

**Qualitätssicherung schriftlicher Prüfungen in der Psychologie: Ein wichtiges Thema mit einigen
offenen Fragen**

(Quality assurance of written examinations in psychology: An important topic with some open
questions)

zur Publikation angenommen in der Zeitschrift Psychologische Rundschau (2021)

Veit Kubik^a, Tamara Marksteiner^b & Tobias Richter^c

(für die Sprechergruppe der Fachgruppe Pädagogische Psychologie der DGPs)

^a Universität Bielefeld, Abteilung für Psychologie (Arbeitseinheit Bildungspsychologie)

^b Universität Mannheim, Professur für Pädagogische Psychologie

^c Universität Würzburg, Lehrstuhl für Psychologie IV

Korrespondenzadresse:

Tobias Richter

Universität Würzburg, Lehrstuhl für Psychologie IV

Röntgenring 10

97070 Würzburg

E-Mail: tobias.richter@uni-wuerzburg.de

Qualitätssicherung schriftlicher Prüfungen in der Psychologie: Ein wichtiges Thema mit einigen offenen Fragen

Wir begrüßen den Diskussionsbeitrag von Lindner et al. (2021), der die zentrale Stellung von Prüfungen und die Notwendigkeit einer Abstimmung von Fachinhalten, Lehrzielen und Berufsanforderungen im Hochschulstudium der Psychologie herausstellt (s. auch Spinath et al., 2018). Trotz ihrer immensen Bedeutung und der hohen Expertise, die die psychologische Diagnostik (und dabei nicht zuletzt auch die pädagogisch-psychologische Diagnostik) hier einbringen kann, sind Hochschulprüfungen und ihre Qualitätssicherung bislang erstaunlich selten Gegenstand fachinterner Diskussionen.

Lindner et al. (2021) stellen evidenzbasierte Empfehlungen für die effiziente und qualitativ hochwertige Erstellung und Auswertung von schriftlichen Prüfungsformaten in Präsenz- und Digitalform zur Diskussion und berücksichtigen dabei besonders standardisierte, geschlossene Aufgaben im Antwort-Wahl-Format, wie sie in universitären (Massen-)Prüfungen besonders verbreitet sind. Aus der Perspektive der Fachgruppe Pädagogische Psychologie unterstützen wir die Vorschläge zur Sicherung einer guten Prüfungspraxis im Fach Psychologie – wie die Etablierung von Kooperationsstrukturen sowie die Einrichtung von Korrekturzirkeln und Aufgabendatenbanken. Allerdings ist die konkrete Umsetzung der Vorschläge sicherlich nicht einfach. So sind vielfältige Probleme zu bewältigen, die von der Schaffung der nötigen Infrastruktur und der Beteiligung einer kritischen Masse von Kolleg(inn)en bis hin zu Fragen der Datensicherheit und des Prüfungsrechts reichen. Zur Lösung dieser Probleme ist eine instituts- und fachübergreifende Anstrengung erforderlich. Die Pädagogische Psychologie beteiligt sich gerne an einer solchen Initiative.

Drei Aspekte, die in dem Papier nur gestreift werden, berühren wichtige offene Fragen und verdienen aus unserer Sicht eine differenziertere Diskussion:

1. Lindner et al. (2021) weisen zurecht darauf hin, dass *Aufgaben mit offenem Antwortformat* für bestimmte diagnostische Zwecke besser geeignet sein können als Aufgaben mit geschlossenem Antwortformat. Im Papier werden als Beispiele etwa die Messung kreativ-schöpferischer Leistungen oder von Problemlösekompetenzen genannt, die ausschließlich bzw. zumindest einfacher mit offenen Antwortformaten erfasst werden können. Hinzufügen könnte

