Journal of Scientific Exploration
Anomalistics and Frontier Science

RESEARCH ARTICLE

# Strategies for Teaching Critical Thinking about Replication Based on A Study of Introductory Psychology Textbooks

**Jeffrey S. Reber**
University of West Georgia
jreber@westga.edu
orcid.org/0000-0002-3528-124X

**Debra L. Katz**
California institute of Human Sciences

**Jacob H. Tengelsen**
Northwest Missouri State University

**Peder Schillemat**
University of West Georgia

SUBMITTED July 6, 2024
ACCEPTED August 22, 2024
PUBLISHED October 15, 2025

https://doi.org/10.31275/20253485



## ABSTRACT

**Background:** The replication crisis highlights the importance of teaching students scientific and critical thinking. Replication relevant content in the Introduction to Psychology textbook (IPT) can be a valuable resource to instructors interested in this skillset.
**Objective:** To suggest strategies instructors could use to teach scientific and critical thinking based on an examination of replication relevant content in IPTs.
**Method:** Researchers and raters analyzed the treatment of replication and the replication crisis in 12 popular IPTs. They also analyzed authors' descriptions of 30 highly cited replicable studies.
**Results:** Results showed that 11 of the IPTs discussed replication, and half addressed the replication crisis. IPT authors rarely mentioned the replication status of studies they reviewed, but they often used language indicating greater confidence in the validity and reliability of the results, indiscriminate of that status.
**Conclusion:** The treatment of replication in IPTs provides a launching point from which to engage students in activities and tasks that can help them appreciate and think critically about this important pillar of science.
**Teaching Implications:** Scientific and critical thinking can be developed by instructors who use the IPT to educate students about replication, the replication crisis, and the replication status of studies.

## KEYWORDS

replication, replication crisis, critical thinking, introduction to psychology.

## STRATEGIES FOR TEACHING CRITICAL THINKING ABOUT REPLICATION BASED ON INTRODUCTORY PSYCHOLOGY TEXTBOOKS

More than just addressing topics in anomalistics, this *Journal* has regularly published papers about important methodological and analytical issues within broader academia and science (Dobyns et al, 2004; Gernert, 2010; Katz, 2022; Scargle, 2000; Sheldrake, 1998; Sturrock, 1994). This paper follows this tradition by focusing on the topic of replication, which is often touted as the cornerstone of scientific integrity and progress (e.g., National Academies of Sciences, Engineering, Medicine, 2019; Holden & Goodwin; 2018) though its exact definition and relevance are hotly debated (Feest, 2019; Iso-Ahola, 2020; Nosek & Errington, 2020). Indeed, not all questions or phenomena readily lend themselves to repeated observations or testing (see e.g., Stevenson, 1999). Nevertheless, replications generally involve the process of conducting successive studies, following the same or improved methods, to verify the results obtained in prior research. This practice is

especially important for controversial studies or maverick theories due to the following reasons:

1. **Confirmation of Findings:** Controversial studies often challenge established knowledge or present surprising results. Replication is necessary to confirm these findings and ensure they are not outliers or the result of methodological errors. Without replication, it is difficult to discern whether the original results are reliable.
2. **Reducing Bias and Error:** Controversial topics can be subject to heightened scrutiny and skepticism, which increases the likelihood of bias, both in favor of and against the findings. Replication helps identify and mitigate these biases, ensuring that the results are not influenced by individual researchers' expectations or methodological flaws.
3. **Building Consensus:** Scientific consensus is not achieved through a single study but through a body of evidence that consistently supports a particular finding. In controversial areas, replication is crucial for building this consensus, as repeated confirmation of results helps to establish a shared understanding and acceptance within the scientific community.
4. **Informing Policy and Public Opinion:** Controversial studies often have significant implications for public policy, health, and societal norms. Reliable, replicated results are essential for informing policy decisions and guiding public opinion. Without replication, policies based on controversial findings may be misguided or harmful.
5. **Enhancing Credibility:** Controversial findings are often met with skepticism. Replication by independent researchers enhances the credibility of these findings, demonstrating that they are not the result of a single laboratory's methods or biases but are generalizable and robust.
6. **Encouraging Scientific Rigor:** Knowing that controversial findings will be subject to replication encourages researchers to use the most rigorous methods possible. This helps to ensure that the initial study is conducted with high standards, reducing the likelihood of errors and increasing the reliability of the findings.
7. **Clarifying Ambiguities:** Controversial studies may produce ambiguous or mixed results. Replication helps clarify these ambiguities by providing additional data points and perspectives. This can lead to a better understanding of the conditions under which the findings hold true or fail.
8. **Public Trust in Science:** Public confidence in scientific research hinges on its reliability and honesty. Replication builds trust by demonstrating that scientific findings are not just isolated incidents but are reproducible and reliable truths. In an era where misinformation can spread rapidly, demonstrating the replicability of scientific studies is crucial for maintaining public trust.

