





# Social Psychological Bulletin

## Evading Open Science: The Black Box of Student Data Collection

Tobias Ludwig<sup>1§</sup> , Marlene Sophie Altenmüller<sup>1§</sup> ,

Leonhard Falk Florentin Schramm<sup>1</sup> , Mathias Twardawski<sup>1</sup> 

[1] *Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany.*

§*These authors contributed equally to this work.*

Social Psychological Bulletin, 2023, Vol. 18, Article e9411, <https://doi.org/10.32872/spb.9411>

**Study plan received:** 2022-04-30 • **Study plan accepted (IPA):** 2023-04-16 • **Full paper received:** 2023-07-31 • **Full paper accepted:** 2023-08-09 • **Published (VoR):** 2023-11-17



**Handling Editors:** Simine Vazire, Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Australia; Brian Nosek, University of Virginia, Charlottesville, VA, USA

**Corresponding Author:** Marlene Sophie Altenmüller, Ludwig-Maximilians-Universität München, Department of Psychology, Leopoldstrasse 13, 80802 Munich, Germany. +49(0)89-21805285. E-mail: [marlene.altenmueller@psy.lmu.de](mailto:marlene.altenmueller@psy.lmu.de)

**Related:** This article is part of the SPB Special Topic "Is Psychology Self-Correcting? Reflections on the Credibility Revolution in Social and Personality Psychology", Guest Editors: Simine Vazire & Brian Nosek, Social Psychological Bulletin, 18, <https://doi.org/10.32872/spb.v18>

**Supplementary Materials:** Data, Materials, Preregistration [see [Index of Supplementary Materials](#)]



### Abstract

While Open Science has arguably initiated positive changes at some stages of the research process (e.g., increasing transparency through preregistration), problematic behaviors during data collection are still almost impossible to detect and pose a great risk to the validity and integrity of psychological research—especially, when researchers use data collected by others (e.g., students). Exploring students' and supervisors' perspectives, the present registered report enlightens this "black box" of student data collection, focusing on questionable research practices and research misconduct (QRP/M). The majority of students did not report having engaged in any problematic behaviors during data collection, but some QRP/M—ranging from somewhat questionable to highly fraudulent—seem quite common (e.g., telling participants the hypothesis beforehand, participating in one's own survey). We provide an overview of students' reported and supervisors' suspected data collection QRP/M, explore potential drivers for these behaviors based on the fraud triangle model (including pressures, opportunities, and rationalizations), and report how students and



This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), CC BY 4.0, which permits unrestricted use, distribution, and reproduction, provided the original work is properly cited.

supervisors perceive the eligibility of student data for further uses (e.g., scientific publications). Moreover, we explore the role of the student-supervisor relationship (e.g., communication and expectations) and Open Science practices in student projects. In summary, our findings suggest the potential scientific value of data from student projects. Fostering transparent communication regarding expectations, experiences, and intentions between supervisors and students might further contribute to strengthening this prospect.

## Keywords

questionable research practices, research misconduct, data collection, supervision, students

### Highlights

- From the student and supervisor perspectives, we investigated questionable research practices and research misconduct during data collection in student projects.
- While 64% of students did not report any problematic data collection practices, some behaviors seem not uncommon among students: 8% participated in their own study, and 26% let participants take part despite them knowing the hypothesis.
- Reducing pressures, opportunities, and rationalizations—together with making Open Science a central element of teaching—could be important starting points for interventions.
- Transparent student-supervisor communication regarding expectations, experiences, and intentions is likely vital for ensuring good quality of data collected by students.

Over the last decade, the ideas of Open Science have arguably brought many changes for the better in psychological research. For example, preregistering studies and openly sharing one's materials and data are becoming more and more standard, or even expected, procedures (e.g., Nosek et al., 2022; Nuijten et al., 2017; Tedersoo et al., 2021). Crucially, this development has not only reached the current research community but has also found its way to the source of the next generation of researchers: current bachelor and master students. Several universities have started to include Open Science practices in their curriculum for psychology students (see Loenneker et al., 2022; Schönbrodt et al., 2018). In fact, in a recent study (Brachem et al., 2022), the majority of psychology students stated that the replication crisis was discussed in at least one of their courses. These developments in teaching apparently raised students' awareness and yielded arguably desirable consequences. For example, most students indicated that the topic "Replication crisis and Open Science" is important, that they have positive attitudes towards Open Science practices (e.g., power analysis) and negative attitudes towards questionable research practices (QRPs), and that they actively engage in using Open Science practices themselves (Brachem et al., 2022; Krishna & Peter, 2018).

In this sense, students seem more and more aware of the existence and potential problems associated with degrees of freedom in the research process: Researchers have to make multiple decisions at every step of the research process—from more or less arbitrary to very substantial choices (i.e., researcher degrees of freedom; Simmons et al., 2011). These decisions can be (intentionally or unintentionally) exploited in an opportunistic and questionable manner (e.g., for achieving desired results), which can increase the probability of false positive results and may inflate effect sizes (e.g., Simmons et al., 2011; Wicherts et al., 2016). Wicherts et al. (2016) summarized such QRPs in an extensive list. Examples include: selecting only specific dependent variables from several alternative outcome measures, using alternative inclusion or exclusion criteria for including participants in analyses, or presenting exploratory results as confirmatory (i.e., hypothesizing after results are known; HARKing). Such QRPs may be considered “gray-area” practices. They lie on a continuum (Steneck, 2006) between responsible conduct of research and research misconduct—with the latter being defined as “deliberate or grossly negligent infringements defined in a set of regulations” (Deutsche Forschungsgemeinschaft, 2022, p. 22). Examples for research misconduct include falsification and fabrication of data and plagiarism (Deutsche Forschungsgemeinschaft, 2022). In sum, both, QRPs and even more severe practices like research misconduct (hereafter abbreviated as QRP/M<sup>1</sup>), are problematic research behaviors that pose a severe risk to the validity of scientific results.

## The Triangle Model of Problematic Research Behavior

Open Science practices carry the hope that they lead to a decrease in QRP/M and, indeed, they might tackle important drivers of such problematic research behaviors. These drivers of QRP/M may be conceptualized in terms of *pressure*, *opportunity*, and *rationalization*, constituting the three key elements in the fraud triangle model (Abdullahi & Mansor, 2015; based on Cressey, 1953). While the model originally focused on fraudulent financial behaviors, it can also be applied to problematic behaviors in science (Malgwi & Rakovski, 2009). Open Science developments have arguably contributed to effectively reducing the impact of these three key elements leading to QRP/M: (i.) The general *pressure* to use such practices has been reduced (e.g., the importance of finding and reporting statistically significant results has arguably shifted towards the importance of “sound” scientific methods, including more publications reporting null results), (ii.) there are fewer *opportunities* for engaging in QRP/M (e.g., preregistrations, open data, and open analysis scripts arguably discourage *p*-hacking or HARKing, for example, by making it easier to be discovered), and (iii.) engaging in *rationalizations* for using QRP/M has become more difficult (e.g., increased awareness for these problematic practices invalidates excuses like “This is how I was taught to do it”). Consequently, by reducing

---

1) We acknowledge that QRPs and research misconduct may differ significantly in their severity. In this paper, we use the umbrella term of QRP/M to refer to this broad range of problematic scientific behaviors.

the impact of the three triangle model elements, these and other developments accompanying Open Science practices might have contributed to a decreased prevalence of QRP/M in psychological science, including student research projects. In line with this reasoning, feeling informed about the topic “Replication crisis and Open Science” is negatively associated with engagement in QRPs (Brachem et al., 2022).

However, one could also argue that the increasing requirement to engage in Open Science practices could have also had unintended negative consequences regarding the three key elements of the triangle model: For example, aiming for larger samples may have (i.) increased pressure on researchers/students during data collection (e.g., having to reach the a-priori calculated and preregistered sample size). Similarly, (ii.) other opportunities to receive desired results that are not (yet) prevented or discouraged (e.g., by increasing its detectability) by Open Science practices may be exploited more often (e.g., telling participants the hypothesis beforehand). Lastly, (iii.) it is plausible to assume that new rationalizations for QRP/M may have emerged (e.g., “Sample size is more important than a perfect data collection process”). Therefore, the use of some QRP/M (e.g., HARKing) might have decreased over the last years but other forms might still be in place or have even increased. However, to be able to assess any such speculative developments for the better or worse and how these might be related to the three elements of the triangle model, we first need more detailed data regarding the prevalence of QRP/M—especially in the data collection process.

## QRP/M in the Data Collection Process

Until now, QRP/M during the data collection process have mostly been considered in terms of falsification and fabrication (Fanelli, 2009; John et al., 2012; Xie et al., 2021). *Falsification* can be defined as “manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record” and *fabrication* as “making up data or results and recording or reporting them” (see para. 2 and 3, United States Department of Health and Human Services Office of Research Integrity, n.d.). In a recent meta-analysis (Xie et al., 2021), the *self-admitted* prevalence for fabrication and falsification was 1.9% and 3.3%, respectively, while the *observed* prevalence was even 12.4% and 10.3%. These results are comparable to those of another meta-analysis, published 12 years earlier (Fanelli, 2009). Asking about the *perceived* prevalence paints an even more severe picture: In a study by Stürmer et al. (2017), 44.7% of respondents perceived data invention to be at least slightly prevalent, while 58.9% perceived active manipulation/faking of data to be at least slightly prevalent.

Moreover, QRP/M during data collection are arguably particularly problematic because data collection is still a “black box.” Even clear rules regarding open data (e.g., how to prepare and upload data for reuse), for example, by the *German Psychological Society* (Gollwitzer et al., 2021), are mostly ineffective if this data is (un)intentionally manipulated during its collection process. Consequently, QRP/M during data collection

are often very difficult to detect and it is, currently, hard or even impossible to tell if (open) data is of good quality.

