

Fachgruppe Allgemeine Psychologie

Eine Umfrage unter Studierenden – ein Appell an die Lehrenden

Dirk Wentura¹, Andreas Eder², Carina G. Giesen³, Christina U. Pfeuffer⁴ und Marian Sauter⁵

¹Fachrichtung Psychologie, Universität des Saarlandes

²Institut für Psychologie, Universität Würzburg

³Institut für Psychologie, Friedrich-Schiller-Universität Jena

⁴Institut für Psychologie, Albert-Ludwigs-Universität Freiburg

⁵Institut für Psychologie und Pädagogik, Universität Ulm

Brachem et al. (2022) legen einen bemerkenswerten Bericht zu fragwürdigen Forschungspraktiken in studentischen Projekten vor. Wir danken Brachem et al. für ihren verdienstvollen Bericht zum Stand von Open Science in der Lehre.

Der Bericht ist in seiner Interpretation taktvoll gegenüber den Lehrenden: Ein naiver Leser könnte meinen, dass *Studierende* von sich aus fragwürdige Forschungspraktiken bei ihren Arbeiten anwenden und empfehlenswerte Techniken wie offene und reproduzierbare Forschungspraktiken nicht nutzen, obwohl sie es vor dem Hintergrund ihres bisherigen Studiums besser wissen sollten. Tatsächlich ist uns allen klar: Insbesondere der Bericht über fragwürdige Praktiken bei frühen studentischen Arbeiten (Empirisches bzw. Experimentelles Praktikum) ist ein Spiegel dessen, was durch die Betreuung der *Lehrenden* eingebracht wurde.

Wir wollen zu bedenken geben, dass einige der von Studierenden berichteten fragwürdigen Forschungspraktiken – zu einem vermutlich geringen Anteil – auch auf Missverständnissen beruhen können. Wer $p = .058$ auf $p = .05$ abrundet, der mogelt; das ist unmissverständlich klar (diese Praktik wurde aber nur äußerst selten berichtet). Bei den anderen fragwürdigen Praktiken kann man sich dagegen prinzipiell vorstellen, dass hier mitunter Missverständnisse dahinterstehen könnten, die durch mangelnde Kommunikation von seiten der Lehrenden verursacht werden. Vielleicht ist die einzelne Studie, die

mit den Studierenden gemeinsam durchgeführt wird, eingebettet in ein Forschungsprogramm, das aus Kontinuitätsgründen die Aufnahme bestimmter Variablen nahelegt, die aber für das konkrete Experiment der Studierenden keine Rolle spielen und daher in deren Bericht unerwähnt bleiben. Oder die konzise Darstellung auf einem Poster verlangt Kürzungen, die aus anderer Perspektive wie ein absichtsvolles Verschweigen von Details wirken könnten. Zu welchem Grad derartige Missverständnisse in der Kommunikation zu fragwürdigen Praktiken beigetragen haben, ist unklar. In jedem Fall können und sollten Lehrende derartige Misskommunikationen durch transparente Erklärungen ihrer Entscheidungen und Vorschläge begegnen und sollten sich im Austausch mit den Studierenden diesbezüglich rückversichern, dass ihre Erklärungen und Vorschläge auch für die Studierenden nachvollziehbar waren. Eine Lösung könnte auch sein, bei der Datenanalyse klar und explizit zwischen konfirmatorischen und explorativen Analysen zu unterscheiden. Empfehlenswert und zudem didaktisch geboten, allein schon um Missverständnisse frühzeitig zu erkennen und strukturiert aufzulösen, erweist sich auch eine Präregistrierung der jeweiligen Studie durch die Studierenden (auf die wir später noch einmal zu sprechen kommen) als hilfreich, insbesondere wenn diese zuvor gemeinsam besprochen wird

Abseits aller denkbaren Missverständnisse wird aber in Brachem et al. (2022) den Lehrenden auch der Spiegel

vorgehalten. Ob zum Beispiel eine Stichprobenplanung oder eine Präregistrierung stattfand oder nicht, darüber kann es wohl wenig Unklarheit geben.

Stichprobenplanung

Die geringe Rate von durchgeführten Stichprobenplanungen bei den empirischen Praktika ist ein bemerkenswerter Punkt. Man hätte ja gerade hier eine besonders hohe Quote erwartet, da die Stichprobenplanung seit Jahrzehnten (und nicht erst seit der Reproduktionskrise) ein Kernelement *guter Lehre* darstellt und damit auch zur tatsächlichen Lehre gehören sollte.

Aber tatsächlich deckt sich das Ergebnis hier mit Beobachtungen, die man als Reviewer oder Editor macht: Wie oft fehlt eine Stichprobenplanung gänzlich? Wie oft bleibt sie lediglich darauf beschränkt zu sagen, man rekrutiere genauso viele Personen wie die Vorläuferstudie hatte, auf die man sich beziehe oder rekrutiere die im Labor übliche Gruppengröße?