A number of these reasons were tested following a 2011 publication by social psychologist Daryl Bem that presented controversial findings from a series of experiments that demonstrated the retroactive influence of future events on current behaviors. Bem's article prompted several attempts to directly replicate the results of his studies (see Korbmacher et al., 2023). To encourage such replications, Bem offered "free, comprehensive packages that included detailed instruction manuals for conducting the experiments, computer software for running the experimental sessions, and database programs for collecting and analyzing the data" (Bem et al, 2016, p. 1). Within a period of two years, Bem and colleagues gathered the results of 90 experiments from 33 different laboratories located in 14 different countries. They conducted a meta-analysis on those results that confirmed Bem's earlier findings, noting that "A Bayesian analysis yielded a Bayes Factor of $5.1 \times 10^9$, greatly exceeding the criterion value of 100 for 'decisive evidence' in support of the experimental hypothesis." (Bem et al, 2016, p. 1).

Other researchers' efforts to reproduce Bem's findings, however, proved to be as anomalous as were the retro-causal effects he was studying. Indeed, a large number of attempts to reproduce Bem's results, along with the results of many other controversial studies, consistently failed to repeat the findings of the original studies in just from as many as one-third (e.g., Camerer et al., 2018) to one-half of the attempts (e.g., Open Science Collaboration, 2015). These replication failures occurred even when large sample sizes were used. As researchers looked deeper into these replication failures, they identified trends of "underpowered studies, publication bias, imprecise theories, and inadequate statistical procedures" (Lewandowsky & Oberauer, 2020, p. 1). Wiggins and Christopherson (2019) also note that, "the widespread use of so-called questionable research practices . . . likely dramatically inflate[d] the false positive rate in the psychological literature" (p. 204).

Though not intended, Bem's 2011 study of an anomalous phenomenon initiated what is now widely known as the replication crisis in psychology (Lewandowsky & Oberauer, 2020; Youyou et al., 2023), a crisis that has proven to be critical to psychology's status and development as a scientific discipline. This crisis demonstrates the importance of close adherence to the key principles and practices of empirical research in designing and carrying out original research and conducting and publishing replications. It also shows how psychologists have practiced scientific and critical thinking to uncover issues in psychological research studies that have made replication difficult to achieve and/or to publish, including issues attending the fit of the empirical method and replication to psychological phenomena that may not readily or easily lend themselves to such approaches. This critical thinking has resulted in a number of important positive reforms in replication research (Korbmacher et al., 2023; Shrout & Rodgers, 2018). Ultimately, the replication crisis, rather than dealing a fatal blow to the discipline, demonstrates how scientific and critical thinking can course correct a discipline and improve its knowledge advancement.

This course correction certainly happens in the psychology lab and in the editorial offices of academic journals and textbooks, but it can also take place in the psychology classroom, including, perhaps most importantly, in the Introduction to Psychology classroom. Few students who enter this course would have a comprehensive understanding of the replication crisis in psychology, especially its fascinating origin story in anomalistic research. Fewer still would appreciate the reasons cited above for valuing replication in the context of controversial findings and would likely not be capable of applying scientific and critical thinking skills to such findings. These students need instruction and guidance in the principles of replication in order to understand its importance to psychology generally, but also to know how to think carefully and critically about surprising and controversial results, like those often found in anomalistic studies. The purpose of this article is to suggest several strategies that Introduction to Psychology course instructors can implement to teach students to think scientifically and critically about replication that build upon the treatment of replication and the replication crisis in Introduction to Psychology textbooks (IPT).

## THE INTRODUCTION TO PSYCHOLOGY TEXTBOOK

In 2016, a working group of educators under the direction of the American Psychological Association's Board of Educational Affairs highlighted the popularity and importance of the Introductory Psychology course (Gurung et al., 2016), falling just behind English Composition, with somewhere between 1.2 and 1.6 million students enrolling in the course each year. The group noted that for many of these students, this course will likely be "their only formal exposure to psychological science" (p. 112), making it "psychology's most influential course" (p. 113). The most common feature of the introductory psychology course is the IPT, with as many as 98% of introductory classes requiring an IPT (Miller & Gentile, 1998). The 2016 working group observed that the curriculum of the introductory course tends to closely follow IPT content and instructors rarely assign materials, or cover topics, beyond the textbook. Generally, they also lecture on topics in proportion to the amount of coverage the topics receive in these books (Gurung et al., 2016). Thus, IPT content plays a significant role in the course curriculum and students' perceptions and understanding of psychological science. Collisson et al. (2015) found that even the number of citations in an IPT can influence student perceptions, noting that "psychology was perceived as less scientific, and its findings less agreed upon when [IPTs] contained fewer citations" (p. 6). This suggests that students are attentive to psychology's status as a science, and the content of the IPT, including studies cited and described in the IPT, can enhance or detract from that status.

The IPT is also an important tool for teaching students the core competencies of scientific and critical thinking. This skillset has been identified by many psychologists as essential for students taking psychology courses (e.g., Stanovich, 2007; Sternberg & Halpern, 2020). McBurney (1996) asserts that the chief purpose of psychology courses is to teach students "to apply the same critical thinking skills to human behavior that scientists do" (pp. vii-viii). Many IPT authors have built critical thinking activities and tasks into their textbooks, and a number of psychologists have written critical thinking handbooks to supplement the critical thinking content in IPTs (e.g., Ruscio, 2006; Slife et al., 2005; Sternberg & Halpern, 2020). The APA's Introductory Psychology Initiative (2023) has also included scientific inquiry and critical thinking as one of its key goals for student development. The IPT, then, plays an essential role in educating students not only about the content of psychology but also in educating students about how psychologists use scientific methods and critical thinking skills to investigate that content.