man noch die für die Psychologie zentrale Fähigkeit des wissenschaftlichen Denkens und Argumentierens (in all seinen unterschiedlichen Facetten; Britt et al., 2014; Fischer et al., 2018). Für die Messung dieser Fähigkeit dürften offene Aufgaben in vielen Fällen ebenfalls besser geeignet sein. Hier stellt sich die komplexe Frage, wie offene Aufgaben in universitären Prüfungen gestaltet werden können, damit sie ihre Vorteile ausspielen können, zugleich aber ein befriedigendes Niveau an Objektivität, Validität und Testfairness und eine ökonomische Durchführbarkeit erreicht werden können. Auf diese Frage müssen gute Antworten gefunden werden, weil in manchen Bereichen offene Aufgaben nach wie vor sehr verbreitet sind bzw. das einzige akzeptierte Prüfungsformat darstellen (z.B. in schriftlichen Prüfungen im Rahmen des ersten Staatsexamens in Bayern). Ein vielversprechender Ansatz der Wissensdiagnostik auf Basis offener Antworten ist eine automatisierte, auf künstliche Intelligenz gestützte Auswertung, die aber bei weitem nicht alle diagnostischen Probleme lösen kann und derzeit eher für eine lernbegleitende Diagnostik geeignet scheint als für den Einsatz in schriftlichen Prüfungen (z.B. Seifried et al., 2016).

2. Die *Nutzung von Tests als Lerngelegenheiten* wird derzeit in der Pädagogischen Psychologie und der angewandten Kognitionspsychologie intensiv beforscht (siehe Metaanalysen zum Testeffekt von Adesope et al., 2017 und Schwierien et al., 2017). Wie Lindner et al. anmerken, wird das enorme lernförderliche Potenzial von semesterbegleitenden „formativen Tests“ als didaktisches Mittel bzw. Lernstrategie in der universitären Lehre noch längst nicht ausgeschöpft (Dunlosky & Rawson, 2015). Allerdings haben solche formativen Tests aus unserer Sicht wenig mit schriftlichen Prüfungen zu tun, die primär der Leistungsfeststellung dienen. Damit formative Tests als Lerngelegenheiten und diagnostische Lernstandserhebung im weitgehend selbstregulierten Lernsetting an der Hochschule von den Studierenden akzeptiert werden und lernförderlich wirken können, sollten sie nicht als Prüfungsleistung deklariert und bewertet werden. Bei der Gestaltung von Aufgaben für Tests als Lerngelegenheiten müssen zudem andere Kriterien beachtet werden als bei der Gestaltung von Aufgaben, die der Leistungsfeststellung dienen. Um lernförderliche Effekte zu erzielen, müssen Aufgaben beispielsweise den aktiven Abruf aus dem Langzeitgedächtnis fördern, zugleich aber so einfach

sein, dass in den meisten Fällen auch die richtige Antwort abgerufen werden kann (Greving & Richter, 2018). Auch die Gewährung von Feedback im Zusammenhang mit formativen Tests spielt eine entscheidende Rolle (Kubik et al., 2020), insbesondere wenn die Abrufbarkeit nicht gewährleistet ist und das Feedback mit zusätzlichen Erklärungen angereichert werden kann (Enders et al., 2020). In diesen – und anderen – Hinsichten unterscheiden sich formative Tests von schriftlichen Prüfungen, sodass beide Themen in der Diskussion klar voneinander getrennt werden sollten.

3. Der Beitrag von Lindner et al. liefert zu *digitalen Prüfungsformaten* einen wichtigen Impuls, dessen Relevanz gerade in der aktuellen Situation deutlich wird: Durch die Digitalisierung der Lehre, verstärkt durch die Corona-Pandemie, erlebt auch die Praxis von Hochschulprüfungen einen deutlichen Schub hin zu einem vermehrten Einsatz von digitalen Prüfungen. Lindner et al. bewerten digitale Prüfungen als besonders positiv, da geschlossene Aufgaben automatisiert ausgewertet werden und komplexe Materialien besonders einfach dargestellt werden können. Betrachtet man hier auch asynchrone und synchrone Online-Klausuren, sind allerdings noch weitere Aspekte zu berücksichtigen, die zu einer differenzierteren Bewertung von digitalen Prüfungsformaten führen. So erfordern beide Online-Formate einen hohen technischen Aufwand, um prüfungsrechtlich einwandfrei durchgeführt werden zu können. Darüber hinaus sind die Möglichkeiten der Täuschung bei Online-Klausuren enorm viel größer als bei digitalen und nicht digitalen Prüfungen, die unter der Aufsicht von Prüfenden stattfinden. Befunde von Brothen und Peterson (2012) legen nahe, dass die Möglichkeit zur Täuschung in Online-Prüfungen auch genutzt wird, wobei Unterschiede in der Testleistung nicht ausschließlich auf Täuschungsverhalten zurückgeführt werden konnten. Die Motivation zu Täuschung als einem Aspekt akademischem Betrugsverhaltens ließe sich prinzipiell durch das Abschwächen von Leistungszielen zugunsten von Lernzielen verringern (Daumiller & Janke, 2020). Dies dürfte allerdings im Kontext von Prüfungen, deren Ergebnisse für den weiteren Lebensweg der Studierenden eine Bedeutung haben, nur schwer umsetzbar sein. Positive Effekte scheinen Online-Prüfungen jedoch auf die Prüfungsangst von Studierenden mit hoher Testangst zu haben (Stowell & Bennett, 2010). Es empfiehlt sich also, den Einsatz von Online-Prüfungen umsichtig

