

# Replikationskrise, Open Science und p-hacking: Die Studierenden

PsyFaKo

16.09.2019

In den letzten Jahren hat sich innerhalb der Psychologie eine intensive Auseinandersetzung mit den Auswirkungen der Replikationskrise sowie den daraus nötigen Weiterentwicklung der Disziplin ereignet. Als ein Grund für die mangelnde Replizierbarkeit psychologischer Forschung wurden die Verwendung fragwürdiger Forschungspraktiken (eng. QRPs), identifiziert. Während es umfangreiche Untersuchungen zur Verbreitung von QRPs unter Wissenschaftler und Wissenschaftlerinnen gibt, ist bisher wenig über die Prävalenz dieser Praktiken unter Studierenden bekannt. In der hier vorgestellten Arbeit wurde daher erstmals eine systematische Befragung unter 1398 Psychologie-Studierenden im deutschsprachigen Raum durchgeführt, um die Verbreitung von QRPs in empirischen 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 und das empirische Vorgehen von Studierenden miteinander zusammenhängen. Die Ergebnisse zeigen, dass fragwürdige Forschungspraktiken auch im Studium verbreitet sind, wobei sich die Prävalenz stark zwischen verschiedenen QRPs und verschiedenen Projektentypen unterscheidet. Unsere Daten geben Hinweise darauf, dass die Verbreitung von QRPs über den Studienverlauf abnimmt. Zudem ist ein Großteil der Studierenden bereits mit der Thematik der Replikationskrise in der Lehre in Berührung gekommen. Deren Behandlung findet größtenteils in der Methodenlehre und weniger in inhaltlich spezialisierten Fächern statt. Auf Basis dieser und weiterer Erkenntnisse benennt die Arbeit abschließend Handlungsempfehlungen für die Weiterentwicklung der psychologischen Lehre, in der die Prinzipien der Offenheit, Transparenz und Kollaboration beim Hervorbringen inhaltlich robuster Forschung bereits während Studium im Vordergrund stehen.

## 27 1 Einleitung

28 Die Nachrichten von der sogenannten Replikationskrise in der Psychologie haben innerhalb der  
29 Disziplin in den letzten Jahren hohe Wellen geschlagen. Verschiedene, groß angelegte Projekte  
30 wurden zur Untersuchung der Reproduzierbarkeit psychologischer Forschung durchgeführt: Die  
31 Open Science Collaboration (2015) schätzte in einer ersten systematischen Untersuchung  
32 anhand von 100 Studien aus drei einflussreichen Journals die Reproduzierbarkeit auf einen  
33 Wert zwischen 36 % (erfolgreiche Replikationen nach übereinstimmenden oder kleineren p-  
34 Werten) und 68 % (signifikanter meta-analytischer Effekt) der untersuchten Studien. Im Projekt  
35 Many Labs 2, in dem 28 Effekte durch über 60 internationale Forschungsgruppen repliziert  
36 wurden, fanden Klein et al. (2018) eine Rate von 54 % erfolgreichen Replikationen. In einer  
37 weiteren Untersuchung konnten Camerer et al. (2018) nur 13 von 21 (62 %) in Nature und  
38 Science veröffentlichte psychologische Studien erfolgreich replizieren.

39 Zahlreiche Publikationen in der Psychologischen Rundschau, darunter zur Inflation von falsch-  
40 positiven Befunden (Ulrich et al., 2016), zum Umgang mit Forschungsdaten (Schönbrodt,  
41 Gollwitzer, & Abele-Brehm, 2017) und schließlich das Themenheft Replizierbarkeit (Klauer,  
42 2018), spiegeln die Relevanz des Themas für unsere Disziplin wider. Auch innerhalb der  
43 Studierendenschaft gewann das Thema jüngst an Relevanz: 2018 wurde auf der 27. PsyFaKo  
44 (Psychologie-Fachschaften-Konferenz<sup>1</sup>) in Würzburg ein Positionspapier zur Replikationskrise  
45 der Psychologie (Konferenzrat der Psychologie-Fachschaften-Konferenz, 2018) verabschiedet,  
46 in dem die Studierenden unter anderem forderten, dass die Themen Replikationskrise und Open  
47 Science einen größeren Stellenwert in der Lehre erhalten. Die Studierenden von heute werden  
48 das Fach Psychologie in Zukunft prägen und entsprechend wichtig ist es, dass das Studium

---

<sup>1</sup> 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 Autor\*innen dieses Berichts sind Teil der Arbeitsgruppe "Open Science", die auf der 27. PsyFaKo in Würzburg (2018) gegründet wurde.

durch eine sorgfältige und zeitgemäße Ausbildung eine angemessene Vorbereitung auf die vielfältigen beruflichen Tätigkeiten von Psycholog\*innen darstellt.

Die vorhandenen Untersuchungen zur Replikationskrise und Open Science beziehen sich primär auf die Art und Weise, wie in der Psychologie Erkenntnisse generiert und welche Forschungsmethoden angewendet werden. Inwiefern diese Themen auch im Rahmen der akademischen Lehre vermittelt werden, stand bisher nicht im Fokus der Debatte. In der hier vorgestellten Arbeit wurde erstmals eine systematische Befragung unter Psychologie-Studierenden im deutschsprachigen Raum durchgeführt, die die Thematisierung der Replikationskrise in der Lehre, die Verwendung fragwürdiger Forschungspraktiken (questionable research practices, QRPs) in empirischen Projekten während des Studiums untersucht. Die gemeinsame Betrachtung der Lehre und des Einsatzes fragwürdiger Forschungspraktiken versprechen Aufschluss darüber, wie die psychologische Lehre und das empirische Vorgehen von Studierenden miteinander zusammenhängen.

Fragwürdige Forschungspraktiken rückten im Rahmen der Replikationskrise als eine möglich Ursache für die mangelnde Replizierbarkeit von psychologischen Effekten in den Fokus (siehe z.B. Simmons, Nelson, & Simonsohn, 2011). Als QRPs gelten Methoden, wie z.B. das nachträgliche Ausschließen von Ausreißern oder das selektive Berichten von Variablen. Diese erhöhen die Wahrscheinlichkeit von falsch-positiven Befunden (Simmons et al., 2011). Die Anwendung von QRPs beeinträchtigt die Verlässlichkeit psychologischer Befunde und unterminiert die empirische Basis der Wissenschaft. Während die Verbreitung von QRPs innerhalb der Wissenschaftsgemeinschaft mehrfach untersucht wurde (u.a. John, Loewenstein, & Prelec, 2012; Fiedler & Schwarz, 2016), ist uns zur Prävalenz von QRPs unter Studierenden nur eine Untersuchung bekannt (Krishna & Peter, 2018). Die darin ermittelten Prävalenzraten für die Anwendung von QRPs in studentischen Abschlussarbeiten sind mit konservativen Schätzungen der Prävalenzraten unter Wissenschaftler\*innen vergleichbar. Ferner wurde die Verwendung von QRPs durch die Einstellungen der Studierenden, die sowohl direkt als auch indirekt mit den Einstellungen der betreuenden Dozierenden zusammenhängt wird, beeinflusst.

Als eine mögliche Ursache für die verbreitete Anwendung von QRPs durch Forschende werden häufig die aktuellen Anreizstrukturen des Wissenschaftsbetriebs, die sich im Schlagwort "Publikationsdruck" manifestieren, genannt (siehe z.B. Bakker, Dijk, & Wicherts, 2012). Umso spannender war für uns die Frage, welche QRPs von Studierenden, die diesem Anreizsystem in

geringerem Maß unterliegen sollten, verwendet werden, und ob sich deren Vorkommen über die im Laufe des Studiums zu absolvierenden empirischen Arbeiten (Seminar-, Bachelor-, Masterarbeit, etc.) verändert.

Die von uns durchgeführte Umfrage besteht aus drei Teilbereichen, deren Ergebnisse einen Einblick in 1) die Verbreitung fragwürdiger und positiver Forschungspraktiken durch Studierende 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.

