

Ansätze zur Ermittlung von Kompetenzniveaus im Fachdidaktischen Wissen

Theoretischer Hintergrund und Motivation

Das Professionswissen von (angehenden) Lehrkräften war in den letzten Jahren Gegenstand intensiver fachdidaktischer Forschung (Kaiser, Bremerich-Vos & König, 2020). Dabei hat sich eine Dreiteilung des Professionswissens in Fachwissen (FW), Fachdidaktisches Wissen (FDW) und Pädagogisches Wissen (PW) durchgesetzt. Diese Modellierung geht auf Arbeiten von Shulman (1986) sowie Analysen von Baumert und Kunter (2006) zurück und wurde von Riese (2009) für die Physik adaptiert. Gramzow, Riese und Reinhold (2013) entwickelten darauf aufbauend ein Modell, welches die interne Struktur des FDW für das Fach Physik genauer beschreibt. Sie schlagen eine zweidimensionale Konzeptualisierung vor, nach der sich das FDW wesentlich anhand von acht Facetten *Instruktionsstrategien, Schülervorstellungen, Experimente, Fachdidaktische Konzepte, Kontext und Interesse, Curriculum Bildungsstandards und Ziele, (Digitale) Medien* sowie *Aufgaben* jeweils bezüglich konkreter physikalischer Fachinhalte charakterisieren lässt. Auf Basis dieses Modells entwickelte Gramzow (2015) ein Testinstrument (offene Kurzantworten und Multiple-Choice- / MC-Aufgaben), wobei sie den physikalischen Fachinhalt auf „Mechanik“ festlegte und sich im Sinne der Zumutbarkeit der Testdurchführung auf die ersten vier der oben genannten Facetten fokussierte. Ähnliche Modellierungen des Professionswissens wurden im Forschungsprojekt KiL (Kleickmann et al., 2014) vorgenommen, welches die Messung der professionellen Kompetenzen in MINT-Lehramtsstudiengängen zum Ziel hatte. Das entsprechende Testinstrument verwenden Schiering et al. (2019), um mithilfe des Scale-Anchoring-Verfahrens (Mullis et al., 2015) erste Niveaubeschreibungen für das FDW im Fach Physik zu entwickeln. Dabei bildet das verwendete Testinstrument teilweise andere Facetten als die oben genannten ab und umfasst unterschiedliche physikalische Inhaltsbereiche. Insgesamt deuten die erhaltenen qualitativen Niveaubeschreibungen darauf hin, dass sich das verfügbare Wissen mit zunehmendem Kompetenzniveau facettenübergreifend zunächst auf reproduktive Aspekte beschränkt, während auf höheren Niveaustufen auch weitere, vernetzende und reflektierende Anteile hinzukommen.

Der vorliegende Beitrag widmet sich in Anlehnung an das Vorgehen von Schiering et al. (2019) der Modellierung von Kompetenzniveaus des FDW aufbauend auf der Operationalisierung von Riese, Gramzow und Reinhold (2017) konkret bezogen auf eine geschlossene Version des oben beschriebenen Testinstruments zum FDW im Bereich Mechanik. Langfristig soll diese Modellierung einen Ausgangspunkt für inhaltlich reichhaltiges Feedback auf Basis von quantitativen Testergebnissen bilden.

Instrument und Design

Das eingesetzte Testinstrument ist eine vollständig geschlossene (Jordans, Zeller, Große-Heilmann & Riese, in diesem Band) Version des oben beschriebenen Testinstruments von Gramzow (2015) und erfasst das Fachdidaktische Wissen in den Facetten *Instruktionsstrategien* (8 Aufgaben), *Schülervorstellungen* (9 Aufgaben), *Experimente* und *Vermittlung eines angemessenen Wissenschaftsverständnisses* (5 Aufgaben) und

