Ein langer Weg zur Qualitätssicherung von Hochschulprüfungen

Stellungnahme zum Diskussionsforum

Marlit A. Lindner¹, Jörn R. Sparfeldt², Olaf Köller¹, Josef Lukas³ und Detlev Leutner⁴

¹IPN – Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik, Kiel

²Universität des Saarlandes, Saarbrücken

³Martin-Luther-Universität Halle-Wittenberg

⁴Universität Duisburg-Essen

Wir freuen uns über die große Resonanz in diesem Diskussionsforum und bedanken uns für die zahlreichen Beiträge, die zusätzliche Perspektiven und wertvolle Anknüpfungspunkte aufzeigen und dabei eine weitgehende Unterstützung unserer Anliegen zur Sicherung der Qualität schriftlicher Prüfungen im Psychologiestudium und

an Hochschulen im Allgemeinen erkennen lassen. Aufgrund der Zeichenbegrenzung können wir nicht auf alle Kommentare und Aspekte in wünschenswertem Umfang eingehen. Wir reflektieren kurz zentrale Diskussionspunkte und beziehen zu zwei stärker kontrovers diskutierten Themen Position: (1) Kollaborative Aufgabenent-

wicklung sowie Aufbau und Nutzung geteilter Aufgabenpools und (2) Innovative Prüfungstechnologien.

Desiderate zur Qualitätssicherung

Ziel unseres Diskussionsbeitrags war es, einen offenen Diskurs zu essenziellen Aspekten der Prüfungsqualität an Hochschulen anzuregen. Wir haben Themen und empirisch gut gestützte Erkenntnisse ins Zentrum gestellt, deren Berücksichtigung man zur Sicherung hochwertiger Leistungsdiagnostik im Rahmen von (schriftlichen) Prüfungen im Hochschulstudium als fundamental ansehen kann. Die in den Kommentaren angesprochenen Themen bilden wichtige Ergänzungen und Vertiefungen. Sie greifen u. a. auch zentrale psychologie-spezifische Punkte auf und geben anregende Impulse für den weiteren Austausch.

Wiederholt benannt und von Busch, Degé, Karbach, Aschersleben und Ebersbach (2021) ausführlich besprochen wurden die strukturellen und vielfältigen inhaltlichen Herausforderungen für Prüfende. Dazu zählen mangelnde personelle und technische Ressourcen bei hohen Studierendenzahlen und teils fehlleitende Anreizsysteme im Spannungsfeld von Lehre und wissenschaftlicher Arbeit. Hinzu kommen mikropolitisch gesteuerte und aus empirischer Sicht teils unsinnige Regelungen in Prüfungsordnungen. Ein hochschulpolitischer, aber fachlich auf Erkenntnissen der Forschung zur Leistungsdiagnostik basierender Dialog ist für die Lösung solcher Probleme unumgänglich.

In diesem Zusammenhang muss auch die Tragweite von Prüfungen und die Notwendigkeit prognostischer Validität für die Vergabe u. a. von Masterstudienplätzen und Berufschancen mehr in den (politischen) Fokus rücken, wie von Schult (2021) differenziert angesprochen. In mehreren Beiträgen wurde in diesem Zusammenhang auch das Problem der Noteninflation in der Psychologie thematisiert, deren engagierter Entgegenwirkung wir ausdrücklich zustimmen. Ebenso teilen wir die Auffassung, dass die Aufklärung bezüglich der Chancen und Grenzen verschiedener Prüfungsformate über Fachgrenzen hinweg entscheidend für eine nachhaltige Verbesserung der allgemeinen Prüfungspraxis ist. Bedarf sehen wir, in Einklang mit Wilhelm, Kretzschmar, Montag und Rauthmann (2021) und Busch et al. (2021), in der Entwicklung und Verbreitung qualitätssichernder Empfehlungen für alle gängigen Prüfungsformate, wobei die Pädagogische Psychologie sowie die Differenzielle Psychologie und die Psychologische Diagnostik federführende Rollen einnehmen sollten.