Strikingly, detailed knowledge about the prevalence of QRP/M during data collection is relatively scarce. Past research has predominantly focused on QRP/M in other stages of the research process like data analysis and reporting (e.g., HARKing, *p*-hacking; e.g., Brachem et al., 2022; Fiedler & Schwarz, 2016; John et al., 2012; Krishna & Peter, 2018). By contrast, behaviors during data collection like falsification or fabrication of data were considered with only a few and very broad items (e.g., Fiedler & Schwarz, 2016; John et al., 2012; Krishna & Peter, 2018; Rajah-Kanagasabai & Roberts, 2015), but there are multiple other or more specific QRP/M in data collection which—to our knowledge—have not been examined in detail. To name a few: Duplicating data rows, creating data from scratch, or changing specific values. Other behaviors might include undisclosed instructions for participants, such as telling participants the hypothesis beforehand and/or having friends, family, or one's (fellow) students participate, who are likely familiar with one's research questions (e.g., to increase the likelihood of finding the hypothesized results). These examples show that, just like in other stages of the research process, QRP/M in data collection consist of practices ranging from questionable conduct to severe misconduct. They might be applied unintentionally, out of carelessness, or intentionally to achieve certain goals (e.g., significant results). In any case, this broad range of behaviors is worth investigating (and differentiating) in more detail than has been done in the past.

## The Student Perspective

Most studies about QRP/M focus on researchers and not students (e.g., Agnoli et al., 2017; Fiedler & Schwarz, 2016; John et al., 2012). We define “researchers” as professional academics whose occupation is conducting research (among other activities like teaching), including all career levels (ranging from doctoral students to professors). In contrast, the label “students” refers here to bachelor and master students (based on the European higher education system). These students are often actively involved in data collection processes as part of their studies (e.g., for thesis projects) and their behavior during this data collection process is likely of high relevance for the characteristics of these data (e.g., their decisions on who they recruit for participation or how they treat participants during participation in the study). Crucially, students' data may be made publicly available (i.e., as open data), and their supervisors, other students, and/or other researchers might use this data as part of their own work (e.g., as pilot data or for publications in scientific journals). Further, the “next generation” of researchers comes from the student population and even if they do not decide to follow an academic track (i.e., become researchers), students' future workplace behaviors might be similar to their research behaviors during their studies (see also Schönbrodt et al., 2018). Hence, students' behavior

in research projects they conduct during their studies and any engagement in QRP/M is highly relevant.

Just like researchers, prior studies suggest that students are inclined to engage in QRP/M. In general, students clearly use QRPs, with prevalence rates in individual projects ranging from 1.9% (“Stopping data collection after achieving the desired result”; p. 9; Krishna & Peter, 2018) to 23.1% (only reporting specific dependent variables; Brachem et al., 2022). Again, studies about students’ engagement in QRP/M have not focused on data collection behaviors, except for broad items referring to falsification and fabrication: Krishna and Peter (2018) report prevalence rates for falsification of 2.9% in individual student projects. Regarding fabrication, studies reveal prevalence rates of 14.6%–35.1% on the individual student—not project—level (i.e., proportion of students, who have shown this behavior at least once; Hard et al., 2006; Rajah-Kanagasabai & Roberts, 2015). Thus, a more detailed look at students’ engagement in the various types of QRP/M during data collection (as outlined above) is warranted.

There are many reasons why the drivers of QRP/M may be different for students and researchers. While researchers are, for example, motivated by competition for tenured positions and publications in high-impact journals (e.g., Stürmer et al., 2017), students may rather be faced with the conflict of finishing their project (within a certain time period) in order to graduate while reaching a (preregistered) sample size—and often without the prospect of contributing to the scientific literature. However, there is little research regarding potential drivers of students’ engagement in QRP/M during data collection. Initial evidence points to the theoretical value of the triangle model to understand such behaviors (Gopalakrishna et al., 2022; Krishna & Peter, 2018; Moran et al., 2022): As drivers for students to engage in QRP/M in general (e.g., *p*-hacking), past research revealed (i.) *pressure* by supervisors and other pressures (e.g., to graduate) as important. (ii.) Regarding *opportunity*, past research revealed the ease of manipulation and the low likelihood of detection as crucial. Lastly, (iii.) alluding to *rationalizations*, past research revealed the possibility of saving resources, QRP-attitudes of student supervisors, and even the recommendation to use these practices as fundamental. Thus, the triangle model may be useful to get an idea regarding the underlying potential drivers of students’ engagement in QRP/M during the data collection process.

In the present research, we want to focus on student data collection behaviors. First and foremost, we aim to assess the prevalence of students’ engagement in QRP/M during data collection in detail, but we will also explore what might drive such behaviors (e.g., the relevance of the three elements of the fraud triangle model), and to examine how student data is used beyond student projects (e.g., by supervisors).

## The Supervisor Perspective

Students’ engagement in QRP/M during data collection is particularly problematic if other researchers intend to use their data (e.g., for their own research). For data collected

by students, supervisors (i.e., researchers who accompany students conducting research) have the role of “gate keepers”. As such, they (at least partly) carry the responsibility of ensuring that only “good quality” data ends up as pilot data, as open data, or in published literature.

There is no consensus among researchers regarding the eligibility of student data for inclusion in the scientific discourse: Some researchers doubt the integrity of research done by students (Agnoli et al., 2017; John et al., 2012), but others may underestimate the actual prevalence of students’ academic dishonesty (e.g., students falsifying research results; Brimble & Stevenson-Clarke, 2005). Both underestimating and overestimating the prevalence of QRP/M in student research projects may be problematic: *Underestimating* this prevalence could lead to supervisors unknowingly using (un)intentionally manipulated data for their own work (or sharing it with other researchers), *overestimating* this prevalence could lead to an unethical waste of resources (e.g., participants’ time, publicly funded research materials, etc.) because eligible data ends up in the file drawer. Importantly, to our knowledge, there is no literature on how student data is actually used beyond student projects and if supervisors actually check their students’ data regarding its eligibility. Thus, it is necessary to investigate supervisors’ perspectives and what they expect from and how they handle data originating from student research projects.

## The Present Research

Taken together, while past research has already investigated the prevalence of QRP/M in various stages of student research projects (e.g., Brachem et al., 2022; Krishna & Peter, 2018), we propose taking a more detailed look at QRP/M related to data collection. Here, “data collection” is very broadly defined and refers to several steps in the research process including: (i.) study preparation (e.g., changing study materials over the course of the data collection), (ii.) actually collecting data (e.g., instructing participants to answer in a certain way), and (iii.) data preprocessing prior to analyses (e.g., claiming raw data that has, in fact, already been processed). While this concept may be rather broad, it specifically includes aspects which are arguably the hardest to check for other researchers in hindsight (given the current standards of study reporting and documentation). For example, it is impossible to infer from open data, open materials, or preregistrations whether participants were systematically instructed in a way that benefited the hypothesis. To better understand students’ engagement in QRP/M during data collection, we explore the relation with students’ perceived pressures, opportunities, and rationalizations (i.e., the triangle model) regarding such behaviors and their data collection. Moreover, supervisors may be seen as the gate keepers for student data to enter the scientific literature. Consequently, it is crucial to assess their expectations and actual use of student data.

In the present research project, we conducted two parallel studies.<sup>2</sup> First, we were interested in the student perspective and investigated whether and how students actually (un)intentionally engage in QRP/M during their data collection (e.g., for thesis projects),

and explored which drivers—based on the fraud triangle model—are related to these behaviors. Second, we were interested in the supervisor perspective on students' engagement in QRP/M and students' drivers related to these behaviors. We also examined whether supervisors use data collected by students for their own research projects and how often they check the data themselves prior to using it. In total, we investigated the extent to which problematic data resulting from student research projects may end up in the published literature.

For both perspectives, most measures had to be constructed specifically for the present research purpose. Thus, we pretested all measures with the “think aloud”-method (Presser et al., 2004). We conducted cognitive interviews with three students (i.e., for the student questionnaire) and two supervisors (i.e., for the supervisor questionnaire). These interviews were guided by suggestions by Willis (1999): While filling out the questionnaire, interviewees were asked to speak out loud what they think. In accordance with the responses given by the interviewees, some measures were revised (e.g., item wording).

The final measures including all study materials, other supplementary information (“SI”), as well as the analysis scripts and anonymized data are provided on the Open Science Framework (see [Supplementary Materials](#)). Both questionnaires and their pretests were programmed with SoSci Survey (Leiner, 2022) and all analyses were conducted in R (R Core Team, 2023). We declare that, in this Stage 2 manuscript, we did not deviate from the methods registered in Stage 1 and that our analyses followed the registered plan as closely as possible (see Footnotes 10 and 12 regarding small deviations).

## Method

### Student Participants

#### Sampling Procedure

We recruited bachelor and master students as participants. Our sample size considerations were informed by Krishna and Peter (2018;  $N = 207$  participants) and Brachem et al. (2022;  $N = 1146/1397$  participants) who also looked at student QRPs. In line with our registered sampling strategy, we collected data for eight weeks<sup>3</sup> by contacting psychology programs at German public universities. However, we originally registered to draw nine<sup>4</sup>

---

2) Regarding the structure of the manuscript, we slightly deviate from the Stage 1 registered report, by reporting the methods and results for the student perspective and the supervisor perspective together instead of consecutively as Study 1 and Study 2. Importantly, the content reported in this Stage 2 manuscript reflects the content of our Stage 1 registered report.