Warum gibt es immer noch diese große Zurückhaltung beim Thema Stichprobenplanung? Das hat sicherlich mehrere Gründe. Diejenigen, die mit Paradigmen arbeiten, in denen in der Regel große Effekte auftreten, machen in der Tat keinen Fehler, solange die von ihnen als üblich angesehene Stichprobengröße das aufgrund einer Planung anzustrebende N deutlich übersteigt. Aber auch hier gilt: Dieser Tatbestand lässt sich doch leicht kommunizieren. Es bleibt unverständlich, warum das nicht in ausreichendem Maße geschieht.

Andere argumentieren, dass man den zu erwartenden Effekt häufig nicht gut beziffern könne. Da man daher mit Daumenregeln arbeiten müsse, könne man diese direkt in Stichprobengrößen ausdrücken („Nehmen wir doch unsere üblichen $N = 40$ “, o.ä.). Hier sollte man entgegenhalten: Wenn schon Daumenregeln, dann doch bitte auf der Ebene von Effektgrößen (z.B. die Annahme eines „mittleren Effektes“ oder das von Brysbaert, 2019, empfohlene $d = .40$, wenn keine Effektstärken aus Vorgängerstudien vorliegen), da man dann wenigstens die Nicht-Linearität der Effekt-Stichprobengröße-Beziehung bei leichten Veränderungen der angenommenen Effektgröße deutlich bemerken wird („Ach du je! Da benötigen wir ja gleich 100 Leute mehr!“).

Im Zuge der Replikationskrise wurde mitunter darauf hingewiesen, dass insbesondere Studien, die auf Zwischen-Versuchspersonen-Designs beruhen, schlecht replizierbar sein. Kein Wunder. Insbesondere wenn man postuliert, dass ein experimenteller Effekt (z.B. ein Stroop-Effekt) in einer Experimentalgruppe (die z.B. zuvor bereits eine mental anstrengende Aufgabe absolviert hat: „ego deple-

tion“) nur etwas größer ausfallen sollte als in einer Kontrollgruppe, wird man sich die Augen ob der erforderlichen Versuchspersonenanzahl reiben, sofern die Vermutung über die Effektdifferenz halbwegs plausibel ist.

Wir vermuten aber auch, dass hinter der geringen Rate von Stichprobenplanung immer noch Unkenntnis auf Seiten der Lehrenden steckt. Es ist mitunter sicherlich nicht trivial, für die verschiedensten Studiendesigns zu planen. Unglücklicherweise verbirgt sich hier noch ein anderes Problem: Es gibt mittlerweile zahlreiche publizierte Studien, die eine Stichprobenplanung beinhalten, aber aufgrund von Fehlern eine viel zu geringe Power aufweisen (vgl. z.B. Lakens, 2013, zur Nutzung von *partiellen Eta-Quadrat*). Hier zeigt sich schlimmstenfalls ein sich perpetuierendes Problem: An der Stelle, an der an einem ganz konkreten Studienbeispiel die Stichprobenplanung geübt werden kann – beim empirischen Praktikum – lassen manche Lehrenden diesen Aspekt ausfallen (ob nun aus Unkenntnis, Bequemlichkeit, Zeitmangel oder pragmatischen Gründen). Dies hat zur Folge, dass die Studierenden – und somit die nächste Generation von Forschenden – diesem dann fälschlicherweise ebenfalls wenig Relevanz beimisst. Wir plädieren dafür, dass in den Fachbereichen verbindliche Vereinbarungen getroffen werden, in allen studentischen empirischen Arbeiten Stichprobenplanungen durchzuführen. Natürlich wird man damit je nach Forschungsfrage mitunter unbequem hohe Stichprobenschätzungen erzielen, deren faktische Rekrutierung gerade bei Experimentalpraktika oder empirischen Abschlussarbeiten schwer bis unmöglich ist. Dies liegt dann aber in der Natur der Forschungsfrage und muss von den Lehrenden explizit thematisiert und in der Lehre im Rahmen von Experimentalpraktika und Abschlussarbeiten beispielsweise durch post-hoc Powerberechnungen situationsangemessen und methodisch adäquat adressiert und eingeordnet werden und sollte den Studierenden nicht als Defizit gewertet werden.