## REPLICATION AND THE IPT

Replication is an important feature of scientific and critical thinking for psychology and other sciences (Fidler

& Wilcox, 2021; Schmidt, 2009). A recent editorial (*Nature*, 2021) reminds journal readers of replication's foundational role in science, noting that "the entire scientific community must recognize that replication is not for replication's sake, but to gain an assurance central to the progress of science: that an observation or result is sturdy enough to spur future work" (para. 13). Despite its importance, research by Anglin and Edlund (2019) found that psychology course instructors tend to discuss replication in their courses only briefly, if at all, and this is particularly true in the Introduction to Psychology course. This is likely a consequence of the limited amount of textbook space and classroom time that instructors can dedicate to any topic in a content-heavy survey course. Anglin and Edlund support this conclusion, finding that instructors of content courses spent less time on replication and related topics than instructors of research methods and statistics courses. As a corrective measure, they suggest that instructors consider replication less a topic of content and more a tool for developing students' "critical scientific thinking skills" (p. 72). The replication crisis is particularly apropos of this purpose and could be used as a case study along with other replication-relevant IPT content to help Introduction to Psychology instructors teach this important skillset.

To that end, knowing the typical treatment of replication in introductory textbooks would be helpful. A few relevant preliminary inquiries can be found. Schmidt (2009), for example, looked at the treatment of replication across a variety of disciplinary materials, including textbooks. He concluded that despite the importance and value placed on replication, it is infrequently addressed in the literature. More recently, Griggs and Whitehead (2014, 2015) reviewed Introductory Social Psychology textbooks and found that the replication status of some famous social psychological studies (e.g., Milgram's obedience study) was noted, but the textbook authors did not discuss methodological issues that impacted the replicability of these studies. Diener and Biswas-Diener (2024) suggest that replication issues with these famous studies, as well as with many lesser-known studies, may also be omitted or misrepresented in IPTs.

These findings, though suggestive, provide little substantial information about the treatment of replication and the replication crisis in IPTs. This study is designed to provide Introduction to Psychology course instructors a better understanding of the general treatment of replication and the replication crisis in and across commonly used IPTs, so instructors can build lesson plans and activities based on that IPT content that will help their students practice and develop their scientific and critical thinking competencies. Three research questions guided the investigation. First, do IPT authors discuss replication, and if so, how do they typically teach it? Second, do IPT authors bring up the replication crisis, and if so, what do they generally say about it? Finally, when IPT authors describe replicable studies, do they note the replication status of the studies, and do they use language that expresses greater or lesser confidence in the validity and reliability of the findings according to the replication status of the studies? The answers to these questions, depending on their consistency, might suggest a number of specific strategies that instructors could take to teach replication and promote scientific and critical thinking.

## METHODS

### Textbook Selection And Data Preparation

Researchers first compiled a comprehensive and representative list of the most often used IPTs using published reports of top-selling books, following a precedented procedure (see Chandler, 2019; Warne et al., 2018). Selected texts were included in lists of the top 30 IPTs for both Amazon and Google. To be included, IPTs were published between 2017 and 2018. There are two reasons for our focus on these dates. First, we initiated this research in 2018, a time when the replication crisis was receiving a lot of public attention and psychology textbooks would likely be most sensitive to replication and the issues attending it. As a check on the current popularity of the selected IPTs, we reviewed the best-selling lists again in 2024 and reconfirmed that these editions remain among the top-selling IPTs on Amazon and Google. The second reason for limiting the inclusion of IPT editions to the same two years is that we wanted to maintain a consistency of comparison across textbooks. This is also the reason why open-source textbooks were excluded. We wanted a sample that was homogeneous with regard to time of publication and format. A final criterion for inclusion was that the IPTs had to be available via eBook or as searchable online versions. Researchers queried publishers as to the availability of electronic, searchable copies of the textbooks. Table 1 lists the 12 IPTs that appeared most often at the top of these reports and were made available to researchers by textbook publishers.

