die Antwortalternativen und Flächen zum Ankreuzen enthält. Im Anschluss erfolgte die Digitalisierung aller Aufgaben unter Nutzung der Freeware LimeSurvey. In Abb. 1 wird eine Aufgabe der Facette *Schülervorstellungen* und der kognitiven Anforderung *Reproduzieren* des online-gestützten, geschlossenen Formats dargestellt.

Aufgabe 6

| Handelt es sich bei den folgenden Aussagen um typische, fachlich nicht korrekte Vorstellungen zu den physikalischen Konzepten der Geschwindigkeit und Beschleunigung? | | |
|---|-----------------------|-----------------------|
| | ja | nein |
| Geschwindigkeit und Beschleunigung zeigen immer in dieselbe Richtung. | <input type="radio"/> | <input type="radio"/> |
| Für einen bestimmten Zeitpunkt kann man keine Geschwindigkeit bestimmen. | <input type="radio"/> | <input type="radio"/> |
| Beschleunigung ist nur eine Erhöhung der Geschwindigkeit. | <input type="radio"/> | <input type="radio"/> |
| Geschwindigkeit und Beschleunigung sind proportional zueinander. | <input type="radio"/> | <input type="radio"/> |
| Geschwindigkeit ist eine Eigenschaft eines Körpers und kein Zustand. | <input type="radio"/> | <input type="radio"/> |

Abb. 1: Darstellung von Aufgabe 6 der Facette *Schülervorstellungen* und der kognitiven Anforderung *Reproduzieren* des online-gestützten, geschlossenen Testinstruments.

Zur Bewertung der geschlossenen Aufgaben wurden neue Kodierregeln festgelegt, bei denen es sich um eine an *Kprim* angelehnte Schwellenbepunktung (0, 1 oder 2 Punkt/e) handelt (z.B. Krebs, 2004), welche eine Reduzierung des durch Raten erreichbaren Scores zur Folge hat. Für die in Abb. 1 dargestellte Aufgabe ergibt sich somit zum Beispiel, dass bei weniger als drei richtig gesetzten Kreuzen 0 Punkte, bei genau drei richtigen 1 Punkt und bei mehr als drei richtigen Kreuzen 2 Punkte erreicht werden. Erste Ergebnisse einer Pilotierung des geschlossenen Testinstruments mit $N = 133$ Lehramtsstudierenden ergeben Aufgabenschwierigkeiten zwischen 0,21 und 0,72, was nach Bortz & Döring (2006) einem annehmbaren Bereich entspricht. Für die interne Konsistenz und somit die Reliabilität der Gesamtskala ergibt sich ein zufriedenstellendes Cronbach-Alpha von $\alpha = 0,67$. Weitere Untersuchungen der Stichprobendaten bspw. hinsichtlich ihrer Reliabilitäten mittels Rasch-Modellierung sind bei Zeller, Jordans & Riese (in diesem Band) zu finden.

Vergleich der beiden Testformate

Um die Aussagekraft des geschlossenen, online-gestützten Tests im Vergleich zum offenen Test zu untersuchen, wurden u. a. bivariate Korrelationen für die jeweiligen Probandenscores der beiden Testformate für eine Teilstichprobe ($N = 24$; $N = 10$ bei der Facette Fachdidaktische Konzepte) untersucht. Bei Betrachtung der Korrelationen in Abhängigkeit der vier Facetten ergeben sich nur zum Teil erhoffte hohe Korrelationen zwischen $r = 0,373$ (Experimente) und

$r = 0,808^{**}$ (Fachdidaktische Konzepte). Zur Einordnung dieser Ergebnisse wurden die Zusammenhänge der Probandenscores separat für die vier MC-Aufgaben betrachtet, welche im offenen und geschlossenen Testinstrument identisch eingebunden und für diese Betrachtung gleich bepunktet wurden (vgl. Tab. 1).

Tab. 1: Bivariate Korrelationen der Probandenscores bei Bearbeitung der unveränderten MC-Aufgaben zwischen den Testformaten (*: $p < 0,05$; **: $p < 0,01$).

| | Aufgabe 5 | Aufgabe 7 | Aufgabe 8 | Aufgabe 19 |
|---------------|-----------|-----------|-----------|------------|
| Pearson's r | 0,690* | 0,521** | 0,047 | 0,888** |

Die abgebildeten Korrelationen fielen geringer aus als erwartet, sodass angenommen werden muss, dass die Fähigkeiten der Studierenden nicht stabil sind. Beide Testformate scheinen daher das Konstrukt des FDW nicht identisch abzubilden, was eine Wahl des Testformats je nach gewünschtem Einsatzbereich nahelegt.

