Replikationskrise, Open Science und *p*-hacking: Die Sicht der Studierenden

Johannes Brachema, Maximilan Frank1, b Tatiana Kvetnayac, Leonhard F. F. Schrammd, Leonhard Volze

18.05.2020

aGeorg-August-Universität Göttingen, bTechnische Universität München, cGoethe-Universität Frankfurt, dLudwigs-Maximilians-Universität München, eUniversität Wien

1Korrespondenzauthor ([maximilian.frank@psy.lmu.de](mailto:maximilian.frank@psy.lmu.de))

# Zusammenfassung

In den letzten Jahren gab es innerhalb der Psychologie eine intensive Auseinandersetzung mit den Auswirkungen der Replikationskrise sowie dem hieraus entstandenen Diskurs über die Weiterentwicklung der Disziplin. Als ein Grund für die mangelnde Replizierbarkeit psychologischer Forschung wurde die Verwendung fragwürdiger Forschungspraktiken (eng. QRPs) identifiziert. Während es umfangreiche Untersuchungen zur Prävalenz von QRPs unter Wissenschaftler\*innen gibt, ist bisher wenig über die Verbreitung dieser Praktiken unter Studierenden bekannt. Mit der hier vorgestellten Arbeit wurde erstmals eine größere Befragung unter 1397 Psychologie-Studierenden im deutschsprachigen Raum durchgeführt, um die Verbreitung von QRPs in studentischen Projekten sowie den aktuellen Stand der akademischen Lehre in Bezug auf die Replikationskrise und Open Science zu erheben. Die gemeinsame Betrachtung der Lehre und des Einsatzes fragwürdiger Forschungspraktiken versprechen Aufschluss darüber, wie die psychologische Lehre mit dem empirischen Vorgehen der Studierenden zusammenhängt. Die Ergebnisse zeigen, dass QRPs auch in studentischen Projekten vorkommen, wobei große Unterschiede in der Verbreitung einzelner QRPs bestehen. Auch zwischen den verschiedenen Projekttypen zeigten sich Unterschiede, so war die Anwendung von QRPs in Experimentalpraktika am stärksten und in Masterarbeiten am schwächsten ausgeprägt. Unsere Daten weisen insgesamt darauf hin, dass die selbstberichtete Verbreitung von QRPs über den Studienverlauf abnimmt. Zudem scheint ein Großteil der Studierenden bereits mit der Thematik der Replikationskrise in der Lehre in Berührung gekommen zu sein. Deren Behandlung findet größtenteils in der Methodenlehre und weniger in inhaltlich spezialisierten Lehrveranstaltungen statt. Wir geben abschließend Impulse zur Weiterentwicklung der psychologischen Lehre, in denen die Prinzipien der Offenheit, Transparenz und Kollaboration beim Hervorbringen inhaltlich robuster Forschung bereits während des Studiums im Vordergrund stehen.

# Abstract

In recent years, there has been an intensive debate within psychology about the conclusions that should be drawn as a consequence of the replication crisis. The use of questionable research practices (QRPs) was identified as one reason for the problems in replicability. While there are extensive studies on the prevalence of QRPs among scientists, little is known about the occurence of these practices among students. The work presented here is a first large-scale survey among 1397 psychology students in German-speaking countries to investigate the occurrence of QRPs in student projects and the current state of academic teaching with regard to the replication crisis and open science. The joint examination of teaching and the use of questionable research practices promises to provide information on how psychological teaching is related to the empirical approach of students. The results show that questionable research practices do occur in student projects, with large differences in the occurence of specific QRPs. There were also differences in the incidence of QRPs between different project types: qrp usage was most frequent in empirical internships and least frequent in master’s theses. Our data suggest that the extent of reported QRPs generally decreases as students make progress in their study programs. In addition, the majority of students seem to have already come into contact with the replication crisis in teaching, mostly in methodology classes. Finally, we provide impulses for the further development of psychological teaching. The principles of openness, transparency, and collaboration play an important role in these recommendations, which are aimed at teaching and producing robust research from early on.

# Einleitung

Die Nachrichten von der sogenannten Replikationskrise in der Psychologie haben innerhalb der Disziplin in den letzten Jahren hohe Wellen geschlagen.

Erfolgreiche Replikationen sind für die Aussagekraft wissenschaftlicher Erkenntnisse unabdingbar, da Forschende nur durch das wiederholte Replizieren eines Effekts zu einer hohen Sicherheit gelangen können, dass ein solcher Effekt wirklich existiert – und dieser nicht durch zufällige Variabilität, Verzerrungen in Forschungsdesigns, oder andere Fehlerquellen in Erscheinung tritt (Amrhein et al., 2019; Cohen, 1994). Auf dieser Sicherheit können Forschende dann aufbauen und den gefundenen Effekt näher oder in verschiedenen Kontexten beleuchten, um umfassenderes Wissen zu generieren.

Viele grundlegende Effekte, die häufig als vermeintlich erwiesen bereits Eingang in die Literatur gefunden haben, konnten in Replikationsstudien nicht erneut gefunden werden. Dies stellt die Zuverlässigkeit vieler Ergebnisse in der psychologischen Forschung in Frage. Verschiedene groß angelegte Projekte wurden zur Untersuchung der Replizierbarkeit anerkannter Effekte psychologischer Forschung durchgeführt und kommen zu eher ernüchternden Ergebnissen. So schätzte etwa die Open Science Collaboration (2015) in einer ersten systematischen Untersuchung anhand von 100 Studien aus drei einflussreichen Journals die Replizierbarkeit psychologischer Studien auf einen Anteil zwischen 36 % (erfolgreiche Replikationen mit *p* < .05 in die Richtung des Originaleffekts) und 68 % (signifikanter Effekt *p* < .05 bei Kombination der Originalstudie und Replikation)“ der untersuchten Studien. . Im Projekt Many Labs 2, in dem 28 Effekte von über 60 internationalen Forschungsgruppen repliziert wurden, fanden Klein et al. (2018) eine Rate von 54 % erfolgreichen Replikationen. Ähnliches zeigten auch Camerer et al. (2018) mit einer Replikationsrate von 62 %.

Zahlreiche Publikationen in der Psychologischen Rundschau, darunter zur Inflation von falsch-positiven Befunden (Ulrich et al., 2016), zum Umgang mit Forschungsdaten (Schönbrodt et al., 2017) und schließlich das Themenheft “Replizierbarkeit” (Klauer, 2018), spiegeln die Relevanz dieses Themas für unsere Disziplin wider.

Replizierbarkeit ist nicht das höchste Ziel psychologischer Forschung und nicht jeder Replikationsversuch wird die Ergebnisse einer Originalstudie exakt widerspiegeln können. Replikationen und ihre direkte Aussagekraft selbst sind vielen methodologischen Fragen unterworfen (Fiedler & Prager, 2018; Munafò & Davey Smith, 2018), bis hin zu der Debatte, was eine Replikation überhaupt ist (siehe z.B. Nosek & Errington, 2020). Die oben beschriebenen Replikationsraten sind dadurch aber nicht automatisch unproblematisch, sondern weisen sehr deutlich auf mangelnde Robustheit im Forschungsprozess hin. In der Psychologie scheinen publizierte Ergebnisse schnell als gesicherte Grundlage angesehen zu werden, auf der Folgestudien geplant werden können - ohne dass deren Belastbarkeit vorher ausreichend überprüft wird, beispielsweise durch Replikation oder Triangulation (siehe z.B. Munafò & Davey Smith, 2018).

Dabei wird durchaus vielfältig diskutiert, welche Schlüsse aus der Replikationskrise gezogen werden sollen. Manche Stimmen hinterfragen die Sichtweise, nach der mit der hohen Anzahl an fehlgeschlagenen Replikationen eine Krise situiert ist (Gilbert et al., 2016; Stanley & Spence, 2014; Stroebe & Strack, 2014), andere fordern ein grundsätzliches Überdenken der analytischen Paradigmen, die in der psychologischen Forschung angewandt werden, z.B. in Form einer Abkehr vom Null-Hypothesen-Signifikanz-Testen (NHST; z.B. Devezer et al., 2019; Dienes & Mclatchie, 2018; Mcshane et al., 2019; Wagenmakers et al., 2017). Trotz der Kritik, die in der Wissenschaftsgemeinschaft an der Praxis des NHST bereits geäußert worden ist, ist es weiterhin das am weitesten verbreitete Verfahren zur Erkenntnisgewinnung in der Psychologie. So werden mögliche Ursachen und darauf aufbauende Lösungsansätze für die Replikationskrise vor allem im Rahmen dieses Paradigmas diskutiert.

Als eine solche Ursache für die mangelnde Replizierbarkeit von psychologischen Effekten rückten fragwürdige Forschungspraktiken (QRPs, questionable research practices) in den Fokus der wissenschaftlichen Öffentlichkeit. Die Anwendung von QRPs, z.B. das nachträgliche Ausschließen von Ausreißern oder das selektive Berichten von Variablen, unterläuft die Annahmen des NHST und erhöht dadurch die Wahrscheinlichkeit von falsch-positiven Befunden (Bakker et al., 2012; Simmons et al., 2011). Zugleich legen andere Untersuchungen nahe, dass solche QRPs durchaus nicht ungewöhnlich in der wissenschaftlichen Praxis sind (Agnoli et al., 2017; Banks et al., 2016; Fiedler & Schwarz, 2016; John et al., 2012). Der Grad, zu dem QRPs in der Psychologie angewendet werden, wird auch aufgrund methodischer Mängel der zugrundeliegenden Studien noch diskutiert. So kritisieren z.B. Fiedler & Schwarz (2016), dass der u.a. von John et al. (2012) verwendete Anteil von Forschenden, die von einmaliger Verwendung einer QRP während ihrer Laufbahn berichten, keine valide Schätzung der Prävalenz dieser QRP darstellt. Nichtsdestotrotz zeigen solche Untersuchungen, dass in psychologischen Studien QRPs angewendet werden, womit die Belastbarkeit und Replizierbarkeit von Forschungsergebnissen in Frage gestellt wird.

Auch in der universitären Lehre wird wissenschaftliche Literatur rezipiert, deren Belastbarkeit zweifelhaft ist, und dient als Grundlage für das den Studierenden vermittelte Wissen. Studierende werden dadurch von eben dieser wissenschaftlichen Literatur für Forschungs- und Praxistätigkeit inspiriert. Es sollte in der Lehre kritisch behandelt werden, dass hier ein nicht ignorierbarer Teil der Literatur wenig belastbar ist. Gerade hier ist zu erwarten, dass auch in Folge der Verbreitung von QRPs in wissenschaftlichen Publikationen den im NHST gemachten Annahmen besondere Aufmerksamkeit geschenkt wird und die akkurate Anwendung der Verfahren in der Lehre einen zentralen Platz einnimmt.

Auch innerhalb der Studierendenschaft gewann das Thema jüngst an Relevanz: Im Jahr 2018 wurde auf der 27. PsyFaKo (Psychologie-Fachschaften-Konferenz)[[1]](#footnote-1) in Würzburg ein Positionspapier zur Replikationskrise der Psychologie (Konferenzrat der Psychologie-Fachschaften-Konferenz, 2018) verabschiedet, in dem die Studierenden unter anderem forderten, dass die Themen Replikationskrise und Open Science einen größeren Stellenwert in der Lehre erhalten. Die Studierenden von heute werden das Fach Psychologie in Zukunft prägen und entsprechend wichtig ist es, dass das Studium durch eine sorgfältige und zeitgemäße Ausbildung eine angemessene Vorbereitung auf die vielfältigen beruflichen Tätigkeiten von Psycholog\*innen darstellt.

Inwiefern diese Themen auch tatsächlich im Rahmen der akademischen Lehre vermittelt werden und in diesem Zuge auch in studentische Arbeiten einfließen, stand bisher allerdings nicht im Fokus von Untersuchungen. Zudem werden die aktuellen Anreizstrukturen des Wissenschaftsbetriebs als ein relevanter Einflussfaktor auf die verbreitete Anwendung von QRPs durch Forschende ausgemacht, die sich vor allem im Schlagwort “Publikationsdruck” manifestieren (siehe z.B. Bakker et al., 2012). Uns stellte sich die Frage, in welchem Ausmaß QRPs von Studierenden in empirischen Arbeiten (Seminar-, Bachelor-, Masterarbeit, etc.) verwendet werden, die diesem Anreizsystem in geringerem Maß unterliegen, und ob sich die Anwendung von QRPs im Laufe des Studiums verändert.

