

Fachdidaktisches Wissen Physik - Validierungsstudien

Ziele und Theoretischer Hintergrund

Obwohl das Fachdidaktische Wissen (FDW) in der Literatur häufig als relevant für unterrichtliches Handeln von Lehrkräften erhofft wird, existiert bislang weder ein Konsens über seine innere Struktur noch ein Testinstrument, welches eine solche Struktur empirisch abzubilden vermag. Ausgehend von dieser Problematik wurde ein Modell fachdidaktischen Wissens entworfen sowie ein Testinstrument entwickelt und pilotiert, welches Analysen auf der Ebene von Subskalen ermöglichen soll. Mit Hilfe dieses Instruments sollen im Rahmen des Verbundprojekts Profile-P (vgl. Kulgemeyer et al., 2012) die Zusammenhänge zwischen den inneren Strukturen des Fachdidaktischen Wissens, des Fachwissens (Gigl et al., in diesem Band) und der Tätigkeit des Erklärens (Tomczyszyn et al., in diesem Band) aufgeklärt werden. Im vorliegenden Beitrag werden die Ergebnisse der Pilotierung sowie Studien zur Prüfung der Validität der Testwertinterpretation vorgestellt.

Testinstrument und Pilotstudie

Das Modell Fachdidaktischen Wissens wurde deduktiv auf der Grundlage von Analysen verschiedener Konzeptualisierungen von FDW und dem pedagogical content knowledge (PCK) entwickelt. Es ordnet sich in die gängigen heuristischen Modelle professioneller Handlungskompetenz als Teil der kognitiven Disposition in Anlehnung an Baumert und Kunter (2006) ein. Der Schwerpunkt liegt auf das an der Universität erwerbbar Wissen. Dazu wurde zunächst ein breit angelegtes Modell entwickelt und anschließend für die Testentwicklung gekürzt (vgl. Gramzow, Riese & Reinhold, 2013). Der physikalische *Inhaltsbereich* wurde auf die Mechanik beschränkt sowie die vier fachdidaktischen *Facetten* „Instruktionsstrategien“, „Schülervorstellungen“, „Experimente und Vermittlung eines angemessenen Wissenschaftsverständnisses“ sowie „Fachdidaktische Konzepte“ ausgewählt.

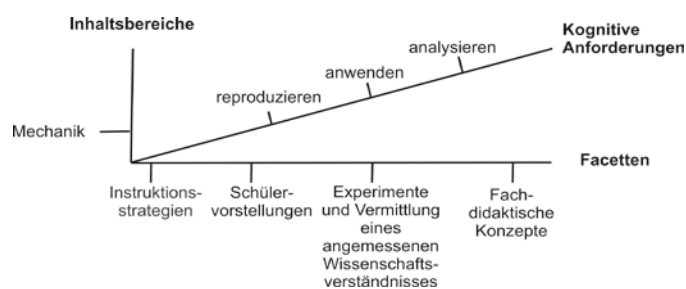


Abb.1: Modell für die Itementwicklung

Des Weiteren wurde dem Modell für die Itementwicklung (Abb.1) als dritte Dimension die *Kognitive Aktivität* hinzugefügt, die die Anforderungen der jeweiligen Items in die Bereiche „Reproduzieren“, „Analysieren“ und „Anwenden“ unterteilt (vgl. Anderson und Krathwohl, 2001). Anhand dieses Modells wurden 50 Aufgaben mit insgesamt 91 Items entwickelt, die in einer ersten Erhebung mithilfe von 2 Testheften pilotiert werden konnten. Dazu wurden 216 Lehramtsstudenten der Physik aus 12 Universitäten und Fachhochschulen befragt. Die Charakteristik der Stichprobe findet sich in der folgenden Tabelle:

Tab.2: Charakteristik der Pilotstichprobe. Angegeben sind jeweils Mittelwerte.

