Weitere Kommentare zum Positionspapier

Konstruktvalide Prüfungssitems tragen zur Testfairness bei

Anja Leue und Lara Aylin Petersen

Christian-Albrechts-Universität zu Kiel

Der Artikel von Lindner, Sparfeldt, Köller, Lukas und Leutner (2021) beschreibt umfassend und vielseitig Qualitätsanforderungen an schriftliche Prüfungen. Dabei wird auch die Kriteriums- und Konstruktvalidität von Prüfungssitems zum Thema gemacht. Der vorliegende Kommentar fokussiert auf Beispiele wie die Konstruktvalidität von Prüfungssitems in einem Facettenansatz gestärkt werden kann. Der Kommentar versteht sich nicht als normative Setzung, sondern als Vorschlag zu einer optimierten „Best-Practice“-Strategie schriftliche Prüfungssitems zu entwickeln. Zur Gestaltung schriftlicher konstruktvalider Prüfungssitems können verschiedene kognitive Modelle als konzeptionelle Orientierungspunkte angesehen werden. Das Modell zum Arbeitsgedächtnis (Baddeley, 2003; Oberauer, Süß, Wilhelm & Wittmann, 2003) differenziert die visuell-räumliche Notiztafel und die phonologische Schleife, zwischen denen die zentrale Exekutive vermittelt. In der visuell-räumlichen Notiztafel werden vor allem real präsentierte figurale Informationen verarbeitet und in der phonologischen Schleife vor allem akustisch wahrnehmbare verbale Informationen, rekonstruierbare numerische und imaginierbare figurale Infor-

mationen. Während in schriftlichen Prüfungen vor allem die visuelle Notiztafel aktiviert werden sollte, aktivieren mündliche Prüfungen eher die phonologische Schleife. Ein Curriculum, das daher schriftliche wie mündliche Teilleistungen in einem Studienfach erfasst, kann konzeptionell als testfair(er) betrachtet werden, weil exekutive Funktionen aus den verschiedenen Teilbereichen leistungsrelevant sind. Des Weiteren kann erwartet werden, dass schriftliche Prüfungssitems, die die visuelle Notiztafel ansprechen, vor allem das Wiedererkennen oder den freien Abruf von Wissen erfordern. Demgegenüber lassen mündliche Prüfungssitems erwarten, dass vor allem der Abruf von Informationen und Wissen im Mittelpunkt steht. Prüfungssitems können somit auch in Bezug auf die geforderten kognitiven Funktionen systematisiert werden, wobei eine Gleichverteilung ebenfalls im Sinne von Testfairness wirksam werden sollte. Eine weitere konzeptionelle Analogie für die Entwicklung von Prüfungssitems lässt sich aus dem Inventar zur Gedächtnisdiagnostik ableiten (IGD; Baller, Brand, Kalbe & Kessler, 2006). Das Modul A des IGD beinhaltet beispielweise Untertests wie verbales Gedächtnis, Reproduktion: Text, Rekognition Fi-

guren. Da schriftliche Prüfungsaufgaben zumeist verbal konzipiert sind, wird in schriftlichen Prüfungen vor allem das verbale Gedächtnis beansprucht. Aber auch numerische und figurale Aufgaben ergänzen oft die Items in schriftlichen Prüfungsformaten, wodurch es zur Rekognition von figuralen Informationen kommt oder zum Abruf numerischen Wissens. Für eine Systematik zur Wissensdiagnostik siehe Beauducel und Süß (2011).