## 2 Methode

Die Umfrage wurde auf dem "Open Science Framework" (OSF) präregistriert ([osf.io/k85r4/](https://osf.io/k85r4/)). Über diese Plattform sind auch unsere Primärdaten in anonymisierter Form zu finden ([osf.io/t3mak/](https://osf.io/t3mak/)). Die Online-Umfrage wurde mit der Open-Source Software formR (Arslan, Walther, & Tata, 2019) erstellt und die Daten in R mit Hilfe der Pakete lme4, lmerTest and der tidyverse-Pakete (Bates, Mächler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2017; R Core Team, 2019; Wickham, 2017) aufbereitet und analysiert. Im Vorfeld wurde eine Pilotbefragung mit 22 Psychologie-Studierenden verschiedener Fachsemester durchgeführt, um die Verständlichkeit und Durchführbarkeit der Befragung zu testen. Hierbei hatten die Befragten die Gelegenheit, mittels offener Kommentarfelder ausführliches Feedback abzugeben, das in der finalen Version der Umfrage berücksichtigt wurde.

**Akquise.** Die Bewerbung der Umfrage erfolgte durch Kontaktpersonen aus den Fachschaften der deutschsprachigen Universitäten. Vereinzelt wurden auch Dozierende angesprochen, die in ihren Veranstaltungen auf die Umfrage hinwiesen.

**Ausschlüsse.** Im präregistrierten Erhebungszeitraum vom 22.11.2018 bis 28.02.2019 klickten 3558 Personen auf den Link zu unserer Umfrage, hiervon bearbeiteten 1503 Personen (42,2 %) alle Fragen. Insgesamt wurden 96 Personen anhand vorab präregistrierter 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 für uns nicht mehr zu aktuell oder kürzlich Studierenden, und 69 Personen waren keine Psychologie-Studierenden oder -Absolvent\*innen. Abweichend von der Präregistrierung schlossen wir auch 9

Personen aus, die ein unplausibles Alter angegeben hatten (16 oder jünger, trotz fortgeschrittenen Studiums). Zudem berücksichtigten wir eine\*n Teilnehmer\*in von einer niederländischen Universität nicht in der Auswertung, da sich unsere Umfrage explizit auf den deutschsprachigen Raum richtete. Für die Analyse unserer Fragen zum Einsatz von QRPs wurden außerdem 251 Teilnehmer\*innen 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 1. oder 2. Bachelor-Semester. Wir bezeichnen diese Teilstichprobe im Folgenden als EE (Empirische Erfahrung)-Teilstichprobe. So blieben 1397 Personen in der vollen Stichprobe und 1146 Personen in der EE-Teilstichprobe.

(( Abbildung 1 ungefähr hier platzieren. ))

**Stichprobenbeschreibung (volle Stichprobe).** Unsere Umfrageteilnehmer\*innen (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 Masterstudierende 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 und Universitäten vertreten. Für 23 Hochschulen lagen uns 30 oder mehr Datenpunkte vor.

**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 “Hast du schon einmal ein empirisches<sup>2</sup> Forschungsprojekt in deinem gesamten bisherigen Studienverlauf (d.h. inkl. Bachelor und Master) durchgeführt oder begonnen (z.B. Expra/Empra, Projektarbeit, Bachelorarbeit, Masterarbeit)?” (Antwortmöglichkeiten: “Ja” / “Nein”). Bei Beantwortung mit “Nein” sprang die Umfrage zum nächsten Fragenblock (kritische Thematisierung von QRPs in

---

<sup>2</sup> Zusätzlich angezeigter Info-Text bei Mouse-Over: “D.h. es wurden Daten erhoben oder ausgewertet. Nicht empirisch wäre z.B. eine reine Literaturarbeit.”

137 der Lehre). Bei Beantwortung mit “Ja” folgten die Fragen zur Verwendung spezifischer  
138 Forschungspraktiken: “Welche der unten genannten Forschungspraktiken hast du selbst bzw.  
139 hat deine Seminargruppe schon mindestens einmal bei einem empirischen Projekt an deiner  
140 Universität unter Anleitung einer Lehrperson angewendet oder wendest sie aktuell an?”  
141 (Antwortmöglichkeiten: “Expra / Empra”, “Projektarbeit”<sup>3</sup>, “Bachelorarbeit”, “Masterarbeit”,  
142 “Anderes”<sup>4</sup>, “Keines”). Zu jeder Praktik gab es einen Info-Text mit einem erklärenden Beispiel.  
143 Die abgefragten Praktiken und die Info-Texte sind in Tabelle 1 aufgelistet. Zusätzlich wurde  
144 gefragt, welche Plattformen die Teilnehmenden mindestens einmal für Präregistrierungen  
145 verwendet hatten (Antwortmöglichkeiten: “Lehrperson nahm Präregistrierung entgegen”,  
146 “aspredicted.org”, “osf.io”, “bei einem Journal”, “Sonstiges”, “Ich habe noch nie präregistriert”).

147 Für neun QRPs (siehe Tabelle 1) wurde abgefragt, ob diese Praktik an der aktuellen Hochschule  
148 in der Lehre kritisch thematisiert wurde (Antwortmöglichkeiten: “Ja”, “Nein”, “Weiß nicht”).  
149 Außerdem fragten wir die Teilnehmenden, ob (Antwortmöglichkeiten: “Ja”, “Nein”, “Weiß nicht”)  
150 und in welchen Veranstaltungen die Replikationskrise an ihrer aktuellen Hochschule thematisiert  
151 wurde (Antwortmöglichkeiten: Auswahl 14 üblicher psychologischer Fächer und “Sonstige”,  
152 genaue Liste siehe Online-Materialien ( <https://osf.io/br2zf/>). Anschließend beantworteten die  
153 Teilnehmenden die drei Fragen “Wie sehr interessiert dich der Themenkomplex fragwürdige  
154 Forschungspraktiken, Open Science und die Replikationskrise?”, “Wie wichtig ist es deiner  
155 Meinung nach, dass der Themenkomplex fragwürdige Forschungspraktiken, Open Science und  
156 Replikationskrise im Studium gelehrt wird?” und “Wie gut fühlst du dich über den  
157 Themenkomplex fragwürdige Forschungspraktiken, Open Science und die Replikationskrise  
158 informiert?”, jeweils auf einer Skala von 1 (“Gar nicht”) bis 5 (“Sehr”). Es folgten noch wenige  
159 zusätzliche Fragen, die wir aus Gründen der Prägnanz hier nicht berichten. Sie können in der  
160 Präregistrierung und den zusätzlichen Online-Materialien nachvollzogen werden  
161 (<https://osf.io/br2zf/>). Die Umfrage schloss mit der Kontrollfrage “Hast du ernsthaft an dieser  
162 Umfrage teilgenommen?” (“Ja” / “Nein”) und einem optionalen freien Kommentarfeld.

---

<sup>3</sup> Zusätzlich angezeigte Begriffserklärung: “Bspw. Arbeit an einer Studie im Rahmen eines Seminars”

<sup>4</sup> Zusätzlich angezeigte Begriffserklärung: “Bspw. Forschungspraktikum oder Hiwi-Job”

(( Tabelle 1 ungefähr hier platzieren. ))

## 3 Ergebnisse

### 3.1 Deskriptive Ergebnisse

#### 3.1.1 Verbreitung fragwürdiger Forschungspraktiken

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

**Unterschiede zwischen Projekten.** Aus den Angaben der Teilnehmenden ergibt sich ein mittlerer Einsatz von 1,26 (SD = 1,39) QRPs pro Projekt. 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 Expra verwendet zu werden, während Masterarbeiten mit durchschnittlich 0,52 (SD = 0,84) verwendeten QRPs den niedrigsten Wert aufweisen. Bachelor- und Projektarbeiten sowie andere Projekte (durchschnittlich jeweils 1,02 (SD = 1,33), 1,30 (SD = 1,51) und 0,85 (SD = 1,37) QRPs pro Projekt) liegen dazwischen. Weitere Details (Median, Minimum und Maximum) sind in den Online-Materialien (<https://osf.io/br2zf/>) zu finden. Betrachtet man die Projekte, die sich eindeutig nach Studienfortschritt ordnen lassen, also Expra, Bachelorarbeit und Masterarbeit, so zeigt sich ein deutlicher Trend: Mit fortschreitendem Studium wurden immer weniger fragwürdige Forschungspraktiken 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 dargestellt. Die am häufigsten genannten Praktiken waren “fehlende Stichprobenplanung” und “selektives Berichten abhängiger Variablen” mit einem Anteil von 34,3 % , 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 fragwürdigen wurden als positive Forschungspraktiken auch abgefragt, ob in den jeweiligen Projekten eine Stichprobenumfangsplanung (basierend auf einer a-priori Poweranalyse) und eine

Präregistrierung durchgeführt wurden. Laut Angaben der Teilnehmenden wurde für 33,7 % der erfassten Projekte eine Power-Analyse zur Stichprobenplanung durchgeführt, und 22,4 % 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, Loewenstein und Prelec (2012) der Anteil der Teilnehmenden zu sehen, die die jeweilige Forschungspraktik mindestens einmal in einem empirischen Projekt eingesetzt hatten.