3) The stopping rule for data collection was based on three criteria: (i.) we will stop after four weeks of data collection as soon as we reach a sample size of  $N > 1000$  (sample size similar to Brachem et al., 2022), (ii.) after six weeks of data collection, we will stop as soon as we reach a sample size of  $N > 500$ , and (iii.) after eight weeks,



German universities by chance from a list of all accredited German public universities with respective psychology degree programs (i.e., bachelor and master), keeping track of how many students were contacted at each university to estimate response rates. As this strategy did not result in sufficient responses, we first increased the random selection to 20 universities and, shortly after, opened the sampling efforts to all German-language universities, because we were informed that the survey link had been broadly forwarded by others to students from different universities (e.g., shared in German wide messenger groups). Participation was incentivized using voucher raffles: independent of the final sample size, five participants were drawn by chance from the pool of participants who signed up for the lottery after completing the study and were awarded a 50 Euro voucher each. Participants had to meet the following criteria to participate in the study: at least 18 years old, good German language skills, currently studying or having studied psychology, and having already been involved in data collection during a student research project in the context of their studies.

### Sample Description

This recruitment procedure resulted in 483 participants who finished the survey and reached the last survey page. As preregistered, ten further participants had to be excluded as they stated that their data should not be used. These final 473 participants were 18 to 53 years old ( $M = 24.34$ ,  $SD = 4.44$ ) and largely female (87%, 12% male, < 1% other).<sup>5</sup> Most participants were currently studying psychology (88%, 12% had studied psychology in the past) and had started their psychology degree after 2015 (94%). In our survey, approximately half of the participants reported their experiences related to their thesis projects (42% bachelor theses, 10% master theses, 38% experimental seminar “Empira/Expra”, 8% seminar work, 2% other). The majority of projects ended only recently (67%, in or after 2022) and was conducted in groups ranging from 0 to 30 other involved students ( $M = 3.36$ ,  $Mdn = 2$ ,  $SD = 4.87$ ; 35% of projects were conducted alone).<sup>6</sup> A third (35%) of these projects was supervised by pre-doctoral researchers, 31% by post-doctoral researchers, and 26% by professors (14% indicated “other” and 5% did not know their supervisor’s position).

---

regardless of how long it takes, we will stop as soon as we have at least  $N = 200$  participants (sample size similar to Krishna & Peter, 2018).

4) Based on an estimated response rate of 20%, which lies on the lower end of those of similar studies (Agnoli et al., 2017; Fiedler & Schwarz, 2016; John et al., 2012). We conservatively calculated with around 300 eligible students per university.

5) This gender distribution is only slightly above the German average for studying psychology which has an approximate 8:2 female to male ratio (Bühner, 2023).

6) Four participants indicated unusually large group sizes (56, 60, 100, 150). We disregarded those values (“NA”) in our descriptive report in the main text. Including them, the average group size was 4.10 ( $Mdn = 2$ ,  $SD = 10.05$ ).

## Supervisor Participants

### Sampling Procedure

In line with our registered sampling strategy, we collected data for 4 weeks.<sup>7</sup> We contacted researchers from all chairs of the psychology departments at the same universities which were randomly selected in the first phase of recruitment for the student perspective and, additionally, we used academic mailing lists of all 17 divisions of the German Psychological Society. Again, participation was incentivized using voucher raffles (2 x 50 Euro). Participants had to meet the following criteria to participate in the study: at least 18 years old, psychology as field of study, currently working as researchers in academia, and having supervised at least one student project including data collection.

### Sample Description

Following this sampling procedure, 205 participants finished the survey and reached the last questionnaire page. As preregistered, we further excluded 6 participants who stated that their data should not be used, resulting in a final sample of  $N = 199$ . Participants were between 24 and 71 years old ( $M = 35.97$ ,  $SD = 9.34$ ) and mostly female (74%, 27% male, 1 NA).<sup>8</sup> About half (49%) finished their psychology degree (master or diploma) after 2015. Career-wise, 40% were pre-doctoral and 37% post-doctoral researchers, 20% were professors, and 3% stated another position (e.g., junior professor). On average, they supervised 16.20 student projects in the last five years at their current location ( $Mdn = 11$ ,  $SD = 15.35$ ), ranging from 1 to 75.<sup>9</sup>

## Measures

Below, we describe the measures used for the student perspective alongside the measures used for the supervisor perspective (see [Supplementary Materials SIs](#) “Procedure”

---

7) Our stopping rule was based on three criteria (in line with our sampling plan for the student perspective): (i.) we will stop after four weeks of data collection as soon as we reach a sample size of  $N > 200$ , (ii.) we will stop after six weeks of data collection as soon as we reach a sample size of  $N > 100$ , and (iii.) after eight weeks, regardless of how long it takes, we will recruit at least  $N = 100$  participants. We based our sample size rationale on the assumption that the target population of supervisors at German universities will largely consist of pre- and postdoctoral researchers. We roughly estimated that they may, on average, have supervised around five student projects with data collection. Therefore, we assumed a ratio of 5:1 (i.e., a target sample of  $N = 100$  researchers for the supervisor perspective compared to a sample of 500 students for the student perspective). Note, that supervisors in our present sample indicated a much higher number of supervised projects ( $M = 21.22$ ,  $Mdn = 12$ ,  $SD = 41.51$ , range: 1–414) and spanned across all career levels (pre-doc to professor).

8) This is higher than the typical female-to-male ratio in psychology in German: In 2020, there were 66% female (vs. 34% male) research associates (pre- and postdoctoral) and 43% female (vs. 57% male) professors (Bühner, 2023).

9) Four participants indicated unusually high numbers (i.e., 100, 250, 300, 414), so we disregarded those values (“NA”) in our descriptive report in the main text. Including these values, the average number of supervised projects was 21.22 ( $Mdn = 12$ ,  $SD = 41.51$ ).

and “Study Materials” for information on the complete study flow). First, we obtained informed consent and assessed demographics. In a short introduction, we informed students that all subsequent questions referred to one specific, completed research project they conducted while studying psychology (i.e., not as part of a job, for example, as a student research assistant), in which they had the most responsibility regarding data collection (i.e., recruiting and instructing participants, data collection, preparing data before analyses, etc.). We asked about the project type (e.g., bachelor or master thesis), the year in which they finished it, the number of other students involved in that project, and which position their supervisor had (i.e., pre-doc/doctoral student, post-doc, professor, other, don’t know).

We informed supervisors in the introduction that all subsequent questions (if not stated otherwise) referred to completed projects that they had supervised as primary supervisor in the last five years, at their current university, and in which students collected data as part of their studies. As noted for each measure, some referred to the specific *number* of supervised projects (e.g., in how many QRP/M were used), and some to the *majority* of supervised projects (e.g., perceptions of drivers in the majority of projects). We asked participants how many of these projects they had supervised and which position they had.

### QRP/M: Use and Drivers

We assessed the engagement in QRP/M with 17 items based on John et al. (2012), Stürmer et al. (2017), and Wicherts et al. (2016). We asked students whether they had used any of these research practices in their project (answer options: “yes” and “no”) and we asked supervisors about their beliefs about students’ engagement in QRP/M (i.e., the number of their supervised student projects in which they suspected students to have used the specific practice). Example items reflecting the broad scope of data collection behaviors we were interested in are: “Knowingly letting participants take part in the study, while being aware that they know the hypotheses from conversations unrelated to their study participation (e.g., friends, family, or fellow students)”, “Changing the study material (e.g., items, stimuli, manipulations) over the course of the data collection without disclosure”, and “Adding pilot data to the data collected for the main study without disclosure.”

To measure the potential drivers for engaging in QRP/M, we used 24 items theoretically derived from the fraud triangle model (Abdullahi & Mansor, 2015; based on Cressey, 1953) and partly based on recent research on related topics (Gopalakrishna et al., 2022; Krishna & Peter, 2018; Moran et al., 2022). We provided participants with an introductory text presenting the QRP/M items once more, neutrally referring to them as “listed behaviors”. We asked students to state how much they agreed with the 24 items regarding their experiences during their project in general and more specifically, regarding these “listed behaviors” during data collection. We asked supervisors how much they believed their students experienced these drivers in the majority of their supervised projects. Example

items are “I felt pressure to reach a certain sample size” (i.e., pressure), “Some ‘listed behaviors’ were or would have been easy to carry out” (i.e., opportunity), and “I believe that some ‘listed behaviors’ correspond to the ‘correct’ way of doing it (or at least I believed so at the time of the project)” (i.e., rationalization). We used a 6-point Likert scale ranging from 1 = “*strongly disagree*” to 6 = “*strongly agree*”.

### **Supervision: Expectations and Goals**

We assessed students’ perceived supervisor expectations (in their project) and supervisors’ actual expectations (regarding the majority of supervised projects) with eleven items using a 6-point Likert scale ranging from 1 = “*not important*” to 6 = “*very important*” (e.g., “Data should be eligible for publication in a scientific journal”). Further, we asked only supervisors how important different supervision goals (i.e., teaching, research, obligation) were to them in the majority of their supervised projects with three items (e.g., “Teaching [i.e., help students to learn skills and achieve a good education]”) and an additional option to type in other goals and measured responses on a 6-point Likert scale ranging from 1 = “*not important*” to 6 = “*very important*”.

### **Data Use: Communicated, Actual, and Expected Use, and Perceived Data Eligibility**

We asked students what they were told about how their supervisors would use their data with one multiple-choice question: “Before starting with the data collection: Did you know whether your supervisor planned to use the data further and, if so, how (e.g., for their own projects)?”. We asked supervisors what they had communicated in the majority of their supervised projects. We provided participants with seven possible response options: using the data for publications, as pilot data, as open data, and other uses (other than research, unclear purpose, no interest, don’t know).

