

Kevin Schmitt¹
Verena Spatz¹

¹Technische Universität Darmstadt

Skalierung eines physikalischen Vorwissenstests für Physik-Nebenfachstudierende mittels Item-Response Theory

Forschungsstand

Wiederholt zeigen Studien bezüglich der Entwicklung von Studienabbruchquoten an deutschen Hochschulen und Universitäten, dass der hohe prozentuale Anteil von Studienabbrechenden in den letzten Jahren stagniert (Heublein et al., 2020; Heublein et al., 2022). Neben den insgesamt höchsten Anteilen an Studienabbruchzahlen von bis zu 50% in Mathematik und Naturwissenschaften, verzeichnet sich im Absolvent:innenjahrgang 2020 in ingenieurwissenschaftlichen Studiengängen eine Studienabbruchquote von 35% bezüglich der Studienanfänger:innen aus den Jahren 2016/2017 (Heublein et al., 2022).

An diese Datengrundlage knüpfen unter anderem empirische Studien an, die den Zusammenhang zwischen fachspezifischem Vorwissen und dem Studienerfolg untersuchen (Müller et al., 2016). Bei Studienanfänger:innen in MINT-Studiengängen zeigt sich, dass neben mathematischem Vorwissen auch das Vorwissen in anderen Grundlagenfächern, wie z.B. Physik, Chemie oder Biologie, ein valider Prädiktor für die Bewältigung von Studienanforderungen ist (Binder et al., 2021; Buschhüter et al., 2017b; Müller, 2018; Sternal & Walliser, 2020).

Diese Ergebnisse können insbesondere im Hinblick auf solche naturwissenschaftlich-technischen Studiengänge relevant sein, bei denen in einer frühen Studienphase u.a. Physik-Nebenfachveranstaltungen belegt werden müssen, die häufig als besonders herausfordernd gelten.

Studienziel und Testkonstruktion

Vor diesem Hintergrund wird am Fachbereich Physik der TU Darmstadt untersucht, ob und wie stark das physikalische Vorwissen von Studierenden ihren Erfolg in Physik-Nebenfachveranstaltungen beeinflusst. Zu diesem Zweck wurde ein Messinstrument für das physikalische Vorwissen der betreffenden Zielgruppe von Nebenfachstudierenden entwickelt. Innerhalb dieses Vorwissenstests wird das physikalische Vorwissen, orientiert am Wissensmodell nach Hailikari (Hailikari et al., 2007), in drei Wissensbereiche unterteilt und daraus folgend in drei Testteilen erhoben: Faktenwissen, Konzeptwissen und Anwendungswissen. Zusätzlich werden in allen Wissensbereichen Themen aus den Inhaltsfeldern Mechanik, Elektrizitätslehre und Optik behandelt. Die Themenauswahl wurde durch einen Abgleich mit dem Schulcurriculum Physik für das Gymnasium in Hessen (Hessisches Kultusministerium, 2016) und den Lerninhalten der betreffenden Physik-Nebenfachveranstaltungen getroffen. Bei der Auswahl der Fragenformate für die einzelnen Testteile sowie der Fragenkonstruktion fand eine Orientierung an bereits bestehenden Vorwissenstests von Binder et al. (Binder et al., 2019) und Buschhüter et al. (Buschhüter et al., 2017a) statt.

Zwei Iterationsschleifen

Das Messinstrument wurde im Wintersemester 21/22 in vier betreffenden Lehrveranstaltungen pilotiert und in einem ersten Iterationsschritt weiterentwickelt (Schmitt & Spatz, im Druck). Dabei wurde eine vollständige Rasch-Analyse der Daten vorgenommen, um das Testinstrument anhand der berechneten Personenfähigkeiten und Itemschwierigkeiten zu skalieren. Daraus resultierend konnten Fragen überarbeitet und die Testlänge deutlich reduziert werden. Auf Grund des zu hohen Schwierigkeitsgrades für die Stichprobe im Anwendungswissen, wurde das Fragenformat dieses Testteils konzeptionell überarbeitet (Schmitt & Spatz, im Druck). Eine anschließende Erhebung im Sommersemester 2022 in einer weiteren Physik-Nebenfachveranstaltung mit dem überarbeiteten Messinstrument bietet nun die Datengrundlage für eine weitere statistische Analyse nach der probabilistischen Testtheorie und eine zweite Iterationsschleife, die in den folgenden Abschnitten zusammenfassend dargestellt wird.