Präregistrierungen fanden mit 282 Nennungen (24,6 % der Teilnehmenden) am häufigsten bei einer Lehrperson statt<sup>5</sup>. Es folgten in absteigender Reihenfolge als Plattformen für Präregistrierungen das Open Science Framework (OSF, 101 mal gewählt, 8,8 % der Teilnehmenden), aspredicted.org (87 mal gewählt, 7,6 % der Teilnehmenden), „andere“ (65 mal gewählt, 5,7 % der Teilnehmenden) und ein Journal (11 mal gewählt, 1,0 % der Teilnehmenden). Von den Teilnehmenden gaben Personen ( % der Teilnehmenden) an, nie eines ihrer empirischen Projekte präregistriert zu haben. Bei dieser Frage war die Auswahl mehrerer Antworten erlaubt.

### 3.1.2 Zusätzliche Ergebnisse

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 75,4 % der Teilnehmenden an, dass die Replikationskrise an ihrer jeweils aktuellen Universität in Lehrveranstaltungen thematisiert wurde, 20,6 % gaben an, dass sie nicht thematisiert wurde („Weiß nicht“: 4,1 %). Unter den Studienanfänger\*innen (Studierende im ersten oder zweiten Bachelor-Semester) 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. 77,6 % positiven Antworten.

---

<sup>5</sup> Möglicherweise bezogen sich die Antworten unserer Teilnehmer\*innen hier auch auf die üblichen Exposés, was die hohen Zahlen erklären würde.



**Betrachtung nach Lehrveranstaltungen.** Tabelle 3 enthält eine Übersicht der Ergebnisse. Die Behandlung der Replikationskrise fand größtenteils in der Methodenlehre (inkl. Extra) statt: 54,5 % aller Teilnehmenden gaben an, dass hier die Replikationskrise thematisiert wurde. In der Lehre der Sozialpsychologie, der am stärksten im Fokus der Kritik stehenden Teildisziplin (siehe z.B. Earp & Trafimow, 2015), wurde am zweithäufigsten über die Replikationskrise berichtet (27,4 %). In den 13 übrigen Fachbereichen (inkl. "Andere") bewegt sich die Abdeckung zwischen 19,0 % (Allgemeine Psychologie) und 0,8 % (Forensische Psychologie). Dabei sollte beachtet werden, dass hochspezialisierte Module wie Forensische Psychologie wohl nur von einem Bruchteil der Befragten belegt wurden. Für solche spezialisierten Module sind die obigen Zahlen daher mit der Häufigkeit der Belegung konfundiert und sollten nur sehr zurückhaltend interpretiert werden. Für allgemeinere Module wie z.B. Sozialpsychologie gilt das nicht in gleichem Maße, da sie in der Regel Teil jedes Psychologie-Studiums sind.

**Behandlung fragwürdiger Praktiken.** Die Ergebnisse (s. Tabelle 4) legen nahe, dass Studierende im Studium durchaus eine kritische Thematisierung einzelner fragwürdiger Forschungspraktiken erleben. So gaben 74,0 % der Teilnehmenden an, eine Problematisierung von HARKing (Hypothesizing After the Results are Known, das Aufstellen von Hypothesen nach Kenntnis der Ergebnisse) erfahren zu haben. HARKing stellte sich als die am häufigsten in der Lehre thematisierte fragwürdige Forschungspraktik heraus. Auf den folgenden Plätzen finden sich fehlende Stichprobenplanung (65,3 %) und selektives Berichten von Variablen (64,1 %). Am seltensten wurden Flexibilität in der Datenanalyse (50,7 %) und das Abrunden von p-Werten (39,9 %) kritisch thematisiert.

**Einstellungen von Studierenden.** Abbildung 4 zeigt die Einschätzung der Wichtigkeit des Themenkomplexes "Open Science und Replikationskrise" durch die Teilnehmenden, ihr Interesse daran und ihre Einschätzung dazu, wie gut sie über das Thema informiert sind. Deutlich wird, dass das Thema "Replikationskrise und Open Science" von der überwältigenden Mehrheit der Teilnehmenden für sehr wichtig oder wichtig gehalten wird, dass eine Mehrheit der Teilnehmenden Interesse an dem Thema hat, und dass sich ein substanzieller Teil der Teilnehmenden als nicht gut informiert einschätzte.

## 3.2 Explorative Analysen

Zur näheren Untersuchung der Daten führten wir eine Reihe von explorativen Analysen durch, von denen wir die wichtigsten hier berichten. Alle weiteren oder nur kurz erwähnten Analysen

sind in den Online-Materialien (<https://osf.io/br2zf/>) abrufbar. Die Analysen wurden in R mit dem Paket “lme4” ausgeführt (Bates et al., 2015; R Core Team, 2019). Für alle Analysen wurden Bonferroni-korrigierte Alpha-Niveaus zugrunde gelegt.

### 3.2.1 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 fragwürdiger Forschungspraktiken zu modellieren. 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 ( $SD = 1,25$ ,  $\chi^2(1) = 1058,34$ ,  $p < .001$ ), für die Universität, an der das Projekt durchgeführt wurde ( $SD = 0,52$ ,  $\chi^2(1) = 42,67$ ,  $p < .001$ ), und für die spezifische Forschungspraktik ( $SD = 1,03$ ,  $\chi^2(1) = 1538,87$ ,  $p < .001$ ). Die numerischen Prädiktoren wurden z-standardisiert. Die Schätzungen der fixen Koeffizienten und die jeweiligen Odds Ratios sind in Tabelle 5 dargestellt. Zugrunde lagen die Daten aus der EE-Teilstichprobe.

Einzelne Beobachtungen wurden wegen inkonsistenter Antworten oder unklarer Zuordnung des jeweiligen Projektes zu einer Universität ausgeschlossen. Ersteres geschah in Fällen, in denen Teilnehmende bei der Frage nach der Anwendung einer spezifischen Forschungspraktik sowohl die Antwortmöglichkeit “Keines” (Praktik wurde in keinem Projekt verwendet), als auch mindestens ein oder mehrere Projekte auswählten (bei QRPs  $n = 91$  Beobachtungen). Zweiteres geschah bei den Projekten “Projektarbeit” und “Anderes”, wenn die Teilnehmenden ihre Universität zwischen Bachelor und Master gewechselt hatten, oder bei Alumni/Alumnae, da deren Bachelor-Universität nicht erfasst wurde ( $n = 495$  Projekte). Zwischen beiden Kriterien gab es Überschneidungen. Insgesamt wurden so 4537 Datenpunkte zu 495 Projekten von 83 Teilnehmenden in der Analyse nicht berücksichtigt.

Das Effektstärkemaß Odds Ratio gibt die multiplikative Änderung im Verhältnis der Odds (dt.: Chancen) von “Forschungspraktik angewandt” und “Forschungspraktik nicht angewandt” bei Änderung des Prädiktors um eine Einheit, bzw. bei kategorialen Prädiktoren im Vergleich zur Referenzgruppe an. Ein Odds Ratio größer als 1 bedeutet, dass die Anwendung der Forschungspraktik wahrscheinlicher wird, ein Odds Ratio unter 1, dass sie weniger wahrscheinlich wird. Zunächst lässt sich beobachten, dass QRPs umso weniger eingesetzt wurden, je eher Teilnehmende den Themenkomplex “Replikationskrise und Open Science” als

wichtig ansahen und je eher sie sich als gut informiert einschätzten (Odds Ratios von 0,80 und 0,72).

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 fragwürdige Forschungspraktik 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 5).

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 mit geringerer Wahrscheinlichkeit QRPs angewandt werden.