Präregistrierungen

Solche Vereinbarungen sollten auch für Präregistrierungen gelten. Der Bericht von Brachem et al. (2022) zeigt eindeutig, dass die Rate von Präregistrierungen noch ausbaufähig ist. Gerade in der Lehre scheint uns dieses Mittel besonderes geboten. Es verhindert zum einen vermutlich weitgehend die oben angesprochenen Missverständnisse. Zum anderen verhindert es fragwürdige Praktiken. Letztlich dienen Präregistrierungen der Selbstvergewisserung in der Planungsphase: Haben wir alles richtig durchdacht und können a priori genau angeben, welche Analyse wir rechnen werden und auf welchen Effekt man somit die

Stichprobenplanung ausrichten muß? Ein positiver Nebeneffekt der Präregistrierung: Die Schwierigkeiten, die Studierende mit den verschiedenen Schritten einer wissenschaftlichen Arbeit haben, treten dadurch direkt zu Anfang auf und lassen sich damit sowohl für Lehrende als auch Studierende effizienter adressieren.

Ein Hinderungsgrund scheint uns ein Missverständnis über den Status von Präregistrierungen zu sein. Präregistrierungen sind nicht „in Stein gemeißelte“ Gesetze; sie sind kein unumstößliches Regelwerk. Denn tatsächlich gibt es Fälle des „Hinterher [d.h. nach Betrachtung der tatsächlichen Daten] ist man schlauer“, da man in der komplexen Welt der Psychologie nicht alles an Eigenarten von Daten (z.B. Nichtlinearität) vorhersehen kann. Wie schon zuvor (also in der Prä-Präregistrierungszeit), kommt letztlich alles auf die Nachvollziehbarkeit und Überzeugungskraft von Begründungen an. Der einzige Unterschied ist nun, dass die Begründungen einen transparenten Anker haben: Die Präregistrierung. Man darf schreiben: „Wir weichen an dieser Stelle von der Präregistrierung ab, da wir Folgendes nicht bedacht hatten: xxx. Aus Transparenzgründen sei hier berichtet, dass der vermutete Effekt bei der präregistrierten Analyse mit $p = .13$ assoziiert ist.“ Es kommt dann auf die Überzeugungskraft von „xxx“ an, ob die Leserinnen und Leser das Ergebnis der als angemessener dargestellten Analyse akzeptieren.

Und natürlich sollten explorative Analysen, sofern sie transparent als solche gekennzeichnet sind, nicht negativ konnotiert sein. Sie sind dagegen explizit erwünscht und können durchaus zum Erkenntnisgewinn und zur Hypothesengenerierung für Folgestudien beitragen. Insbesondere wenn diese mit innovativen Methoden durchgeführt werden. Es könnte sein, dass man in Präregistrierungen mitunter etwas konservativ in der Analyseform plant, da man sich in dieser Vorfestlegung nicht auf unerprobtes Terrain begeben will. Aber wie schon Wilcox (1998) im Titel eines Artikels fragte: „How many discoveries have been lost by ignoring modern statistical methods?“ Man sollte ob der Replikationskrise nicht vergessen, dass es zwei Fehlertypen gibt. Zusammenfassend, erscheint es uns in der Lehre wichtig, dass Lehrende den Studierenden gegenüber beispielsweise klar zwischen konfirmatorischen (hypothesentestenden) und explorativen (hypothesengenerierenden) Analysen unterscheiden.

Forderungen für das empirische Praktikum

Es erscheint uns sinnvoll, den von Brachem et al. (2022) vorgelegten Bericht einmal mehr zum Anlass zu nehmen, Präregistrierungen und a priori Stichprobenplanungen als verbindliche Elemente von empirischen Praktika und studentischen Abschlussarbeiten zu vereinbaren. Darüber hinaus ist eine transparente und nachvollziehbar konservierte Kommunikation, insbesondere die Entscheidungsfindung bezüglich Hypothesen, Stichprobe, Studiendesign, geplanten Analysen und Kriterien zum Umgang mit Ausreißern zentraler Bestandteil einer exzellenten Lehre, sowohl in Empirischen Praktika, als auch bei der Betreuung von Abschlussarbeiten.

Literatur

- Brachem, J., Frank, M., Kvetnaya, T., Schramm, L. F. F. & Volz, L. (2022). Replikationskrise, p -hacking und Open Science. Eine Umfrage zu fragwürdigen Forschungspraktiken in studentischen Projekten und Impulse für die Lehre. *Psychologische Rundschau*, 73, 1–17. <https://doi.org/10.1026/0033-3042/a000562>
- Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, 2 (1), 1–38. <https://doi.org/10.5334/joc.72>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t -tests and ANOVAs. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00863>
- Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist*, 53, 300–314.

Danksagung

Wir danken Anand Krishna für Anregungen zu diesem Kommentar.

Prof. Dr. Dirk Wentura

Universität des Saarlandes
Campus
66123 Saarbrücken
wentura@mx.uni-saarland.de

<https://doi.org/10.1026/0033-3042/a000572>