Eine weitere Systematisierung von Prüfungssitems lässt sich mit Bezug zum Intelligenz-Struktur-Test 2000 R (Liepmann, Beauducel, Brocke & Amthauer, 2007) aufzeigen. Das Modul zum schlussfolgernden Denken wie auch der Wissenstest sind im Hinblick auf verbale, numerische und figurale Iteminhalte ausgewogen konzipiert. Daraus lässt sich ableiten, dass es auch in Prüfungssitems erstrebenswert ist, diese drei Inhaltsbereiche möglichst gleichhäufig als Itemkontext zu präsentieren. Verbunden mit den Überlegungen zur gleichverteilten Anzahl von Inhalten in Prüfungssitems, empfiehlt es sich eine Facettenstruktur bei der Konstruktion von Prüfungssitems anzustreben (vgl. Guttman & Greenbaum, 1998; Hackett, 2014; Shye, Elizur & Hoffman, 1994; Süß & Beauducel, 2005). Eine weitere vielsprechende Systematik zur inhaltlichen Facettierung schriftlicher Prüfungssitems kann dem Wilde-Intelligenz-Test 2 (WIT-2; Kersting, Althoff & Jäger 2008) entnommen werden. In Prüfungsfragen mit offenen Antworten wird die Reproduktion (gelernten) Textes gefordert. Bei der Vervollständigung von Grafiken wird die Rekognition von Grafiken aus den Lehrmaterialien gefordert. Prüfungssitems zu Jahreszahlen oder Kennwerten erfordern die Rekognition numerischer Inhalte. Im Sinne eines facettentheoretischen Ansatzes könnte somit eine Gleichverteilung von Prüfungssitems aus den Strukturen verbal, numerisch und figural in zukünftigen Prüfungsformaten angestrebt werden. Eine weitere Facettenstruktur in Prüfungssitems bietet sich für die geforderten kognitiven Funktionen an (z. B. Abruf, Wiedererkennen, Supervidieren im Sinne des Ergänzens fehlender Informationen, vgl. Oberauer et al., 2003). Daraus ergibt sich für die Entwicklung von Prüfungssitems beispielsweise eine explizierbare Facettenstruktur für 3 Inhaltsbereiche (verbal, numerisch, figural) x 3 Kognitive Funktionen (Abruf, Wiedererkennen, Rekognition). Diese oder ähnliche Facettenstrukturen könnten für verschiedene Teilfächer der Psychologie erprobt werden und hinausgehend über die Konstruktvalidität zur weiteren Förderung

der Testfairness in schriftlichen Prüfungssitems beitragen (Kane & Bridgeman, 2017).

Literatur

- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews*, 4, 829–839.
- Baller, G., Brand, M., Kalbe, E. & Kessler, J. (2006). *Inventar zur Gedächtnisdiagnostik*. Göttingen: Hogrefe.
- Beauducel, A. & Süß, H.-M. (2011). Wissensdiagnostik: Allgemeine und spezielle Wissenstests. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Leistungs-, Intelligenz- und Verhaltensdiagnostik (Enzyklopädie der Psychologie. Themenbereich B, Methodologie und Methoden. Serie II, Psychologische Diagnostik, Bd. 3, S. 235–273)*. Göttingen: Hogrefe.
- Guttman, R. & Greenbaum, C. W. (1998). Facet theory: Ist development and current status. *European Psychologist*, 3, 13–34.
- Hackett, P. M. W. (2014). *Facet theory and the mapping sentence: Evolving philosophy, use and application*. Hampshire: Palgrave Macmillan.
- Kane, M. & Bridgeman, B. (2017). Research on validity theory and practice at ETS. In R. E. Bennett & M. von Davier (Eds.), *Advancing human assessment: The methodological, psychological and policy contributions of ETS*. (chap. 16). Cham: SpringerOpen.
- Kersting, M., Althoff, K. & Jäger, A. O. (2008). *Wilde-Intelligenz-Test 2*. Göttingen: Hogrefe.
- Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R*. Göttingen: Hogrefe.
- Lindner, M. A., Sparfeldt, J. R., Köller, O., Lukas, J. & Leutner, D. (2021). Ein Plädoyer zur Qualitätssicherung schriftlicher Prüfungen im Psychologiestudium. *Psychologische Rundschau*, 72, 93–105. <https://doi.org/10.1026/0033-3042/a000524>
- Oberauer, K., Süß, H.-M., Wilhelm, O. & Wittmann, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, 31, 167–193.
- Shye, S., Elizur, D. & Hoffman, M. (1994). *Introduction to facet theory: Content design and intrinsic data analysis in behavioral research*. Thousand Oaks, CA: Sage.
- Süß, H.-M. & Beauducel, A. (2005). Faceted models of intelligence. In O. Wilhelm & R. Engle, (Eds.), *Understanding and measuring intelligence* (pp. 313–332). London: Sage.

Prof. Dr. Anja Leue

Lehrstuhl für Psychologische Diagnostik,
Differentielle und Persönlichkeitspsychologie
Institut für Psychologie
Christian-Albrechts-Universität zu Kiel
Olshausenstraße 75
24118 Kiel
leue@psychologie.uni-kiel.de

<https://doi.org/10.1026/0033-3042/a000533>