### 3.2.2 Einsatz positiver Forschungspraktiken

Die Ergebnisse bezüglich positiver Forschungspraktiken fassen wir hier nur kurz zusammen. Die vollständigen Tabellen zu allen Modellen können in den Online-Materialien (<https://osf.io/br2zf/>) abgerufen werden. Wir untersuchten hier ebenfalls in generalisierten gemischten linearen Modellen die gleichen Prädiktoren und zufälligen Achsenabschnitte wie bei der Exploration zu QRPs (Fußnote: Mit Ausnahme von "Praktik gelehrt", da diese Frage nur für QRPs gestellt wurde). Signifikante Unterschiede zeigten sich zwischen Master- und Bachelorstudierenden ( $B = -0,38$ ,  $SE = 0,13$ ,  $OR = 0,68$ ,  $95\% \text{ KI} = [0,53; 0,87]$ ,  $p = .003$ ). Master-Studierende wiesen demnach eine geringere Wahrscheinlichkeit für den Einsatz positiver Forschungspraktiken auf als Bachelor-Studierende. 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 Einfluss der subjektiv eingeschätzten Informiertheit der Teilnehmenden, je informierter sie sich sahen, desto wahrscheinlicher war der Einsatz positiver Forschungspraktiken ( $B = 0,36$ ,  $SE = 0,06$ ,  $OR = 1,43$ , 95 %  $KI = [1,29; 1,60]$ ,  $p < .001$ ). In Projektarbeiten ( $B = -0,71$ ,  $SE = 0,13$ ,  $OR = 0,49$ , 95 %  $KI = [0,38; 0,63]$ ,  $p < .001$ ) und “Anderen” Projekten ( $B = -0,87$ ,  $SE = 0,16$ ,  $OR = 0,42$ , 95 %  $KI = [0,31; 0,57]$ ,  $p < .001$ ) wurden jeweils mit signifikant geringerer Wahrscheinlichkeit positive Praktiken eingesetzt als im Extra (Details siehe Online-Materialien, <https://osf.io/br2zf/>).

Ein linearer Trend über den Studienverlauf wie bei QRPs zeigte sich in einer weiteren Analyse nicht, wohl aber ein quadratischer Trend ( $B = 0,34$ ,  $SE = 0,10$ ,  $OR = 1,40$ , 95 %  $KI = [1,16; 1,70]$ ,  $p .001^{**}$ ). Die Unterschiede zwischen Bachelorarbeiten und Expras ( $B = -0,23$ ,  $SE = 0,10$ ,  $OR = 0,79$ , 95 %  $KI = [0,65; 0,96]$ ,  $p = .020$ ) sowie zwischen Masterarbeiten und Expras ( $B = 0,21$ ,  $SE = 0,18$ ,  $OR = 1,24$ , 95 %  $KI = [0,87; 1,76]$ ,  $p = .233$ ) waren in der vorgehenden Analyse statistisch nicht signifikant, die Odds Ratios deuten aber tendenziell darauf hin, dass verglichen mit Expras in Bachelorarbeiten mit geringerer und in Masterarbeiten mit höherer Wahrscheinlichkeit positive Forschungspraktiken angewendet werden. Vor diesem Hintergrund erscheint der quadratische Trend für den Studienverlauf plausibel.

### 3.2.3 Eindruck von Informiertheit

Welche Faktoren hängen damit zusammen, ob Studierende sich über den Themenkomplex gut informiert fühlen?

In einer weiteren explorativen Analyse untersuchten wir den Zusammenhang zwischen den verfügbaren, potentiell informativen 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 ( $SD = 0,38$ ,  $\chi^2(1) = 135,12$ ,  $p < .001$ ). Die Schätzungen der fixen Koeffizienten sind in Tabelle 6 dargestellt.

Die Analyse deutet auf signifikante 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. Ob es de facto eine im Netzwerk der Open Science Initiativen gelistete Open Science Initiative an der jeweiligen Universität gibt, scheint nicht mit der subjektiven Informiertheit der Teilnehmenden zusammenzuhängen, wohl aber das Wissen der Teilnehmenden um eine Open Science Initiative an ihrer Universität: Sowohl Teilnehmende, die sich über diese Frage unsicher sind, als auch solche, die eine Initiative an ihrer Universität kennen (etwas größerer Zusammenhang als bei unsicheren Teilnehmenden), schätzen sich als besser informiert ein, als Teilnehmende, die keine Initiative an ihrer Universität kennen. Schließlich zeigt 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 Studierende, denen dieses Thema nicht in der Lehre begegnet ist.

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.

## 4 Diskussion

In diesem Beitrag stellen wir eine umfassende und systematische Befragung vor, in der die Prävalenz von fragwürdigen und positiven Forschungspraktiken unter Studierenden sowie die Thematisierung dieser Aspekte in der psychologischen Lehre untersucht wurde. Unseren Ergebnissen zufolge sind fragwürdige Forschungspraktiken auch im Studium verbreitet, wobei sich stark unterschiedliche Prävalenzen zwischen verschiedenen QRPs und verschiedenen Projekten zeigten. 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 (75,4 %) 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. Aus diesen Ergebnissen und der Diskussion möglicher Ursachen leiten wir im Folgenden Implikationen für die psychologische Lehre ab.

#### 4.1 Prävalenz von QRPs in studentischen Projekten

Zur Prävalenz von QRPs unter wissenschaftlich tätigen Psycholog\*innen liegt bereits einige Literatur vor (vgl. Agnoli, Wicherts, Veldkamp, Albiero, & Cubelli, 2017; Banks, Rogelberg, Woznyj, Landis, & Rupp, 2016; Fiedler & Schwarz, 2016; John et al., 2012). Unsere Ergebnisse sind mit dieser Literatur nur bedingt vergleichbar, da Studierende ihre empirischen Projekte unter stark abweichenden Bedingungen anfertigen: Der Lerneffekt ist tendenziell wichtiger als der Erkenntnisgewinn durch die durchgeführten Studien, die Forschung findet unter enger Betreuung statt und es gibt in der Regel keinen unmittelbaren Publikationsdruck. Unter Berücksichtigung dieser Unterschiede zeigt sich im Vergleich mit Stichproben psychologischer Forscher\*innen, dass der Anteil der Personen, die spezifische QRPs mindestens einmal verwendet haben (s. Abb. 3a, rechts), unter den Teilnehmenden unserer Befragung überwiegend geringer ist als unter Forschenden. ~~So zeigen sich bei einer Betrachtung der direkt vergleichbaren QRP "Selektives Berichten von Variablen" Einsatzquoten von 36,5 % in unserer Stichprobe, während diese bei John et al. (2012) zwischen 63,4 % und 66,5 % und bei Fiedler & Schwarz (2016) ca. 34 % betrug. Eine naheliegende Erklärung dafür ist, dass Studierende insgesamt auf eine deutlich geringere Anzahl bearbeiteter Projekte zurückblicken und sie bei ihrer Durchführung engmaschiger angeleitet und betreut werden, als dies bei Forscher\*innen der Fall ist. Die Befragungen von Wissenschaftler\*innen weisen allerdings ohnehin teils recht Unterschiede in der Höhe der Einsatzquoten auf, wie die genannten Zahlen zeigen.~~

Im Vergleich unserer Ergebnisse mit den Befunden von Krishna & Peter (2018) für studentische Abschlussarbeiten zeigen sich für einzelne Forschungspraktiken vergleichbare Prävalenzen, so etwa hinsichtlich der QRPs "Flexibler Ausschluss von Ausreißern" (12,6 % vs 15,5 % bei Krishna & Peter (2018)), "Selektives Berichten von Bedingungen" (9,8 % vs. 7,7 %) wie auch der positiven Forschungspraktik der Poweranalyse (33,7 % 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 untersuchten Projekttypen, den Hochschulen der Befragten als auch auf Unterschiede in der exakten Formulierung der Items (vgl. Fiedler & Schwarz, 2016) zurückzuführen sein.

Ähnlich wie in Studien, die auf Forschenden-Stichproben basieren, zeigt sich, dass bestimmte QRPs scheinbar häufiger angewandt werden als andere. Beispielsweise waren “Selektives Berichten von Variablen” und “Ausschluss von Ausreißern” 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 (49,3 % mind. einmaliger Einsatz) heraus. Damit zusammenhängend berichten 48,2 % der Befragten, im Laufe des Studiums für keines ihrer empirischen Projekte eine Poweranalyse durchgeführt zu haben. Ein noch höherer Anteil der Befragten (62,2 %<sup>6</sup>) hat noch nie eine Präregistrierung durchgeführt. Ein stärkerer Einsatz u.a. dieser Forschungspraktiken ist bereits vielfach empfohlen worden, um QRPs zu reduzieren (Banks et al., 2016; Nosek & Lakens, 2014).