Während es, wie oben behandelt, einige Befunde zur Verbreitung von QRPs bei Forschenden gibt, ist uns dazu unter Studierenden hingegen nur eine Studie bekannt (Krishna & Peter, 2018). Um diese Frage näher zu beleuchten, wurde in der hier vorgestellten Arbeit eine große Befragung unter Psychologiestudierenden im gesamten deutschsprachigen Raum durchgeführt. Die gemeinsame Betrachtung der Lehre und des Einsatzes fragwürdiger Forschungspraktiken versprechen Hinweise darauf, wie die psychologische Lehre und das empirische Vorgehen von Studierenden zusammenhängen.

Studierende führen in aller Regel im Verlauf ihres Studiums nur eine begrenzte Anzahl empirischer Projekte durch. In der Konsequenz ist ein interessanter methodischer Aspekt unserer Umfrage, dass wir uns dem Anteil der von einer QRP tatsächlich betroffenen Projekte in unserer Stichprobe einfacher annähern können, als das bei Forschenden der Fall wäre.

Die von uns durchgeführte Umfrage besteht aus drei Teilbereichen, deren Ergebnisse einen Einblick in 1) die Verbreitung fragwürdiger und positiver Forschungspraktiken unter Studierenden der Psychologie, 2) die Verankerung des Themenkomplexes “Replikationskrise, QRPs und Open Science” in der Lehre sowie 3) die Einstellung der Studierenden zu diesem Themenkomplex, bieten.

# Methode

Der Zeitraum der Datenerhebung und Kriterien zum Ausschluss von Teilnehmenden wurden vor der Durchführung beschrieben (siehe <https://osf.io/k85r4/>) und sind, ebenso wie die anonymisierten Primärdaten, auf dem Open Science Framework (OSF) zugänglich (<https://osf.io/t3mak/>). Dort sind auch der R-Code zur Datenanalyse und die reproduzierbare RMarkdown-Datei (Allaire et al., 2020) zur Erstellung dieses Berichts verfügbar. Die Online-Umfrage wurde mit der Open-Source Software formR (Arslan et al., 2019) erstellt und die Daten in R mit Hilfe der Pakete lme4, lmerTest und der tidyverse-Pakete (Bates et al., 2015; Kuznetsova et al., 2017; R Core Team, 2019; Wickham, 2017) aufbereitet und analysiert. Im Vorfeld wurde eine Pilotbefragung mit 22 Psychologiestudierenden verschiedener Fachsemester durchgeführt, um die Verständlichkeit und Durchführbarkeit der Befragung zu testen. Hierbei hatten die Befragten die Gelegenheit, mittels offener Kommentarfelder ausführliches Feedback abzugeben, welches in der finalen Version der Umfrage berücksichtigt wurde.

## Akquise

Die Bewerbung der Umfrage erfolgte über die Kanäle der PsyFaKo. Hierzu wurden mehrere Bewerbungsmails an die deutschsprachigen Psychologie-Fachschaftsvertretungen versandt mit der Bite, die Umfrage an Studierende weiterzuleiten. Den Fachschaften stand die Wahl der an den jeweiligen Hochschulen benutzten Werbemittel (Aushänge, Newsletter, Mailverteiler, Facebook-Posts) frei, wobei ein einheitlicher Textvorschlag zur Bewerbung bereitgestellt wurde. Außerdem wurde die Umfrage auf der Mitgliederversammlung der 28. PsyFaKo im November 2018 gegenüber den anwesenden Studierenden beworben. Vereinzelt wurden auch Dozierende angesprochenund gebeten, in ihren Lehrveranstaltungen auf die Umfrage hinzuweisen. Die Teilnahme an der Studie war freiwillig und es gab keine monetären Anreize.

## Ausschlüsse

Im festgelegten Erhebungszeitraum vom 22.11.2018 bis 28.02.2019 verzeichneten wir 3558 Klicks auf den Link zu unserer Umfrage und 2171 Personen starteten die Umfrage. Hiervon wiederum bearbeiteten 1503 Personen (69,2 %) alle Fragen. Von diesen 1503 Personen wurden 96 Personen anhand vorab festgelegter Kriterien von der Analyse ausgeschlossen: Acht Personen gaben an, nicht ernsthaft teilgenommen zu haben, 19 Personen hatten ihren Studienabschluss 2016 oder früher gemacht und zählten damit nach unseren Kriterien nicht mehr zu aktuell oder kürzlich Studierenden, und 69 Personen waren keine Psychologiestudierende oder -absolvent\*innen. Abweichend von den im Vorhinein festgelegten Kriterien schlossen wir auch 10 Personen aus, die ein unplausibles Alter angegeben hatten (16 oder jünger, trotz fortgeschrittenen Studiums). Zudem berücksichtigten wir eine Person einer niederländischen Universität nicht in der Auswertung, da unsere Umfrage explizit auf den deutschsprachigen Raum ausgerichtet war. Für die Analyse unserer Fragen zum Einsatz von QRPs wurden außerdem 251 Teilnehmende nicht berücksichtigt, die noch nie ein empirisches Projekt in ihrem Studium durchgeführt hatten. Von den in diesem Schritt ausgeschlossenen Teilnehmenden waren 97,6 % Studierende im ersten oder zweiten Bachelor-Semester. Wir bezeichnen die Teilstichprobe im Folgenden als EE-Teilstichprobe (Empirische Erfahrung-Teilstichprobe). So blieben 1397 Personen in der vollen Stichprobe und 1146 Personen in der EE-Teilstichprobe.

## Stichprobenbeschreibung (volle Stichprobe)

Unsere Befragten (weiblich: 78,2 %; männlich: 20,5 %; divers: 0,2 %; keine Angabe: 1,1 %) waren im Mittel 22,93 Jahre alt (*SD* = 3,94) und Psychologiestudierende in allen Phasen ihres Studiums (s. Abbildung 1). Es nahmen mehr Bachelor- als Master-Studierende teil (63,7 % zu 32,6 %). Der Anteil von Alumni/Alumnae war mit 3,1 % gering. In der Stichprobe sind Teilnehmende von 47 deutschsprachigen staatlichen und privaten Hochschulen vertreten. Für 23 dieser Institutionen lagen uns 30 oder mehr Datenpunkte vor[[2]](#footnote-2). Eine detaillierte Auflistung der Hochschulen und Datenpunkte steht in den Online-Materialien (<https://osf.io/br2zf/>) zur Verfügung.

(( Abbildung 1 ungefähr hier platzieren. ))

## Prozedur und Fragen

Teilnehmende beantworteten zunächst demographische Fragen (Alter, Geschlecht, Studienfortschritt, aktuelle Universität, bei Master-Studierenden auch Bachelor-Universität, ggf. Jahr des Abschlusses). Angezeigte Fragen mussten, wenn hier nicht anders angegeben, als Pflichtfragen beantwortet werden. Anschließend folgten Fragen zur Verwendung von fragwürdigen und positiven Forschungspraktiken in empirischen Projekten während des Studiums. Zunächst beantworteten die Teilnehmenden die Frage, ob sie im bisherigen Studienverlauf schon ein Forschungsprojekt begonnen oder durchgeführt hatten. Bei Beantwortung mit “Nein” sprang die Umfrage zum nächsten Fragenblock (kritische Thematisierung von QRPs in der Lehre). Bei Beantwortung mit “Ja” folgten die Fragen zur Verwendung spezifischer Forschungspraktiken: “Welche der unten genannten Forschungspraktiken hast du selbst bzw. hat deine Seminargruppe schon mindestens einmal bei einem empirischen Projekt an deiner Universität unter Anleitung einer Lehrperson angewendet oder wendest sie aktuell an?” (Antwortmöglichkeiten: “Expra / Empra”, “Projektarbeit”[[3]](#footnote-3), “Bachelorarbeit”, “Masterarbeit”, “Anderes”[[4]](#footnote-4), “Keines”). Zu jeder Praktik gab es einen Info-Text mit einem erklärenden Beispiel. Die abgefragten Praktiken und die Info-Texte sind in Tabelle 1 aufgelistet. Zusätzlich wurde gefragt, welche Plattformen die Teilnehmenden mindestens einmal für Präregistrierungen verwendet hatten.

Für neun QRPs (siehe Tabelle 1) wurde abgefragt, ob diese Praktik an der aktuellen Hochschule in der Lehre kritisch thematisiert wurde. Außerdem fragten wir die Teilnehmenden, ob und in welcher Art von Lehrveranstaltung die Replikationskrise an ihrer *aktuellen* Hochschule thematisiert wurde. Die Teilnehmenden schätzen auf einer 5-Punkt-Likert-Skala ihr Interesse, die Wichtigkeit sowie ihre eigene Informiertheit über den Themenkomplex “Forschungspraktiken, Open Science und die Replikationskrise” ein. Es folgten noch wenige zusätzliche Fragen, die wir aus Gründen der Prägnanz hier nicht berichten. Die Umfrage schloss mit der Kontrollfrage “Hast du ernsthaft an dieser Umfrage teilgenommen?” und einem optionalen freien Kommentarfeld. Der komplette Fragebogen inklusive Formulierungen kann in der Registrierung unseres OSF-Projekts (<https://osf.io/k85r4/>) nachvollzogen werden.

(( Tabelle 1 ungefähr hier platzieren. ))

# Ergebnisse

## Deskriptive Ergebnisse

### Verbreitung fragwürdiger Forschungspraktiken

Den Ergebnissen in diesem Abschnitt liegen die Daten der EE (Empirische Erfahrung)-Teilstichprobe (*N* = 1146) zugrunde.

**Unterschiede zwischen Projekten**. Aus den Angaben der Teilnehmenden ergibt sich ein mittlerer Einsatz von 1,26 (*SD* = 1,40) QRPs pro Projekt[[5]](#footnote-5). Wie in Abbildung 2 zu sehen ist, zeigen sich dabei Unterschiede zwischen den abgefragten Projekten. Besonders viele fragwürdige Praktiken scheinen demnach mit durchschnittlich 1,42 (*SD* = 1,63) Praktiken pro Projekt im Experimentalpraktikum (Expra) verwendet zu werden, während Masterarbeiten mit durchschnittlich 0,50 (*SD* = 0,85) verwendeten QRPs den niedrigsten Wert aufweisen. Bachelor- und Projektarbeiten sowie andere Projekte (durchschnittlich jeweils 1,03, *SD* = 1,34; 1,30, *SD* = 1,52;[[6]](#footnote-6) und 0,83, *SD* = 1,35[[7]](#footnote-7) QRPs pro Projekt) liegen dazwischen. Weitere Details sind in den Online-Materialien (<https://osf.io/br2zf/>) zu finden. Betrachtet man die Projekte, die sich eindeutig nach Studienfortschritt ordnen lassen, also Expra, Bachelor- und Masterarbeit, so zeigt sich ein deutlicher Trend: Mit fortschreitendem Studienverlauf wurden weniger QRPs von den Teilnehmenden eingesetzt.

(( Abbildung 2 ungefähr hier platzieren. ))

**Unterschiede zwischen Praktiken**. Auf der linken Seite von Abbildung 3a ist die Einsatzhäufigkeit der einzelnen abgefragten Forschungspraktiken pro Projekt dargestellt. Die am häufigsten genannten Praktiken waren “fehlende Stichprobenplanung” und “selektives Berichten abhängiger Variablen” mit einem Anteil von 34,2 %, bzw. 23,1 % betroffenen Projekten. Die von den Teilnehmenden am seltensten eingesetzten Praktiken waren das Abrunden von p-Werten und das Nacherheben von Versuchspersonen zum Zweck der Herbeiführung statistischer Signifikanz mit 1,9 %, bzw. 6,4 % betroffenen Projekten. Neben den QRPs wurden auch zwei positive Forschungspraktiken abgefragt. Laut Angaben der Teilnehmenden wurde für 33,6 % der erfassten Projekte eine Stichprobenumfangsplanung (basierend auf einer a priori Poweranalyse), und 22,5 % der erfassten Projekte wurden präregistriert. Eine detailliertere Aufschlüsselung dieser Ergebnisse nach Forschungsprojekt ist in Abbildung 3b dargestellt. Auf der rechten Seite von Abbildung 3a ist analog zur Darstellung von John et al. (2012) der Anteil der Teilnehmenden zu sehen, die die jeweilige Forschungspraktik mindestens einmal in einem empirischen Projekt eingesetzt hatten. Im Vergleich der beiden Größen zeigt sich, dass der Anteil betroffener Projekte in unserer Stichprobe durchweg niedriger ausfällt, als der Anteil der Teilnehmenden, die eine Praxis im Laufe ihres Studiums mindestens einmal angewandt haben: Im Mittel beträgt der Unterschied 8,2 Prozentpunkte bei QRPs (13,2 % zu 21,4 %) und 16,7 Prozentpunkte (28,0 % zu 44,7 %) bei den beiden positiven Forschungspraktiken.