Fachdidaktische Konzepte (9 Aufgaben). Im geschlossenen Testformat sind alle (2. Teilstichprobe), bzw. alle bis auf zwei (1. Teilstichprobe) Aufgaben als elementare MC-Aufgaben mit unterschiedlicher Anzahl an Antwortmöglichkeiten mithilfe der Software LimeSurvey implementiert. Insgesamt lagen Daten zu $N = 125$ Bearbeitungen, dabei $N_1 = 78$ der ersten Testversion, von Studierenden aus insgesamt 6 Universitäten vor. Da die Bearbeitungen der ersten Teilstichprobe zu Beginn dieser Untersuchung bereits abgeschlossen waren, ließen sich einige, fehlende demografische Daten nicht mehr im Detail rekonstruieren. Allerdings kann davon ausgegangen werden, dass die Aufteilung der Studierenden nach Bachelor- und Masterstudiengang (Lehramt Physik) in etwa 1:1 ist und somit a priori eine vergleichsweise hohe Varianz zu erwarten ist.

Methodik

Zur Ermittlung von qualitativen Kompetenzniveaus wurde nach einer Exploration unterschiedlicher möglicher Methoden das Scale-Anchoring Verfahren nach dem Vorbild von Schiering et al. (2019) gewählt. Den ersten Schritt stellt dabei eine Modellierung im Rahmen der Item Response Theorie (IRT) dar. Zwei Items mussten auf Basis dieser Modellierung und ergänzender konfirmatorischer Faktorenanalysen ausgeschlossen werden.

Da das Testinstrument hauptsächlich aus MC-Aufgaben bestand, ergaben sich zunächst für eine IRT-Modellierung über 150 dichotome Antwortoptionen als Einzelitems, die in 31 Aufgabenteilen zusammengefasst sind. Diese Menge ist weder mit der vorhandenen Stichprobengröße verlässlich modellierbar, noch lassen sich bei einer Zerfaserung aller Antwortoptionen im Rahmen des Scale-Anchoring Verfahrens noch sinnvolle, inhaltlich kondensierte Niveaubeschreibungen vornehmen. Zur inhaltlichen Verdichtung wurde daher jede Aufgabe mithilfe eines Kprim-Schwellensystems (Krebs, 2004) umkodiert. Dadurch ergaben sich noch 62 relevante Kodiereinheiten. Die Verwendung von Partial-Credit Modellen erwies sich bei der Verwendung der mit dem Schwellensystem ausgewerteten Daten aufgrund der systematischen Vertauschung der Schwierigkeitsparameter der Aufgaben als problematisch. Stattdessen wurden die Aufgabenschwelle als einzelne Items aufgefasst und die Daten mithilfe von dichotomen Rasch-Modellen untersucht.

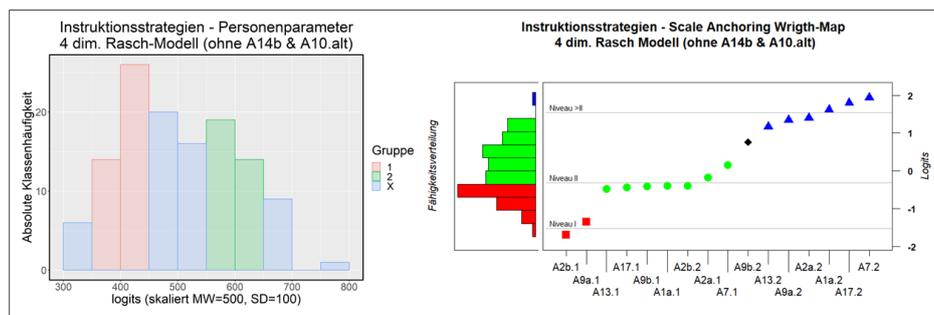


Abb. 1: Illustrationen zum Scale-Anchoring Verfahren exemplarisch für die Facette Instruktionsstrategien. Links die Einordnung der Proband:innen in die Ausgangsgruppen, Rechts die Wright-Map mit Aufgabengruppen und Niveaugrenzen.