Further, we asked students about the actual and future expected data use with two questions (“Except you, who has used your data [to the best of your knowledge]?” and “Beyond that, who do you think will use your data in the future?”) with five multiple-choice options (supervisor, other students, other researchers, other, nobody) each. Deviating from the student perspective, we asked supervisors in how many supervised projects they engaged in certain actual data use behaviors (three items: in own projects, in paper submissions, as pilot data) and behaviors regarding their scrutiny of the eligibility of the data (four items: doubting correctness, checking data if no plans for further use, checking data if plans for further use, using it despite knowing it might be problematic).

Further, we asked students and supervisors (for the majority of projects) to what extent they think the data would be eligible as pilot data, for publication, and as open data (i.e., perceived data eligibility) with a 6-point Likert scale ranging from 1 = “*not at all*” to 6 = “*absolutely*”.

## Open Science Practices

We asked about the use of five Open Science practices: preregistration, power considerations, open data, open materials, and open analysis script. For students, the answer options for each practice were “Yes (e.g., by me, by my supervisor)”, “No”, and “I don’t know”. For supervisors, we asked in how many projects these practices were applied.

## Preliminary Data-Treatment

Prior to analyzing the data, we calculated an overall QRP/M score for every student by counting the activities (out of 17) they indicated they engaged in. For every supervisor, we calculated the prevalence of the responses for suspected engagement of students’ in QRP/M, for data use behaviors, and for applying the five Open Science Practices by dividing supervisors’ respectively indicated number of projects by their total number of projects.

Furthermore, we calculated mean values across the three items measuring perceived data eligibility, across ten of the eleven items measuring (perceived) supervisor expectations (reflecting expectations towards a high scientific standard in student projects)<sup>10</sup>, and across the 24 items measuring QRP/M drivers. Additionally, we calculated separate means for each of the three theoretically derived elements (i.e., pressure, opportunity, and rationalization) of QRP/M drivers.<sup>11</sup>

# Results

## Descriptive Analyses

### QRP/M Prevalence Rates

Similar to John et al. (2012), we report a relative QRP/M self-reported prevalence for each QRP/M by counting all students who chose the answer option “yes” and dividing this number by the total number of participants. For supervisors, we report the mean suscep-

---

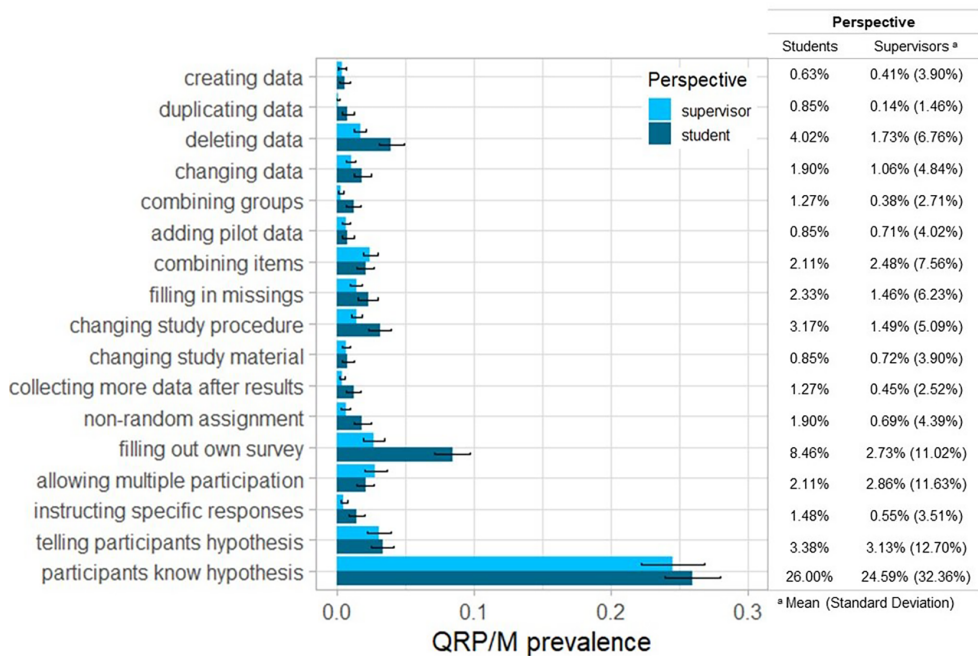
10) Deviating from our Stage 1 registration, we do not include the item “Results should be statistically significant” in the composite variable of supervisor expectations because it is arguably the only item in that scale that does not reflect expectations towards a high scientific standard. However, the overall pattern of results is very similar when including it as registered. The mean of students’ perceptions of their supervisors’ expectations is 3.96 ( $SD = 0.92$ ) in the registered and 4.08 ( $SD = 0.98$ ) in the new composite. Similarly, the mean of supervisors’ reported expectations is 4.22 ( $SD = 0.73$ ) in the registered and 4.45 ( $SD = 0.76$ ) in the new composite.

11) We validated our theoretical categorization of items by further providing three student assistants (i.e., the student perspective) and two researchers not involved in the project (i.e., the supervisor perspective) with definitions for the three elements of the fraud triangle and had them categorize each item. Our analyses are based on the majority solution for item categorization (item 21, stating the behavior was due to an instruction by the supervisor, was not included in any of the three elements due to achieving no clear majority solution and only used in the overall average).

ted prevalence rates of QRP/M in student projects (each supervisors' number of projects with suspected QRP/M divided by their total number of projects) and their standard deviations. Figure 1 displays these self-reported and suspected QRP/M prevalence rates. Overall, 64% of students did not indicate any of the listed behaviors, 24% indicated one, and 12% indicated 2 or more of the behaviors. On average, students indicated engagement in 0.63 QRP/M ( $SD = 1.25$ ). Overall, supervisors reported 1352 QRP/M across 4223 projects (note that these behaviors could also cumulate within single projects); 35% of supervisors did not suspect any of the listed behaviors in any of the student projects they supervised.

**Figure 1**

*Students' QRP/M Prevalence Rates as Reported by Students and Suspected by Supervisors*



*Note.* QRP/M prevalence rates as reported by students (i.e., the average of responses “no” = 0 and “yes” = 1 to having engaged in the respective behavior) and suspected by supervisors (i.e., average of each supervisors' number of student projects in which QRP/M were suspected divided by their total number of projects). Error bars indicate the standard error of the mean, prevalence rates and standard deviations (for supervisors' suspected QRP/M) are reported as numbers on the right side of the figure.

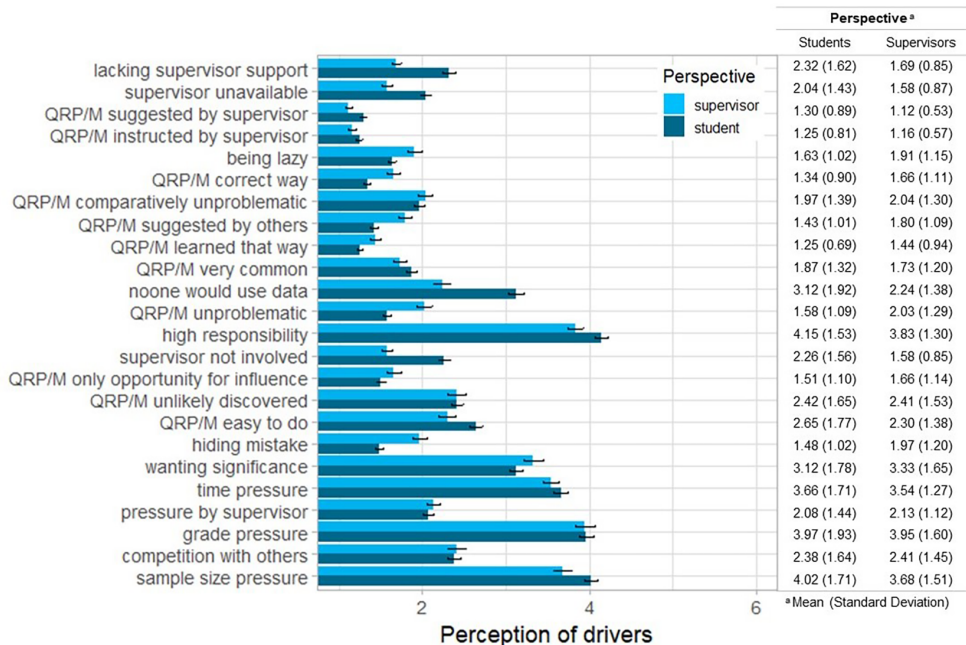
**QRP/M Drivers**

Figure 2 displays the means, standard errors, and standard deviations for all (perceived) QRP/M drivers separately. All drivers averaged, students' responses ranged from 1.00 to

4.58 ( $M = 2.28$ ,  $SD = 0.70$ ) on a scale from 1 (“strongly disagree”) to 6 (“strongly agree”). Breaking this down into the three key elements derived from the fraud triangle, students agreed most to the items capturing experiences of pressure ( $M = 2.96$ ,  $SD = 1.08$ ), a bit less to the items capturing opportunity ( $M = 2.60$ ,  $SD = 0.96$ ), and least to the items capturing rationalizations ( $M = 1.81$ ,  $SD = 0.69$ ). Quite similarly, supervisors’ overall responses ranged from 1.00 to 4.92 ( $M = 2.22$ ,  $SD = 0.70$ ) and, regarding the three key elements based on the fraud triangle, they again agreed most to the items capturing experiences of pressure ( $M = 3.00$ ,  $SD = 0.98$ ), a bit less to the items capturing opportunity ( $M = 2.36$ ,  $SD = 0.88$ ), and least to the items capturing rationalizations ( $M = 1.75$ ,  $SD = 0.71$ ).