Explizit hervorheben möchten wir nochmals, dass die adäquate Auswahl von *Aufgabeninhalten* essenzieller Kern valider Hochschulprüfungen ist. Dies spiegelt sich auch in den Anregungen zu einer systematischen kognitionspsychologischen Betrachtung von Aufgabenanforderungen und zur sorgsam Auswahl inhaltlicher Facetten wider (vgl. Wilhelm et al., 2021; Leue & Petersen, 2021). Diagnostisch gesehen definiert das Universum der Aufgaben das zu messende Konstrukt im Sinne eines formativen Messmodells. Dies unterscheidet sich von psychologisch-diagnostischen Situationen, denen ein reflektives Messmodell und die Annahme der Eindimensionalität zugrunde liegen. Kompetenzorientierte Prüfungen müssen für eine valide Messung die Tiefe und die Breite eines Faches widerspiegeln. In diesem Sinne ist auch die Auswahl der Prüfungsaufgaben und -anforderungen vorzunehmen, was sich nicht unbedingt mit einer rein „psychometrisch“ orientierten Testzusammenstellung deckt, wie sie beispielsweise im Rahmen computerbasierter adaptiver Prüfungen vorgeschlagen wird (vgl. Fink, Nauermann & König, 2021). In adaptiven Prüfungen liegt einer individualisierten Aufgabenauswahl die (in früheren Kohorten bestimmte) Schwierigkeit, nicht jedoch eine inhaltliche Überlegung der Prüfungszusammenstellung zugrunde.

Eine anstehende Herausforderung für die Psychologie ist neben der kompetenzorientierten Ausrichtung von Prüfungen die zusätzliche Prüfung von handlungsorientierten Kompetenzen im Rahmen beispielsweise der psychotherapeutischen Qualifizierung mit dem Ziel der Approbation, wie von Rief (2021) dargestellt. Nicht nur hier besteht großes Potenzial für digital gestützte Aufgabenformate und gleichzeitig entsprechend viel Forschungsbedarf zur Validierung solcher Verfahren.

Kollaborative Aufgabenentwicklung

Bei unserem Vorschlag zum Aufbau und zur Nutzung gemeinsamer Aufgabenpools haben wir uns missverständlich ausgedrückt und möchten unsere Position klarstellen. Unser Vorstoß war weder gemeint im Sinne der Einrichtung eines übergeordneten, externen Qualitätsmanagements nach Vorbild des *Instituts für Medizinische und Pharmazeutische Prüfungsfragen* (IMPP) noch im Sinne einer notwendigen Einigung auf fachspezifische Kerncurricula, um eine Vergleichbarkeit zwischen Standorten zu erreichen (vgl. die Diskussion von Gollwitzer, Gleibs, Häfner, Lange & Schneider, 2021; Spinath et al., 2021; Wilhelm et al., 2021).

Gemeint von uns war vor allem das Potenzial einer systematischen Steigerung der Effizienz für Prüfende, in-

dem Kooperation nicht nur auf Ebene der Forschung, sondern auch auf Ebene der Lehre und der Prüfungs-gestaltung aktiv gefördert wird. Die gemeinsame Speisung von fachspezifischen Aufgabenpools, die einer freiwilligen Pflege und Nutzung im Sinne von Open Educational Resources unterliegen sollten, wäre vor diesem Hintergrund eine Option für alle teilnahmewilligen Prüfenden, um sich durch Aufgaben anderer inspirieren zu lassen. Im besten Falle wäre die Nutzung verbunden mit der Bereitstellung eigener Aufgaben und einer Rückmeldung basaler psychometrischer Parameter, was mittelfristig empirisch fundierte Einschätzungen der Qualität von Aufgaben und ihrer Schwierigkeiten erlauben würde. Der DGPs (bzw. den Fachgruppen) könnte in diesem Sinne vor allem eine organisatorisch vermittelnde Rolle zukommen, beispielsweise bei der Einrichtung geeigneter Repositorien.

Allgemein stimmen wir der Einschätzung zu, dass die Qualitätssicherung von Prüfungen vor allem selbstbestimmt an den einzelnen Standorten und dabei vorrangig auf konzeptionell-diagnostischer Ebene ablaufen sollte. Wie von Spinath et al. (2021) aufgebracht, wäre dem niederländischen Vorbild folgend eine fachübergreifende prüfungsdidaktische Beratungsstelle an Hochschulen in Deutschland besonders zu begrüßen.

Die Freiheit der Lehre ist ein hohes Gut, und es besteht aus unserer Sicht keine Notwendigkeit, in diese durch gemeinsame Curricula und Prüfungskriterien einzugreifen, wie es auch Gollwitzer et al. (2021) darlegen. Es wird sicher immer ein gewisses Dilemma geben im Hinblick auf den Wunsch der Vergleichbarkeit von Studierendenleistungen über Standorte hinweg und der anzustrebenden Kongruenz spezifischer Lehrinhalte mit entsprechenden Prüfungsinhalten. Dennoch könnte eine Verständigung auf abstrakte Kerne von Fachcurricula hilfreich sein, um Standards im Sinne des kleinsten gemeinsamen Nenners zu definieren. So könnte, basierend auf einem geteilten (kompetenzorientierten) Fachverständnis, die Vergleichbarkeit von Studienleistungen verbessert werden, während die thematische Ausgestaltung selbstverständlich weiterhin ganz in der Hand der einzelnen Lehrenden liegen würde. Denkbar wäre darüber hinaus die gemeinsame Erstellung von Tests für die Vergabe von Masterstudienplätzen (vgl. Schult, 2021) als potenzielle Ergänzung zur Bachelornote.