Stichprobe

Aus der Erhebung im Sommersemester 2022 konnten insgesamt $n=113$ gültig ausgefüllte Vorwissenstests in der Veranstaltung „Physik für Umwelt- und Bauingenieurwesen“ erfasst werden (50 weiblich, 63 männlich). Die rein digitale Erhebung fand zwischen der ersten und dritten Woche nach Vorlesungsbeginn mit einem zeitlichen Umfang von ca. 45 Minuten statt. Im Gegensatz zu den Lehrveranstaltungen der ersten Iterationsschleife, die für das erste („Physik für Elektrotechnik“ und „Physik für Biologiestudierende“) bzw. dritte Fachsemester („Physik für Maschinenbau“ und „Physik für Chemiestudierende“) vorgesehen sind, ist die Veranstaltung im Sommersemester im zweiten Fachsemester angesiedelt. Eine Besonderheit dieser Veranstaltung ist zudem, dass die Bestehensquoten in der Abschlussprüfung häufig besonders niedrig ausfallen (teilweise nur ca. 35% Bestehensquote).

Statistische Datenanalyse

Die Daten werden nach der probabilistischen Testtheorie anhand des Rasch-Analyseverfahrens ausgewertet. Dafür wird die Software „R“ mit den dazugehörigen Paketen „mirt“ und „WrightMap“ verwendet. Als eine Grundvoraussetzung für das statistische Verfahren gilt die Unidimensionalität des zu analysierenden Tests (Wright, 1997). Daher werden die drei verschiedenen Testteile statistisch getrennt voneinander betrachtet.

Zur Auswahl des IRT-Modells wird der M2-Modelfit nach Maydeu-Olivares (2013) berechnet (Chen et al., 2018). Als passendstes Modell zum Antwortverhalten der Studierenden zeigt sich das zweiparametrische Graded Response Model (Samejima, 1969) mit p - und RMSEA-Werten von 0.01/0.05 (Faktenwissen), 0.26/0.04 (Konzeptwissen) und 0.47/0.00 (Anwendungswissen). Im Gegensatz zum klassischen Rasch-Modell, bei dem alle Items eine einheitliche Trennschärfe zugeordnet bekommen, wird im gewählten Modell neben der Itemschwierigkeit die Trennschärfe jedes einzelnen Items berechnet. Zusätzlich können in diesem Modell alle drei Testteile analysiert werden, da sowohl dichotom bewertete Items (Fakten- und Anwendungswissen) als auch polytom bewertete Items (Konzeptwissen) analysiert werden können.

Bei der vollständigen Datenanalyse werden außerdem die Differential Item Functioning (DIF) für männliche und weibliche Studierende, Personen- und Itemparameter, In- und Outfit-Werte

für Item- und Personfit, sowie die Personenreliabilität des Tests - analog zur Bestimmung der Testreliabilität in der klassischen Testtheorie (Wright & Masters, 1982) - bestimmt.

Ergebnisse der Datenanalyse

Für eine überblicksartige Darstellung der Testteile werden im Folgenden die Mittelwerte der statistischen Werte wiedergegeben.

Im Testteil Faktenwissen weisen Itemschwierigkeit ($M=-0.03$, $SD=1.85$) und Personenfähigkeit ($M=0.05$, $SD=1.32$) ähnliche Mittelwerte auf. Die Items der anderen Testteile, Konzeptwissen ($M=0.34$, $SD=0.83$) und Anwendungswissen ($M=3.30$, $SD=3.98$) sind nach dem Wissensmodell komplexer, sodass die Personenfähigkeiten (Konzeptwissen: $M=0.13$, $SD=0.93$; Anwendungswissen: $M=0.28$, $SD=1.42$) auch stärker vom jeweiligen Mittelwert der Itemschwierigkeiten abweichen.