## 4.2 Unterschiede zwischen Projekttypen im Studienverlauf

Zwischen den QRP-Prävalenzen und den einzelnen 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. Umso wichtiger scheint auch bei der Behandlung der QRPs im Studium die Rolle der empirischen Projekte und der Betreuenden dieser Projekte. Durch ihre Praxis- und Anwendungsnahe sind diese zur Vermittlung methodischer Kenntnisse eine wichtige Ergänzung zu rein theoretischen Veranstaltungen. Die Expras dienen dabei als Raum, um theoretisch

---

<sup>6</sup> 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 %. Die leicht unterschiedlichen Zahlen sind erklärbar durch minimale Inkonsistenzen im Antwortverhalten der Teilnehmenden, da es sich um zwei unterschiedliche Fragen handelte.

erlernte Methoden erstmals auszuprobieren und dabei aus entstandenen Fehlern sowie deren Korrektur durch den Betreuer 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. Zusammenfassend lässt sich feststellen, dass Studierenden QRPs in vielen studentischen Projekten angewandt werden. Die beobachtete Verbreitung von QRPs ist bedenklich und sollte Anlass zu Überlegungen geben, wie ihr Einsatz reduziert werden kann.

### 4.3 Mögliche Ursachen für QRPs in studentischen Projekten

Die möglichen Ursachen für die Verwendung von QRPs unter Wissenschaftlern und Wissenschaftlerinnen (Wissenschaftler\*innen) lassen sich in mehrere Bereiche einteilen: Unzureichende methodische Kenntnisse der Forschenden, die Anreizstruktur des akademischen Betriebs, insbesondere der Publikationsdruck, sowie kognitive Verzerrungen (Edwards & Roy, 2017; Krishna & Peter, 2018; Sijtsma, 2016). Die naheliegenden Ursachen bei den Studierenden unterscheiden sich davon: Studierende sind im Gegensatz zu den Forscher\*innen keinem direkten Publikationsdruck ausgesetzt, sondern befinden sich stattdessen in einem Anreizsystem, das den zügigen Erwerb von ECTS-Punkten mit möglichst guten Bewertungen durch die Dozierenden fördert. 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 indirekt, welche QRPs Studierende während der Datenanalyse und der Berichterstellung verwenden.

Bei der Einordnung der Ergebnisse müssen weitere Besonderheiten empirischer Projekte im studentischen Rahmen beachtet werden. Der beschränkte Zeithorizont (i.d.R. bis zu sechs Monate für eine Abschlussarbeit) könnte erklären, wieso bestimmte QRPs unter Studierenden im Vergleich zu Forschenden weniger häufig (z.B. Nacherheben von Versuchspersonen) und



andere häufiger sind (z.B. fehlende Stichprobenplanung). Parallel zu den Einflüssen in der Wissenschaft ist auch davon auszugehen, 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 wurden weniger QRPs eingesetzt als in Extras (s. Abbildungen 2 und 3b). Dieser lineare Trend konnte auch durch die explorativen Analysen untermauert werden. Eine naheliegende Erklärung für die Abnahme der QRP-Prävalenz im Studienverlauf ist, dass die Befragten im Laufe ihres Studiums mehr methodische Kompetenzen erwerben und für die Probleme, die mit dem Einsatz fragwürdiger Forschungspraktiken verbunden sind, stärker sensibilisiert werden. Für die Verwendung der zwei abgefragten positiver Forschungspraktiken zeigte sich dieser zeitliche Verlauf nicht: Aktuelle Bachelorstudierende wandten mit einer höheren Wahrscheinlichkeit positive Forschungspraktiken an, 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 nicht möglicherweise nicht in gleichem Umfang gegeben war. Sowohl die Verwendung fragwürdiger als 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 fragwürdiger Forschungspraktiken assoziiert. **COMMENT: Hier noch Interpretation dieses Befundes (JB).** 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 einer nachträglich aufgestellten Hypothese als von vornherein erwartet).

#### 4.4 Fragen zur universitären Lehre

**Thematisierung der Replikationskrise.** Die aktuelle wissenschaftliche Debatte um die fehlende Replizierbarkeit psychologischer Befunde scheint erfreulicherweise schon überwiegend Einzug in die Lehre gefunden zu haben. Ein substanzieller Anteil der Teilnehmenden (75,4 %) ist im Laufe des Psychologiestudiums bereits mit dem Themenkomplex Replikationskrise, QRPs

und Open Science in Berührung gekommen. Verwunderlich ist aber, dass 45,5 % der Befragten in methodischen Veranstaltungen noch nichts von der Replikationskrise erfahren haben. Dies wäre in der Methodenlehre eigentlich als selbstverständlich zu erachten. Abseits der Methodenlehre kamen lediglich 27,4 % der Studierenden in Veranstaltungen der Sozialpsychologie mit dem Thema der Replikationskrise in Kontakt. Das ist insofern hervorzuheben, als dass die Sozialpsychologie im Zuge der Replikationsdebatte besonders in die Kritik geraten ist (siehe z.B. Earp & Trafimow, 2015). Andererseits wird die Thematik in den weiteren Anwendungsfächern der Psychologie - von Allgemeiner Psychologie, über klinische Psychologie bis hin zu forensischer Psychologie - noch seltener thematisiert (Anteile 19,0 %, und 0,8 %), obwohl Probleme der Replizierbarkeit auch hier bereits in der Literatur diskutiert wurden (Button et al., 2013; Tackett et al., 2017). Insgesamt kann daher nicht davon gesprochen werden, dass die Replikationskrise abseits der Methodenveranstaltungen systematisch Eingang in die Lehre gefunden hat. 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 der wahrgenommenen Wichtigkeit dieses Themas. **COMMENT: Hier wären noch 1-2 Sätze zur Interpretation der Ergebnisse wichtig. Sonst liest sich der Abschnitt fast genau wie der Ergebnisteil (JB).** Positiv ist hier hervorzuheben, dass die Behandlung der Replikationskrise in der Lehre einen messbaren positiven Zusammenhang mit der Informiertheit der Studierenden aufweist.

**Behandlung fragwürdiger Forschungspraktiken.** Die kritische Betrachtung konkreter QRPs scheint durchaus Eingang in die Lehre gefunden zu haben. Jedoch variieren die Zahlen je nach betrachteter Praktik: Die Spanne reicht von % (Abrunden von p-Werten) bis 74,0 % (HARKing) der Befragten, in deren Lehrveranstaltungen die jeweiligen QRPs kritisch behandelt worden sind. HARKing, die am häufigsten thematisierte fragwürdige Forschungspraktik, gehörte zu den in empirischen Projekten am seltensten angewandten Praktiken. Das häufigere Auftreten von anderen QRPs wie "Keine Stichprobenumfangsplanung" oder "Selektives Berichten von Variablen/Bedingungen/Hypothesen" bei gleichzeitig häufiger Betrachtung in Lehrveranstaltungen könnte dagegen 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 der einzelnen Praktik in der Lehre scheint für deren Einsatz maßgeblich zu sein, sondern die generelle subjektiv eingeschätzte Informiertheit der Studierenden über den Themenkomplex "Replikationskrise und Open Science". Zentral scheint daher nicht primär zu sein, ob spezifische

Forschungspraktiken in der Lehre vorkommen, sondern dass Studierende sich über den Themenkomplex der Replikationskrise sowie QRPs informiert fühlen und für die Problematik ihrer Verwendung sensibilisiert werden.

## 4.5 Limitationen

Obwohl wir für die vorgestellte Umfrage eine umfangreiche Stichprobe mit Studierenden aus verschiedenen Semestern und Universitäten befragt haben, existieren hinsichtlich des Studiendesigns sowohl konzeptionelle als auch inhaltliche Einschränkungen, die wir im Folgenden ausführen wollen.

Da die Teilnahme an der Umfrage freiwillig war, unterliegt die Stichprobe möglicherweise Verfälschungen durch Selbstselektion. 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 könnten, im Rahmen dieser Umfrage nicht eindeutig beantwortet werden.

Da unsere Befragung komplexe Anforderungen an die Erinnerung unserer Teilnehmenden stellte, könnten auch fehlerhafte oder verzerrte Erinnerungen die Qualität unserer Daten beeinträchtigen. So ist vorstellbar, dass selbstwertdienliche Erinnerungs- sowie Antwortmuster von Studierenden dazu führen könnten, dass der QRP-Einsatz in der Befragung tendenziell unterschätzt wird.

Eine geläufige Kritik an Studien zur Prävalenz von fragwürdigen Forschungsmethoden bezieht sich auf den Einfluss der sprachlichen Formulierung der jeweiligen QRPs. 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 Forschungspraktik angezeigt werden konnten, entgegenzuwirken (s. Tabelle 1).

