



30.11.2021

Anonymisierung von Prüfungsleistungen

Dr. Sebastian Tillmann

**Referat für Gleichstellung, Familienförderung und
Diversity**

Hinweis zur Methode

Vorab möchte ich auf Folgendes hinweisen. Die hier verfasste Zusammenstellung stellt keinen systematischen Überblick der Forschungslage dar und erhebt nicht den Anspruch, die gesamte Forschung zu dieser Thematik erfasst zu haben. Vielmehr handelt es sich um einen kurzen Überblick relevanter Befunde, die allerdings, wo immer möglich, auf umfassende systematische Reviews und Meta-Analysen gestützt werden.

Anonymisierung von Prüfungsleistungen

Die menschliche Fähigkeit große Datenmengen aufzunehmen und zu interpretieren ist sehr eingeschränkt, unsere Aufmerksamkeit ist begrenzt und die Anzahl von Dingen auf die wir uns gleichzeitig konzentrieren können ist extrem begrenzt (Marois & Ivanoff, 2005). Um mit dieser Informationsflut fertig zu werden besitzen Menschen eine Reihe von adaptiven kognitiven Mechanismen, um ihre Umgebung zu interpretieren, die auch als Heuristiken bezeichnet werden (vom griechischen „*heurísko*“, d.h. „ich finde“). Dabei handelt es sich um Faustregeln oder Short-Cuts, die es erlauben Berücksichtigung mancher, aber nicht aller verfügbaren Informationen Entscheidungen mit hoher Genauigkeit zu treffen (Gigerenzer & Brighton, 2009).

Während diese Heuristiken im alltäglichen Leben zumeist sehr zuverlässig funktionieren, gibt es besondere Kontexte, in denen dieser Prozess nicht immer ausreichend ist oder von anderen Störfaktoren beeinflusst werden kann. Welche Informationen zur Entscheidung herangezogen werden, unterliegt einer Reihe von Verzerrungen, die zwar evolutionär adaptiv waren, aber im heutigen Kontext oftmals maladaptiv sein können, wie z.B. die natürliche Bevorzugung von Personen, mit denen man verwandt ist. Das ist unter anderem dann von Bedeutung von Bewertungen von Prüfungsleistungen vorgenommen werden müssen, die ja den Anspruch haben so objektiv wie möglich und vor allem fair zu sein.

Hierbei können eine ganze Reihe kognitiver Verzerrungen (engl. „*Biases*“) zum Tragen kommen, die dazu führen, dass Bewertungen an Objektivität verlieren und gleichzeitig die faire Bewertung aller Studierenden nicht mehr gewährleistet ist. Nachfolgend stelle ich exemplarisch einige diese Biases und ihre möglichen Auswirkungen kurz vor:

Bestätigungsfehler (Hart et al., 2009; Nickerson, 1998): Die Tendenz, aktiv nach Informationen zu suchen, die das eigene Weltbild bestätigen. Dies kann dazu führen, dass Prüfungsleistungen nicht für sich genommen betrachtet werden, sondern im Kontext der bisherigen Leistungen von Studierenden, z.B. wenn eine vorhergegangene Präsentation ausgezeichnet war, besteht die Gefahr, dass Dozierende in späteren Prüfungsleistungen eher nach positiven Elementen suchen, die das vorhergehende Bild bestätigen.

Heiligenschein/Teufelshörnereffekt (Gräf & Unkelbach, 2018; Nisbett & Wilson, 1977): Eine Person wird besser/schlechter bewertet weil von bekannten Eigenschaften (z.B. verbaler Schlagfertigkeit, freundliches und professionelles Auftreten) auf unbekannte Eigenschaften (z.B. Qualität von schriftlichen Hausarbeiten) geschlossen wird, obwohl eine direkte kausale Verbindung zwischen beiden Eigenschaften nicht anzunehmen ist.

In-Group Bias (Efferson, Lalive & Fehr, 2008; Taylor & Doria, 1981): Wir bevorzugen Personen unserer eigenen Gruppe deutlich gegenüber Personen einer anderen

Gruppe (in- vs. out-group). Studierende, die Gemeinsamkeiten mit der dozierenden Person haben, könnten daher unbewusst besser behandelt werden („Ist ja auch eine Mutter, da kann ich gut nachvollziehen warum die Prüfungsleistung für sie schwieriger war“; „War auf der gleichen Schule wie ich, vermutlich hatte er einfach einen schlechten Tag“.)

Mere-Exposure Effekt (Serenko & Bontis, 2011): Eine wiederholte Wahrnehmung einer Sache führt zu einer positiveren Bewertung, d.h. wir mögen Dinge, die wir schon gut kennen und die uns öfter präsentiert wurden. Studierende, mit denen man öfter Kontakt hatte könnten deshalb bevorzugt werden, weil sie vertrauter sind.

Es ist ausgesprochen schwierig, diese Biases durch nur direktes Verhalten und Achtsamkeit zu umgehen. Stattdessen empfiehlt es sich Informationen, die zu Bias führen können, gar nicht erst verfügbar zu machen oder im Rahmen von Prozeduren zu filtern. Daher empfiehlt es sich Prüfungsleistungen zu anonymisieren, wenn dies praktisch möglich ist.