(( Abbildung 3 ungefähr hier platzieren. ))

Präregistrierungen fanden mit 280 Nennungen (24,4 % der Teilnehmenden) am häufigsten bei einer Lehrperson statt[[8]](#footnote-8). Es folgten in absteigender Reihenfolge als Plattformen für Präregistrierungen das Open Science Framework (OSF, *n* = 97, 8,5 % der Teilnehmenden), aspredicted.org (*n* = 87, 7,6 % der Teilnehmenden), “andere” (*n* = 62, 5,4 % der Teilnehmenden) und ein Journal (*n* = 11, 1,0 % der Teilnehmenden). Von den Teilnehmenden gaben 696 Personen (60,7 % der Teilnehmenden) an, nie eines ihrer empirischen Projekte präregistriert zu haben. Bei dieser Frage war die Auswahl mehrerer Antworten erlaubt.

### Behandlung in der Lehre und Einstellungen von Studierenden

Den Ergebnissen in diesem Abschnitt liegen die Daten der vollen Stichprobe (*N* = 1397) zugrunde. Die einfachen Korrelationen zwischen den numerischen Variablen finden sich in Tabelle 2. Zu den hier dargestellten Ergebnissen finden sich ausführliche Tabellen mit allen Daten in den Online-Materialien (<https://osf.io/br2zf/>).

**Thematisierung der Replikationskrise**. Insgesamt gaben 75,7 % der Teilnehmenden an, dass die Replikationskrise an ihrer jeweils aktuellen Universität in Lehrveranstaltungen thematisiert wurde, 20,4 % gaben an, dass sie nicht thematisiert wurde (“Weiß nicht”: 4,1 %). Unter den Studienanfänger\*innen (Studierende im ersten oder zweiten Bachelor-Semester) gaben 60,2 % der Teilnehmenden an, in Lehrveranstaltungen von der Replikationskrise gehört zu haben. Für alle Studierenden im Bachelor und im Master zeigte sich im Allgemeinen jeweils ein ähnliches Bild mit 75,3 % bzw. 77,4 % positiven Antworten.

**Betrachtung nach Lehrveranstaltungen**. Für diese Auswertung verwendeten wir nur die Daten von Studierenden, die sich zum Zeitpunkt der Umfrage im Master-Studium befanden (*n* = 455) und werteten nur Antworten in Bezug auf Lehrveranstaltungen aus, die nach den Empfehlungen der DGPs (Deutsche Gesellschaft für Psychologie, 2005) in Bachelorstudiengängen Psychologie als Pflichtveranstaltungen eingestuft werden. So ist die Annahme berechtigt, dass in dieser Teilgruppe alle Teilnehmenden die betrachteten Veranstaltungen besucht haben. Die Behandlung der Replikationskrise fand größtenteils in der Methodenlehre (inkl. Expra) statt: 55,2 % der Master-Studierenden unter den Teilnehmenden gaben an, dass hier die Replikationskrise thematisiert wurde. In Diagnostik-Lehrveranstaltungen wurde am zweithäufigsten über die Replikationskrise berichtet (22,4 %).

**Behandlung fragwürdiger Praktiken**. Die Ergebnisse legen nahe, dass Studierende im Studium durchaus eine kritische Thematisierung einzelner QRPs erleben. So gaben 73,9 % der Teilnehmenden an, eine Problematisierung des Aufstellens von Hypothesen nach Kenntnis der Ergebnisse (HARKing, Hypothesizing After the Results are Known) erfahren zu haben, womit sich HARKing als die am häufigsten in der Lehre thematisierte QRP herausstellte. Auf den folgenden Plätzen finden sich fehlende Stichprobenplanung (65,1 %) und selektives Berichten von Variablen (64,0 %). Am seltensten wurden Flexibilität in der Datenanalyse (50,8 %) und das Abrunden von *p*-Werten (39,7 %) kritisch thematisiert.

**Einstellungen von Studierenden**. Anhand von Abbildung 4 wird deutlich, dass das Thema “Replikationskrise und Open Science” von der überwältigenden Mehrheit der Teilnehmenden (94,2 %) für sehr wichtig oder wichtig gehalten wird, dass eine Mehrheit der Teilnehmenden (68,5 %) Interesse an dem Thema hat, und dass sich ein substanzieller Teil der Teilnehmenden (34,1 %) als nicht gut informiert einschätzt.

(( Abbildung 4 ungefähr hier platzieren. ))

## Explorative Regressionsanalysen

Zur näheren Untersuchung der Daten führten wir eine Reihe von explorativen Regressionsanalysen durch, von denen wir die wichtigsten hier berichten[[9]](#footnote-9). Alle weiteren oder nur kurz erwähnten Analysen sind in den Online-Materialien abrufbar (<https://osf.io/br2zf/>). Für alle Analysen wurden Bonferroni-korrigierte Alpha-Niveaus zugrunde gelegt.

### Einsatz fragwürdiger Forschungsmethoden

In einem generalisierten linearen gemischten Modell für dichotome abhängige Variablen mit logit-Linkfunktion nahmen wir alle verfügbaren, potentiell informativen Kovariaten auf, um die Odds für den Einsatz von QRPs zu modellieren (siehe Tabelle tab.glmm.qrp. Die abhängige Variable war ein Indikator dafür, ob eine spezifische Forschungspraktik in einem spezifischen Projekt eingesetzt wurde (0 = “Nein”, 1 = “Ja”). Das Modell enthält zufällige Achsenabschnitte für den/die jeweilige/n Teilnehmer\*in (*SD* = 1,25, (1) = 1058,34, *p* <.001), für die Universität, an der das Projekt durchgeführt wurde (*SD* = 0,52, (1) = 42,67, *p* <.001), und für die spezifische Forschungspraktik (*SD* = 1,03, (1) = 1538,87, *p* <.001). Die numerischen Prädiktoren wurden z-standardisiert. Die Schätzungen der fixen Koeffizienten und die jeweiligen Odds Ratios sind in Tabelle 3 dargestellt. Zugrunde lagen die Daten aus der EE-Teilstichprobe.

Einzelne Beobachtungen schlossen wir wegen inkonsistenter Antworten oder unklarer Zuordnung des jeweiligen Projektes zu einer Universität aus, z.B. bei Projektarbeiten, wenn ein Universitätswechsel zwischen Bachelor und Master erfolgte (siehe Online-Materialien, <https://osf.io/br2zf/>).

Darüber hinaus lassen sich Unterschiede zwischen den einzelnen Projekttypen erkennen, wobei alle Projekttypen im Vergleich zur Referenzkategorie “Expra” eine geringere Wahrscheinlichkeit des Einsatzes von QRPs aufweisen. Die Effektstärken variieren dabei von *OR* = 0,27 (Masterarbeiten) bis *OR* = 0,77 (Projektarbeit). Bei einer Masterarbeit ist die Chance, dass eine QRP eingesetzt wird, also weniger als ein Drittel der Chance, dass sie in einem Expra eingesetzt wird. Die übrigen untersuchten Kovariaten wiesen keinen signifikanten Zusammenhang mit dem Einsatz von QRPs auf (siehe Tabelle 3).

In einem nächsten Schritt untersuchten wir im Detail, ob sich Hinweise für eine systematische Veränderung der Einsatzwahrscheinlichkeit von QRPs im Verlauf des Studiums finden. Dazu betrachteten wir ausschließlich Expras, Bachelorarbeiten und Masterarbeiten, da diese Projekte eindeutig verschiedenen Phasen des Studiums zugeordnet werden können. Die Analyse wurde auch hier mit einem generalisierten linearen gemischten Modell für dichotome abhängige Variablen mit logit-Linkfunktion mit der gleichen abhängigen Variable und, abgesehen vom Projekttyp, den gleichen Kovariaten wie im vorherigen Modell durchgeführt. Der Projekttyp wurde nun als geordneter Faktor in das Modell mit aufgenommen, so dass ein linearer Trend über die drei eingeschlossenen Projekttypen untersucht werden konnte. Tatsächlich zeigte sich ein linearer Trend, *OR* = 0.4 [0,32; 0,50], p <.001, der darauf hindeutet, dass mit fortschreitendem Studium, bzw. in anspruchsvolleren Projekten von unseren Teilnehmenden mit geringerer Wahrscheinlichkeit QRPs angewandt wurden.

### Einsatz positiver Forschungspraktiken

Die Ergebnisse bezüglich positiver Forschungspraktiken fassen wir hier nur kurz zusammen. Wir untersuchten hier ebenfalls in generalisierten gemischten linearen Modellen die gleichen Prädiktoren und zufälligen Achsenabschnitte wie bei der Exploration zu QRPs (mit Ausnahme von “Praktik gelehrt”, da diese Frage nur für QRPs gestellt wurde). Deutliche Unterschiede zeigten sich zwischen Master- und Bachelor-Studierenden (*B* = -0,38, *SE* = 0,13, *OR* = 0,68, 95 % *KI* = [0,53; 0,87], *p* = .003). Master-Studierende wiesen demnach eine geringere Wahrscheinlichkeit für den Einsatz positiver Forschungspraktiken auf. Diese Beobachtung zeigte sich noch deutlicher, wenn nur Expras und Bachelorarbeiten untersucht wurden, so dass ein etwas direkterer Vergleich zwischen länger zurückliegenden Projekten (Expras und Bachelorarbeiten von aktuellen Master-Studierenden) und kürzer zurückliegenden Projekten (von aktuellen Bachelor-Studierenden) gezogen werden kann (siehe Online-Materialien, <https://osf.io/br2zf/>).

Auch in diesem Modell zeigte sich ein positiver Zusammenhang zwischen der subjektiv eingeschätzten Informiertheit der Teilnehmenden und dem Einsatz positiver Forschungspraktiken (*B* = 0,36, *SE* = 0,06, *OR* = 1,43, 95 % *KI* = [1,29; 1,60], *p* <.001).

### Eindruck von Informiertheit

In einer weiteren explorativen Analyse untersuchten wir den Zusammenhang zwischen den in Modell aufgenommenen Kovariaten und der subjektiven Informiertheit unserer Teilnehmenden über den Themenkomplex “Replikationskrise und Open Science”. Für die Analyse nutzten wir ein lineares gemischtes Modell mit einem zufälligen Achsenabschnitt für die aktuelle (bei Alumni/Alumnae: letzte) Universität (*SD* = 0,38, (1) = 135,12, *p* <.001). Die Schätzungen der fixen Koeffizienten sind in Tabelle 4 dargestellt.

Die Analyse deutet auf positive Zusammenhänge der subjektiven Informiertheit mit der Anzahl von durchgeführten oder begonnenen Projekten und der Wichtigkeit, die dem Themenkomplex “Replikationskrise und Open Science” beigemessen wird, hin. Schließlich zeigte sich auch ein deutlicher Zusammenhang zwischen einer berichteten Behandlung der Replikationskrise in der Lehre und der subjektiven Informiertheit der Teilnehmenden: Mit einem Beta-Gewicht von 0,63 ist dies sogar der stärkste in dieser Analyse beobachtete Zusammenhang. Demnach schätzten sich Studierende, die über eine Behandlung der Replikationskrise in der Lehre an ihrer Universität berichteten, im Mittel um 0,63 Standardabweichungen besser informiert ein, als solche, denen dieses Thema nicht in der Lehre begegnet war..

Weitere explorative Analysen führten wir zur Behandlung der Replikationskrise in der Lehre und zur subjektiv eingeschätzten Wichtigkeit des Themenkomplexes “Replikationskrise und Open Science” durch unsere Teilnehmenden durch. Die Ergebnisse können in den zusätzlichen Online-Materialien (<https://osf.io/br2zf/>) abgerufen werden.

# Diskussion

In diesem Beitrag stellen wir eine Befragung vor, in der die Verbreitung von fragwürdigen und positiven Forschungspraktiken unter Studierenden sowie die Thematisierung dieser Aspekte in der psychologischen Lehre untersucht wurde. Unseren Ergebnissen zufolge sind QRPs auch im Studium verbreitet. Hierbei zeigten sich große Unterschiede in der Verbreitung zwischen verschiedenen QRPs und verschiedenen Projekten. Konkret liefern unsere Daten Hinweise darauf, dass in fortgeschritteneren Projekten wie Masterarbeiten weniger QRPs angewandt werden als in Projektarbeiten zu Studienbeginn, wie bspw. Expras. Zudem ist ein Großteil der Studierenden (75,7 %) bereits mit der Thematik der Replikationskrise in der Lehre in Berührung gekommen. Ihre Behandlung findet größtenteils in der Methodenlehre und weniger in inhaltlich spezialisierten Fächern statt. Studierende sehen den Themenkomplex “Replikationskrise und Open Science” mit überwältigender Mehrheit als wichtig und interessant an, fühlen sich aber unterschiedlich gut informiert. Mit Blick auf diese Ergebnisse und die Diskussion möglicher Ursachen geben wir abschließend Impulse für die Verbesserung der psychologischen Lehre.