**Figure 2**

*Experiences of Students’ Possible QRP/M Drivers as Reported by Students and Perceived by Supervisors*



*Note.* Experiences of students’ possible QRP/M drivers as reported by students and perceived by supervisors. Both response scales ranged from 1 = “strongly disagree” to 6 = “strongly agree”. Error bars indicate the standard error of the mean, means and standard deviations are reported as numbers on the right side of the figure.

### (Perceived) Supervision Expectations and Goals

Overall, students’ average response to the items measuring their perceptions of their supervisor’s expectations towards a high scientific standard in their projects ranged from 1.10 to 6.00 ( $M = 4.08$ ,  $SD = 0.98$ ) on a scale from 1 = “not important” to 6 = “very impor-

*tant*". Considerably higher, supervisors' average responses of their actual expectations ranged from 1.80 to 6.00 ( $M = 4.45$ ,  $SD = 0.76$ ) on the same scale. Table 1 displays the means and standard deviations for all items individually.

**Table 1**

*(Perceived) Supervisor Expectations as Perceived by Students and Reported by Supervisors*

(Perceived) Supervisor Expectation	Perspective $M$ ( $SD$ )	
	Students	Supervisors
Data should be eligible for a scientific publication.	3.07 (1.81)	3.80 (1.66)
Data should be eligible as pilot data.	3.19 (1.72)	4.13 (1.53)
Data should be eligible as open data.	3.09 (1.76)	3.32 (1.69)
Data quality should be high.	4.75 (1.28)	5.27 (1.00)
Results should be statistically significant.	2.72 (1.41)	1.91 (1.18)
Data collection should be spotless.	4.21 (1.38)	4.48 (1.36)
Data collection should be in line with highest scientific standards.	4.91 (1.18)	5.45 (0.80)
Study should be preregistered.	3.59 (2.10)	3.44 (1.88)
Sample size should be determined by power analysis.	4.34 (1.84)	4.41 (1.56)
A certain sample size should be reached.	4.64 (1.43)	4.55 (1.26)
Decisions during the scientific process should be made transparent.	5.04 (1.27)	5.65 (0.84)

Note. Response scales ranged from 1 = "not important" to 6 = "very important".

Regarding their supervision goals (scale from 1 = "not important" to 6 = "very important"), supervisors agreed most and very strongly that their goal in supervising student projects was teaching (i.e., helping students to develop their skills;  $M = 5.55$ ,  $SD = 0.71$ ). The agreement to the goal of contributing to one's own research was rather moderate and much more variable ( $M = 4.24$ ,  $SD = 1.45$ ). Finally, the agreement that supervising projects was to fulfill an obligation was even lower, but still moderate and very variable ( $M = 3.47$ ,  $SD = 1.59$ ).

### Data Use and Data Eligibility

Table 2 displays students' reports of what data use was communicated to them prior to data collection, how the data was actually used, and how they expected the data to be used in the future, as well as supervisors' answers regarding communicated data use in the majority of their supervised projects. Regarding supervisors' actual data use behaviors, we again calculated relative frequencies (number of projects divided by total number of projects): On average, supervisors reported that they actually used the data from 44% ( $SD = 33\%$ ) of student projects (see Table 2). Further, they submitted a paper with data from 16% ( $SD = 20\%$ ) of projects and used the data as pilot data from 26%



( $SD = 28\%$ ) of projects. They doubted the correctness / integrity of the data in 10% of projects ( $SD = 20\%$ ).<sup>12</sup>

**Table 2**

*Frequencies of Data Use as Reported by Students and by Supervisors*

Data use	Perspective	
	Students	Supervisors
<b>Communicated use</b>		
for scientific publication	19%	52%
as pilot data	11%	43%
as open data	9%	19%
for other use (e.g., test data)	9%	13%
for unknown future use	19%	50%
no interest in use	26%	22%
“don’t know”	29%	6%
<b>Actual use</b>		
by supervisor	40%	44%
by other students	37%	—
by other researchers (not supervisor)	11%	—
by others	3%	—
by no one (except student)	39%	—
<b>Expected use</b>		
by supervisor	42%	—
by other students	20%	—
by other researchers (not supervisor)	15%	—
by others	3%	—
by no one	48%	—

*Note.* For students: frequencies based on selection in multiple choice question regarding their project. For supervisors: communicated use based on selection in multiple choice question regarding majority of projects supervised in the last five years at their current institution; actual use based on number of projects divided by total number of projects.

Regarding perceptions of data eligibility, students perceived their data to be moderately eligible for further use ( $M = 3.96$ ,  $SD = 1.24$ ). Interestingly, they agreed most that their

12) We further asked in how many student projects supervisor participants checked the data if planning to use it ( $M = 49\%$ ,  $SD = 39\%$ ) and if not planning to use it ( $M = 39\%$ ,  $SD = 42\%$ ) and whether they used the data despite knowing it was problematic ( $M = 1\%$ ,  $SD = 6\%$ ). As these represent conditional frequencies (e.g., dependent on use or not), it was suboptimal to assess them within the same format as the unconditional questions (e.g., actual data use). Thus, we refrain from interpreting these.

data was eligible as open data ( $M = 4.25$ ,  $SD = 1.56$ ), only then as pilot data ( $M = 4.09$ ,  $SD = 1.46$ ), and least as data for a scientific publication ( $M = 3.54$ ,  $SD = 1.55$ ). Slightly higher, supervisors perceived the data in the majority of their supervised projects to be moderately eligible for further use ( $M = 4.26$ ,  $SD = 0.94$ ). Different to students, they agreed most and strongly that their data was eligible as pilot data ( $M = 5.09$ ,  $SD = 1.08$ ), then, but more moderately, as open data ( $M = 3.91$ ,  $SD = 1.37$ ), and least as data for a scientific publication ( $M = 3.79$ ,  $SD = 1.18$ ).

## Open Science Practices

Table 3 displays the average frequencies of projects in which each Open Science practice (i.e., preregistration, power analysis, open data, open materials, and open analysis script) was used. When averaging these responses (for students: “yes” = 1, “no/don’t know” = 0) into a general “Open Science” indicator, students’ mean was 0.33 ( $SD = 0.29$ ) and supervisors’ mean was 0.29 ( $SD = 0.30$ ).

**Table 3**

*Frequencies of Open Science Practices as Reported by Students and Supervisors*

Open Science practice	Perspective		
	Students		Supervisors
	yes	don’t know	yes ( $SD$ )
preregistration	47%	18%	37% (59%)
power analysis	70%	7%	65% (57%)
open data	19%	26%	17% (28%)
open materials	16%	29%	14% (27%)
open analysis script	12%	28%	12% (25%)

*Note.* For students: remaining percentage = “no”. For supervisors: average frequency based on each supervisor’s number of projects in which each practice was used divided by each supervisor’s total number of projects. Number in brackets represent the standard deviation.

## Correlational Analyses

### Correlational Overview

Table 4 provides an overview of correlations between the general descriptive indicators in this study from both the students’ and the supervisors’ perspectives. We report and interpret the size of the correlation coefficients following conventions in the literature (Funder & Ozer, 2019).<sup>13</sup>

**Table 4***Correlations of (Suspected) QRP/M With Several Indicators for the Students' and Supervisors' Perspectives*

Variable	1	2	3	4	5
1. QRP/M	—	.56 [.45, .65]	-.17 [-.30, -.03]	-.19 [-.32, -.05]	-.11 [-.24, .03]
2. QRP/M drivers	.44 [.37, .51]	—	-.09 [-.23, .05]	-.35 [-.46, -.22]	.01 [-.13, .14]
3. Supervisor expectations	-.18 [-.27, -.09]	-.20 [-.28, -.11]	—	.41 [.29, .52]	.35 [.22, .46]
4. Data eligibility	-.17 [-.25, -.08]	-.24 [-.33, -.16]	.46 [.38, .53]	—	.09 [-.05, .22]
5. Open Science	-.06 [-.15, .03]	-.07 [-.16, .02]	.55 [.48, .61]	.28 [.19, .36]	—

*Note.* Correlations with confidence intervals for the student perspective are below the diagonal. Correlations with confidence intervals for the supervisor perspective are above the diagonal. Values in square brackets indicate the 95% confidence interval for each correlation. QRP/M = students' reported QRP/M (number of indicated QRP/M) and supervisors' suspected QRP/M (composite of suspected QRP/M prevalence rates based on each supervisors' number of projects with suspected QRP/M divided by total number of projects); QRP/M drivers = composite of items assessing [supervisors' perception of] students' experiences during their project; Supervisor expectations = composite of items assessing [students' perception of] supervisors' expectations of a high scientific standard in student projects; Data eligibility = composite of items assessing whether students and supervisors believe the student data would be eligible for further use (e.g., for a scientific publication); Open Science = reported Open Science practices in student projects (students: used vs. not used/don't know; supervisors: composite of frequencies for each practice based on number of projects with Open Science practices divided by total number of projects).