abzuwägen und auch hier allgemeine Qualitätsstandards zu entwickeln, mit denen die benannten Probleme gelöst werden können.

Literatur

Adesope, O. O., Trevisan, D. A. & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659–701.

<https://doi.org/10.3102/0034654316689306>

Britt, M. A., Richter, T. & Rouet, J.-F. (2014). Scientific literacy: The role of goal-directed reading and evaluation in understanding scientific information. *Educational Psychologist*, 49(2), 104–122. <https://doi.org/10.1080/00461520.2014.916217>

Brothen, T. & Peterson, G. (2012). Online exam cheating: A natural experiment. *International Journal of Instructional Technology and Distance Learning*, 9(2), 15–20.

Daumiller, M. & Janke, S. (2020). Effects of performance goals and social norms on academic dishonesty in a test. *British Journal of Educational Psychology*, 90(2), 537–559.

<https://doi.org/10.1111/bjep.12310>

Dunlosky, J. & Rawson, K. A. (2015). Practice tests, spaced practice, and successive relearning: Tips for classroom use and for guiding students' learning. *Scholarship of Teaching and Learning in Psychology*, 1(1), 72–78. <https://doi.org/10.1037/stl0000024>

Enders, N., Gaschler, R. & Kubik, V. (2020). Online quizzes with closed questions in formal assessment: How elaborate feedback can promote learning. *Psychology Learning & Teaching*. Advance online publication. <https://doi.org/10.1177/1475725720971205>

Fischer, F., Chinn, C. A., Engelmann, K. & Osborne, J. (Eds.). (2018). *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge*. New York: Routledge.

Greving, S. & Richter, T. (2018). Examining the testing effect in university teaching: Retrievability and question format matter. *Frontiers in Psychology*, 9, Article 2412.

<https://doi.org/10.3389/fpsyg.2018.02412>

Kubik, V., Gaschler, R. & Hausman, H. (2020). Enhancing student learning in research and educational practice: The power of retrieval practice and feedback. *Psychology Learning & Teaching*.

Seifried, E., Lenhard, W., Baier, H. & Spinath, B. (2012). On the reliability and validity of human and LSA-based evaluations of complex student-authored texts. *Journal of Educational Computing Research*, 47(1), 67–92. <https://doi.org/10.2190/EC.47.1.d>

Schwieren, J., Barenberg, J. & Dutke, S. (2017). The testing effect in the Psychology classroom: A meta-analytic perspective. *Psychology Learning & Teaching*, 16(2), 179–196. <https://doi.10.1177/1475725717695149>

Spinath, B., Antoni, C., Bühner, M., Elsner, B., Erdfelder, E., Fydrich, T. et al. (2018). Empfehlungen zur Qualitätssicherung in Studium und Lehre. Verabschiedet vom Vorstand der DGPs am 20. April 2018. *Psychologische Rundschau*, 69, 183–192. <https://doi.org/10.1026/0033-3042/a000408>

Stowell, J. R. & Bennett, D. (2010). Effects of online testing on student exam performance and test anxiety. *Journal of Educational Computing Research*, 42(2), 161–171. <https://doi.org/10.2190/EC.42.2.b>