## Verbreitung von QRPs in studentischen Projekten

### Einfluss der betrachteten Kenngröße

In unserer überwiegend aus Studierenden bestehenden Stichprobe zeigte sich ein deutlicher Unterschied zwischen der geschätzten Häufigkeit einmaliger Anwendung bestimmter fragwürdiger Forschungspraktiken und dem geschätzten Anteil betroffener Projekte; letzterer war im Mittel um 8,2 Prozentpunkte niedriger (s. Abb. 3a). Dabei ist auch zu beachten, dass der Anteil betroffener Projekte je nach Art des Projekts stark schwankte (s. Abb. 2 und 3b, sowie Tabelle 3). Von der Häufigkeit einmaliger Anwendung einer QRP in irgendeinem Projekt kann deshalb nicht auf die generelle Prävalenz dieser QRP in allen Projekttypen geschlossen werden, vielmehr muss der Anteil betroffener Projekte im Zentrum der Analyse stehen und der Projekttyp berücksichtigt werden. Diese Betrachtung der Unterschiede zwischen den Häufigkeiten schließt an eine ähnliche Diskussion unter Forschenden an (Fiedler & Schwarz, 2016). Die Unterscheidung sollte auch bei zukünftigen Untersuchungen zur Prävalenz von Forschungspraktiken beachtet werden.

### Schätzungen zur Verbreitung von QRPs im Vergleich

Zur Verbreitung von QRPs unter wissenschaftlich tätigen Psycholog\*innen liegt bereits einige Literatur vor (vgl. Agnoli et al., 2017; Banks et al., 2016; Fiedler & Schwarz, 2016; John et al., 2012). Im Vergleich zu Stichproben mit Forschenden zeigte sich unter den Teilnehmenden unserer Umfrage überwiegend ein geringerer Anteil von mindestens einmaliger Verwendung spezifischer QRPs. Dieser Vergleich ist allerdings nicht aussagekräftig für die Frage, ob Forschende häufiger QRPs verwenden als Studierende, da Forschende selbstverständlich auf eine wesentlich längere Zeit wissenschaftlicher Tätigkeit und damit eine größere Zahl angefertigter Forschungsarbeiten zurückblicken, als dies bei Studierenden der Fall ist.

Studierende fertigen ihre empirischen Projekte darüber hinaus unter stark abweichenden Bedingungen an: Der Lerneffekt ist tendenziell wichtiger als der Erkenntnisgewinn, die Forschung findet unter enger Betreuung statt und es gibt in der Regel keinen Publikationsdruck. Im Vergleich unserer Ergebnisse mit den Befunden von Krishna & Peter (2018) für studentische Abschlussarbeiten zeigen sich für einzelne Forschungspraktiken vergleichbare Einsatzhäufigkeiten pro Forschungsprojekt, so etwa hinsichtlich der QRPs “Flexibler Ausschluss von Ausreißern” (12,8 % vs. 15,5 % bei Krishna & Peter, 2018), “Selektives Berichten von Bedingungen” (9,8 % vs. 7,7 %) wie auch der positiven Forschungspraktik der Poweranalyse (33,6 % vs. 35,3 %). Bei anderen wiederum zeigen sich Unterschiede: So war das selektive Berichten von Variablen in unserer Stichprobe häufiger (23,1 % vs. 5,8 % bei Krishna & Peter, 2018), das Abrunden von p-Werten (1,9 % vs. 10,4 %) und HARKing (6,2 % vs. 15,0 %) hingegen seltener. Diese Inkonsistenzen könnten sowohl auf systematische Unterschiede zwischen den jeweils untersuchten Projekttypen als auch auf Unterschiede in der exakten Formulierung der Items (vgl. Fiedler & Schwarz, 2016), oder auf die Stichprobe zurückzuführen sein.

Ähnlich wie in Studien basierend auf Stichproben von Forschenden zeigt sich, zeigt sich, dass bestimmte QRPs scheinbar häufiger angewandt werden als andere. Beispielsweise waren “Selektives Berichten von Variablen” und “Ausschluss von Ausreißern”, unabhängig von der betrachteten Kenngröße, in allen Stichproben unter den am häufigsten genannten QRPs, das “Abrunden von p-Werten” dagegen unter den am seltensten genannten. Als am häufigsten berichtete Forschungspraktik stellte sich der Beginn der Datenerhebung ohne konkrete Stichprobenplanung heraus. Damit zusammenhängend berichten 48,3 % der Befragten, im Laufe des Studiums für keines ihrer empirischen Projekte eine Poweranalyse durchgeführt zu haben. Ein noch höherer Anteil der Befragten (62,3 %[[10]](#footnote-10)) hat noch nie eine Präregistrierung durchgeführt. Ein stärkerer Einsatz u.a. dieser Forschungspraktiken ist bereits vielfach empfohlen worden, um den Annahmen der verwendeten statischen Verfahren besser gerecht zu werden (Banks et al., 2016; Nosek & Lakens, 2014). Der beschränkte Zeithorizont studentischer Projekte (i.d.R. bis zu sechs Monate für eine Masterarbeit und drei Monate für eine Bachelorarbeit) könnte erklären, wieso bestimmte QRPs unter Studierenden im Vergleich zu Forschenden weniger häufig (z.B. Nacherheben von Versuchspersonen) und andere häufiger vorkamen (z.B. fehlende Stichprobenplanung).

### Unterschiede zwischen Projekten

Zwischen den QRP-Häufigkeiten bei verschiedenen Projekttypen lassen sich Unterschiede erkennen. So weist bspw. der Projekttyp Expra im Vergleich zu den anderen erhobenen Typen wie Bachelor- oder Masterarbeiten die höchste Wahrscheinlichkeit für die Anwendung von QRPs auf. Durch ihre Praxis- und Anwendungsnähe sind diese zur Vermittlung methodischer Kenntnisse eine wichtige Ergänzung zu rein theoretischen Veranstaltungen. Die Expras dienen dabei als Raum, um theoretisch erlernte Methoden erstmals auszuprobieren und dabei aus entstandenen Fehlern sowie deren Korrektur durch die Betreuenden zu lernen. In späteren Abschlussarbeiten hingegen sind die methodischen Kenntnisse gefestigter und die Qualitätsanforderungen, die Eigenverantwortung sowie die Motivation der Studierenden, eine qualitativ hochwertige Arbeit zu produzieren, höher. Dies könnte einen Teil der Unterschiede zwischen den Projekten erklären.

## Mögliche Ursachen für QRPs in studentischen Projekten

### Unterschiede zwischen Forschenden und Studierenden

Die möglichen Ursachen für die Verwendung von QRPs unter Forschendenlassen sich in mehrere Bereiche einteilen: Unzureichende methodische Kenntnisse, die Anreizstruktur des akademischen Betriebs, insbesondere der Publikationsdruck, sowie kognitive Verzerrungen (Edwards & Roy, 2017; Krishna & Peter, 2018; Sijtsma, 2016). Während methodische Kenntnisse und kognitive Verzerrungen auch unter Studierenden relevante Faktoren sein dürften, sind Studierende anders als Forschende keinem direkten Publikationsdruck ausgesetzt. Stattdessen befinden sie sich in einem Anreizsystem, das den zügigen Erwerb von ECTS-Punkten mit möglichst guten Bewertungen durch die Dozierenden fördert.

### Die Bedeutung von Betreuenden

Um gute Noten zu erhalten, orientieren sich Studierende zumeist an den Vorgaben der Betreuenden. Krishna & Peter (2018) konnten in ihrer Studie keine Hinweise darauf finden, dass Studierende erwarten, durch statistisch signifikante Ergebnisse bessere Noten erzielen zu können. Sie fanden aber, dass die Einstellungen der Betreuenden die Anwendung von QRPs durch die Studierenden auf zwei Arten, je nach Phase des Projekts, beeinflussen (vgl. Wicherts et al., 2016): Zum einen haben Betreuende bei der Ausarbeitung des Studiendesigns direkten Einfluss auf die mögliche Anwendung von QRPs (z.B. Nacherheben von Versuchspersonen zur Herbeiführung von Signifikanz). Zum anderen beeinflussen Aussagen und Verhaltensweisen der Betreuenden aufgrund ihrer Vorbildfunktion, welche QRPs Studierende während der Datenanalyse und der Berichterstellung verwenden.

### Methodische Kenntnisse

Die Vermutung liegt nahe, dass die allgemeine Methodenkenntnis der Studierenden Einfluss auf die Anwendung von QRPs hat. Dieser Einfluss zeigt sich in unserer Studie möglicherweise in der Beobachtung, dass empirische Projekte umso weniger von QRPs betroffen waren, je mehr sie mit einem fortgeschrittenen Studium assoziiert waren: In Masterarbeiten wurden weniger QRPs eingesetzt als in Bachelorarbeiten und in diesen wiederum weniger als in Expras (s. Abbildungen 2 und 3b). Dieser lineare Trend zeigte sich auch in den explorativen Analysen. Eine naheliegende Erklärung für die geringere QRP-Verbreitung im fortgeschrittenen Studienverlauf ist, dass die Befragten im Laufe ihres Studiums mehr methodische Kompetenzen erwerben und für die Probleme, die mit dem Einsatz von QRPs verbunden sind, stärker sensibilisiert werden. Für die Verwendung der zwei abgefragten positiven Forschungspraktiken zeigte sich ein anderes Muster: Aktuelle Bachelor-Studierende wandten mit einer höheren Wahrscheinlichkeit positive Forschungspraktiken an, als aktuelle Master-Studierende. Erklärbar wäre dies durch eine in jüngster Zeit zunehmend prominente Rolle positiver Forschungspraktiken in der Lehre, die zur Zeit des Bachelorstudiums der aktuellen Master-Studierenden möglicherweise nicht in gleichem Umfang gegeben war. Sowohl die Verwendung von QRPs wie auch positiver Forschungsmethoden hängt mit dem durch die Teilnehmenden subjektiv eingeschätzten Grad ihrer Informiertheit zusammen: Eine höher eingeschätzte Informiertheit ist mit einer wahrscheinlicheren Anwendung positiver und weniger wahrscheinlichen Verwendung von QRPs assoziiert. Dies legt nahe, dass sich das theoretisch erworbene Wissen über Forschungspraktiken, Open Science und die Replikationskrise im Laufe des Studiums in der methodischen Herangehensweise an die eigenen Projekten unserer Befragten niederschlägt.

Eine mangelnde konzeptionelle Trennung zwischen explorativer und konfirmatorischer Forschung in Lehrveranstaltungen könnte ebenfalls einen Einfluss auf die Prävalenz von QRPs haben. Falls Studierenden eine explizite Kennzeichnung explorativer Analysen nicht vermittelt wird, könnte es dadurch zum vermehrten Einsatz fragwürdiger Forschungspraktiken kommen (z.B. Berichten eines unerwarteten Befunds als im Vorhinein hypothetisiert).

## Fragen zur universitären Lehre

### Thematisierung der Replikationskrise

Die aktuelle wissenschaftliche Debatte um die fehlende Replizierbarkeit psychologischer Befunde scheint erfreulicherweise schon substanziellen Einzug in die Lehre gefunden zu haben; immerhin berichten 75,7 % unserer Teilnehmenden von einer Behandlung des Themas in der Lehre an ihrer Hochschule. Dabei scheint diese Thematisierung eher in der Methodenlehre als in fachbezogenen Veranstaltungen stattzufinden. Die explorativen Analysen zeigen einen deutlichen Zusammenhang zwischen der subjektiv eingeschätzten Informiertheit der Studierenden über die Replikationskrise und der Behandlung dieses Thema in der Lehre, sowie mit der wahrgenommenen Wichtigkeit dieses Themas. Eine wissensvermittelnde Wirkung der Lehre, die auch das Bewusstsein unserer Befragten für die Bedeutung dieser Thematik geschärft haben könnte, scheint möglich.

### Behandlung fragwürdiger Forschungspraktiken

Die kritische Betrachtung konkreter QRPs scheint ebenfalls durchaus einen Platz in den von unseren Befragten besuchten Lehrveranstaltungen zu haben. Die Zahlen variieren je nach betrachteter Praktik: Die Spanne reicht von 39,7 % (Abrunden von p-Werten) bis 73,9 % (HARKing) der Befragten, in deren Lehrveranstaltungen die jeweiligen QRPs kritisch behandelt wurden. HARKing, die am häufigsten thematisierte QRP, gehörte zu den am seltensten angewandten Praktiken in empirischen Projekten, was für einen positiven Einfluss der Lehre im Bezug auf diese Praktik sprechen könnte. Im Gegensatz dazu könnte das häufigere Auftreten von anderen QRPs wie “Keine Stichprobenumfangsplanung” oder “Selektives Berichten von Variablen/Bedingungen/Hypothesen” bei gleichzeitig häufiger Betrachtung in Lehrveranstaltungen auch darauf hinweisen, dass entweder prävalente QRPs (wünschenswerterweise) auch häufiger thematisiert werden oder die Behandlung in Lehrveranstaltungen nur wenig Einfluss auf die Anwendung der QRPs hat. Nicht die Behandlung einer spezifischen Praktik in der Lehre scheint unter unseren Befragten für deren Einsatz maßgeblich gewesen zu sein, sondern die generelle subjektiv eingeschätzte Informiertheit der Studierenden über den Themenkomplex “Replikationskrise und Open Science”.