### QRP/M and Their Drivers

There was a very large association between QRP/M and the combined drivers (students:  $r = .44$ , supervisors:  $r = .56$ ; see Table 4). This means that students who reported more QRP/M and supervisors who suspected more QRP/M also agreed more strongly to students' experiences reflecting the elements of the fraud triangle model (i.e., pressures, opportunities, and rationalizations). Looking at this more closely, Table 5 displays the associations between the QRP/M and the three driving elements based on the triangle model of fraud: For students, the association with the pressure element was medium ( $r = .21$ ), while the associations with the opportunity ( $r = .35$ ) and rationalization ( $r = .50$ )

13) In the [Supplementary Materials](#) (see SI "Data and Analysis"), we also provide R code to receive all  $p$ -values (two-tailed  $\alpha < .05$ ) of correlation coefficients which are (i.) uncorrected and (ii.) adjusted after Bonferroni for the convenience of the reader in a table but refrain from interpreting these.

elements can be considered large to very large. From the supervisors' perspective, the association with the pressure element can be considered large ( $r = .37$ ) and the associations with opportunity ( $r = .55$ ) and with rationalization ( $r = .55$ ) were both very large.

**Table 5**

*Correlations of (Suspected) QRP/M With Possible QRP/M Drivers for the Students' and the Supervisors' Perspectives*

Variable	1	2	3	4
1. QRP/M	—	.37 [.24, .48]	.55 [.45, .64]	.55 [.44, .64]
2. Pressure	.21 [.12, .29]	—	.58 [.48, .67]	.55 [.44, .64]
3. Opportunity	.35 [.27, .43]	.44 [.36, .51]	—	.76 [.69, .81]
4. Rationalization	.50 [.43, .57]	.45 [.38, .52]	.68 [.62, .72]	—

*Note.* Correlations with confidence intervals for the student perspective are below the diagonal. Correlations with confidence intervals for the supervisor perspective are above the diagonal. Values in square brackets indicate the 95% confidence interval for each correlation. QRP/M = students' reported QRP/M (number of indicated QRP/M) and supervisors' suspected QRP/M (composite of suspected QRP/M prevalence rates based on each supervisors' number of projects with suspected QRP/M divided by total number of projects). Pressure, opportunity, and rationalization are the respective composites of the driver items categorized according to theoretical considerations derived from the fraud triangle model.

In the [Supplementary Materials SI](#), we also report the correlations of the QRP/M score with each individual driver (see [Supplementary Materials SI](#) "Selected Correlations"). Most correlations were small to medium. The highest associations of students' QRP/M were with the items stating that this was how they learnt to do it ( $r = .50$ ), the right way to do it ( $r = .44$ ), and very common among students ( $r = .42$ ), and that this was the only opportunity to influence their results ( $r = .43$ ). For supervisors, the highest associations were with the items stating that students believed these behaviors were unproblematic ( $r = .55$ ), very common ( $r = .54$ ), unlikely discovered ( $r = .54$ ), and less problematic than other behaviors ( $r = .48$ ). We recorded only few negligible or very small correlations: Three drivers for students—wanting to find evidence for an effect ( $r = .05$ ), having high responsibility ( $r = .03$ ), and feeling one's supervisor was not available ( $r = .08$ )—and one for supervisors (that the behavior was suggested by the supervisor,  $r = .08$ ; highly discrepantly, this last item showed one of the strongest associations for students;  $r = .41$ ).

## (Perceived) Supervision Expectations and Goals

For both the students' and supervisors' perspective, the correlations of QRP/M and overall supervisor expectations were negative and rather small (students:  $r = -.18$ , supervisors:  $r = -.17$ ; see Table 4). This means that the more students and supervisors agreed with the items assessing various expectations regarding student projects (i.e., "higher" expectations towards a high scientific standard), the less they reported or suspected engagement in QRP/M. This pattern becomes clearer, when looking at the associations with the individual expectations (see Supplementary Materials SI "Selected Correlations"): The general association seemed to be driven by small to moderate individual associations with expectations that were directly related to the data itself (e.g., "Data should be eligible for a scientific publication," students:  $r = -.13$ , supervisors:  $r = -.21$ ; "Data quality should be high," students:  $r = -.12$ , supervisors:  $r = -.10$ ) and the data collection process (e.g., "Data collection should be in line with the highest scientific standards," students:  $r = -.25$ ; supervisors:  $r = -.19$ ).

While the supervisors' supervision goal of teaching was unrelated to suspected QRP/M ( $r = .03$ ), the goal of contributing to one's own research ( $r = -.12$ ) was small but negatively associated with suspected QRP/M, and agreeing that supervision was an obligation was moderately positively associated with suspected QRP/M ( $r = .24$ ).

## Data Eligibility

QRP/M were negatively associated with perceived data eligibility (students:  $r = -.17$ , supervisor:  $r = -.19$ ; see Table 4). The separate associations with the eligibility for different forms of use (scientific publication, pilot data, open data, see Supplementary Materials SI "Selected Correlations") were, for students, very small (as open data,  $r = -.08$ ) to small (as pilot data,  $r = -.17$ ; for a scientific publication,  $r = -.15$ ). For supervisors, while still small for scientific publications ( $r = -.17$ ), they were negligible for pilot data ( $r = -.04$ ) and moderate for open data ( $r = -.21$ ).

## Open Science Practices

Table 4 shows that the association of QRP/M with the general Open Science indicator was (very) small (students:  $r = -.06$ , supervisors:  $r = -.11$ ). Looking separately at each Open Science practice (see Supplementary Materials SI "Selected Correlations"), for students (used = 1, not used / don't know = 0), associations with QRP/M were all very small to negligible ( $r$  between  $-.08$  and  $-.00$ ), and, for supervisors, they were small ( $r$  between  $-.18$  and  $-.10$ ) or negligible (power analyses:  $r = .03$ ).

We also looked at correlations between the Open Science practices and the key elements of the fraud triangle (see Supplementary Materials SI "Selected Correlations"), which were, for students, negligible for pressure ( $r$  between  $.03$  and  $.04$ ), very small for opportunity ( $r$  between  $-.09$  and  $-.05$ ), and small to very small for rationalizations ( $r$  be-

tween  $-.12$  and  $-.07$ ). For supervisors, we found only very small to negligible associations ( $r$  between  $-.09$  and  $.08$ ).

## Exploratory, Non-Registered Analyses

### Communication of Data Use

We further explored whether students' perceptions of what their supervisors communicated about the data's use prior to data collection might make a difference. Out of all 473 student participants, 260 stated that their supervisor either said that they had no interest in using the data ( $n = 124$ ) or that they did not know what the supervisor wanted to do with the data ( $n = 136$ ). We coded these ("no" or "don't know") as 0, and if any form of use was communicated (e.g., as pilot data or for a scientific publication) as 1. Comparing these groups, we found that the group in which some use was communicated reported less QRP/M ( $M = 0.47$ ,  $SD = 1.08$ ) than the group in which no interest in further use was communicated or students did not know whether the data would be used ( $M = 0.75$ ,  $SD = 1.37$ ). However, the effect was rather small,  $d = 0.22$ , 95%  $CI_d$  [0.04, 0.40].

### Project Type

Moreover, we explored whether the kind of student project makes a difference for QRP/M. Descriptively, the 198 bachelor thesis projects ( $M = 0.57$ ,  $SD = 1.08$ ), 49 master thesis projects ( $M = 0.45$ ,  $SD = 1.29$ ), and 178 empirical seminar (Empra/ Expra) projects ( $M = 0.58$ ,  $SD = 1.02$ ) were quite comparable, while the QRP/M score in the 40 projects done in other seminars ( $M = 1.43$ ,  $SD = 2.33$ ) was more than 2.5 times as high. Looking at the number of involved students, we find a small positive correlation between group sizes and engagement in QRP/M ( $r = .13$ ). We further differentiated between thesis projects in general (i.e., bachelor and master theses; coded as 1,  $n = 247$ ) and other projects (coded as 0,  $n = 226$ ). These comparison groups were descriptively different from each other (thesis projects:  $M = 0.54$ ,  $SD = 1.12$ ; other projects:  $M = 0.72$ ,  $SD = 1.37$ ), yet with a very small effect size,  $d = 0.14$ , 95%  $CI_d$  [-0.04, 0.32].

## Discussion

The present research aimed to enlighten the black box of student data collection behavior, particularly those behaviors that might be considered questionable research practices or even scientific misconduct (QRP/M). We collected data from both students and supervisors to investigate the extent to which students (un)intentionally engage in QRP/M, also exploring what might drive these behaviors and how student data is used. Before going into a detailed discussion of our results, we would like to point out that the following interpretations are limited in two important ways: First, our data is correlational, thus causal claims are severely restricted and the presently detected (or

absent) associations might also be due to confounding, third variables. Second, while we descriptively compare responses from the two perspectives of students and supervisors, we cannot directly match them due to our anonymous sampling procedure. This might have also introduced some sample selection effects and our results cannot be assumed to be representative for the psychology student and supervisor population.

## How Prevalent is QRP/M in Student Data Collection?

First, and most importantly, our data provides a more detailed (albeit not representative) picture of the prevalence rates of students' data collection QRP/M and allow for a comparison with supervisors' suspicions about such QRP/M engagement. Importantly, the majority of students (64%) did not report any QRP/M. However, apparently, there are also not just "a few bad apples" committing multifaceted fraud: Most of the other students indicated only one, maybe two QRP/M (i.e., a univariate, right-skewed distribution). Many behaviors were extremely rare with prevalence rates below 1% (e.g., creating new data from scratch, duplicating data) or rare with rates below 3% (e.g., changing data, filling in missings, combining groups or items, collecting more data after analyses showed undesirable results), but there were also more noteworthy behaviors with prevalence rates ranging up to 26%.