Trotz des kleinen Stichprobenumfangs und der oben beschriebenen Schwierigkeiten ergab sich dabei eine EAP-Reliabilität der Gesamtskala von 0.79. Eine vierdimensionale Rasch-Modellierung mit Unterteilung in die Facettenstruktur wies annehmbare (Adams & Wu, 2007, sowie Yamamoto et al., 2013) Fit-Statistiken auf ($RMSD = 0.11$, $MNSQ\text{-Itemfits} = 0.99 \pm$

0.14) und erwies sich als signifikant ($p < 0.01$) passender als die eindimensionale Modellierung. Die EAP-Reliabilitäten lagen in einem Bereich von 0.56 (Experimente) bis 0.73 (Schülervorstellungen). Dem Scale-Anchoring Verfahren folgend wurden nun auf Facettenebene zwei Personengruppen anhand von Abschnitten auf der Fähigkeitsskala festgelegt, die statistisch signifikant unterscheidbar und zudem von ausreichender Größe (min. 10% der Proband:innen) sind (Abb. 1 links). Darauf aufbauend wurden die Aufgaben gemäß dem Schema von Mullis et al. (2015) in Gruppen eingeordnet. Die Mittelwerte dieser Aufgabengruppen dienen dann als Niveaugrenzen. Das Ergebnis der Analyse stellen vier Wright-Maps wie in Abbildung 1 rechts dar. Aufgaben in der Nähe der Niveaugrenzen werden anschließend zur Niveauformulierung genutzt.

Ergebnisse und Ausblick

Im Rahmen der Erprobung wurden explizite Niveaus für die Facette Instruktionsstrategien formuliert. Dazu wurden die Aufgaben einer Reanalyse unterzogen und die jeweiligen adressierten Kernkompetenzen zusammengefasst. Anschließend wurden für jede Aufgabe kurze Niveaubeschreibungen verfasst, die im letzten Schritt inhaltlich gebündelt wurden. Die Ergebnisse der Niveauformulierungen sind in Tabelle 1 festgehalten. Da die Modellierung der MC-Aufgaben in Schwellen dazu führt, dass diese in den Wright-Maps getrennt auftreten, haben die Niveauformulierungen teils graduellen („erste Eigenschaften“, „viele geeignete Alltagsbeispiele“ und Ähnliches) Charakter.

Aufgrund der vielversprechenden Ergebnisse des Verfahrens für die vorliegende Testversion sind weitere, elaboriertere Untersuchungen mit über 800 bereits vorliegenden Datensätzen des ursprünglichen offenen Testinstruments geplant. Dabei wäre insbesondere die Lösung vom fachlichen Inhalt wünschenswert und die Verbindung zu Testergebnissen weiterer Instrumente wie beispielsweise zum Fachwissen oder zur Planungsperformanz interessant. Schließlich soll untersucht werden, ob sich ein finales Niveaumodell im Längsschnitt als Entwicklungsmodell des FDW im Studienverlauf validieren lässt.

I	Die Studierenden identifizieren erste Eigenschaften, aufgrund derer sich Alltagsbeispiele zur Einführung des Konzepts der Geschwindigkeit als vektorielle Größe eignen und können erste geeignete Alltagsbeispiele zum Fachinhalt <u>Kreisbewegung</u> auswählen. (Aufgabe 2b.1, Aufgabe 9a.1)
II	Die Studierenden beginnen vorteilhafte Handlungsalternativen zum Umgang mit problematischen Schülervorstellungen zu identifizieren. (Aufgabe 1a.1, Aufgabe 17.1) Die Studierenden identifizieren Eigenschaften, aufgrund derer sich Alltagsbeispiele bei unterschiedlichen Fachinhalten zur Verwendung eignen und können erste geeignete Alltagsbeispiele zum Fachinhalt <u>Geschwindigkeiten</u> auswählen. (Aufgabe 2a.1, Aufgabe 2b.2, Aufgabe 9b.1) Die Studierenden beginnen, erste falsche Interpretationen von Zeit-Weg Diagrammen aufgrund von typischen problematischen Vorstellungen zu identifizieren. (Aufgabe 13.1)
>II	Die Studierenden identifizieren vorteilhafte Handlungsalternativen zum Umgang mit problematischen Schülervorstellungen. (Aufgabe 1a.2, Aufgabe 17.2) Die Studierenden können viele geeignete Alltagsbeispiele zu unterschiedlichen Fachinhalten auswählen. (Aufgabe 2a.2, Aufgabe 9a.2)

Tab. 1: Beispielhafte erste Niveauformulierungen für die Facette Instruktionsstrategien.