## Limitationen

Die Stichprobe unterliegt wahrscheinlich Verfälschungen durch Selbstselektion. Die Teilnahme an der Umfrage war freiwillig und ohne monetäre Anreize. So kann nicht ausgeschlossen werden, dass Personen, die bereits ein erhöhtes Vorwissen über bzw. Interesse am Themenkomplex “Replikationskrise und Open Science” besaßen, mit einer höheren Wahrscheinlichkeit an der Befragung teilgenommen haben.

Die vorliegende Befragung ist eine Querschnittsuntersuchung, in der teils über verschiedene, von den Teilnehmern retrospektiv berichtete, Projekt- und Lehrveranstaltungen aggregiert wurde. Aus diesem Grund kann die Frage, welche kausalen Mechanismen die vorliegenden Beobachtungen erklären, im Rahmen dieser Umfrage nicht beantwortet werden. Auch fehlerhafte oder verzerrte Erinnerungen könnten die Qualität unserer Daten beeinträchtigt haben.

Eine geläufige Kritik an Studien zur Prävalenz von QRPs bezieht sich auf den Einfluss der sprachlichen Formulierung der jeweiligen Praktiken. Hierbei können diese entweder recht allgemein, wie bei John et al. (2012) oder Agnoli et al. (2017), oder eher spezifisch, wie bei Fiedler & Schwarz (2016), beschrieben werden, was naheliegenderweise einen Einfluss auf die erhobene Prävalenzrate haben könnte. Entsprechende Einflüsse sind auch bei der vorliegenden Umfrage nicht auszuschließen. Durch unseren Begleittext zum “HARKing” (s. Tabelle 1) gaben wir z.B. eine spezifischere Handlung vor, als es in vergleichbaren Studien durch die Formulierung “claiming to have predicted an unpredicted finding” (deutsch: “behaupten, ein unerwartetes Ergebnis sei vorhergesagt worden”) der Fall war. Insgesamt kann nicht ausgeschlossen werden, dass Befragte an manchen Stellen die Beschreibungen der QRPs konzeptuell missverstanden haben. Beispielsweise könnten Teilnehmende bei “flexibler Datenanalyse” auch an zulässige Exploration denken. Dieser Umstand wurde auch von einzelnen Befragten im freien Kommentarfeld kommuniziert. Auch bei der Abfrage bisheriger Präregistrierungen durch die Studierenden könnten sich Verständnisprobleme ergeben haben, inwiefern Exposés, die von vielen Betreuenden vor dem Beginn einer empirischen Arbeit erwünscht werden, bereits dem Anspruch einer Präregistrierungen genügen oder nicht. Solchen Verständnisproblemen versuchten wir durch die Bereitstellung von Erklärungstexten, die zu jeder abgefragten Forschungpraktik angezeigt werden konnten, entgegenzuwirken (s. Tabelle 1).

Bei den abgefragten Kategorien “Projektarbeiten” und “anderen Projekten” bleibt anhand unserer Daten unklar, ob die Antworten der Teilnehmenden sich auf ein oder mehrere Projekte beziehen. Wenn ein nicht zu vernachlässigender Anteil unserer Teilnehmenden mehr als eine “Projektarbeit” und/oder mehr als ein “anderes Projekt” durchgeführt hat, könnten der hier berichtete mittlere Einsatz von QRPs pro Projekt, der jeweilige mittlere Einsatz von QRPs in den Projekten “Projektarbeit” und “andere Projekte” und der Anteil von pro QRP betroffenen Projekten etwas überschätzt sein. Unsere Daten zur Behandlung der Replikationskrise in einzelnen Lehrveranstaltungen sind sehr grob aufgelöst und können daher allenfalls als grobe thematische Einordnung interpretiert werden.

## Fazit

Wie auch die Psychologie-Fachschaften-Konferenz (PsyFaKo) in ihrem Positionspapier (Konferenzrat der Psychologie-Fachschaften-Konferenz, 2018) darlegte, stellt die Replikationskrise in der Geschichte der psychologischen Forschung eine Zäsur dar, die viele fruchtbare Diskussionen über die zukünftige Entwicklung der psychologischen Wissenschaft hervorgebracht hat. Die Empfehlungen zur Steigerung der Robustheit von Forschungsergebnissen konzentrieren sich u.a. auf die Änderung der Anreizstrukturen, die Anwendung von positiven Forschungspraktiken und eine erhöhte Sensibilisierung für QRPs und deren Konsequenzen für die Replizierbarkeit psychologischer Befunde (Banks et al., 2016; Sijtsma, 2016; Simmons et al., 2011). Um diesen gewünschten Wissenszuwachs nachhaltig zu verankern, ist eine weiter vertiefte Thematisierung des Themenkomplexes - auch außerhalb der methodischen Veranstaltungen - in der universitären Lehre unerlässlich. Dies spiegelt sich auch in der Meinung unserer Teilnehmenden wider, die diesen Themen eine große Bedeutung beimessen. Zudem lässt sich feststellen, dass QRPs in vielen studentischen Projekten angewandt werden. Die beobachtete Verbreitung von QRPs sollte Anlass zu Überlegungen geben, wie ihr Einsatz reduziert werden kann.

## Impulse für die psychologische Lehre

### Handlungsempfehlungen

Auf der Grundlage der hier vorgestellten Ergebnisse möchten wir folgende Impulse für die inhaltliche Ausgestaltung der psychologischen Lehre geben:

1. Die Replikationskrise, ihre Ursachen und Lösungsansätze sollten in den methodischen Veranstaltungen im Psychologiestudium flächendeckend als fester Bestandteil thematisiert werden. Das Ziel sollte es sein, dass sich alle Studierenden gut über den Themenkomplex der Replikationskrise, QRPs und Open Science informiert fühlen.
2. Auch in den Grundlagen- und Anwendungsfächern sollte eine kritische Bestandsaufnahme der Replizierbarkeit von anerkannten Befunden erfolgen. Lehrmaterialien in Grundlagen- und Anwendungsfächern sollten vor dem Hintergrund der Replikationskrise hinsichtlich spezifischer Theorien und Modelle, deren Replizierbarkeit fraglich ist, fortlaufend kritisch überarbeitet werden. Anstelle des Erlernens und Weitertragens möglicherweise falsch-positiver Befunde kann so ein kritischer Umgang mit aktuellen Forschungsergebnissen angeregt werden.
3. Empirische Projekte im Rahmen des Studiums sollten genutzt werden, um positive Forschungspraktiken zu vermitteln und die Problematik fragwürdiger Forschungspraktiken in der Praxis zu thematisieren. Das Ziel sollte sein, dass Studierende bereits früh Methoden einer offenen, reproduzierbaren Arbeitsweise beherrschen, die sie zur erfolgreichen Bewältigung aktueller Herausforderungen der Disziplin befähigen.
4. Sowohl explorative als auch konfirmatorische Ansätze sollten in der Lehre behandelt und voneinander abgegrenzt werden. Exploratives wissenschaftliches Arbeiten sollte als notwendig wertgeschätzt und von Studierenden und Forschenden in empirischen Projekten transparent kenntlich gemacht werden. Darin sehen wir nicht nur eine methodische Notwendigkeit, sondern auch eine Chance, die Psychologie als Wissenschaft weiterzuentwickeln. Dadurch kann der Bildung fundierter Theorien auf der Grundlage umfassender Exploration in Verbindung mit sorgfältigen konfirmatorischen Tests mehr Raum gegeben werden (Banks et al., 2016; Glöckner, Fiedler, & Renkewitz, 2018).
5. Die Psychologie als vielfältige Wissenschaft braucht einen Methodenpluralismus, dessen Grundlagen bereits in der akademischen Lehre gelegt werden sollte. Während NHST für viele Fragestellungen ein informativer Ansatz sein kann, könnten für andere Untersuchungen andere Verfahren besser geeignet sein (bspw. bayesianische Hypothesentestung, Wagenmakers et al., 2017; Quantitative Modellierung, Lewandowsky & Farrell, 2011; oder qualitative Ansätze, Hammarberg et al., 2016). Auch die theoretischen Annahmen des NHST sollten bereits im Studium eine Rolle spielen.
6. Studierende haben für die methodisch korrekte Arbeitsweise bei der Anfertigung von Abschlussarbeiten und ähnlichen Projekten eine hohe Eigenverantwortung. Ein ideales Curriculum kann demnach allein keine Garantie für eine vollständige Reduktion von QRPs in studentischen empirischen Projekten sein. Deswegen ist es wichtig, bereits im Studium ein nachhaltiges Bewusstsein für Eigenverantwortung zu schaffen (siehe auch Fiedler & Schwarz, 2016). Das Ziel sollte es sein, Studierende zu Mitwirkenden an einer wissenschaftlichen Kultur auszubilden, in der die Prinzipien der Offenheit, Transparenz und Kollaboration beim Hervorbringen inhaltlich robuster Forschung im Vordergrund stehen.

### Praxisbeispiele

Wie diese Impulse in die Realität umgesetzt werden können, kann anhand bereits umgesetzter Lehrkonzepte nachvollzogen werden. So können Thematiken wie die Replikationskrise, fragwürdige und offene Forschungspraktiken, oder der Publikationsprozess im Rahmen eines dezidierten Kurse an Studierende vermittelt werden, wie z.B. an der Universität Amsterdam („Good Research Practices“, Sarafoglou et al., 2020) oder der Universität Cambridge (Orben, 2020). Chopik et al. (2018) konnten zeigen, dass Studierende durch solche Kurse einen informationellen Zugewinn und eine höhere Sensibilisierung für aktuelle Probleme der Disziplin erfahren. In Bezug auf empirische Projekte im Studium wurde beispielsweise an der Ludwig-Maximilians-Universität München bereits ein methodisches Kern-Curriculum für Expras (Schönbrodt et al., 2018, S. 42) festgelegt, das Studierende mittels der Elemente 1) explizite Stichprobenplanung, 2) Präregistrierung, 3) offene Daten sowie 4) reproduzierbare Analyseskripte an eine von Transparenz geleitete Arbeitsweise heranführt. In solchen empirischen Projekte können auch Replikationen zur Einübung von Open Science-Prinzipien eingesetzt werden, wie in einer Initiative der Fern-Universität Hagen (Jekel et al., 2020) oder im Rahmen des *Collaborative Replications and Education Project* (Wagge et al., 2019). Eine grundlegende didaktische Neuausrichtung der methodischen Schwerpunkte in der Lehre vollzog sich an der School of Psychology in Glasgow. Anstelle des Erlernens vieler statistischer Verfahren wurde der Fokus auf die Kompetenzen gelegt, Daten kritisch analysieren und auf reproduzierbare Weise berichten zu können (u.a. in Bewertungskriterien des Kursprojekts). In diesem Zuge erhielt auch das Arbeiten mit offener und Reproduzierbarkeit unterstützender Software, z.B. RMarkdown oder Git, einen zentralen Stellenwert (Barr et al., 2019). Diese Beispiele[[11]](#footnote-11) zeigen, wie solche Inhalte auf verschiedene Weisen Einfluss in die psychologische Lehre finden können, um eine ganzheitlich solide Ausbildung zu gewährleisten.

### Ausblick

Heutige Studierende werden als zukünftige Wissenschaftler\*innen die Entwicklung der akademischen Psychologie maßgeblich prägen. Deswegen ist es von besonderer Bedeutung, bereits im Studium offene und transparente Forschungspraktiken konsequent umzusetzen, um die psychologische Forschung der Zukunft belastbarer und offener zu gestalten. Mit der vorliegenden Arbeit möchten wir einen Beitrag auf dem Weg zu diesem Ziel leisten. Wir hoffen, dass wir als studentische Open-Science-Initiative der PsyFaKo eine Auseinandersetzung mit dem aktuellen Stand der Lehre in der Psychologie und deren Weiterentwicklung anregen können.