Innovative Prüfungstechnologien

Die Zukunft der (schriftlichen) Hochschulprüfungen ist digital und von technischen Innovationen geprägt, davon sind wir überzeugt. Wir sehen jedoch auch, dass aktuell noch grundlegende didaktische und strukturelle Heraus-

forderungen zu bearbeiten sind, bevor computerbasierte Prüfungsformate und komplexe technische Tools den hochgesteckten Erwartungen an Effizienz- und Validitätsverbesserungen in der Breite gerecht werden können. Wenngleich unsere Empfehlungen die Zukunft der Prüfungspraxis begleiten sollten, haben wir uns daher bewusst dagegen entschieden, vor allem Zukunftsmusik zu diskutieren.

Sicherlich werden datengetriebene Systeme und künstliche Intelligenzen in den kommenden Jahren eine wichtige Rolle bei der Prüfungserstellung, -administration und -auswertung spielen (z.B. automatisierte Aufgabenkonstruktion, Vorkorrektur offener Antworten). Eine einfache Computerisierung von Papierprüfungen ist schon vielfach leicht möglich. Zunehmend wird aber beispielsweise die Einbindung von interaktiven Aufgabenformaten, multimedialen Elementen, automatisiertem Feedback und Gamifizierung (u.a. in virtuellen Realitäten) spannende Neuerungen bringen, die für sich genommen eigene Beiträge und insbesondere auch mehr Forschung verdienen. Denn vor allem sollte technische Innovation in klug durchdachte didaktische Konzepte eingebunden sein und, wie auch Wilhelm et al. (2021) betonen, nicht nur der Technik wegen eingesetzt werden. Zudem bringen Formate wie Distanzprüfungen administrative und juristische Hürden mit sich, beispielsweise im Hinblick auf die Vermeidung und Aufklärung von Täuschungsversuchen.

Nach wie vor können wir der überaus positiven Sicht auf computerbasiertes adaptives Testen im Rahmen summativer Prüfungen an Hochschulen, wie von Fink et al. (2021) und Frey, Spoden und Schultze (2021) vertreten, nicht zustimmen. Grund ist, dass Hochschulprüfungen der Logik formativer Messmodelle folgen, in denen der gesamte Aufgabenpool und nicht eine interindividuell variierende Auswahl daraus das zu erfassende Merkmal (im Sinne einer Kompetenz) repräsentiert. Für formative Verfahren, die der semesterbegleitenden Unterstützung erfolgreicher Lernprozesse dienen, sehen wir dagegen durchaus Vorteile adaptiver Testverfahren. Im Einklang mit der Forschung zum Testing-Effekt können diese z.B. helfen, Aufgaben mit individuell passenden Schwierigkeiten vorzulegen, die eine optimale Lerngrundlage darstellen (vgl. Kubik, Marksteiner & Richter, 2021).

Letztlich können computerbasierte Prüfungen aktuell selbst an Leuchtturmstandorten aufgrund von Kapazitätsgrenzen noch nicht flächendeckend in summativen Prüfungen eingesetzt werden. Während also die Zukunft der Hochschulprüfungen mancherorts sicher begonnen hat, so ist dies gemäß unserer Beobachtung bisher noch keine hinreichend geteilte Zukunft. Der Weg zur selbstverständlichen, effizienten Nutzung digitaler Prüfungsformate ist noch mit vielen Herausforderungen gepflastert, die nicht unterschätzt werden sollten.

Ausblick

Allgemein zeichnen sich in diesem Forum die Vielfalt und Breite der Themenfelder, Probleme und potenziellen Lösungsansätze im Zusammenhang mit Hochschulprüfungen ab. Wir teilen die Auffassung, dass die Psychologie hier als Schlüsseldisziplin anzusehen ist und großen Einfluss auf angrenzende Fächer, insbesondere auch auf die Ausbildung angehender Lehrkräfte (vgl. Drechsel, Dörfler, Berthold & Dutke, 2021), haben kann und auch haben sollte. Diese Schlüsselrolle beinhaltet unter anderem die didaktisch-diagnostische Beratung von Hochschulen sowie zukunftsweisende Forschung im Bereich der Leistungsmessung und zugehöriger Phänomene – Forschung, die sich nicht zuletzt auch auf Chancen und Grenzen digitaler Prüfungsformate beziehen sollte.