#### 4.6 Implikationen für die psychologische Lehre

Wie bereits 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. Im Mittelpunkt der Diskussion stand die hohe Prävalenz fragwürdiger Forschungspraktiken. 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 Thematisierung des Themenkomplexes in der universitären Lehre unerlässlich. Dies spiegelt sich auch in der Meinung der Studierenden wider, da der Themenkomplex “Replikationskrise, QRPs und Open Science” von nahezu allen Befragten als wichtig oder sehr wichtig wahrgenommen wurde. Auf der Grundlage unserer Ergebnisse möchten wir folgende Handlungsempfehlungen für die inhaltliche Ausgestaltung der psychologischen Lehre vorschlagen:

1. Die Replikationskrise und ihre Ursachen sollten in den methodischen Veranstaltungen im Psychologiestudium flächendeckend als fester Bestandteil thematisiert werden. Im Zentrum

stehen sollte dabei, dass sich Studierende ausreichend über den Themenkomplex der Replikationskrise, QRPs und Open Science informiert fühlen. Dies sehen wir als Grundlage dafür, dass positive Forschungspraktiken effektiv erlernt und eingesetzt werden können.

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 Forschungsentwicklungen angeregt werden.

3. Empirische Projekte im Rahmen des Studiums sollten genutzt werden, um positive Forschungsmethoden zu vermitteln und die Problematik fragwürdiger Forschungspraktiken in der Praxis zu thematisieren. Ziel soll 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 solches wertgeschätzt und von Studierenden und Forscher\*innen in empirischen Projekten transparent kenntlich gemacht werden. Darin sehen wir nicht nur eine methodische Notwendigkeit, sondern auch eine Chance, die Psychologie als Wissenschaft weiterzuentwickeln, indem der Bildung fundierter Theorien auf der Grundlage umfassender Exploration in Verbindung mit sorgfältigen konfirmatorischen Tests mehr Raum gegeben wird (Banks et al., 2016; Glöckner, Fiedler, & Renkewitz, 2018).

5. Studierende haben für die methodisch korrekte Arbeitsweise bei der Anfertigung von Abschlussarbeiten und ähnlichen Projekten eine hohe Eigenverantwortung. Ein ideales Curriculum allein kann demnach 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 (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.

Wie diese Implikationen in die Realität umgesetzt werden können, wird anhand mehrerer, bereits umgesetzter Lehrkonzepte beispielhaft deutlich. So wird die Thematik der Replikationskrise an der Universität Amsterdam seit 2015 im Rahmen eines dezidierten “Good Research Practices”-Kurses an Studierende vermittelt (Sarafoglou, Hoogeveen, Matzke, & Wagenmakers, 2019). Auch von Chopik, Bremner, Defever, & Keller (2018) konnte gezeigt werden, dass Studierende durch solche Kurse einen informationellen Zugewinn und eine höhere Sensibilisierung für aktuelle Probleme der Disziplin erfahren. In Bezug auf empirische Projekte wurde an der Ludwig-Maximilians-Universität München bereits ein methodisches Kern-Curriculum für Expras (Schönbrodt, Maier, Heene, & Bühner, 2018, S. 42) festgelegt, das mittels der Elemente 1) expliziter Stichprobenplanung, 2) Präregistrierung, 3) offenen Daten sowie 4) reproduzierbarer Analyseskripte Studierende an eine von Forschungstransparenz geleitete Arbeitsweise heranführt. Auch an der School of Psychology in Glasgow vollzog sich eine grundlegende, paradigmatische Neuausrichtung der methodischen Schwerpunkte in der Lehre. Anstelle des Erlernens möglichst vieler statistischer Verfahren wurde der Fokus auf die Kernkompetenzen gelegt, Daten eigenständig kritisch analysieren und auf reproduzierbare Weise berichten zu können. Dies schlug sich u.A. in Bewertungskriterien des präregistrierten, empirischen Kursprojektes nieder. In diesem Zuge erhielt auch Arbeiten mit offener und reproduzierbarer Software, z.B. RMarkdown und Git, einen zentralen Stellenwert (Barr et al., 2019).

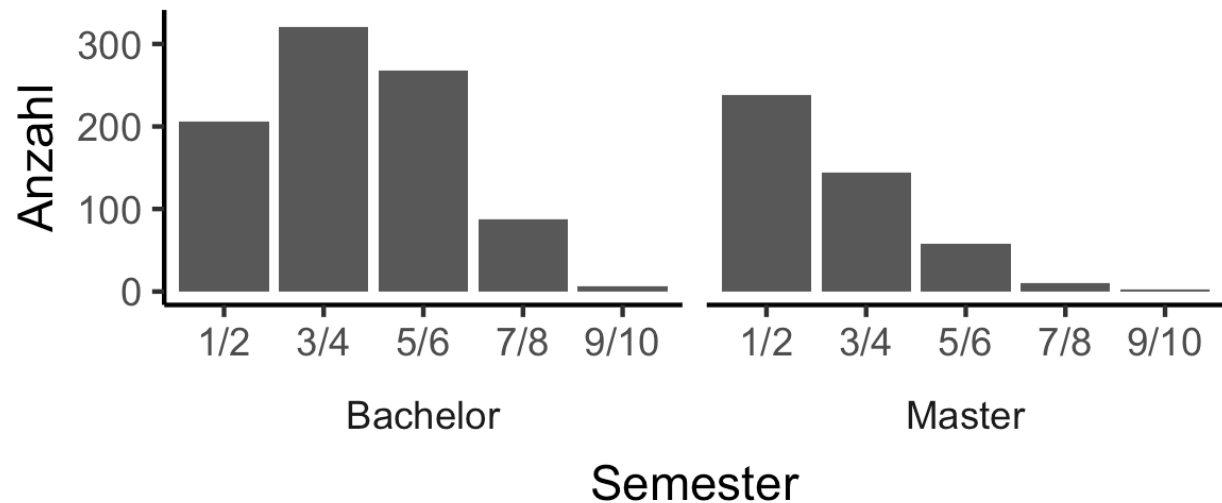
Unsere hier vorgestellte Umfrage gibt einen ersten Überblick über die Prävalenz von QRPs in verschiedenen studentischen empirischen Projekten sowie über die Behandlung der Replikationskrise und ihrer Ursachen in Form von QRPs in der Lehre. Es wurden mögliche Ursachen für die Verwendung von QRPs diskutiert, die als Anhaltspunkt zukünftiger Forschung dienen können. **COMMENT: Steht hier auch etwas unvermittelt/deplatziert. Idee für eine Überleitung, an den Anfang der Diskussion oder ganz raus? (TK)**

Heutige Studierende werden als zukünftige WissenschaftlerInnen die Entwicklung der akademischen Psychologie maßgeblich prägen. Deswegen ist es bereits im Studium von besonderer Bedeutung, 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 damit eine weitere

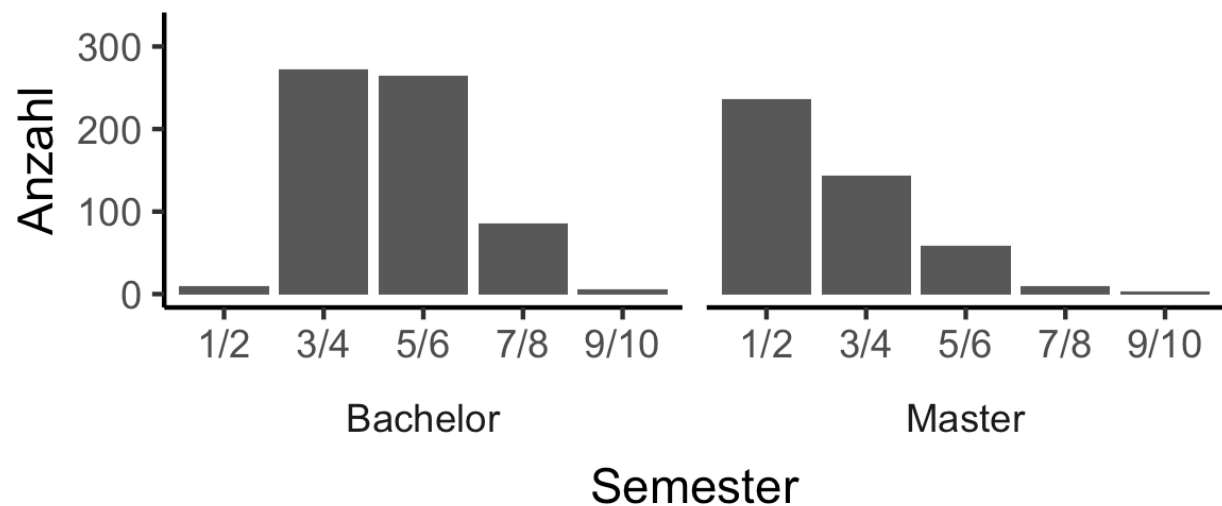
643 Auseinandersetzung mit dem aktuellen Stand der Lehre in der Psychologie und deren  
644 Weiterentwicklung anregen können.