# Anmerkungen

**Danksagung**  
Wir danken den folgenden Mitgliedern der PsyFaKo Open Science AG für ihre Unterstützung bei der Erarbeitung der Umfrage, die diesem Artikel zugrunde liegt: Alina Eiberger, Luise Heyde-Schulte, Pauline Hallmann, Marieke Hill, Nora Güngerich, Jannika Pyttlich und Florian Grünendahl. Besonderen Dank möchten wir Marieke Hill und Alina Eiberger für ihre Beiträge zum Erstentwurf des Manuskripts aussprechen. Zudem danken wir Dr. Felix Schönbrodt, Anne M. Scheel, Prof. Dr. Mario Gollwitzer, Dr. Tanja Gerlach und Dr. Thomas Schultze für ihr wertvolles Feedback zu frühen Fassungen des Manuskripts.

**Autorenbeiträge**  
Konzeptualisierung, Methode: JB MF TK LS MH AE  
Datenaufbereitung, statistische Analyse und Visualisierung: JB  
Manuskript - eingereichte Fassung: JB MF TK LS MH AE  
Manuskript - finale Fassung: JB MF TK LS LV

# Abbildungen

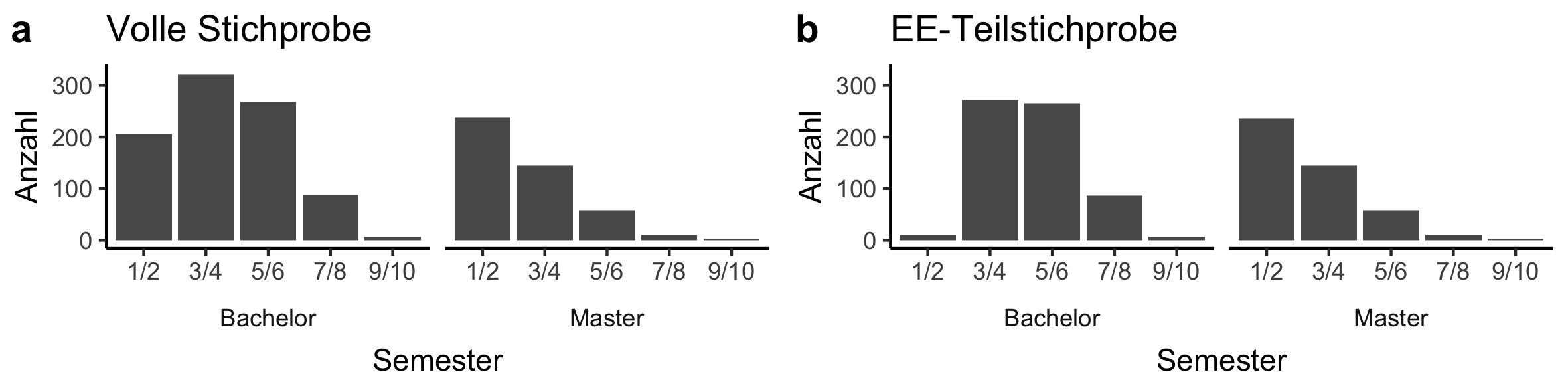


Abbildung 1: Aufschlüsselung der Teilnehmenden nach Studienfortschritt. (a) Volle Stichprobe. (b) EE-Teilstichprobe.

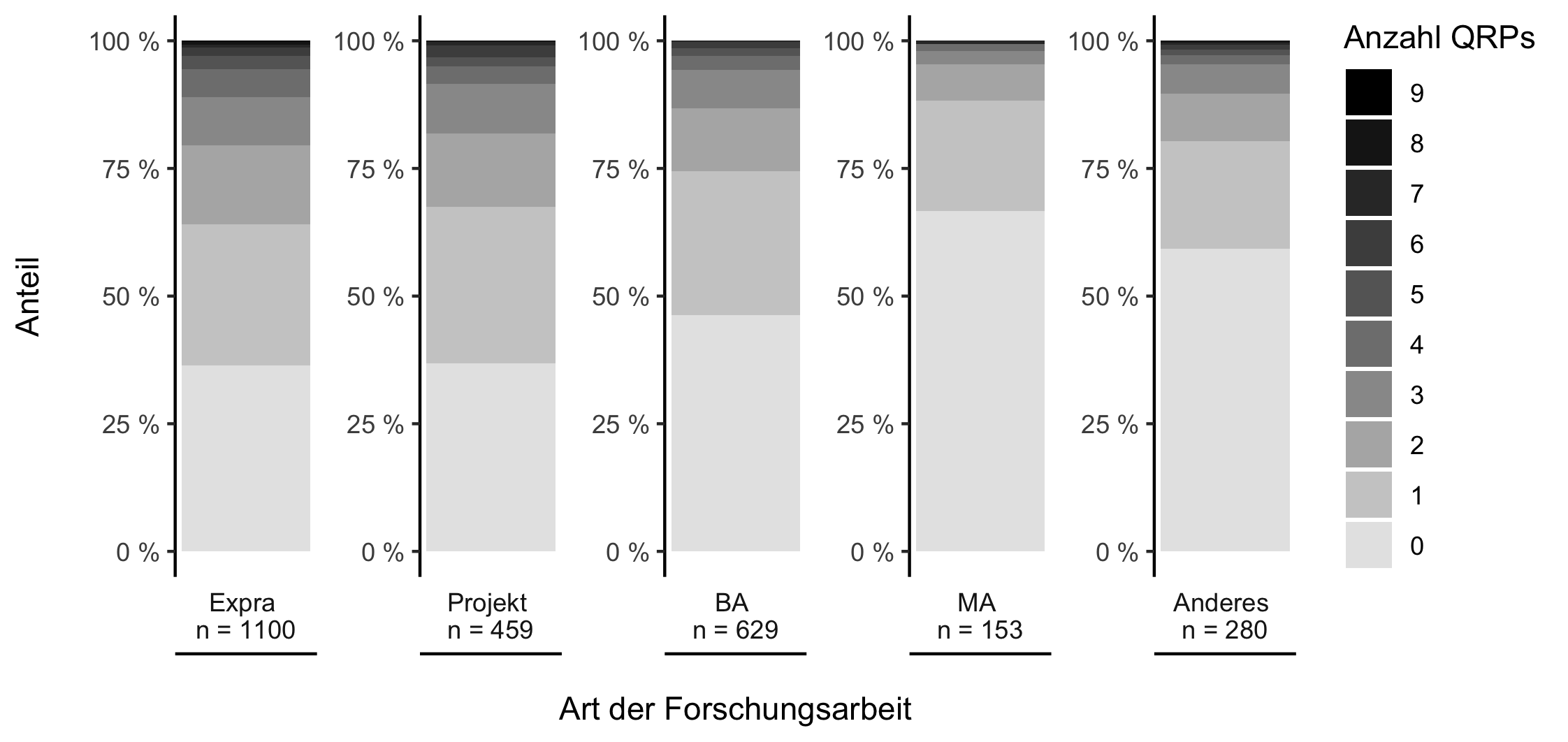


Abbildung 2: Anzahl eingesetzter fragwürdiger Forschungspraktiken (QRPs) aufgeschlüsselt nach Art des Projekts. BA = Bachelorarbeit, MA = Masterarbeit, Expra = Experimentalpraktikum.

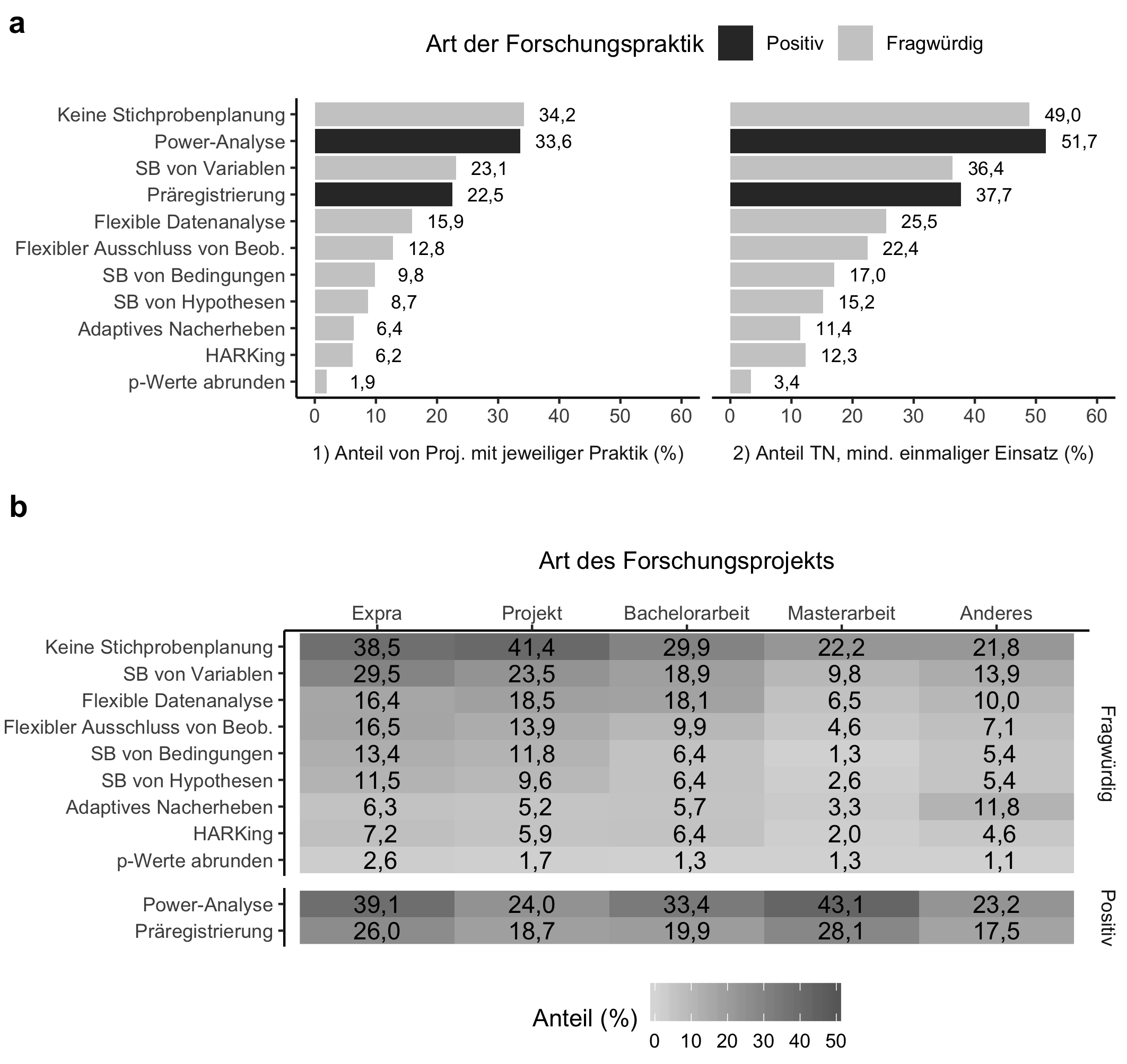


Abbildung 3: a) Anteil von Projekten, in denen die jeweilige Forschungspraktik eingesetzt wurde (links); Anteil von Teilnehmenden, die die jeweilige Forschungspraktik mindestens einmal in einem empirischen Projekt eingesetzt haben (rechts). b) Anteil von Projekten, in denen die jeweilige Forschungspraktik eingesetzt wurde, aufgeteilt nach Projekt. S.B. = Selektives Berichten.

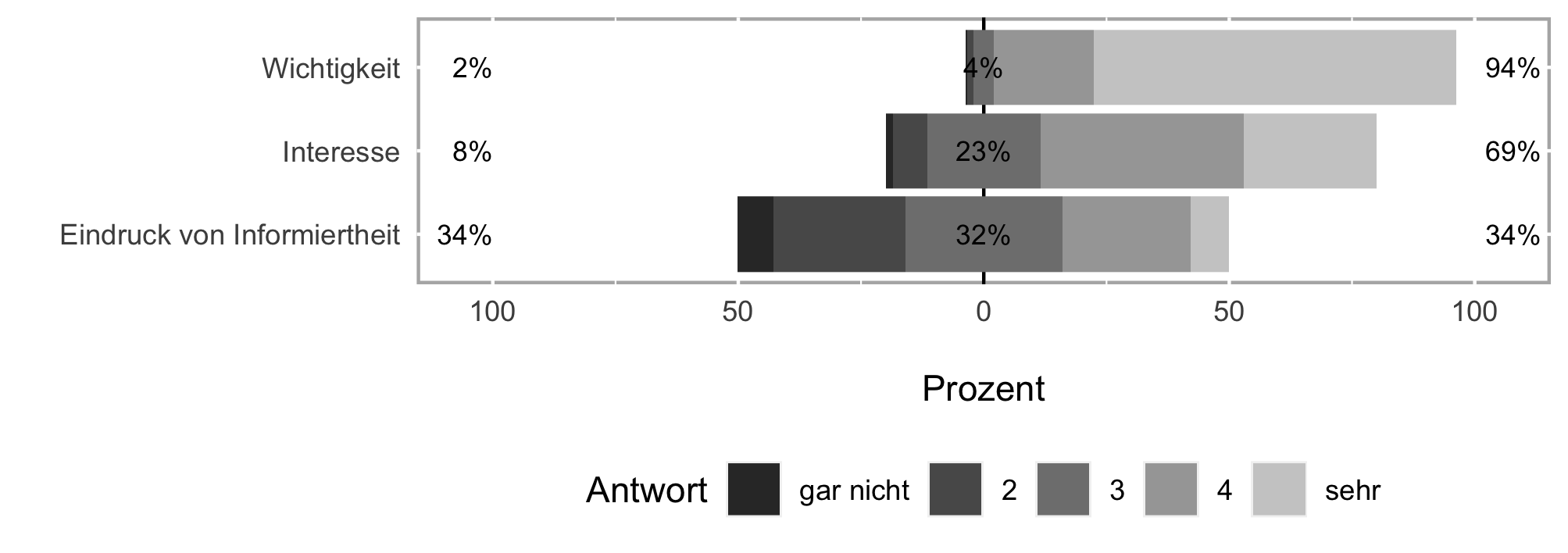


Abbildung 4: Einstellungen der Teilnehmenden zum Themenkomplex “Open Science und Replikationskrise” auf einer 5-Punkte Likert-Skala von “gar nicht” bis “sehr”. Bedeutung der Prozentzahlen: “gar nicht” & “2” (links), “3” (mittig), “4” & “sehr” (rechts).