	N	Note Abitur	Weiblich	Fach- semester	Erworbene Creditpoints (CP) in physikdidaktischen Veranstaltungen
HR Lehramt	71	2,5	48 %	4,6	8,9
Gym Lehramt	127	2,0	39 %	5,2	10,0
Gesamt	216	2,2	44 %	5,0	9,3

Die auf Grundlage einer Raschanalyse erzeugte Wrightmap zeigt, dass der Test zu schwer und die Varianz zu gering ist (.39). Der Itemfit ist bei 86 von 91 Items jedoch gut ($0.8 < \text{MNSQ} < 1.2$ und $-1.9 < T < 1.9$), ebenso die EAP-Reliabilität für den gesamten Test (.84).

Validitätsbegriff

In Anlehnung an z. B. Blömeke (2013) und Kane (2013) wird Validität nicht als Eigenschaft des Testinstruments, sondern als eine der Interpretation definiert. In diesem Zusammenhang ist es zunächst relevant, die intendierten Interpretationen festzulegen. Hartig, Frey und Jude (2012) folgend soll das hier entwickelte Testinstrument dazu verwendet werden, Leistungen zu erklären. Hierzu sind theoretische Annahmen über das Konstrukt nötig, deren Validität überprüft werden muss. Die Annahmen gliedern sich in drei Ebenen:

- (1) FDW lässt sich zu anderen Wissensbereichen (insbesondere Fachwissen) abgrenzen.
- (2) FDW ist an der Universität erwerbbar.
- (3) FDW lässt sich in vier fachdidaktischen Facetten gliedern.

Untersuchungen zur Konstruktvalidierung

Zunächst kann im Sinne der Verortung des Konstrukts in einem nomologischen Netzwerk (vgl. z. B. Hartig & Jude, 2007; Schaper, 2009) untersucht werden, ob Testwerte Indikatoren für Ausprägungen eines nicht direkt beobachtbaren Konstrukts sind. Sofern die Testwerte mit einer theoretischen Vorhersage übereinstimmen, kann dies als Indiz der Bestätigung des theoretischen Konstrukts angesehen werden. In Bezug auf die Ebenen (1) und (2) ist anzunehmen, dass die erreichten Creditpoints in fachdidaktischen Lehrveranstaltungen höher mit den FDW-Personenfähigkeiten korrelieren als die erreichten Creditpoints in Fachveranstaltungen. Diese Annahme wird durch die Testwerte unterstützt. Es zeigt sich (jeweils unter Kontrolle der Fachsemester) eine Person-Korrelation von $r=.375$ der FDW-Personenfähigkeiten mit den CP Physikdidaktik und eine Person-Korrelation von $r=.281$ der FDW-Personenfähigkeiten mit den CP Physik. Im Regressionsmodell ist jedoch ausschließlich der Zusammenhang zu den CP Physikdidaktik relevant.

Des Weiteren ist es möglich, eine statistische Prüfung der inneren Struktur des Wissensbereichs mithilfe von Modellvergleichen vorzunehmen (Ebene 3). Hier zeigt ein χ^2 -Test, dass ein 4D-Rasch-Modell, welches in vier fachdidaktischen Facetten unterteilt, eine hochsignifikant bessere Passung als ein 1D-Rasch-Modell aufweist.

Ebenfalls bezogen auf Ebene (3) wurde eine Prüfung der Modellpassung der Itemkonstruktionsregeln vorgenommen. Hierfür wurden zwei fachdidaktische Experten herangezogen, die anhand eines Manuals zur Beschreibung des Modells die einzelnen Items jeweils einer Facette zuordnen sollten. Es zeigt sich eine hohe Übereinstimmung der Zuordnungen (Cohens κ zwischen .73 und .89).

Untersuchungen zur Inhaltsvalidierung

Bezogen auf Ebene (1) der Annahmen über das Konstrukt wurde eine Expertenbefragung an den Verbundstandorten durchgeführt, um die inhaltliche Passung der Items hinsichtlich der universitären Curricula zu ermitteln. Mithilfe einer fünfstufigen Rating-Skala sollte jedes Item dahingehend bewertet werden, ob Lernenden im Laufe des Fachdidaktik-Studiums die

Möglichkeit haben, die nötigen Kompetenzen zu erwerben, um die entsprechende Aufgabe korrekt zu lösen (1=Ja, absolut; 5=Nein, in keinem Fall). Eine Mittelwertbildung der Einschätzungen liefert Werte im Bereich von 1.0 bis 3.2. Der über alle Items berechnete Mittelwert liegt bei 2.1, seine Standardabweichung bei 0.6. Daraus folgt, dass einige Items eine weniger hohe curriculare Passung aufweisen und daher für eine Testoptimierung entfernt werden sollten.

