

David Woitkowski
Josef Riese
Peter Reinhold

Universität Paderborn

Konstruktion von Niveaus des fachlichen Wissens Physik

Mehrere Studien der empirischen Bildungswissenschaften nutzen die Angabe von Kompetenzniveaus zur Kommunikation wesentlicher Testresultate. Dabei handelt es sich um inhaltliche Beschreibungen der Kenntnisse oder Fähigkeiten der jeweiligen Probandengruppen. Im Rahmen des vorliegenden Projektes wurde für das fachliche Wissen in Physik ein Niveau-modell erstellt (Woitkowski, 2015). Das Verfahren basiert auf einem schwierigkeiterzeugenden Aufgabenmerkmal (der Aufgabenkomplexität), welches im Rahmen des zugrunde gelegten Strukturmodells beschrieben wurde (Woitkowski, Riese, Reinhold, 2011).

Verfahren zur Niveaufinition in der Literatur

In der Literatur werden drei Verfahren zur Definition von Kompetenzniveaus vorgestellt, die jeweils auf einem Rasch-skalierten Testinstrument basieren. Dabei wird die Fähigkeits-Skala aus der Rasch-Analyse in mehrere Abschnitte, *Niveaus*, aufgeteilt. Dann werden die Items am unteren Rand des jeweiligen Niveaus zur inhaltlichen Charakterisierung herangezogen, da diese von den jeweiligen Probanden auf dem beschriebenen Niveau hinreichend wahrscheinlich gelöst werden können, von denjenigen auf dem Niveau darunter aber nicht.

Das *Scale-Anchoring-Verfahren* (Beaton & Allen, 1992) nutzt ein externes Kriterium zur Einteilung von Probandengruppen, bei PISA (Prenzel et al., 2007) und TIMSS (Klieme, 2000) wurde z. B. die Fähigkeits-Skala einfach in gleich breite Abschnitte eingeteilt, die dann durch die Analyse der Items am unteren Rand inhaltlich charakterisiert wurden.

Bei den Bildungsstandards Physik wurde die Bookmark-Methode (Mitzel et al., 2001) angewandt, bei der stärker von den Testitems selbst ausgegangen wird. Diese wurden in einem *Ordered Item Booklet* nach Schwierigkeit geordnet und einer Expertenkommission vorgelegt, die *Cut-Scores* dort festlegen sollte, wo ein fachdidaktisch begründeter Schritt zwischen den Itemanforderungen festgestellt werden konnte.

Ein drittes Verfahren, z. B. bei DESI (Hartig, 2007) und MT21 (Blömeke et al., 2008) nutzt die im Kompetenzstrukturmodell beschriebenen schwierigkeiterzeugenden Aufgabenmerkmale. Dieses Verfahren wird im vorliegenden Projekt zur Niveaufinition eingesetzt und im Folgenden anhand der Fachwissens-Gesamtskala von Woitkowski (2015) exemplarisch skizziert.

Regressionsanalytische Definition von Kompetenzniveaus

Im Rahmen dieses Verfahrens wird zunächst eine Regressionsanalyse durchgeführt, die die Item-Schwierigkeit auf das schwierigkeiterzeugende Aufgabenmerkmal (hier: die hierarchische Komplexität, vgl. Woitkowski et al., 2011) zurückführt. Die Regressionsparameter für die FW-Skala zeigt Tab. 1. Die Varianzaufklärung von $R^2 = 46.4\%$ kann als akzeptabel gelten, eine Auftrennung der Skala nach Fach-Stufen (Schul-, vertieftes, universitäres Wissen) liefert für die Teilskalen z. T. wesentlich höhere Werte bis zu 70%.

Aus den Regressionsparametern b_i lässt sich für jede Itemgruppe (also jede Komplexitätsstufe) ein durchschnittlicher Item-Parameter, also ein für die Gruppe *typisches Item* bestimmen. Diese werden in die Wright-Map als vertikale Linien eingezeichnet (Abb. 1). Die

Prädiktoren	$b_i(p)$
Konstante	-1.75***
Komplexität: II	1.39***
Komplexität: III	2.36***
Komplexität: IV	3.88***
Var.-Aufkl. R^2	46.4 %
Signifikanz p	***