While students rather rarely (3.38%) told their participants the hypothesis directly before starting the study, more than a quarter of students in our survey (26.00%) let participants take part in their study despite being aware that they knew the hypothesis from other conversations (e.g., friends, family, fellow students), making this by far the most common QRP/M. Depending on the respective research questions and methodologies, it could be debated how severe this specific QRP/M actually is (e.g., when using physiological methods, reaction times, or working with babies), like some participants also justified in an open text field in our study. However, demand characteristics and how participants themselves think about the assumed hypothesis is a highly influential factor in psychological research, potentially producing false positives and considerably in- and deflating effect sizes (Coles & Frank, 2023). Thus, this very high prevalence should not be judged lightly.

The second highest QRP/M reported by students (8.46%), and, noteworthy, one that supervisors seemed to not have on their radar as much (suspected: 2.73%), was students taking part in their own survey despite this being recorded as if it was a "real" participant. This QRP/M is especially problematic since it is unambiguous data fraud (i.e., data fabrication) and, arguably, one of the hardest to detect for supervisors and other researchers (e.g., no traces of data set manipulations, no "untouched" original data; see Simonsohn et al., 2023). For example, even having full access to the initially recorded data and the survey tool hosting the questionnaire would hardly allow supervisors to detect students filling out their own survey "under the participant guise" while data collection was ongoing. But, still, having access to the original data might at least help

to detect other problematic behaviors (e.g., deleting data, 4.02%—the third most prevalent QRP/M).

Interestingly, supervisors, on average, had a quite similar perception of students' data collection behaviors: The suspected prevalence rates matched the actual reported rates quite closely (with notable exceptions as discussed above). However, supervisors were quite variable in their responses (see standard deviations/errors, [Figure 1](#); this is also true for most other supervisor responses in our survey), possibly reflecting different experiences and supervision practices and goals (e.g., regarding supervising student projects mostly as obligation vs. a chance for scientific contributions). Only a third of supervisors did not suspect any QRP/M in any of their supervised student projects. Thus, dealing with or, at least, suspecting QRP/M in student projects seemed quite common. Interestingly, only in 10% of projects did supervisors have doubts about data integrity.

Most prevalence rates of students' QRP/M are in a similar range (between 1% and 3%) as other self-reported data-related QRP/M among researchers ([Fanelli, 2009](#); [Xie et al., 2021](#)) and among students ([Krishna & Peter, 2018](#)). As in previous research ([Brachem et al., 2022](#)), exploratory analyses suggest that the project type might play an important role for QRP/M, with students seeming to be particularly prone to such behaviors in seminar projects. This might be explained by these projects often involving many other students and little freedom regarding own research interests, potentially leading to decreased feelings of individual responsibility and commitment.

Crucially, we assume that the present prevalence rates are likely underestimating the real numbers. For example, presumably, only highly motivated students clicked on the survey link and filled out the questionnaire (e.g., as indicated by the low response rate in our sampling procedure). Additionally, these prevalence estimates are largely based on students from German universities within the European higher education system and might not generalize to students from other education systems.

## What Drives Students' QRP/M Behavior?

In the present project, we also investigated potential drivers of students' engagement in QRP/M. Indeed, we find that diverse experiences related to the three key elements of the fraud triangle model ([Abdullahi & Mansor, 2015](#); [Malgwi & Rakovski, 2009](#)) are positively associated with students' problematic behaviors. Thus, our findings corroborate previous research (e.g., [Brachem et al., 2022](#); [Gopalakrishna et al., 2022](#); [Krishna & Peter, 2018](#); [Moran et al., 2022](#)), also identifying pressures (e.g., time pressure or pressure to reach a certain grade), opportunities (e.g., QRP/M were unlikely to be discovered), and rationalizations (e.g., claiming to have learned it that way) as relevant drivers for QRP/M.

Comparing students' experiences and supervisors' perceptions of these, we again find that, on average, supervisors seemed to have a good intuition about how their students think and feel during their projects. Notable differences emerged mainly in experiences directly related to the student-supervisor-relationship: Students seemed to perceive their



supervisors as less involved, available, and supportive than the supervisors themselves thought their students perceived them. Note that, as mentioned above, we cannot match students with their supervisors and the present discrepancies might also be explained by sample selection effects (e.g., highly student- and teaching-oriented researchers might have been especially motivated to participate in our supervisor survey). One other important difference was that students indicated more strongly than supervisors that they [the students] thought their data would not be used (i.e., a rationalization for QRP/M). We will come back to this later.

It should be noted that, while supporting the broader theoretical idea of pressures, opportunities, and rationalizations playing a role in QRP/M, these elements are not intended—or should be treated—as distinct psychological constructs. In fact, many specific experiences assessed here likely contribute to two or more driving elements at once (e.g., time pressure might increase feelings of pressure and work as a rationalizing argument). Investigating the precise psychological mechanisms underlying the drivers of QRP/M is an important avenue for future research, and also calls for a thorough reworking of the measures applied here.

## How Does Supervisor-Student Communication Affect QRP/M?

Comparing students' and supervisors' reports of what was communicated about how the data would be used beyond the specific student project suggests a large discrepancy: Consistently, students reported much lower communicated uses than supervisors (e.g., 19% vs. 52% stated it was communicated that the data would be used for a scientific publication), with a striking 29% of students not knowing what data use was intended (vs. supervisors claiming 5% of students did not know; see also the possible rationalization regarding data use discussed above). Beyond this, almost half of the students (48%) expected no future use of their data. This reveals a problematic communication gap. Results from exploratory analyses stress potential consequences of this: Those students who knew of any data use (vs. no use) reported less QRP/M.

Further, inspecting the experiences of the supervisor-student relationship more closely, we also looked at (perceived) supervisor expectations. Students seemed to underestimate how much supervisors expect a high scientific standard in student projects (but, again, this might be due to sample differences). Interestingly, high expectations were negatively associated with QRP/M. Here, expectations regarding the quality of the data and the data collection process seem particularly influential.

These findings demonstrate the importance of explicitly communicating the priority of clean data collection in line with the highest scientific standard as data might be used beyond the students' projects. Clear supervisor-student communication is likely an important buffer against questionable or even fraudulent behaviors.

## What Role Does Open Science Play?

We further speculated that Open Science practices might decrease or, particularly during data collection, even increase (e.g., by increasing pressure to reach a certain sample size) the prevalence of QRP/M among students. However, for supervisors, implementing Open Science practices (e.g., open data and analysis scripts) was only weakly associated with less suspected QRP/M; for students this association was even smaller. Thus, we tentatively conclude that Open Science likely has no unintended negative consequences on data collection behaviors. And, maybe, it could even be a (very) small buffer for QRP/M (e.g., by reducing perceived opportunities and discounting possible rationalizations; see also Brachem et al., 2022: feeling informed about Open Science correlated negatively with QRPs).

Moreover, Open Science seems to become more and more of a standard procedure in student projects. Compared to Brachem et al.'s (2022) survey (data collected in 2018/2019), students in our survey reported more Open Science practices: Preregistrations (47% vs. 23%) were applied in almost half and power analyses (70% vs. 34%) in the majority of projects. Sharing practices (open data, materials, and analysis scripts) were less common (12%–19%); arguably, because these files are often only made available upon publication.

## How Is Student Data Used?

In line with the speculation that low sharing practices might be due to low publication rates of student data, supervisors indeed indicated that only 16% of student projects ended up as part of a scientific paper submission. But, otherwise, it was not uncommon for supervisors to further use student data: On average, supervisors had used data from almost half of the projects in some way for their own work (44%), which fits the 40% of students who stated that their supervisor had used their data. Still, the perceived eligibility for further use seems to reflect some—albeit, not grave—concerns: Supervisors perceived student data as moderately to highly eligible—even a bit higher than the students themselves.

## Conclusion

Opening up the black box of student data collection reveals that some questionable or even fraudulent data collection behaviors are not uncommon among students. For example, many students let participants take part in their study despite being aware that they know the hypothesis, some students even participate in their own surveys. However, most students reported *not* having engaged in any of the QRP/M we listed. And, so far, only relatively few student projects seem to end up in the published literature. Thus, supervisors should maybe consider how empirical student projects may not only be an opportunity for teaching but also for research. To gather high-quality student data,

university education, particularly research-focused teaching, should work on reducing perceived pressures, opportunities, and rationalizations for problematic behaviors. This might even mean changing the curriculum and its underlying incentive structure regarding how we train, test, and grade our students (e.g., collaborative replication projects instead of individual empirical projects, [Button et al., 2020](#); [Creaven et al., 2021](#)). Making Open Science a central element of teaching could provide fertile ground for promoting good scientific practices, including appropriate data collection behaviors ([Pownall et al., 2023](#)). In this spirit, and maybe most importantly, students and supervisors need to communicate more clearly and transparently about their expectations and experiences in student projects, discussing how student data might seriously contribute to the scientific discourse.

---

**Funding:** This research is funded by a grant awarded by the German Research Foundation (Deutsche Forschungsgemeinschaft DFG) under the Priority Program "META-REP" (project no. 467852570).

---

**Acknowledgments:** The authors have no additional (i.e., non-financial) support to report.

---

**Competing Interests:** The authors have declared that no competing interests exist.

---

**Author Contributions:** *Tobias Ludwig*—Idea, conceptualization | Design planning | Data management (storage, curation, processing, etc.) | Data analysis | Writing | Feedback, revisions. *Marlene Sophie Altenmüller*—Idea, conceptualization | Design planning | Data management (storage, curation, processing, etc.) | Data analysis | Data collection | Visualization (data presentation, figures, etc.) | Writing | Feedback, revisions | Project coordination, administration | Supervision, mentoring. *Leonhard Falk Florentin Schramm*—Idea, conceptualization | Design planning | Feedback, revisions. *Mathias Twardawski*—Idea, conceptualization | Design planning | Data collection | Feedback, revisions | Project coordination, administration | Supervision, mentoring.