645 5 Abbildungen

**a** Volle Stichprobe

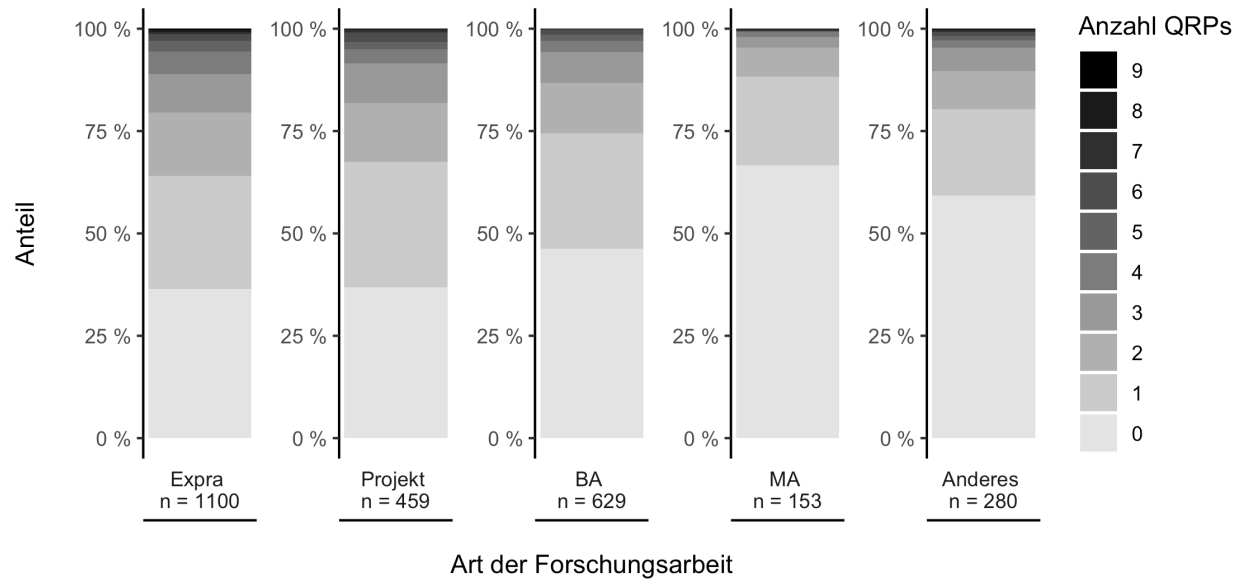


**b** EF-Teilstichprobe



646  
647 *Figure 1: Aufschlüsselung der Teilnehmenden nach Studienfortschritt. (a) Volle Stichprobe. (b)*  
648 *EE-Teilstichprobe.*

## Replikationskrise & Open Science: Die Studierenden

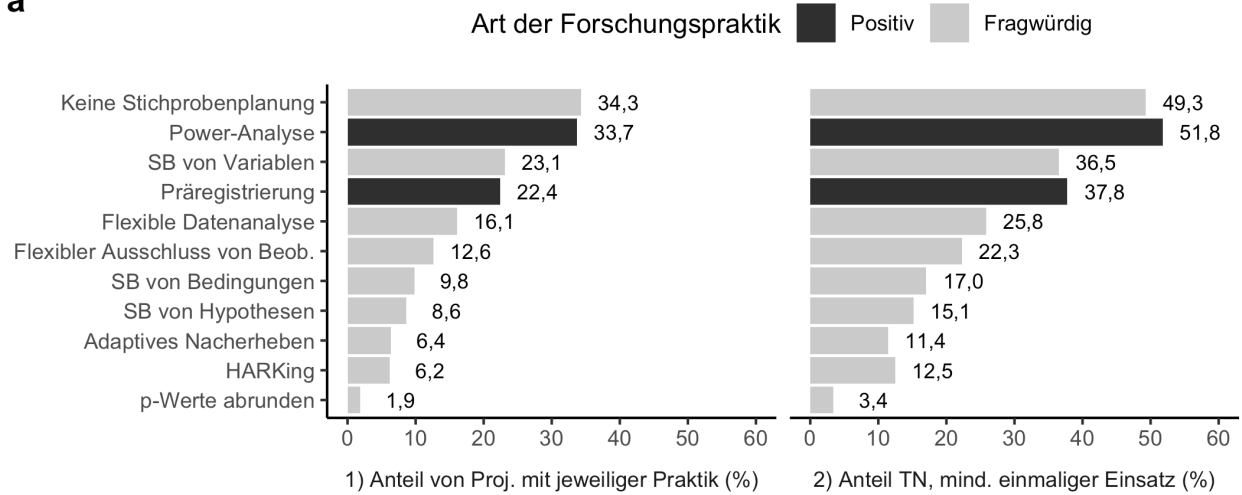


649

650 *Figure 2: Anzahl eingesetzter fragwürdiger Forschungspraktiken aufgeschlüsselt nach Art des*  
651 *Projekts. BA = Bachelorarbeit, MA = Masterarbeit*



a



b

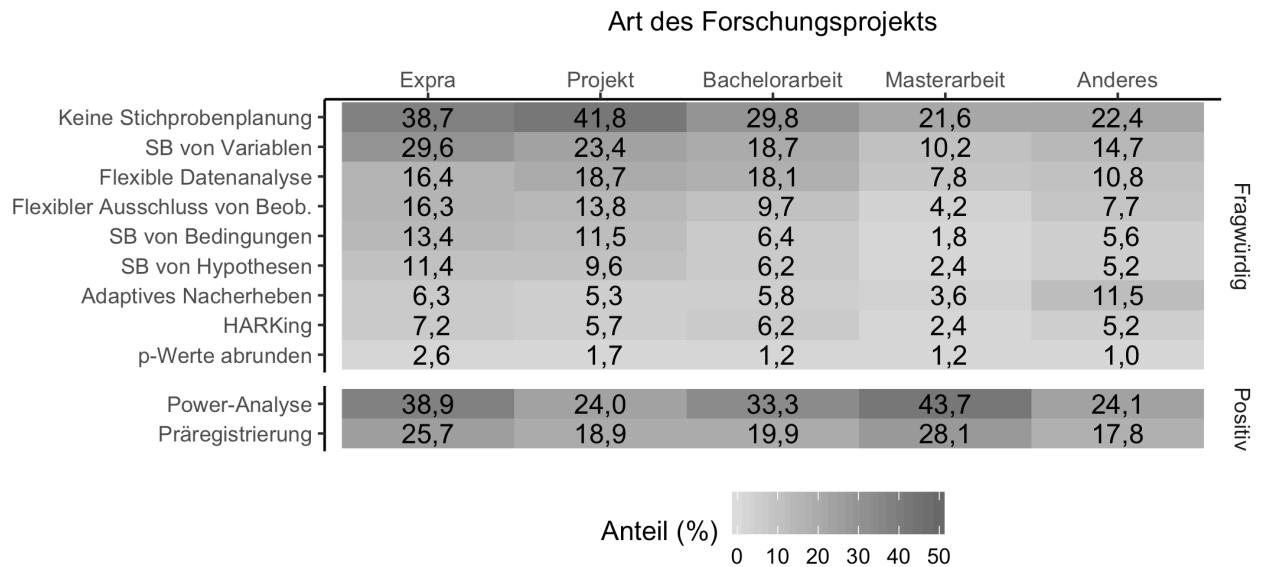


Figure 3: Anteil von Projekten, in denen die jeweilige Forschungspraktik eingesetzt wurde. S.B. = Selektives Berichten.