Unterschiedliche Studien zeigen auf, dass nicht-anonymisierte Prüfungsleistungen zu verzerrten Ergebnissen führen. In einer Studie in Schweden wies Bygren (2020) mit 17.235 Studierenden einen leicht positiven Bias zu Gunsten von Studierenden mit Immigrationshintergrund auf, was einer positiven Diskriminierung entspricht. Ähnlich zeigten Malouff, Stein, Bothma, Coulter und Emmerton (2014) auf, dass Dozierende die vorher eine sehr gute Präsentation einer Studierenden sahen, dazu neigten eine spätere schriftliche Arbeit dieser Studierenden deutlich besser zu bewerten, als in einer vergleichbaren Experimentalkondition in der die Präsentation deutlich schlechter ausfiel. Anders gesagt: Vorhergehende exzellente Leistungen beeinflussen über den Heiligenscheineffekt spätere Bewertungen. Die unterschiedlichen Prüfungsleistungen, die formal unabhängig voneinander bewertet werden sollten, werden aber eher gemeinsam gesehen.

Abgesehen von diesen Einzelbefunden zeigt auch eine Zusammenfassung von 20 Studien in der Meta-Analyse von Malouff und Thorsteinsson (2016), dass die Präsenz von irrelevanten Informationen über Studierende zu einer Verzerrung von Bewertungen führen kann. Davon betroffen waren in dieser Zusammenfassung besonders Studierende marginalisierter ethnischer Gruppen, Studierende aus bildungsfremden Hintergründen, Studierende die weniger attraktiv waren und Studierende die vorhergehend schlechtere Leistungen erbracht haben. Ein weiteres wichtiges Fazit der Meta-Analyse war zudem, dass auch erfahrende Dozierende und solche die im Bewerten von Prüfungsleistungen trainiert wurden, Verzerrungen in ihren Bewertungen aufweisen, wenn auch geringere als neue oder untrainierte Prüfer*innen. Erfahrung und Übung allein sind somit kein ausreichender Behelf gegen diesen Effekt.

Fazit

Die Kombination aus normalen menschlichen Biases und der Schwierigkeit, diese bewusst zu kontrollieren kann zu erheblichen Verzerrungen bei der Bewertung von Prüfungsleistungen führen. Ein Weg, um dies zu umgehen, ist irrelevante Informationen (z.B. Namen) in Prüfungsformaten, die dies zulassen, zu eliminieren und entsprechend eine Anonymisierung von Prüfungsleistungen durchzuführen.

Literaturverzeichnis

- Bygren, M. (2020). Biased grades? Changes in grading after a blinding of examinations reform. *Assessment & Evaluation in Higher Education*, 45(2), 292–303. <https://doi.org/10.1080/02602938.2019.1638885>
- Efferson, C., Lalive, R. & Fehr, E. (2008). The coevolution of cultural groups and ingroup favoritism. *Science (New York, N.Y.)*, 321(5897), 1844–1849. <https://doi.org/10.1126/science.1155805>
- Gigerenzer, G. & Brighton, H. (2009). Homo heuristicus: why biased minds make better inferences. *Topics in Cognitive Science*, 1(1), 107–143. <https://doi.org/10.1111/j.1756-8765.2008.01006.x>
- Gräf, M. & Unkelbach, C. (2018). Halo effects from agency behaviors and communion behaviors depend on social context: Why technicians benefit more from showing tidiness than nurses do. *European Journal of Social Psychology*, 48(5), 701–717. <https://doi.org/10.1002/ejsp.2353>
- Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J. & Merrill, L. (2009). Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychological Bulletin*, 135(4), 555–588. <https://doi.org/10.1037/a0015701>
- Malouff, J. M., Stein, S. J., Bothma, L. N., Coulter, K. & Emmerton, A. J. (2014). Preventing halo bias in grading the work of university students. *Cogent Psychology*, 1(1), 988937. <https://doi.org/10.1080/23311908.2014.988937>
- Malouff, J. M. & Thorsteinsson, E. B. (2016). Bias in grading: A meta-analysis of experimental research findings. *Australian Journal of Education*, 60(3), 245–256. <https://doi.org/10.1177/0004944116664618>
- Marois, R. & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*, 9(6), 296–305. <https://doi.org/10.1016/j.tics.2005.04.010>
- Nickerson, R. S. (1998). Confirmation bias. A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Nisbett, R. E. & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of personality and social psychology*, 35(4), 250–256. <https://doi.org/10.1037/0022-3514.35.4.250>
- Serenko, A. & Bontis, N. (2011). What's familiar is excellent: The impact of exposure effect on perceived journal quality. *Journal of Informetrics*, 5(1), 219–223. <https://doi.org/10.1016/j.joi.2010.07.005>
- Taylor, D. M. & Doria, J. R. (1981). Self-serving and group-serving bias in attribution. *The Journal of Social Psychology*, 113(2), 201–211. <https://doi.org/10.1080/00224545.1981.9924371>