Für den Itemfit sind Outfit- (Faktenwissen: $M=0.94$, $SD=0.17$; Konzeptwissen: $M=0.81$, $SD=0.17$; Anwendungswissen: $M=0.84$, $SD=0.22$) und Infit-Werte (Faktenwissen: $M=0.97$, $SD=0.06$; Konzeptwissen: $M=0.91$, $SD=0.09$; Anwendungswissen: $M=0.90$, $SD=0.15$) im guten, bzw. sehr guten Bereich (Boone et al., 2014).

Ähnliche Ergebnisse liefert die Berechnung der Personen Outfit- (Faktenwissen: $M=0.94$, $SD=0.32$; Konzeptwissen: $M=0.81$, $SD=0.46$; Anwendungswissen: $M=0.84$, $SD=0.63$) und Infit-Werte (Faktenwissen: $M=0.99$, $SD=0.19$; Konzeptwissen: $M=0.92$, $SD=0.44$; Anwendungswissen: $M=0.91$, $SD=0.49$). Aus den berechneten Personenfähigkeiten kann zudem die Personenreliabilität bestimmt werden. Dafür ergibt sich für das Faktenwissen ein Wert von 0.76, für das Konzeptwissen 0.83 und für das Anwendungswissen 0.66.

Fazit und Ausblick

Auf Grundlage der statistischen Werte kann das Testinstrument für die Hauptstudie im Wintersemester 22/23 bzw. im Sommersemester 23 finalisiert werden. Items mit Kennwerten die nicht im guten bzw. sehr guten Bereich liegen werden dabei überarbeitet. Zusätzlich können insgesamt drei Items beim Faktenwissen, ein Item beim Konzeptwissen und zwei Items im Anwendungswissen gestrichen und dafür jeweils ein Item im Konzept- und im Anwendungswissen aus der ersten Pilotierung hinzugefügt werden. Es bleibt zu überprüfen, ob die geänderten Items weiterhin im akzeptablen bzw. guten Bereich bleiben.

Die Personenreliabilitäten der einzelnen Testteile zeigen für das Faktenwissen eine akzeptable und für das Konzeptwissen eine gute innere Konsistenz. Im Bereich Anwendungswissen weist der Wert von 0.66, analog zur Interpretation von Cronbachs Alpha in der klassischen Testtheorie, auf eine fragwürdige innere Konsistenz hin (Streiner, 2003). Dies ist auf den sehr hohen Schwierigkeitsgrad der Items im Vergleich zu den Personenfähigkeiten zurückzuführen (Linacre, 2022).

Allgemein reproduzieren die Daten außerdem den steigenden Komplexitätsgrad zwischen den Wissensbereichen aus dem zugrunde liegenden Wissensmodell. Dies bildet sich beim Vergleich der Personenfähigkeiten und Itemschwierigkeiten zwischen den verschiedenen Testteilen ab.

Als ein Ausblick auf die Haupterhebung deuten die hohen Werte der Standardabweichungen für die Mittelwerte der Personenfähigkeiten bereits darauf hin, dass die Stichprobe der Nebenfachstudierenden im Hinblick auf ihr physikalisches Vorwissen besonders heterogen zu sein scheint. Neben dieser Betrachtung von Unterschieden im physikalischen Vorwissen

zwischen den Studiengängen, soll des Weiteren eine Untersuchung des Zusammenhangs von physikalischem Vorwissen und dem Erfolg in der Abschlussprüfung der Lehrveranstaltungen erfolgen.