# Tabellen

Tabelle 1. Abgefragte Forschungspraktiken und Begleittexte.

| Bezeichnung (kurz) | Begleittext |
| --- | --- |
| **Fragwürdige Forschungspraktiken** | |
| Datenerhebung ohne konkrete Stichprobenplanung beginnen (Keine Stichprobenplanung) | Es wird eine willkürliche Zahl von Versuchspersonen erhoben, oder es werden Daumenregeln angewandt (z.B. mindestens 20 VP pro Bedingung). Im Vorfeld eines Projekts findet keine Power-Analyse statt, um die Stichprobengröße zu bestimmen. |
| Über zusätzliche Erhebung entscheiden, nachdem die Ergebnisse bekannt sind (Adaptives Nacherheben) | Beispiel: Bei einem knapp nicht signifikanten p-Wert 10 zusätzliche Versuchspersonen rekrutieren (z.B. in der Hoffnung, doch noch ein signifikantes Ergebnis zu erzielen) |
| Flexible Datenanalyse ohne Kennzeichnung (Flexible Datenanalyse) | Beispiel: Verschiedene, im Vorfeld nicht genau festgelegte statistische Tests für ein und dieselbe Hypothese durchführen oder den Einfluss von Kovariaten (z.B. Geschlecht) untersuchen, über die es ursprünglich keine spezifischen Hypothesen gab. Die Analyse wird nicht als explorativ gekennzeichnet. |
| Flexibler Ausschluss von Ausreißern (Flexibler Ausschluss von Beob.) | Beispiel: Der ursprüngliche Test einer Hypothese ergab einen p-Wert von p = .13. Nach dem Ausschluss von vier Ausreißern (±3 Standardabweichungen über dem Mittelwert) wird ein p-Wert von p = .043 erreicht. Daraufhin wird entschieden, die Ausreißer auszuschließen, um den p-Wert signifikant zu machen. Im Bericht wird nur die zweite Analyse berichtet. |
| Hypothesen an Ergebnisse anpassen (HARKing) | Beispiel: Vor der Analyse der Daten gab es die Hypothese, dass eine untersuchte Therapie-Methode wirksam ist. Der Effekt wird nicht signifikant, aber es zeigt sich eine Interaktion mit dem Geschlecht der Patient\*innen. Deshalb wird die Hypothese dieser Studie geändert, die Rolle des Geschlechts wird in die Hypothese mit aufgenommen. |
| Selektives Berichten von Hypothesen (SB von Hypothesen) | Beispiel: Von den ursprünglichen fünf Hypothesen lieferten zwei signifikante Ergebnisse. Nur diese beiden Hypothesen werden berichtet. |
| Selektives Berichten von experimentellen Bedingungen (SB von Bedingungen) | Beispiel: Von drei Bedingungen (Kontrollgruppe, Experimentalgruppe 1 und Experimentalgruppe 2) nur die ersten beiden im Artikel/Bericht beschreiben, zwischen denen es einen signifikanten Unterschied gab. |
| Selektives Berichten von Variablen | Beispiel: Von zehn erhobenen abhängigen Variablen nur die drei Variablen im Bericht/Artikel beschreiben, die statistisch signifikant mit der unabhängigen Variable zusammenhingen. |
| p-Werte abrunden | Beispiel: Der Test für die zentrale Hypothese ergibt einen p-Wert von p = .058, im Bericht wird p = .05 geschrieben. |
| **Positive Forschungspraktiken** | |
| Poweranalyse durchführen (Power-Analyse) | Beispiel: Vor der Datenerhebung wurde ausgerechnet, dass die Stichprobe aus mindestens 60 Personen bestehen muss, um den zu untersuchenden Effekt mit einer Wahrscheinlichkeit von z.B. 80% finden zu können. |
| Ein Experiment oder Teile davon präregistrieren (Präregistrierung) | Beispiel: Vor der Datenerhebung wurde schriftlich festgehalten, welche Daten erhoben, welche Hypothesen untersucht und welche statistischen Tests verwendet werden. |

Tabelle 2. Korrelationen zwischen numerischen Variablen.

|  | Alter | Semester | Wichtigkeit | Subj. Inform. | Interesse | Anz. Projekte |
| --- | --- | --- | --- | --- | --- | --- |
| Alter | 1,00 |  |  |  |  |  |
| Semester | 0,15\*\*\* | 1,00 |  |  |  |  |
| Wichtigkeit | 0,10\*\* | 0,07\* | 1,00 |  |  |  |
| Subj. Inform. | 0,06 | 0,10\*\*\* | 0,27\*\*\* | 1,00 |  |  |
| Interesse | 0,14\*\*\* | 0,04 | 0,49\*\*\* | 0,19\*\*\* | 1,00 |  |
| Anz. Projektea | 0,31\*\*\* | 0,23\*\*\* | 0,13\*\*\* | 0,18\*\*\* | 0,13\*\*\* | 1,00 |
| Markierungen: \* p < 0,0083, \*\* p < 0,0017, \*\*\* p < 0,0002 (adjustierte Alpha-Niveaus nach Bonferroni) | | | | | | |
| aAnzahl der von der/dem jeweiligen Teilnehmenden begonnenen oder durchgeführten Projekte. | | | | | | |

Tabelle 3. Modell zur Untersuchung der Zusammenhänge von Kovariaten mit dem Einsatz fragwürdiger Forschungspraktiken. Schätzungen der fixen Effekte. Abhängige Variable: Forschungspraktik angewendet (0 – Nein, 1 – Ja).

| Prädiktor | B | SE | OR | KI (95.00 %) | z | p |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | -2,13 | 0,39 | 0,12 | [0,06; 0,26] | -5,43 | <.001\*\*\* |
| Alter | 0,06 | 0,06 | 1,06 | [0,95; 1,19] | 1,05 | .292 |
| Geschlecht - Männlich | 0,06 | 0,11 | 1,06 | [0,85; 1,32] | 0,49 | .626 |
| Geschlecht - Divers | -0,13 | 1,00 | 0,88 | [0,12; 6,19] | -0,13 | .895 |
| Geschlecht - Keine Antwort | -0,38 | 0,49 | 0,68 | [0,26; 1,78] | -0,78 | .433 |
| SF - Studiere nicht (mehr) | 0,71 | 0,31 | 2,04 | [1,11; 3,74] | 2,30 | .022 |
| SF - Master | 0,18 | 0,13 | 1,20 | [0,93; 1,55] | 1,41 | .158 |
| SF - Andere | -0,87 | 1,15 | 0,42 | [0,04; 3,98] | -0,76 | .450 |
| Anzahl emp. Projekte | -0,07 | 0,06 | 0,93 | [0,83; 1,04] | -1,28 | .200 |
| Interesse | 0,03 | 0,05 | 1,03 | [0,93; 1,14] | 0,57 | .566 |
| Wichtigkeit | -0,22 | 0,06 | 0,80 | [0,72; 0,90] | -3,96 | <.001\*\* |
| Eindruck von Informiertheit | -0,33 | 0,06 | 0,72 | [0,65; 0,80] | -5,86 | <.001\*\*\* |
| OSI an der Uni - Ja | 0,14 | 0,16 | 1,15 | [0,85; 1,57] | 0,91 | .363 |
| OSI bekannt - Unsicher | 0,04 | 0,12 | 1,04 | [0,82; 1,33] | 0,34 | .731 |
| OSI bekannt - Ja | -0,28 | 0,12 | 0,76 | [0,60; 0,97] | -2,23 | .026 |
| RK gelehrt - Unsicher | 0,54 | 0,26 | 1,72 | [1,03; 2,89] | 2,06 | .040 |
| RK gelehrt - Ja | -0,08 | 0,13 | 0,92 | [0,71; 1,18] | -0,66 | .512 |
| Praktik gelehrt - Ja | -0,08 | 0,06 | 0,92 | [0,82; 1,04] | -1,36 | .175 |
| Projekt – Anderes | -0,79 | 0,11 | 0,45 | [0,37; 0,56] | -7,29 | <.001\*\*\* |
| Projekt – Projektarbeit | -0,27 | 0,08 | 0,77 | [0,66; 0,89] | -3,38 | .001\* |
| Projekt – Bachelorarbeit | -0,46 | 0,07 | 0,63 | [0,55; 0,72] | -6,87 | <.001\*\*\* |
| Projekt – Masterarbeit | -1,31 | 0,17 | 0,27 | [0,19; 0,37] | -7,84 | <.001\*\*\* |
| \*Anmerkungen\*. Datenpunkte insg. = 19 052, Projekte = 2 126, Teilnehmende = 1 063 | | | | | | |
| Referenzkategorien: Weiblich (Geschlecht), Bachelor (Studienfortschritt), Nein (OSI an der Uni), Nein (OSI bekannt), Nein (RK gelehrt), Nein (Praktik gelehrt), Expra (Projekt). | | | | | | |
| Abkürzungen: B – Beta-Gewicht, SE – Standard Error (Standardfehler), OR – Odds Ratio (Chancenverhältnis), KI - Konfidenzintervall, RK - Replikationskrise, OSI - Open Science Initiative, QRP – questionable research practice (Fragwürdige Forschungspraktik), SB – Selektives Berichten, SF - Studienfortschritt | | | | | | |
| Markierungen: \* p < 0,00238, \*\* p < 0,00048, \*\*\* p < 0,00005 (adjustierte Alpha-Niveaus nach Bonferroni) | | | | | | |

Tabelle 4. Modell zur Untersuchung von Zusammenhängen der Kovariaten mit dem Eindruck der Teilnehmenden, wie gut sie über den Themenkomplex ‘Replikationskrise und Open Science’ informiert sind. Schätzungen der fixen Effekte. Abhängige Variable: Eindruck von Informiertheit (numerisch, min: 1 (gar nicht), max: 5 (sehr), z-standardisiert).

|  | B | SE | t | p |
| --- | --- | --- | --- | --- |
| (Intercept) | -0,87 | 0,08 | -11,35 | <.001\*\*\* |
| Alter | -0,02 | 0,03 | -0,82 | .410 |
| Geschlecht - Männlich | 0,13 | 0,06 | 2,30 | .022 |
| Geschlecht - Divers | -0,30 | 0,49 | -0,61 | .540 |
| Geschlecht - Keine Antwort | 0,09 | 0,22 | 0,39 | .697 |
| SF - Studiere nicht (mehr) | -0,15 | 0,15 | -0,98 | .326 |
| SF - Master | 0,00 | 0,07 | 0,01 | .988 |
| SF - Andere | 0,10 | 0,29 | 0,34 | .736 |
| Anzahl emp. Projekte | 0,10 | 0,02 | 4,16 | <.001\*\*\* |
| Interesse | 0,07 | 0,03 | 2,63 | .009 |
| Wichtigkeit | 0,14 | 0,03 | 5,27 | <.001\*\*\* |
| OSI an der Uni - Ja | 0,25 | 0,11 | 2,34 | .027 |
| OSI bekannt - Unsicher | 0,26 | 0,06 | 4,58 | <.001\*\*\* |
| OSI bekannt - Ja | 0,41 | 0,06 | 6,73 | <.001\*\*\* |
| RK gelehrt - Unsicher | 0,11 | 0,12 | 0,90 | .367 |
| RK gelehrt - Ja | 0,63 | 0,06 | 10,47 | <.001\*\*\* |
| Anmerkungen. \*N\* = 1 397 | | | | |
| Referenzkategorien: Weiblich (Geschlecht), Bachelor (Studienfortschritt), Nein (OSI an der Uni), Nein (OSI bekannt), Nein (RK gelehrt). | | | | |
| Abkürzungen: B – Beta-Gewicht, SE – Standard Error (Standardfehler), OR – Odds Ratio (Chancenverhältnis), KI - Konfidenzintervall, SF - Studienfortschritt, OSI - Open Science Initiative, RK - Replikationskrise | | | | |
| Markierungen: \* p < 0,00333, \*\* p < 0,00067, \*\*\* p < 0,00007 (adjustierte Alpha-Niveaus nach Bonferroni) | | | | |