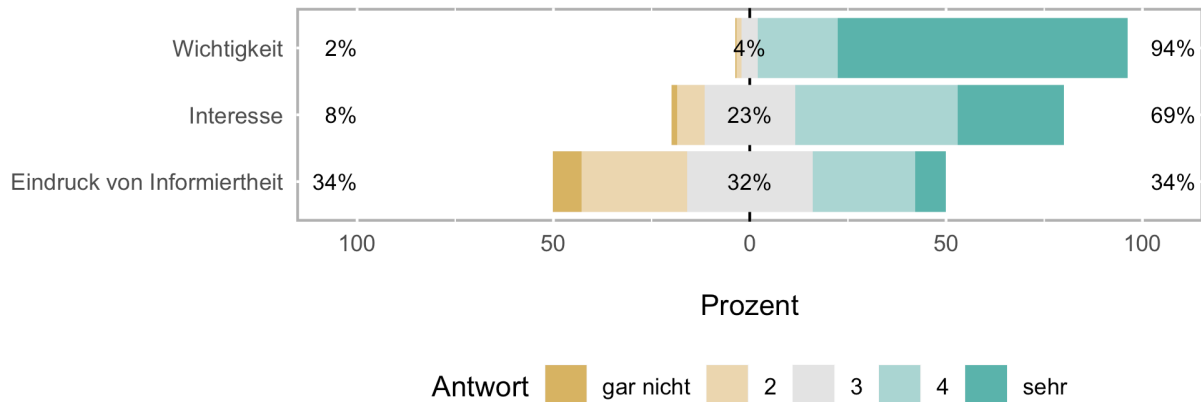


Figure 4: Einstellungen der Teilnehmenden zum Themenkomplex "Open Science und Replikationskrise" auf einer 5-Punkte Likert-Skala von "gar nicht" bis "sehr". Bedeutung der Prozentzahlen: "gar nicht" & "2" (links), "3" (mittig), "4" & "sehr" (rechts).

3

## 6 Tabellen

Tabelle 1. Abgefragte Forschungspraktiken und Begleittexte.

Bezeichnung (kurz)	Begleittext
<b>Fragwürdige Forschungspraktiken</b>	
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)

Bezeichnung (kurz)	Begleittext
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 ( $\pm 3$ 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 zusammenhängen.
p-Werte abrunden	Beispiel: Der Test für die zentrale Hypothese ergibt einen p-Wert von $p = .058$ , im Bericht wird $p = .05$ geschrieben.
<b>Positive Forschungspraktiken</b>	
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.

Bezeichnung (kurz)	Begleittext
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.

662

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. Projekte <sup>a</sup>	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)

<sup>a</sup>Anzahl der von der/dem jeweiligen Teilnehmenden begonnenen oder durchgeführten Projekte.

663

Table 3. Anteil von Teilnehmenden, die in der jeweiligen Lehrveranstaltung an ihrer aktuellen  
 (Alumni: letzten) Universität eine Thematisierung der Replikationskrise erfahren haben (N =  
 1397).

664

665

Lehrveranstaltung	n	Anteil (%)
Methodenlehre, Statistik, Expra	770	54,5
Sozialpsychologie	387	27,4
Allgemeine Psychologie	268	19,0
Einführungsveranstaltung	263	18,6
Diagnostik	244	17,3
Differentielle Psychologie	197	13,9
Andere	164	11,6
Klinische Psychologie	121	8,6

Lehrveranstaltung	n	Anteil (%)
Biologische / Neuropsychologie	106	7,5
Arbeits-, Betriebs- und Organisationspsychologie	92	6,5
Pädagogische Psychologie	71	5,0
Entwicklungspsychologie	62	4,4
Gesundheitspsychologie	16	1,1
Medien- / Kommunikationspsychologie	13	0,9
Forensische Psychologie	11	0,8

666 *Table 4. Anteil von Teilnehmenden, die an ihrer aktuellen (bei Alumni: letzten) Universität in*  
667 *mindestens einer Lehrveranstaltung eine kritische Betrachtung der jeweiligen Forschungspraktik*  
668 *erfahren haben (N = 1397).*

Fragwürdige Forschungspraktik	n	Anteil (%)
HARKing	1046	74,0
Keine Stichprobenplanung	923	65,3
S.B. von Variablen	906	64,1
S.B. von Hypothesen	864	61,1
Flexible Stichprobengröße	854	60,4
Flexibler Ausschluss von Beob.	847	59,9
S.B. von Bedingungen	832	58,9
Flexible Datenanalyse	717	50,7
Runden von p-Werten	564	39,9
Nicht sicher	130	9,2
Keine	57	4,0

*Tabelle 5. 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

Prädiktor	B	SE	OR	KI (95.00 %)	z	p
-----------	---	----	----	--------------	---	---

Referenzkategorien: Weiblich (Geschlecht), Bachelor (Studienfortschritt), Nein (OSI an der Uni), Nein (OSI bekannt), Nein (RK gelehrt), Nein (Praktik gelehrt), Extra (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 6. 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***

	B	SE	t	p
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)

## 7 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>
- 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., ... Stack, N. (2019). *Redesigning methods curricula for reproducibility*. Abgerufen von <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>
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376.  
<https://doi.org/10.1038/nrn3475>
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., ... 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>
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, 6, 621.  
<https://doi.org/10.3389/fpsyg.2015.00621>
- 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., & Schwarz, N. (2016). Questionable Research Practices Revisited. *Social Psychological and Personality Science*, 7(1), 45–52.  
<https://doi.org/10.1177/1948550615612150>
- Glöckner, A., Fiedler, S., & Renkewitz, F. (2018). Belastbare und effiziente Wissenschaft. *Psychologische Rundschau*, 69(1), 22–36. <https://doi.org/10.1026/0033-3042/a000384>
- 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., ... Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and

- 725 Settings. *Advances in Methods and Practices in Psychological Science*, 1(4), 443–490.  
726 <https://doi.org/10.1177/2515245918810225>
- 727 Konferenzrat der Psychologie-Fachschaften-Konferenz. (2018). *Positionspapier der*  
728 *Psychologie-Fachschaften-Konferenz (PsyFaKo) zum Thema „Replikationskrise und*  
729 *Open Science in der Psychologie“*. Abgerufen von  
730 <https://psyfako.org/index.php/positionspapiere/>
- 731 Krishna, A., & Peter, S. M. (2018). Questionable research practices in student final theses –  
732 Prevalence, attitudes, and the role of the supervisor’s perceived attitudes. *PLoS ONE*,  
733 13(8). <https://doi.org/10.1371/journal.pone.0203470>
- 734 Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). {lmerTest} Package: Tests in  
735 Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26.  
736 <https://doi.org/10.18637/jss.v082.i13>
- 737 Nosek, B. A., & Lakens, D. (2014). Registered Reports. *Social Psychology*, 45(3), 137–141.  
738 <https://doi.org/10.1027/1864-9335/a000192>
- 739 Open Science Collaboration. (2015). Estimating the reproducibility of psychological science.  
740 *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- 741 R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Abgerufen von  
742 <https://www.r-project.org/>
- 743 Sarafoglou, A., Hoogeveen, S., Matzke, D., & Wagenmakers, E.-J. (2019). Teaching Good  
744 Research Practices: Protocol of a Research Master Course. *Psychology Learning &*  
745 *Teaching*. <https://doi.org/10.1177/1475725719858807>
- 746 Schönbrodt, F. D., Maier, M., Heene, M., & Bühner, M. (2018). Forschungstransparenz als  
747 hohes wissenschaftliches Gut stärken. *Psychologische Rundschau*, 69(1), 37–44.  
748 <https://doi.org/10.1026/0033-3042/a000386>
- 749 Schönbrodt, F., Gollwitzer, M., & Abele-Brehm, A. (2017). Der Umgang mit Forschungsdaten im  
750 Fach Psychologie: Konkretisierung der DFG-Leitlinien. *Psychologische Rundschau*,  
751 68(1), 20–35. <https://doi.org/10.1026/0033-3042/a000341>
- 752 Sijtsma, K. (2016). Playing with Data—Or How to Discourage Questionable Research Practices  
753 and Stimulate Researchers to Do Things Right. *Psychometrika*, 81(1), 1–15.  
754 <https://doi.org/10.1007/s11336-015-9446-0>
- 755 Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed  
756 flexibility in data collection and analysis allows presenting anything as significant.  
757 *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>

- Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., ...  
Shrout, P. E. (2017). It's Time to Broaden the Replicability Conversation: Thoughts for  
and From Clinical Psychological Science. *Perspectives on Psychological Science*, 12(5),  
742–756. <https://doi.org/10.1177/1745691617690042>
- Ulrich, R., Erdfelder, E., Deutsch, R., Strauß, B., Brüggemann, A., Hannover, B., ... 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>
- 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'*. Abgerufen von  
<https://cran.r-project.org/package=tidyverse>