Tab. 1: Regressionsanalyse der Itemschwierigkeiten der Items der FW-Gesamtskala

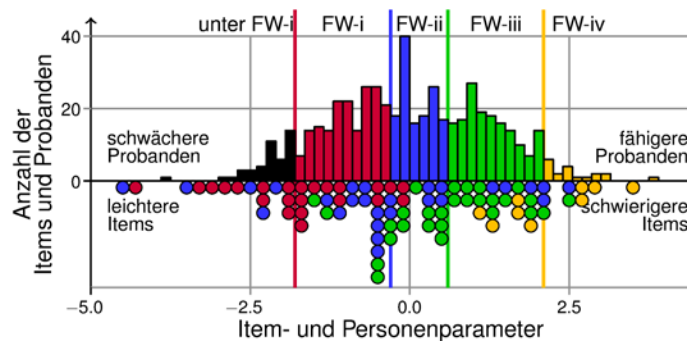


Abb. 1: Konstruktion von Niveaus der Fachwissens-Gesamtskala anhand der Wright-Map. Die Items sind nach Komplexität eingefärbt: (I) rot; (II) blau; (III) grün; (IV) gelb

Farben ordnen dabei jeweils die Items den Komplexitätsstufen zu. Man sieht deutlich, wie die *Fakten-Items* (rot) eher im leichteren Bereich, die *Multivariate-Interdependenz-Items* (gelb) deutlich im schwierigeren Bereich streuen.

Zur Sicherung der Qualität werden nun zwei Qualitätskriterien überprüft: Das **Signifikanzkriterium** besagt, dass sich zwei Itemgruppen im t-Test signifikant voneinander unterscheiden müssen. Dieses Kriterium wird jeweils mit $p < .01$ erfüllt. Damit streuen die Items einer Gruppe nicht zu stark und überlappen sich nicht zu sehr.

Das **Abstandskriterium** besagt, dass ein Proband, der auf der Höhe eines *typischen Items* liegt, für das nächst schwierigere *typische Item* eine Lösungswahrscheinlichkeit von maximal $P = 25\%$ haben sollte, um die Niveaus nicht zu nah aneinander zu rücken. Das entspricht einem Abstand von 1.1 Logit auf der Rasch-Skala. Dieses Kriterium wird von den Itemgruppen II und III (Blau und Grün) verletzt, diese liegen nur 0.97 Logit auseinander (entspricht $P = 27\%$). Während im Falle der FW-Gesamtskala diese Abweichung toleriert wird, führt sie bei anderen Skalen (ebenso wie eine Verletzung des Signifikanzkriteriums) zur Zusammenlegung zweier Itemgruppen und somit zu einer Reduzierung der Anzahl der gebildeten Niveaus.

Die vier so gebildeten und überprüften *typischen Items* dienen nun als untere Grenze für die jeweiligen Niveaus. So können die rot markierten Probanden das *typische rote Item* hinreichend sicher lösen (nämlich mit $P > 50\%$), das *typische blaue Item* aber nicht. Bei der inhaltlichen Beschreibung der Niveaus kann nun also auf die Beschreibung der Komplexitätsstufen aus dem Strukturmodell zurückgegriffen werden, was eine wesentlich übersichtlichere Niveaubeschreibung ermöglicht, als z. B. bei den Bildungsstandards, bei denen bis zu 10 nicht deutlich zusammenhängende Merkmale zur Beschreibung einzelner Niveaus genutzt werden (vgl. IQB, 2013, S. 33–37). Die Niveaubeschreibung der FW-Skala zeigt Tab. 2.

Diskussion

Das eingesetzte Verfahren produziert für diese Skala 5 Niveaus, wobei sich eine Probandengruppe als *unter Niveau i* bezeichnen lässt, über diese können keine positiven Könnensausagen gemacht werden, da sie unterhalb des leichtesten *typischen Items* liegen. Bei anderen Skalen wurden Itemgruppen zusammengelegt und somit weniger (im Extremfall nur 3) Niveaus gebildet. Welche Itemgruppen dabei sinnvoll zusammengefasst werden, muss jeweils im Einzelfall anhand statistischer und inhaltlicher Überlegungen argumentiert werden.

Aufgrund des starken Bezugs zum Kompetenzstrukturmodell kann eine modellgeleitete Übertragbarkeit auf andere Itembestände für möglich gehalten werden (vgl. Hartig, 2007). Auch ein Vergleich mit anderen Niveaumodellen und eine Assoziation einzelner Niveaus

mit relevanten Bildungszielen wurden durchgeführt. Eine Ausführliche Darstellung findet sich bei Woitkowski (2015), Kapitel 17.