**Table 1.** Twelve Top Selling IPTs.

| Author(s) | Year | Trials | Hits | Hits % |
|---|---|---|---|---|
| Ciccarelli & White | 2017 | Psychology: An Exploration | 5th | Pearson |
| Coon et al. | 2018 | Introduction to Psychology: Gateways to Mind and Behavior | 15th | Cengage Learning. |
| Feldman | 2018 | Understanding Psychology | 14th | McGraw-Hill Education. |
| Griggs | 2017 | Psychology: A Concise Introduction | 5th | Worth Publishers |
| Hockenbury & Nolan | 2018 | Psychology | 8th | Worth Publishers |
| Kalat | 2017 | Introduction to Psychology. | 11th | Cengage Learning |
| Lilienfeld et al. | 2017 | Psychology: From Inquiry to Understanding | 4th | Pearson |
| Myers & DeWall | 2018 | Exploring Psychology in Modules | 11th | Worth Publishers |
| Rathus et al. | 2017 | Psych: Introduction to Psychology. | 3rd | Cengage Learning |
| Schacter et al. | 2018 | Psychology | 4th | Worth Publishers |
| Wade et al. | 2017 | Invitation to Psychology | 7th | Pearson |
| Weiten | 2017 | Psychology: Themes and Variations | 10th | Cengage Learning |

Criteria for the selection of articles to evaluate in and across the textbooks included that they had to have been cited in at least 3 of the 12 IPTs, and they had to be studies that had the potential to be replicated, thus eliminating historical, biographical, or theoretical articles. To assess the author's treatment of replication, researchers conducted a search of the body of each IPT using the term "replication" and any of its roots and derivatives to identify where replication was mentioned by the author. A similar search was conducted using the combination of the terms "replication" and "crisis" and their synonyms. Passages that contained these words were copied and pasted into a spreadsheet.

## Treatment of Replication, The Replication Crisis, and Replication Status

Researchers conducted a content analysis of all IPT excerpts addressing replication and the replication crisis. The analysis included examining how much attention was generally dedicated to these topics by IPT authors and discerning any common and differing themes that emerged from the texts. To examine the replication status of studies cited in IPTs, a sample of 1/3 of the articles that were original replicable studies and were referenced in at least 3 of the 12 textbooks was randomly selected for analysis. To ascertain any impact of replication status on the IPT authors' descriptions of the studies, researchers first determined whether the replicable studies had been directly or conceptually replicated. Direct replications repeat the protocols of the original study but with a different sample. Conceptual replications use a different methodology to test the same hypothesis as the original study (see Nosek & Errington, 2017).

Researchers determined whether the sample of studies had been directly or conceptually replicated using Google Scholar (see Zientek et al., 2018). After locating each sampled study in Google Scholar, researchers selected the option to view all available published articles, dissertations, and theses that cited the study. This provided a list of hundreds and, in some cases, thousands of sources. Researchers then used Google Scholar's advanced search to examine the text of each source to see if it clearly qualified or disqualified as being either a direct or conceptual replication. As a matter of quality control, the sample articles were divided evenly among researchers for examination, with six of the articles assigned to more than one examiner to ensure the consistency of results.

## Descriptions of Replicable Studies

Researchers recruited two psychology graduate students, one in the PhD program and one in the MA program at UWG, who volunteered to serve as raters. The raters were blind to the purposes of the research. They only knew that they would be examining textbook excerpts in which the authors described psychological research. Two researchers conducted a training session with the raters in which they walked them through sample excerpts and showed them how to identify and distinguish between explicit, implied, and non-reported references to the replication status of the studies following previously documented coding protocols (e.g., Makel et al., 2012). An example of an explicit mention that was used to train the raters was taken from the Rathus et al. (2017) IPT, which reads, "The Schachter

and Singer study has been replicated, but with different results…" (p. 225). An example excerpt that was used to train raters on identifying implied references to replication was taken from Myers and DeWall's (2018) IPT, who, after describing a study by Savic et al. (2005), wrote that "Other studies of brain responses to sex-related sweat odors and to pictures of male and female faces have found similar gay-straight differences" (p. 412).

Raters were also trained to examine the extent to which IPT authors used language in their descriptions of the studies that indicated strong confidence in the reliability and validity of the study findings. Recall that successful replications increase confidence in the knowledge gained from a study, whereas the knowledge gained from unreplicated studies is more tentative (National Academies of Sciences, Engineering, and Medicine, 2019). Statements that generalize findings beyond the sample and use more factual verbs, like "these findings show that people are. . ." indicate strong confidence in the veracity of the findings, whereas a phrase signifying weaker confidence would avoid inferences to the population and use tentative verbs, like "these findings suggest that some people who are like participants in this study's sample, may. . .". One IPT excerpt used in training raters that illustrates confident language reviews the results of an oft-cited study of short-term memory (Peterson & Peterson, 1959), after which the IPT authors (Wade et al., 2017) conclude that "when people are prevented from rehearsing, the contents of their working memories quickly fade" (p. 239). An illustration of less confident language is used by Schacter et al. (2018) in citing the results of the same study. They draw the more reserved conclusion that "these results suggest that information can be held in the short-term memory store for about 15 to 20 seconds" (p. 222).

Raters practiced examining IPT excerpts for explicit mentions of replication, implied mentions, or no mention of replication, as well as indications of direct and/or conceptual replications in the study description, and language connoting strong confidence in the reliability and validity of the findings, until they achieved consistent assessments, with little variation, across multiple examples. Following their training, raters were provided with access to a Google form that presented each excerpt, followed by questions that asked for the raters to indicate if the study description made explicit reference, implied reference, or no reference to replication, whether the reference, if present, implicated direct, conceptual replication, or both, and whether the language used in describing the study indicated strong confidence in the validity and reliability of the study findings. Each rater independently and anonymously evaluated each excerpt.