Literatur

- Adams, R. J., & Wu, M. L. (2007). The mixed-coefficients multinomial logit model. A generalized form of the Rasch model. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (p. 55-76). New York: Springer.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469-520.
- Gramzow, Y. (2015). Fachdidaktisches Wissen von Lehramtsstudierenden im Fach Physik: Modellierung und Testkonstruktion. In H. Niedderer, H. Fischler & E. Sumfleth (Hrsg.), *Studien zum Physik- und Chemielernen* (Bd. 181). Berlin: Logos Verlag.
- Gramzow, Y., Riese, J., & Reinhold, P. (2013). Modellierung fachdidaktischen Wissens angehender Physiklehrkräfte. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 7-30.
- Jordans, M., Zeller, J., Große-Heilmann, R., & Riese, J. (in diesem Band). Weiterentwicklung eines physikdidaktischen Tests zum Online-Assessment. In S. Habig (Hrsg.), *Unsicherheit als Element von naturwissenschaftsbezogenen Bildungsprozessen, Tagungsband der GDCP Jahrestagung 2021*. Essen: Universität Duisburg-Essen.
- Kaiser, G., Bremerich-Vos, A., & König, J. (2020). Professionswissen. In C. Cramer, J. König, M. Rothland & S. Blömeke (Hrsg.), *Handbuch Lehrerinnen- und Lehrerbildung* (S. 811–818). Bad Heilbrunn: Klinkhardt.
- Kleickmann, T., Großschedl, J., Harms, U., Heinze, A., Herzog, S., Hohenstein, F., Köller, O., Kröger, J., Lindmeier, A., Loch, C., Mahler, D., Möller, J., Neumann, K., Parchmann, I., Steffensky, M., Taskin, V. & Zimmermann, F. (2014). Professionswissen von Lehramtsstudierenden der mathematisch-naturwissenschaftlichen Fächer-Testentwicklung im Rahmen des Projekts KiL. *Unterrichtswissenschaft*, 42(3), 280–288.
- 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. Mai 2021.
- Mullis, I. V. S., Cotter, K. E., Centurino, V. A. S., Fishbein, B. G. & Liu, J. (2015) Using Scale Anchoring to Interpret the TIMSS 2015 Achievement Scales. In M. O. Martin, I. V. S. Mullis & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (p. 14.1-14.47).
- Riese, J. (2009). Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften. In H. Niedderer, H. Fischler & E. Sumfleth (Hrsg.), *Studien zum Physik- und Chemielernen* (Bd. 97). Berlin: Logos Verlag.
- Riese, J., Gramzow, Y., & Reinhold, P. (2017). Die Messung fachdidaktischen Wissens bei Anfängern und Fortgeschrittenen im Lehramtsstudiengang Physik. *Zeitschrift für Didaktik der Naturwissenschaften*, 23, 99–112.
- Schiering, D., Sorge, S., Petersen, S., & Neumann, K. (2019) Konstruktion eines qualitativen Niveaumodells im fachdidaktischen Wissen von angehenden Physiklehrkräften. *Zeitschrift für Didaktik der Naturwissenschaften*, 25, 211-229.
- Shulman, L. S. (1986): Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.
- Yamamoto, K., Khorramdel, L., & von Davier, M. (2013). Scaling PIAAC cognitive data. In OECD (Eds.), *Technical Report of the Survey of Adults Skills (PIAAC)* (Chap. 17). Paris: OECD.