Aufbau des Rückmeldeformats

Durch die Schließung und Digitalisierung der Testaufgaben sowie der Festlegung neuer Bepunktungsregeln ist die Basis für die Bereitstellung (teil-)automatisierter Rückmeldungen geschaffen. Nach aktuellem Stand bestehen die fünfseitigen Rückmeldedokumente aus einer Titelseite mit kurzer Einleitung und Kontaktdaten. Darauf folgt eine Erläuterung der Testkonzeption, bevor die individuellen Ergebnisse der Studierenden bzw. der Lerngruppe auf Gesamt- und Skalenebene dargestellt werden. In den Dokumenten für Studierende werden die persönlichen Ergebnisse in die der Lerngruppe und einer Referenzgruppe mit ähnlichem Studienfortschritt (bspw. 50 Masterstudierende) grafisch eingeordnet. Dozierende erhalten eine Rückmeldung, in der die Ergebnisse ihrer Lerngruppe denen einer Referenzgruppe gegenübergestellt werden. Am Ende des Dokuments werden zum besseren Verständnis die Facetten des FDW erläutert und Literaturempfehlungen gegeben. Bisher ist noch keine instantane Rückmeldung direkt nach der Testbearbeitung möglich, aber die mit LimeSurvey erhobenen Datensätze können mithilfe von RStudio so ausgewertet werden, dass im Anschluss die individuellen Rückmeldedokumente automatisch erzeugt werden. Diese werden den Studierenden über eine Webseite auf Servern der RWTH Aachen anonymisiert zur Verfügung gestellt. Um eine Einordnung der Ergebnisse in eine immer genauer werdende Referenzgruppe zu ermöglichen, ist eine Erweiterung der Baseline durch weitere Testerhebungen und -einsätze erforderlich. Weiterhin besteht das Ziel, durch die Entwicklung von Kompetenzniveaus mit inhaltlicher Beschreibung die Rückmeldung mit individuellen Fähigkeitsprofilen anreichern zu können (s. Zeller, Jordans & Riese, in diesem Band).

Literatur

- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. Zeitschrift für Erziehungswissenschaft, 9(4), 469-520.
- Bortz, J. & Döring, N. (2006). Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler (4. Aufl.). Heidelberg: Springer Verlag.
- Gramzow, Y. (2015). Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik. Modellierung und Testkonstruktion. In Niedderer, H., Fischler, H. & Sumfleth, E. (Hrsg.), Studien zum Physik- und Chemielernen, 181. Berlin: Logos Verlag.
- Krebs, R. (2004). Anleitung zur Herstellung von MC-Fragen und MC-Prüfungen für die ärztliche Ausbildung. https://www.iml.unibe.ch/attachment/7/download/mc_anleitung.pdf. Zugegriffen: 04. August 2021.

- Kröger, J., Neumann, K. & Petersen, S. (2013). Messung professioneller Kompetenz im Fach Physik. In S. Bernholt (Hrsg.), *Inquiry-based Learning - Forschendes Lernen*, Kiel: IPN, 533-535.
- Kulgemeyer, C., Borowski, A., Buschhüter, D., Enkrott, P., Kempin, M., Reinhold, P., Riese, J., Schecker, H., Schröder, J. & Vogelsang, C. (2020). Professional knowledge affects action-related skills: The development of preservice physics teachers' explaining skills during a field experience. *Journal of Research in Science Teaching*, 57, 1-29.
- Kulgemeyer, C., Kempin, M. & Weißbach, A. (2021). Entwicklung von Professionswissen und Reflexionsfähigkeit im Praxissemester. In S. Habig (Hrsg.), *Naturwissenschaftlicher Unterricht und Lehrerbildung im Umbruch?*. GDCP online Jahrestagung 2020. Tagungsband Universität Duisburg-Essen, 262-265.
- Riese, J., Kulgemeyer, C., Zander, S., Borowski, A., Fischer, H. E., Gramzow, Y., Reinhold, P., Schecker, H., & Tomczyszyn, E. (2015). Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik. In S. Blömeke & O. Zlatkin - Troitschanskaia (Hrsg.), *Kompetenzen von Studierenden: 61. Beiheft der Zeitschrift für Pädagogik*. Weinheim: Beltz, 55-79.
- Riese, J., Gramzow, Y., & Reinhold, P. (2017). Die Messung fachdidaktischen Wissens bei Anfängern und Fortgeschrittenen im Lehramtsstudiengang Physik. *Zeitschrift für Didaktik der Naturwissenschaften*, 23(1), 99–112.
- Riese, J., Vogelsang, C., Schröder, J., Borowski, A., Kulgemeyer, C., Reinhold, P. & Schecker, H. (eingereicht). Entwicklung von Unterrichtsplanungsfähigkeit im Fach Physik: Welchen Einfluss hat Professionswissen?. *Zeitschrift für Erziehungswissenschaften*.
- Shulman, L. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15 (2), 4-14.
- Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S., Leutner, D., Neuhaus, B. J., Sandmann, A., Sumfleth, E., Thillmann, H., & Wirth, J. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 7-28.
- Vogelsang, C., Borowski, A., Buschhüter, D., Enkrott, P., Kempin, M., Kulgemeyer, C., Reinhold, P., Riese, J., Schecker, H. & Schröder, J. (2019). Entwicklung von Professionswissen und Unterrichtsperformanz im Lehramtsstudium Physik – Analysen zu valider Testwertinterpretation. *Zeitschrift für Pädagogik*, 65 (4), 473-491.
- Zeller, J., Jordans, M. & Riese, J. (in diesem Band). Ansätze zur Ermittlung von Kompetenzniveaus im Fachdidaktischen Wissen.