Excerpts were assigned to raters in a fully crossed design, allowing for systematic rater bias to be identified and controlled. As the data was nominal, Cohen's Kappa was selected as the IRR measure. Following McHugh (2012), only Kappas at .60 and above would be considered adequate to place confidence in rater agreement. If Kappas met this cutoff for each measure, discrepancies between raters would be resolved by researchers who would identify systematic bias (i.e., patterns of ratings that were inconsistent with training) and then confirm the presence of the bias with the relevant rater. Any scores from a single rater identified and confirmed as biased would be excluded from further analysis. If both raters' scores on an excerpt were inconsistent with training, then they would rescore the excerpt following retraining by the researchers.

## RESULTS

### Mentions of Replication

Researchers found that in 11 of the 12 IPTs, replication was usually discussed in a section of the first chapter that reviewed the scientific method. On average, IPT authors wrote approximately three paragraphs on the topic of replication, but that quantity ranged somewhat widely. One IPT (Coon et al., 2022) mentioned replication only one time in the entire text in one reference to a cited study. Another IPT (Kalat, 2017) discussed replication in detail in a two-page section. In reviewing replication, almost all the authors describe its basic assumptions, how it is conducted directly and conceptually, and how it strengthens confidence in the original study findings. Statements about replication like this one from Lilienfeld et al. (2017) were common: "Replicability means that a study's findings can be duplicated consistently. Replication is the cornerstone of a dependable science" (p.26). Ciccarelli and White's (2017) treatment of replication was also representative of most IPT authors: "If others can replicate your research (meaning, do exactly the same study over again and get the same results), it gives much more support to your findings" (p. 20). Hockenbury and Nolan (2018) reflect a common concern expressed by IPT authors about studies that have not been successfully replicated, stating, "If the replication of a study fails to produce the same basic findings, confidence in the original findings is reduced" (p. 18).

## Mentions of the Replication Crisis

The replication crisis was discussed in six of the 12 IPTs, typically in a paragraph or two immediately following the general discussion of replication in the scientific method section of the book. In these reviews of the crisis, authors typically acknowledge concerns with replication in the discipline. Ciccarelli and White (2017), for example, state that "some evidence suggests that editors of peer-reviewed journals have tended to publish positive research results overall rather than embrace direct replications or 'old' knowledge or negative results studies" (p. 22). Kalat (2017) similarly acknowledges that "in spite of the agreed importance of replication, not many psychologists try to replicate someone else's findings" (p. 47). A number of IPT authors cite the Open Science Collaboration (2015) study's failure to replicate 64 of 100 studies and other similar replication studies (e.g., Ciccarelli & White, 2017; Myers & DeWall, 2018). IPT authors who discuss the crisis also regularly cite concerns about researchers misrepresenting their findings and call for "more replications and greater transparency as researchers increasingly disclose their detailed methods and data" (Myers & DeWall, 2018, p.20)." Interestingly, none of the authors discussed the origin of the crisis in 2011 and the role of Bem's controversial research findings in it.

Some IPT authors note the positive changes that the replication crisis helped to instigate within the discipline. Kalat (2017), for example, writes that "More and more scientists have been calling for attempts to replicate a result, using exactly the same procedure as the original study (Simons, 2014), and several journals have agreed to publish the results, whatever they may be" (p. 47). However, by and large, the IPT authors who address the crisis downplay its severity. Some diffuse the crisis by noting that "the same problem occurs in other fields also" (Kalat, 2017, p. 47). Others state that the media and a number of psychologists have exaggerated the issues and focused too narrowly and unfairly on psychology (e.g., Myers & DeWall, 2018). Authors also point to weaknesses in recent studies that have attempted to replicate some of the most well-known studies in the discipline, casting doubt on the credibility of the attempted replications (Hockenbury & Nolan, 2018; Kalat, 2017; Rathus et al., 2017). Myers and DeWall (2018) take this last point to a meta-level, asserting that those studies that have attempted to replicate previous studies must themselves be replicated before any confidence can be placed in their findings.

## The Replication Status of Highly Cited Studies

Researchers identified all the studies that were cited by no fewer than three IPT authors and then screened those studies as to their capacity to be replicated. This resulted in a total of 90 replicable articles. One-third of those studies were randomly selected and analyzed to determine whether they had been directly and/or conceptually replicated. Table 2 lists the author(s) and year of each study and indicates its replication status. Researchers found that 10 of the 30 studies had been directly replicated, and 23 of the 30 studies had been conceptually replicated (note that these are not mutually exclusive categories). Four of the studies had not been replicated either directly or conceptually.

Table 2 also shows the number and ratings of excerpts across the 12 IPTs in which the authors cited and described the results of the 30 selected studies. It is important to note that in no case were the scores from both raters on an excerpt identified and confirmed as biased, so all excerpts are represented by at least one rating that was consistent with training. Ratings indicated that 39 of the 172 excerpts (23%; Cohen's Kappa = .70) used language that in some way implicated the replication status of the studies, with 9% of the excerpts explicitly indicating replication status (Cohen's Kappa = .80). The raters identified language connoting strong confidence in the reliability and validity of the findings in 81% of the excerpts (Cohen's Kappa = .62).