Niveau FW-i (Probandenfähigkeiten im Bereich $-1.75 \leq \theta < -0.36$)
Probanden auf diesem Niveau können Anforderungen bewältigen, die die Kenntnis und Wiedergabe von unverknüpften Fakten des physikalischen Fachwissens wie Merksätzen, Formeln, Definitionen oder Verweise auf einfache Tatsachen des Inhaltsbereichs Mechanik ohne weitere Arbeit damit erfordern. Dazu gehören auch Anforderungen, deren Lösungen gut ausgebildeten Studenten üblicherweise auswendig bekannt sind.
Niveau FW-ii (Probandenfähigkeiten im Bereich $-0.36 \leq \theta < 0.61$)
Probanden auf diesem Niveau können Anforderungen bewältigen, die die Kombination weniger Fakten des physikalischen Fachwissens im Inhaltsbereich Mechanik ohne Rechnung erfordern. Dies kann die Nutzung eines Diagramms oder die Beschreibung von Prozessen oder Vorher-Nachher-Zusammenhängen in Form von Worten oder Skizzen sein, bei denen keine weitere Begründung nötig ist.
Niveau FW-iii (Probandenfähigkeiten im Bereich $0.61 \leq \theta < 2.13$)
Probanden auf diesem Niveau können Anforderungen bewältigen, die einschrittige Begründungen oder Rechnungen im Bereich des physikalischen Fachwissens im Inhaltsbereich Mechanik verlangen. Diese umfassen Rechnungen, die das Auswählen einer passenden Formel, Umstellen und Einsetzen von Zahlenwerten oder etwa das Aufstellen eines Kräftegleichgewichts oder einer Energie- oder Impulsbilanz erfordern. An nicht-rechnerischen Begründungen werden lineare Begründungen der Form „Weil x, darum y“ oder die Berücksichtigung maximal eines wesentlichen Einflussfaktors beherrscht.
Niveau FW-iv (Probandenfähigkeiten im Bereich $2.13 \leq \theta$)
Probanden auf diesem Niveau können Anforderungen bewältigen, die die gleichzeitige Operation mit mehreren Einflussfaktoren oder physikalischen Zusammenhängen oder den Umgang mit komplexen Begründungslinien im Inhaltsbereich Mechanik erfordern.

Tab. 2: Beschreibung der gebildeten Niveaus der FW-Skala (Woitkowski, 2015, S. 255)

Literatur

- Beaton, A. E. & Allen, N. L. (1992). Interpreting Scales Through Scale Anchoring. *Journal of Educational Statistics*, 17(2), 191-204.
- Blömeke, S., Lehmann, R., Seeber, S. et al. (2008). Niveau- und institutionenbezogene Modellierungen des fachbezogenen Wissens. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer*. (S. 105–134). Münster: Waxmann.
- Hartig, J. (2007). Skalierung und Definition von Kompetenzniveaus. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung* (S. 83–99). Weinheim u.a: Beltz.
- IQB (2013). *Kompetenzstufenmodelle zu den Bildungsstandards im Fach Physik für den Mittleren Schulabschluss: Kompetenzbereiche „Fachwissen“ und „Erkenntnisgewinnung“*, Humboldt-Universität. Verfügbar unter: http://www.iqb.hu-berlin.de/bista/ksm/KSM_Physik.pdf.
- Klieme, E. (2000). Fachleistungen im voruniversitären Mathematik- und Physikunterricht. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III : Dritte Internationale Mathematik- und Naturwissenschaftsstudie* (Bd. 2, S. 57–128). Opladen: Leske + Budrich.
- Mitzel, H. C., Lewis, D. M., Patz, R. J. & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizek (Hrsg.), *Setting performance standards. Concepts, methods and perspectives* (S. 249–281). Mahwah, NJ: Erlbaum.
- Prenzel, M., Schöps, K., Rönnebeck, S. et al. (2007). Naturwissenschaftliche Kompetenz im internationalen Vergleich. In M. Prenzel, C. Artelt, J. Baumert et al. (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 63–106). Münster: Waxmann.
- Woitkowski, D., Riese, J. & Reinhold, P. (2011). Modellierung fachwissenschaftlicher Kompetenz angehender Physiklehrkräfte. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 289-313.
- Woitkowski, D. (2015). *Fachliches Wissen Physik in der Hochschulausbildung: Konzeptionalisierung, Messung, Niveaubildung*. Berlin: Logos.