Ferner wurde eine Lautes-Denken-Studie (n=15) implementiert, bei der mithilfe einer Inhaltsanalyse nach Mayring (2010) Wissensbestände identifiziert werden konnten, auf die die Probanden bei der Bearbeitung der Aufgaben zurückgreifen. Insbesondere bezogen auf Ebene (2) konnten so wichtige Hinweise dahingehend gewonnen werden, ob Aufgaben ausschließlich mit Fachwissen lösbar sind oder nicht. In dieser Hinsicht problematische Aufgaben sollten für die Haupterhebung entfernt werden.

Zusammenfassung und Ausblick

Es konnte ein Test entwickelt werden, der eine modellkonforme innere Struktur des Fachdidaktischen Wissens in Physik empirisch abbilden kann. Des Weiteren wurden Studien zur Untersuchung der Validität der intendierten Testwertinterpretation durchgeführt. Diese kann jedoch nur falsifiziert (z. B. durch die Identifikation problematischer Items) und nicht verifiziert werden. Trotzdem scheint der Test mit Einschränkungen für die o. g. Ziele des Verbundprojekts ProfiLe-P geeignet. Für die Haupterhebung, die seit dem Wintersemester 13/14 läuft, wurde eine Kürzung des Instruments auf 65 Minuten vorgenommen. Dazu wurden im ersten Schritt Hinweise aus den Validierungsstudien und im zweiten Schritt statistische Merkmale des Tests bzw. der Items herangezogen. Die Haupterhebung des Verbundprojekts ProfiLe-P ist im Sommer 2015 abgeschlossen. Zu diesem Zeitpunkt können voraussichtlich erste differenzierte Zusammenhangsanalysen zwischen dem Fachdidaktischen Wissen, dem Fachwissen und der Tätigkeit des Erklärens vorgestellt werden.

Literatur

- Anderson, L. W.; Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives. New York: Longman.
- Baumert, J., Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Blömeke, S. (2013). Validierung als Aufgabe im Forschungsprogramm "Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor": (KoKoHS Working Papers, 2). Berlin & Mainz: Humboldt-Universität & Johannes Gutenberg-Universität.
- Gramzow, Y.; Riese, J.; Reinhold, P. (2013). Modellierung fachdidaktischen Wissens angehender Physiklehrkräfte. *ZfDN (Zeitschrift für Didaktik der Naturwissenschaften)*, 19, 7-30
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion*. Berlin: Springer, 143-173.
- Hartig, J. & Jude, N. (2007). Empirische Erfassung von Kompetenzen und psychometrische Kompetenzmodelle. In J. Hartig & E. Klieme (Hrsg.), *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik*, Bonn and Berlin: Bundesministerium für Bildung und Forschung, 17-36
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50 (1), 1-73.
- Kulgemeyer, C.; Borowski, A.; Fischer, H.; Gramzow, Y.; Reinhold, P.; Riese, J.; Schecker, H.; Tomczyszyn, E.; Walzer, M. (2012). ProfiLe-P – Professionswissen in der Lehramtsausbildung Physik. Vorstellung des Forschungsprojekts. *PhyDid B – Beiträge zur DPG-Frühjahrstagung*.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (11. Aufl.). Weinheim: Beltz.
- Schaper, N. (2009). Aufgabenfelder und Perspektiven bei der Kompetenzmodellierung und -messung in der Lehrerbildung. In N. Schaper, A. H. Hilligus & P. Reinhold (Hrsg.), *Kompetenzmodellierung und -messung in der Lehrerbildung*, Landau/Pfalz: Empirische Pädagogik, 166-199.