Only four of the thirty sampled studies have not yet been directly or conceptually replicated (Archer, 2004; Gottesman, 1991; Hirst et al., 2015; Trahan et al., 2014). This speaks well to the practice and prevalence of replication in psychology, at least for more highly cited studies. However, none of the IPT authors explicitly noted the replication status of these unreplicated studies, though the raters did detect implied references to replication in two of the study descriptions. The ratings also indicate that most of the IPT authors who cited and described these non-replicated studies used language connoting strong confidence in the veracity and generalizability of the findings in their descriptions. On the other hand, raters detected three studies that have been at least directly replicated (Caspi et al., 2002; Ferguson, 2015; Freedman & Fraser, 1966) that were not explicitly or implicitly identified by IPT authors as having been replicated and that most IPT authors described using more tentative language (i.e., language connoting weaker confidence in the veracity of the results).

**Table 2.** Replication Status of Thirty Studies.

| Article | Replication Status | No. of IPT excerpts citing the study | No. of IPT excerpts explicitly mentioning replication status | No. of IPT excerpts implying replication status | No. of IPT excerpts with no indication of replication status | No. of IPT excerpts using confident language |
|---|---|---|---|---|---|---|
| Allport & Odbert (1936) | Conceptual | 5 | 0 | 3 | 2 | 5 |
| Archer (2004) | Not Replicated | 4 | 0 | 0 | 4 | 3 |
| Asch (1956) | Direct & Conceptual | 6 | 3 | 2 | 1 | 5 |
| Bandura (1986) | Conceptual | 8 | 1 | 3 | 4 | 8 |
| Bartlett (1932) | Direct & Conceptual | 7 | 0 | 0 | 7 | 7 |
| Burger (1999) | Conceptual | 6 | 2 | 0 | 4 | 4 |
| Caspi et al. (2002) | Direct & Conceptual | 3 | 0 | 0 | 3 | 1 |
| Cepeda et al. (2006) | Conceptual | 5 | 2 | 1 | 2 | 5 |
| Clancy (2005) | Conceptual | 6 | 0 | 0 | 6 | 3 |
| Duckworth & Seligman (2005) | Direct & Conceptual | 6 | 3 | 1 | 2 | 6 |
| Eagly et al. (1991) | Conceptual | 5 | 0 | 1 | 4 | 5 |
| Ferguson (2015) | Direct & Conceptual | 4 | 0 | 0 | 4 | 0 |
| Freedman & Fraser (1966) | Direct | 7 | 0 | 0 | 7 | 1 |
| Gottesman (1991) | Not Replicated | 8 | 0 | 1 | 7 | 3 |
| Hirst et al. (2015) | Not Replicated | 5 | 0 | 1 | 4 | 5 |
| Hyde (2005) | Direct | 4 | 0 | 0 | 4 | 4 |
| Karau & Williams (1993) | Conceptual | 6 | 2 | 0 | 4 | 5 |
| Killingsworth & Gilbert (2010) | Conceptual | 3 | 0 | 0 | 3 | 2 |
| Kirsch (2010) | Direct & Conceptual | 4 | 0 | 0 | 4 | 3 |
| Latané et al. (1979) | Conceptual | 5 | 0 | 3 | 2 | 4 |
| Laumann et al. (1994) | Conceptual | 4 | 0 | 0 | 4 | 4 |
| Milgram (1963) | Direct & Conceptual | 8 | 1 | 0 | 7 | 5 |
| Pavlov (1927) | Conceptual | 9 | 1 | 3 | 5 | 9 |
| Peterson & Peterson (1959) | Conceptual | 10 | 0 | 1 | 9 | 9 |
| Rosenhan (1973) | Direct | 7 | 0 | 0 | 7 | 5 |
| Savic et al. (2005) | Conceptual | 4 | 2 | 1 | 1 | 4 |
| Tolman & Honzik (1930) | Conceptual | 8 | 0 | 0 | 8 | 8 |
| Trahan et al. (2014) | Not Replicated | 3 | 0 | 1 | 2 | 3 |
| Tversky & Kahneman (1973) | Conceptual | 4 | 0 | 0 | 4 | 4 |
| Watson & Rayner (1920) | Conceptual | 8 | 0 | 0 | 8 | 8 |
| Totals | | 172 | 17 | 22 | 133 | 139 |

## DISCUSSION

The results of this study suggest a few common trends across the content of traditionally published IPTs like those examined in this study. First, there will likely be some discussion of replication, which is typically found in the first chapter of the book, where the principles of science are reviewed. There may also be some coverage of the replication crisis, though the authors may not share the context of its origin or treat it as a crisis. Additionally, it is unlikely that replication or the replication crisis will be mentioned again elsewhere in the book, including as it pertains to the replication status of the studies cited in the book. Finally, the replication status of studies is not only unlikely to be reported or implied, but in most cases, IPT authors will use language in their descriptions of the studies that connotes strong confidence in the validity and reliability of the studies, whether the studies have been replicated or not.