# Literaturverzeichnis

Agnoli, F., Wicherts, J. M., Veldkamp, C. L., Albiero, P., & Cubelli, R. (2017). Questionable research practices among Italian research psychologists. *PLoS ONE*, *12*(3). <https://doi.org/10.1371/journal.pone.0172792>

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2020). *Rmarkdown: dynamic documents for R*. <https://github.com/rstudio/rmarkdown>

Amrhein, V., Greenland, S., & Mcshane, B. B. (2019). Statistical significance gives bias a free pass. *European Journal of Clinical Investigation*, *49*(12). <https://doi.org/10.1111/eci.13176>

Arslan, R. C., Walther, M. P., & Tata, C. S. (2019). Formr: a study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-019-01236-y>

Bakker, M., Dijk, A. van, & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*(6), 543–554. <https://doi.org/10.1177/1745691612459060>

Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. (2016). Editorial: Evidence on questionable research practices: the good, the bad, and the ugly. *Journal of Business and Psychology*, *31*(3), 323–338. <https://doi.org/10.1007/s10869-016-9456-7>

Barr, D., Woods, H. C., DeBruine, L., Lai, R., McAleer, P., McNee, S., Nordmann, E., Paterson, H., & Stack, N. (2019). *Redesigning methods curricula for reproducibility*. <https://psyteachr.github.io/sips2019/>

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using {lme4}. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. <https://doi.org/10.1038/s41562-018-0399-z>

Chopik, W. J., Bremner, R. H., Defever, A. M., & Keller, V. N. (2018). How (and whether) to teach undergraduates about the replication crisis in psychological science. *Teaching of Psychology*, *45*(2), 158–163. <https://doi.org/10.1177/0098628318762900>

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, *49*(12), 997–1003. <https://doi.org/10.1037/0003-066X.49.12.997>

Deutsche Gesellschaft für Psychologie. (2005). *Empfehlungen der DGPs zur Einrichtung von B.Sc.-M.Sc.-Studiengängen in Psychologie an den Universitäten*. [https://www.dgps.de/index.php?id=143{\&}tx{\\_}ttnews{\%}5Btt{\\_}news{\%}5D=993{\&}cHash=03097507021fb333802bbeaf76963033](https://www.dgps.de/index.php?id=143%7b\&%7dtx%7b\_%7dttnews%7b\%25%7d5Btt%7b\_%7dnews%7b\%25%7d5D=993%7b\&%7dcHash=03097507021fb333802bbeaf76963033)

Devezer, B., Nardin, L. G., Baumgaertner, B., & Buzbas, E. O. (2019). Scientific discovery in a model-centric framework: Reproducibility, innovation, and epistemic diversity. *PLoS ONE*, *14*(5), 1–23. <https://doi.org/10.1371/journal.pone.0216125>

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin and Review*, *25*(1), 207–218. <https://doi.org/10.3758/s13423-017-1266-z>

Edwards, M. A., & Roy, S. (2017). Academic research in the 21st century: maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environmental Engineering Science*, *34*(1), 51–61. <https://doi.org/10.1089/ees.2016.0223>

Fiedler, K., & Prager, J. (2018). The regression trap and other pitfalls of replication science—illustrated by the report of the Open Science Collaboration. *Basic and Applied Social Psychology*, *40*(3), 115–124. <https://doi.org/10.1080/01973533.2017.1421953>

Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological and Personality Science*, *7*(1), 45–52. <https://doi.org/10.1177/1948550615612150>

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science". *Science*, *351*(6277). <https://doi.org/10.1126/science.aad7243>

Hammarberg, K., Kirkman, M., & Lacey, S. D. (2016). *Qualitative research methods: when to use them and how to judge them*. *31*(3), 498–501. <https://doi.org/10.1093/humrep/dev334>

Jekel, M., Fiedler, S., Torras, R. A., Mischkowski, D., Dorrough, A. R., & Glo, A. (2020). How to teach open science principles in the undergraduate curriculum - the Hagen cumulative science project. *Psychology Learning & Teaching*, *19*(1), 91–106. <https://doi.org/10.1177/1475725719868149>

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. <https://doi.org/10.1177/0956797611430953>

Klauer, K. C. (2018). Themenheft Replizierbarkeit. *Psychologische Rundschau*, *69*(1), 1–2. <https://doi.org/10.1026/0033-3042/a000385>

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., … Nosek, B. A. (2018). Many Labs 2: investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–490. <https://doi.org/10.1177/2515245918810225>

Konferenzrat der Psychologie-Fachschaften-Konferenz. (2018). *Positionspapier der Psychologie-Fachschaften-Konferenz (PsyFaKo) zum Thema „Replikationskrise und Open Science in der Psychologie“*. <https://psyfako.org/index.php/positionspapiere/>

Krishna, A., & Peter, S. M. (2018). Questionable research practices in student final theses – Prevalence, attitudes, and the role of the supervisor’s perceived attitudes. *PLoS ONE*, *13*(8). <https://doi.org/10.1371/journal.pone.0203470>

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). {lmerTest} package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: principles and practice*. SAGE. <https://doi.org/10.4135/9781483349428>

Mcshane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. (2019). Abandon statistical significance. *The American Statistician*, *73*(S1), 235–245. <https://doi.org/10.1080/00031305.2018.1527253>

Munafò, M. R., & Davey Smith, G. (2018). Robust research needs many lines of evidence. *Nature*, *553*(7689), 399–401. <https://doi.org/10.1038/d41586-018-01023-3>

Nosek, B. A., Ebersole, C. R., Dehaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>

Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLoS Biology*, *18*(3), 1–8. <https://doi.org/10.1371/journal.pbio.3000691>

Nosek, B. A., & Lakens, D. (2014). Registered reports. *Social Psychology*, *45*(3), 137–141. <https://doi.org/10.1027/1864-9335/a000192>

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. <https://doi.org/10.1126/science.aac4716>

Orben, A. (2020). *Psychology as a robust science [Course Syllabus]*. <https://www.amyorben.com/docs/syllabus.pdf>

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>

Sarafoglou, A., Hoogeveen, S., Matzke, D., & Wagenmakers, E.-J. (2020). Teaching good research practices: protocol of a research master course. *Psychology, Learning & Teaching*, *19*(1), 45—–59. <https://doi.org/10.1177/1475725719858807>

Schönbrodt, F. D., Maier, M., Heene, M., & Bühner, M. (2018). Forschungstransparenz als hohes wissenschaftliches Gut stärken. *Psychologische Rundschau*, *69*(1), 37–44. <https://doi.org/10.1026/0033-3042/a000386>

Schönbrodt, F., Gollwitzer, M., & Abele-Brehm, A. (2017). Der Umgang mit Forschungsdaten im Fach Psychologie: Konkretisierung der DFG-Leitlinien. *Psychologische Rundschau*, *68*(1), 20–35. <https://doi.org/10.1026/0033-3042/a000341>

Sijtsma, K. (2016). Playing with data—or how to discourage questionable research practices and stimulate researchers to do things right. *Psychometrika*, *81*(1), 1–15. <https://doi.org/10.1007/s11336-015-9446-0>

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>

Stanley, D. J., & Spence, J. R. (2014). Expectations for replications: are yours realistic? *Perspectives on Psychological Science*, *9*(3), 305–318. <https://doi.org/10.1177/1745691614528518>

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, *9*(1), 59–71. <https://doi.org/10.1177/1745691613514450>

Ulrich, R., Erdfelder, E., Deutsch, R., Strauß, B., Brüggemann, A., Hannover, B., Tuschen-Caffier, B., Kirschbaum, C., Blickle, G., Möller, J., & Rief, W. (2016). Inflation von falsch-positiven Befunden in der psychologischen Forschung. *Psychologische Rundschau*, *67*(3), 163–174. <https://doi.org/10.1026/0033-3042/a000296>

Wagenmakers, E.-J., Verhagen, J., Ly, A., Matzke, D., Steingroever, H., Rouder, J. N., & Morey, R. (2017). The need for Bayesian hypothesis testing in psychological science. In S. O. Lilienfeld & I. Waldman (Hrsg.), *Psychological science under scrutiny: recent challenges and proposed solutions*. John Wiley; Sons.

Wagge, J. R., Brandt, M. J., Lazarevic, L. B., Legate, N., Christopherson, C., Wiggins, B., & Grahe, J. E. (2019). *Publishing research with undergraduate students via replication work: the collaborative replications and education project*. *10*. <https://doi.org/10.3389/fpsyg.2019.00247>

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Aert, R. C. van, & Assen, M. A. van. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid P-hacking. *Frontiers in Psychology*, *7*(NOV), 1–12. <https://doi.org/10.3389/fpsyg.2016.01832>

Wickham, H. (2017). *Tidyverse: easily install and load the ’tidyverse’*. <https://cran.r-project.org/package=tidyverse>

1. Die PsyFaKo bildet die Interessenvertretung der deutschen Psychologiestudierenden. Auf der einmal pro Semester stattfindenden Konferenz versammeln sich Fachschaftsdelegationen aus dem deutschsprachigen Raum, um sich zu vernetzen, sich über aktuelle Entwicklungen im Fach Psychologie auszutauschen und um Positionspapiere zu verabschieden. Die Autorinnen und Autoren dieses Berichts sind Teil der Arbeitsgruppe “Open Science”, die auf der 27. PsyFaKo in Würzburg (2018) gegründet wurde. [↑](#footnote-ref-1)
2. Bei Universitätswechsel zwischen Bachelor und Master wurde je ein Datenpunkt für die Bachelor- und die Master-Hochschule gezählt. Das trifft auf 188 Teilnehmende zu. [↑](#footnote-ref-2)
3. Zusätzlich angezeigte Begriffserklärung: “Bspw. Arbeit an einer Studie im Rahmen eines Seminars” [↑](#footnote-ref-3)
4. Zusätzlich angezeigte Begriffserklärung: “Bspw. Forschungspraktikum oder Hiwi-Job” [↑](#footnote-ref-4)
5. Bei den abgefragten Kategorien “Projektarbeiten” und “anderen Projekten” bleibt anhand unserer Daten unklar, ob die Antworten Teilnehmenden sich auf ein oder mehrere Projekte beziehen (siehe Limitationen). Bei Ausschluss dieser Kategorien ergibt sich ein mittlerer Einsatz von 0,99 (*SD* = 1,25) pro Projekt. [↑](#footnote-ref-5)
6. Siehe vorherige Fußnote. [↑](#footnote-ref-6)
7. Siehe vorherige Fußnote. [↑](#footnote-ref-7)
8. Möglicherweise bezogen sich die Antworten unserer Teilnehmenden hier auch auf die üblichen Exposés, was die hohen Zahlen erklären würde. [↑](#footnote-ref-8)
9. Die in dieser Studie verwendeten Begriffe explorativer und konfirmatorischer Forschung sollen hypothesengenerierende Untersuchungen von der Testung im Vorhinein eindeutig festgelegter Hypothesen abgrenzen. Diese Definition entspricht dem vorherrschenden Verständnis der psychologischen Wissenschaftsgemeinschaft (siehe Nosek et al., 2018). Alle Ergebnisse dieser Umfrage sind damit als explorativ anzusehen, nachdem keine Hypothesen im Voraus festgelegt wurden. [↑](#footnote-ref-9)
10. Berechnet durch den Anteil von Teilnehmenden, die angaben, noch nie ein empirischen Projekt präregistriert zu haben. Bei der eigentlich äquivalenten Berechnung als Anteil von Teilnehmenden, die nicht mindestens einmal ein empirischen Projekt präregistriert haben ergibt sich ein Anteil von 60,7 %. Die leicht unterschiedlichen Zahlen sind erklärbar durch kleine Inkonsistenzen im Antwortverhalten der Teilnehmenden, da es sich um zwei unterschiedliche Fragen handelte. [↑](#footnote-ref-10)
11. Weitere Beispiele zu Konzepten von Open-Science-Kursen können auch hier eingesehen werden: <https://osf.io/vkhbt/> [↑](#footnote-ref-11)