---

**Data Availability:** For this article, data is freely available ([Ludwig et al., 2022](#)).

---

## Supplementary Materials

For this article, the following supplementary materials are available (for access see [Index of Supplementary Materials](#) below):

- All study materials (student and supervisor questionnaires), other supplementary information (selected correlations, detailed study procedure), the analysis scripts and anonymized data (see [Ludwig et al., 2022](#))
- The registration of the study plan (Stage 1 manuscript) after in-principle acceptance (IPA) of the current Registered Report (see [Altenmüller et al., 2023](#))

## Index of Supplementary Materials

- Altenmüller, M. S., Ludwig, T., Schramm, L. F. F., & Twardawski, M. (2023). *Evading Open Science: The black box of student data collection* [Preregistration]. OSF Registries. <https://doi.org/10.17605/osf.io/8n5tx>
- Ludwig, T., Altenmüller, M. S., Schramm, L. F. F., & Twardawski, M. (2022). *Evading Open Science: The black box of student data collection* [Data, Analysis, Materials]. OSF. <https://osf.io/8s6x3/>

## References

- Abdullahi, R., & Mansor, N. (2015). Fraud triangle theory and fraud diamond theory: Understanding the convergent and divergent for future research. *International Journal of Academic Research in Accounting Finance and Management Sciences*, 5(4), 54–64. <https://doi.org/10.6007/IJARAFMS/v5-i4/1823>
- Agnoli, F., Wicherts, J. M., Veldkamp, C. L. S., Albiero, P., & Cubelli, R. (2017). Questionable research practices among Italian research psychologists. *PLoS One*, 12(3), Article e0172792. <https://doi.org/10.1371/journal.pone.0172792>
- Brachem, J., Frank, M., Kvetnaya, T., Schramm, L. F. F., & Volz, L. (2022). Replikationskrise, p-hacking und Open Science: Eine Umfrage zu fragwürdigen Forschungspraktiken in studentischen Projekten und Impulse für die Lehre. *Psychologische Rundschau*, 73(1), 1–17. <https://doi.org/10.1026/0033-3042/a000562>
- Brimble, M., & Stevenson-Clarke, P. (2005). Perceptions of the prevalence and seriousness of academic dishonesty in Australian universities. *Australian Educational Researcher*, 32(3), 19–44. <https://doi.org/10.1007/BF03216825>
- Bühner, M. (2023). Zur Lage der Psychologie. *Psychologische Rundschau*, 74(1), 1–20. <https://doi.org/10.1026/0033-3042/a000616>
- Button, K. S., Chambers, C. D., Lawrence, N., & Munafò, M. R. (2020). Grassroots training for reproducible science: A consortium-based approach to the empirical dissertation. *Psychology Learning & Teaching*, 19(1), 77–90. <https://doi.org/10.1177/1475725719857659>
- Coles, N. A., & Frank, M. C. (2023). *A quantitative review of demand characteristics and their underlying mechanisms*. PsyArXiv. <https://doi.org/10.31234/osf.io/uw85a>
- Creaven, A.-M., Button, K., Woods, H., & Nordmann, E. (2021). *Maximising the educational and research value of the undergraduate dissertation in psychology*. PsyArXiv. <https://doi.org/10.31234/osf.io/deh93>
- Cressey, D. R. (1953). *Other people's money: A study of the social psychology of embezzlement*. Free Press.
- Deutsche Forschungsgemeinschaft. (2022). *Guidelines for safeguarding good research practice: Code of conduct*. <https://doi.org/10.5281/ZENODO.6472827>
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*, 4(5), Article e5738. <https://doi.org/10.1371/journal.pone.0005738>

- Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological & Personality Science*, 7(1), 45–52. <https://doi.org/10.1177/1948550615612150>
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168. <https://doi.org/10.1177/2515245919847202>
- Gollwitzer, M., Abele-Brehm, A., Fiebach, C. J., Ramthun, R., Scheel, A., Schönbrodt, F., & Steinberg, U. (2021). Management und Bereitstellung von Forschungsdaten in der Psychologie: Überarbeitung der DGPs-Empfehlungen. [English translation: <https://doi.org/10.31234/osf.io/24ncs>]. *Psychologische Rundschau*, 72(2), 132–146. <https://doi.org/10.1026/0033-3042/a000514>
- Gopalakrishna, G., ter Riet, G., Vink, G., Stoop, I., Wicherts, J. M., & Bouter, L. M. (2022). Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands. *PLoS One*, 17(2), Article e0263023. <https://doi.org/10.1371/journal.pone.0263023>
- Hard, S. F., Conway, J. M., & Moran, A. C. (2006). Faculty and college student beliefs about the frequency of student academic misconduct. *The Journal of Higher Education*, 77(6), 1058–1080. <https://doi.org/10.1353/jhe.2006.0048>
- 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>
- 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), Article e0203470. <https://doi.org/10.1371/journal.pone.0203470>
- Leiner, D. J. (2022). *SoSci Survey* [Computer software]. <https://www.soscisurvey.de>
- Loenneker, H. D., Huber, J. F., Artemenko, C., Heller, J., & Nuerk, H.-C. (2022). Realitätscheck Open Science in der universitären Lehre. *Psychologische Rundschau*, 73(1), 47–49. <https://doi.org/10.1026/0033-3042/a000577>
- Malgwi, C. A., & Rakovski, C. (2009). Behavioral implications of evaluating determinants of academic fraud risk factors. *Journal of Forensic & Investigative Accounting*, 1(2), 1–37.
- Moran, C., Richard, A., Wilson, K., Twomey, R., & Coroiu, A. (2022). I know it's bad, but I have been pressured into it: Questionable research practices among psychology students in Canada. *Canadian Psychology/Psychologie Canadienne*, 12–24. <https://doi.org/10.1037/cap0000326>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73(1), 719–748. <https://doi.org/10.1146/annurev-psych-020821-114157>
- Nuijten, M. B., Borghuis, J., Veldkamp, C. L. S., Dominguez-Alvarez, L., van Assen, M. A. L. M., & Wicherts, J. M. (2017). Journal data sharing policies and statistical reporting inconsistencies in psychology. *Collabra. Psychology*, 3(1), Article 31. <https://doi.org/10.1525/collabra.102>

- Pownall, M., Azevedo, F., König, L. M., Slack, H. R., Evans, T. R., Flack, Z., Grinschgl, S., Elsherif, M. M., Gilligan-Lee, K. A., de Oliveira, C. M. F., GJoneska, B., Kalandadze, T., Button, K., Ashcroft-Jones, S., Terry, J., Albayrak-Aydemir, N., Dëchtërenko, F., Alzahawi, S., Baker, B. J., ...FORRT. (2023). Teaching open and reproducible scholarship: A critical review of the evidence base for current pedagogical methods and their outcomes. *Royal Society Open Science*, 10(5), Article 221255. <https://doi.org/10.1098/rsos.221255>
- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Wiley series in survey methodology* (pp. 1–22). John Wiley & Sons. <https://doi.org/10.1002/0471654728.ch1>
- Rajah-Kanagasabai, C. J., & Roberts, L. D. (2015). Predicting self-reported research misconduct and questionable research practices in university students using an augmented Theory of Planned Behavior. *Frontiers in Psychology*, 6, Article 535. <https://doi.org/10.3389/fpsyg.2015.00535>
- R Core Team. (2023). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Schönbrodt, F. D., Maier, M., Heene, M., & Bühner, M. (2018). Forschungstransparenz als hohes wissenschaftliches Gut stärken: Konkrete Ansatzmöglichkeiten für Psychologische Institute. *Psychologische Rundschau*, 69(1), 37–44. <https://doi.org/10.1026/0033-3042/a000386>
- 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>
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2023). Data Falsificada (Part 1–4). *Data Colada*. <https://datacolada.org/109>
- Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics*, 12(1), 53–74. <https://doi.org/10.1007/s11948-006-0006-y>
- Stürmer, S., Oeberst, A., Trötschel, R., & Decker, O. (2017). Early-career researchers' perceptions of the prevalence of questionable research practices, potential causes, and open science. *Social Psychology*, 48(6), 365–371. <https://doi.org/10.1027/1864-9335/a000324>
- Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K., & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, 8(1), Article 192. <https://doi.org/10.1038/s41597-021-00981-0>
- United States Department of Health and Human Services Office of Research Integrity. (n.d.). *Definition of research misconduct*. <https://ori.hhs.gov/definition-research-misconduct>
- Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, 7, Article 1832. <https://doi.org/10.3389/fpsyg.2016.01832>

- Willis, G. B. (1999). *Cognitive interviewing, a "how to" guide-reducing survey error through research on the cognitive and decision processes in surveys* [Course]. Meeting of the American Statistical Association, Research Triangle Institute. Retrieved August 24, 2023, from <https://www.hkr.se/contentassets/9ed7b1b3997e4bf4baa8d4eceed5cd87/gordonwillis.pdf>
- Xie, Y., Wang, K., & Kong, Y. (2021). Prevalence of research misconduct and questionable research practices: A systematic review and meta-analysis. *Science and Engineering Ethics*, 27(4), Article 41. <https://doi.org/10.1007/s11948-021-00314-9>

**PSPS**  
Polish  
Social  
Psychological  
Society

*Social Psychological Bulletin* (SPB)  
is an official journal of the Polish  
Social Psychological Society  
(PSPS).



leibniz-psychology.org

PsychOpen GOLD is a publishing  
service by Leibniz Institute for  
Psychology (ZPID), Germany.