These trends in the IPTs examined in this study suggest at least three strategies instructors could employ to help teach their students to think scientifically and critically about replication. These strategies are designed so that any instructor can use them regardless of the specific content of their chosen IPT. First, instructors can

supplement the replication relevant content of the IPT with resources and materials that expand upon that content and further develop students' scientific and critical thinking skills. Second, course instructors can create experiential learning opportunities that give students direct exposure to replication research. Finally, instructors can teach students how to investigate the replication status of key studies covered in the course and how that status can and should be reflected in the way the study is presented and described.

## Strategy #1: Supplement IPT Content

### *Replication*

There are many publications at varying levels of sophistication that can be used to add detail to lecture notes and slide decks or can be assigned to students as additional reading on the topic of replication. Instructors who want to emphasize scientific thinking, for example, can assign students book chapters, such as "Guiding Principles for Scientific Inquiry" (National Research Council, 2002). This chapter introduces the philosophy of science supporting replication and describes in detail the impact and importance of reproducibility in advancing knowledge. Chapters like this help students understand the entire scientific enterprise more fully, as well as the role of replication within it. This chapter was also published before the replication crisis, which provides students with a pre-crisis sense of the importance of the concept that can be compared with post-crisis sources.

Journal articles also provide an increased understanding of replication within the context of scientific research, but typically do so with more specificity in terms of the disciplines involved. For example, an editorial written by the editors of *Management Review Quarterly* (Block & Kuckertz, 2018) lists in a clear and straightforward manner seven principles of replication for the social sciences, noting differences from replication in the natural sciences as they go. Their first principle, for example notes that social science research can only account for a small percentage of the variance in the dependent variable and thus can only suggest "quasi-laws" of human behavior (para. 5). Understanding these differences and seeing what journal editors are looking for in potentially publishable replications in a social science field informs students about what really matters to the discipline through the eyes of scholars who serve as quality control experts in their field. Psychology instructors could further narrow this focused approach to psychology journal articles (Cherry, 2023), noting how the practice of replication in psychology compares to other social sciences.

There are also articles that emphasize critical thinking about replication. For example, Goodman et al. (2016) critically examine the "nonstandard and unsettled" (p. 341) ways that research reproducibility is described across disciplines, resulting in misconceptions and, at times, the use of reproducibility terms as a "surrogate for 'truth'" (Goodman et al., 2016, p. 341). Sharing this concern, Rabeyron (2020) calls into question the feasibility and fit of replication to Psi and other phenomena and considers the need for a "post-modern psychology" that appreciates the "complexity of consciousness" and "the entanglement of the observer and the observed" (p. 8). He uses Bem's (2011) research that kicked off the replication crisis and the Ganzfeld protocol to help readers think critically about potential biases in replication that might hinder the advancement of knowledge in this important psychological domain. Instructors could effectively use supplementary articles like these to help students examine the assumptions and implications of replication in psychological science and to consider alternative approaches to reproducibility that might yield meaningful outcomes.

### *The Replication Crisis*

Before looking outside of IPTs for material that takes up the replication crisis in a manner that is accessible to introductory students and can support their scientific and critical thinking about this topic, it is important to note that in light of the crisis, several of the newer editions of the IPTs examined in this study provide greater detail and depth in their treatment of replication and the replication crisis (e.g., Coon et al., 2022; Myers & DeWall, 2020; Schacter et al., 2020; Wade et al., 2020). Indeed, some IPT authors have made this a highlighted feature of their newest editions, noting that they now provide "expanded coverage of replication" (Phelps et al., 2022, p. xiii), with 1-2 pages of discussion on the topics of replication and the replication crisis rather than the 2-3 paragraphs in previous editions. A few authors (e.g., Myers & DeWall, 2020) even note that they have intentionally excluded studies that were not successfully replicated in recent attempts, like the studies that did not replicate in the 2015 Open Science Collaborative study. Others (e.g., Phelps et al., 2022) state that their "textbook features studies that have replicated or would likely replicate based on the rigor of their methods" (p. 18). This expanded coverage of replication and the crisis

provides instructors more material within the IPT itself to work with as they teach students to think critically and scientifically.

Still, earlier editions of IPTs, including those used in this study, remain very popular, especially in cases of university initiatives to provide low or no-cost textbook options for students. Indeed, it has become increasingly common for instructors to use more affordable earlier editions of IPTs (McGowan, 2020). Consequently, supplementary materials may still be needed. Fortunately, there is a bounty of articles written in accessible language that can walk students through the history of the replication crisis, both in science generally (e.g., Baker, 2016; Lehrer, 2010) and in psychology specifically (e.g., Gray et al, 2021). Several of these articles discuss what is being done to improve the methods (e.g., Hedges et al, 2024) and ethics (e.g., Peels & Bouter, 2023) of scientific research to resolve the replication crisis in psychology and elsewhere. These articles can help students understand replication's epistemological importance as well as science's goal and value of self-correction.