Literatur

- Busch, J., Degé, F., Karbach, J., Aschersleben, G. & Ebersbach, M. (2021). Fachgruppe Entwicklungspsychologie. Kommentar zu Lindner, M. A. et al. (2021). Ein Plädoyer zur Qualitätssicherung schriftlicher Prüfungen im Psychologiestudium. *Psychologische Rundschau*, 72, 110–112. <https://doi.org/10.1026/0033-3042/a000527>
- Drechsel, B., Dörfler, T., Berthold, K. & Dutke, S. (2021). Kommission Psychologie in Lehramtsstudiengängen. Qualitätssicherung schriftlicher Prüfungen nicht nur in der, sondern auch durch die Psychologie. *Psychologische Rundschau*, 72, 118–120. <https://doi.org/10.1026/0033-3042/a000530>
- Fink, A., Naumann, P. & König, C. (2021). Computerisierte adaptive Klausuren im Psychologiestudium. *Psychologische Rundschau*, 72, 125–127. <https://doi.org/10.1026/0033-3042/a000532>
- Frey, A., Spoden, C. & Schultze, M. (2021). Fachgruppe Methoden und Evaluation. Die Zukunft der Hochschulklausuren hat bereits begonnen. *Psychologische Rundschau*, 72, 112–116. <https://doi.org/10.1026/0033-3042/a000528>
- Gollwitzer, M., Gleibs, I., Häfner, M., Lange, J. & Schneider, D. (2021). Fachgruppe Sozialpsychologie. Kommentar zu Lindner, M. A. et al. (2021). *Psychologische Rundschau*, 72, 116–118. <https://doi.org/10.1026/0033-3042/a000529>
- Kubik, V., Marksteiner, T. & Richter, T. (2021). Fachgruppe Pädagogische Psychologie – Qualitätssicherung schriftlicher Prüfungen in der Psychologie. Ein wichtiges Thema mit einigen offenen Fragen. *Psychologische Rundschau*, 72, 106–108. <https://doi.org/10.1026/0033-3042/a000525>
- Leue, A. & Petersen, L. A. (2021). Konstruktvalide Prüfungssitemen tragen zur Testfairness bei. *Psychologische Rundschau*, 72, 122–123. <https://doi.org/10.1026/0033-3042/a000533>
- Rief, W. (2021). Kommentar zu Lindner, M. A. et al. (2021). Ein Plädoyer zur Qualitätssicherung schriftlicher Prüfungen im Psychologiestudium. *Psychologische Rundschau*, 72, 124–125. <https://doi.org/10.1026/0033-3042/a000534>
- Schult, J. (2021). Zur prognostischen Funktion von Prüfungsleistungen bei der Master-Studierendenauswahl. Kommentar zu Lindner, M. A. et al. (2021). Ein Plädoyer zur Qualitätssicherung schriftlicher Prüfungen im Psychologiestudium. *Psychologische Rundschau*, 72, 127–128. <https://doi.org/10.1026/0033-3042/a000535>
- Spinath, B., Antoni, C. H., Bermeitinger, C., Bühner, M., Elsner, B., Erdfelder, E. et al. (2021). Kommission „Studium und Lehre“. Kommentar zu Lindner, M. A. et al. (2021). „Ein Plädoyer zur Qualitätssicherung schriftlicher Prüfungen im Psychologiestudium. *Psychologische Rundschau*, 72, 121–122. <https://doi.org/10.1026/0033-3042/a000531>
- Wilhelm, O., Kretschmar, A., Montag, C. & Rauthmann, J. (2021). Fachgruppe Differentielle, Persönlichkeitspsychologie und Diagnostische Psychologie. Weiterentwicklungen psychologischer Prüfungsformate aus diagnostischer Sicht. *Psychologische Rundschau*, 72, 108–110. <https://doi.org/10.1026/0033-3042/a000526>

Dr. Marlit Annalena Lindner

IPN – Leibniz-Institut für die Pädagogik
der Naturwissenschaften und Mathematik
Olshausenstraße 62
24118 Kiel
mlindner@leibniz-ipn.de

<https://doi.org/10.1026/0033-3042/a000536>