Literatur

- Binder, T., Sandmann, A., Sures, B., Friege, G., Theyssen, H. & Schmiemann, P. (2019). Assessing prior knowledge types as predictors of academic achievement in the introductory phase of biology and physics study programmes using logistic regression. *International Journal of STEM Education*, 6(1). <https://doi.org/10.1186/s40594-019-0189-9>
- Binder, T., Waldeyer, J. & Schmiemann, P. (2021). Studienerfolg von Fachstudierenden im Anfangsstudium der Biologie. *Zeitschrift für Didaktik der Naturwissenschaften*. Vorab-Onlinepublikation. <https://doi.org/10.1007/s40573-021-00123-4>
- Boone, W. J., Staver, J. R. & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer Netherlands. <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=1593304>
- Buschhüter, D., Spoden, C. & Borowski, A. (2017a). Physics knowledge of first semester physics students in Germany: a comparison of 1978 and 2013 cohorts. *International Journal of Science Education*, 39(9), 1109–1132. <https://doi.org/10.1080/09500693.2017.1318457>
- Buschhüter, D., Spoden, C. & Borowski, A. (2017b). Studienerfolg im Physikstudium: Inkrementelle Validität physikalischen Fachwissens und physikalischer Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 23(1), 127–141. <https://doi.org/10.1007/s40573-017-0062-7>
- Chen, F., Liu, Y., Xin, T. & Cui, Y. (2018). Applying the M2 Statistic to Evaluate the Fit of Diagnostic Classification Models in the Presence of Attribute Hierarchies. *Frontiers in psychology*, 9, 1875. <https://doi.org/10.3389/fpsyg.2018.01875>
- Hailikari, T., Nevgi, A. & Lindblom-Ylänne, S. (2007). EXPLORING ALTERNATIVE WAYS OF ASSESSING PRIOR KNOWLEDGE, ITS COMPONENTS AND THEIR RELATION TO STUDENT ACHIEVEMENT: A MATHEMATICS BASED CASE STUDY. *Studies in Educational Evaluation*, 33(3-4), 320–337. <https://doi.org/10.1016/j.stueduc.2007.07.007>
- Hessisches Kultusministerium. (2016). *Kerncurriculum gymnasiale Oberstufe*.
- Heublein, U., Huttsch, C. & Schmelzer, R. (2022). *Die Entwicklung der Studienabbruchquoten in Deutschland*. https://doi.org/10.34878/2022.05.DZHW_BRIEF
- Heublein, U., Richter, J. & Schmelzer, R. (2020). DZHW-Brief 03|2020 - Die Entwicklung der Studienabbruchquoten an den deutschen Hochschulen. *DZWH Brief(03)*.
- Linacre, J. M. (2022). *A User's Guide to Winsteps: Rasch-Model Computer Programs*. <https://www.winsteps.com/manuals.htm>
- Maydeu-Olivares, A. (2013). Goodness-of-Fit Assessment of Item Response Theory Models. *Measurement: Interdisciplinary Research & Perspective*, 11(3), 71–101. <https://doi.org/10.1080/15366367.2013.831680>
- Müller, J. (2018). *Studienerfolg im Fach Physik: Zusammenhang zwischen Modellierungskompetenz und Studienerfolg* [Dissertation]. Universität Duisburg-Essen.
- Müller, J., Dammann, E., Fischer, H. E., Lang, M., Borowski, A., Lorke, A. & Menkenhagen, J. (2016). Physikalisch-mathematische Modellierung in der Physik und im Bauingenieurwesen als Prädiktor für Studienerfolg. In Christian Maurer (Vorsitz), *Implementation fachdidaktischer Innovation im Spiegel von Forschung und Praxis: Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Zürich 2016*. Symposium im Rahmen der Tagung von Gesellschaft für Didaktik der Chemie und Physik.
- Samejima, F. (1969). *Estimation of Latent Ability using a response Pattern of graded Scores*. *Psychometric Monograph: Bd. 17*. The William Byrd Press. <https://www.psychometricsociety.org/sites/main/files/file-attachments/mn17.pdf>
- Schmitt, K. & Spatz, V. (im Druck). Physikalisches Vorwissen in Physik-Nebenfachveranstaltungen. In H. Grötzebauch (Hrsg.), *PhyDid B, Didaktik der Physik Beiträge zur virtuelle DPG-Frühjahrstagung 2022*.
- Sternal, O. & Walliser, N.-O. (2020). Physik-Vorkenntnisse von Studienanfängerinnen und -anfängern in MINT-Fächern. *die hochschullehre*(6), 103–118. http://www.hochschullehre.org/wp-content/files/diehochschullehre_2020_Sternal_Walliser_Physik-Vorkenntnisse_Studienanfangerinnen_MINT.pdf
- Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, 80(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18
- Wright, B. D. (1997). *A History of Social Science Measurement*.
- Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Mesa Pr.