### Strategy #2: Provide Experiential Learning Opportunities

Another strategy that can help students learn and practice scientific and critical thinking is to create experiential learning replication opportunities in the Introduction to Psychology course. The Collaboration Replications and Education Project (CREP) was established in 2014 to provide guidelines and opportunities for teaching and mentoring students in conducting direct replications of psychological research (Grahe et al., 2014). The CREP website (https://osf.io/stdgm/) provides a list of studies at varying levels of complexity and technology that can be directly replicated, usually within the course of a semester. CREP provides detailed instructions for carrying out the replication, and a contact person from CREP is available for consultation and direction. It is even possible to publish the results of the replication (Wagge et al., 2019). As noted in our introduction, Bem (2011) created materials to encourage replication of his precognitive studies, which are commonly cited as having ignited the replication crisis. These materials could be used within a classroom setting. Not only would this introduce students to topics related to replication and directly to the crisis, but it could raise awareness of other scientific dilemmas, such as how scientists react when findings fall outside of the current mainstream worldview. It would also expose students to the extensive breadth of anomalistic research that is often excluded or underreported in IPTs as well.

Several instructors who have built replication research into their classes discuss the pedagogical value of practicing replication in the classroom and identify its impact on the objectives of scientific and critical thinking in psychology (e.g., Frank & Saxe, 2012). Student participation in these studies can range from conducting the study as researchers under the instructor's supervision to being research participants. There are several ways to build experiential learning into the introductory course to teach the importance of replication, curb the crisis of replication, and impact student success (e.g., presentations at undergraduate research conferences and possibly publication).

### Strategy #3: Teach Students to Investigate the Replication Status of Studies

Another strategy instructors could integrate into their teaching is to create assignments for students that allow them to investigate the replication status of studies. Following the example of this study, students can investigate the replication status of studies that have impacted the discipline (like those examined in this article). Using the "cited by" feature in Google Scholar and other search engines, students can work in small groups to search the content of articles that cite the key study to see if the study has been replicated, and, if so, whether it was a direct or conceptual replication. Instructors could also assign students to write a description of the study's results in which they identify the replication status of the study and use language connoting more or less confidence in the validity and reliability of the study based on the replication status. This will help students think critically about how studies should be described, not only by textbook and journal authors, but also by the students themselves when they share the results of studies they read about in their Introduction to Psychology class in presentations, publications, and informal conversations.

## LIMITATIONS, IMPLICATIONS, AND APPLICATIONS

One limitation of the present study is that it necessarily reflects a snapshot in time and a delimited scope. The textbooks examined were published on average 6 years after the replication crisis in psychology began and two years after the Open Science Collaboration (2015) study was published, a time when the crisis was in full swing

and receiving a lot of attention. The results also reflect the content of popular IPTs published in traditional formats. It does not include more recently published editions, though we did examine and report on a sample of those newer editions in the discussion section. The study also omits OER textbooks, which do not have page limitations and other restrictions placed on traditionally published books and therefore are able to be more responsive and immediately revisable in light of current trends. Future research could investigate the treatment of replication in OER and newer edition IPTs to flesh out a more comprehensive understanding of the treatment of replication and the replication crisis in IPTs. This would be very helpful and would allow researchers to chronicle any changes over time. However, it would not likely affect the value of the three strategies for teaching critical thinking about replication and the replication crisis described here. We believe that these strategies, which were developed based on the results of this study—supplementing IPT content, giving students opportunities to practice replication, and teaching students to examine the replication status of key studies—can only improve students' scientific and critical thinking skills, regardless of the status of replication in the discipline at any given time and the IPT content that addresses it.

A key implication of the findings of the present study is that they raise important questions about replication in psychology. We address these questions here for the consideration of readers generally, but specifically for readers of this journal who teach replication, whether in an Introduction to Psychology course or in a class on another psychology subject. These questions, even if not answered in a course lecture or discussion, can be very helpful in deepening students' critical and scientific thinking about replication, but also about the philosophy of science that informs it, and can easily be taken for granted (Slife et al., 2005). Clearly, the importance of replication to the discipline and its undergirding philosophy of science has been highlighted in the wake of the replication crisis. Prior to the crisis, however, studies could go unreplicated for years, and yet they would be regularly cited in IPTs and other publications with no mention or implication of their replication status and without using more tentative language in the descriptions of the findings of the studies.

The Milgram (1963) study, for example, is one of the most controversial studies in the history of psychology. It was not replicated for 45 years (Burger, 2009). Yet, its results have been presented in almost every IPT for several decades without mention of its replication status and with language connoting high confidence in the reliability and validity of the findings. This is well illustrated by this excerpt from an IPT published prior to the Burger replication that did not state or imply that the Milgram study had not been replicated and yet confidently generalized the findings of the study beyond the sample of the research participants: "Social conditioning for obeying legitimate authorities is so strongly ingrained that people often lack the words or the ways to do otherwise (Kasschau, 2003, p. 560). Our own study shows that even in the wake of the crisis, and after the Milgram study has been replicated, the replication status of the Milgram study is rarely stated or implied, and there is no corresponding effect on the use of confident language in study descriptions. This trend is consistent with the other controversial studies included in our sample (e.g., Asch, 1956; Rosenhan, 1973).

It is interesting to note that this same trend did not occur in the case of Bem's (2011) study, even though Bem provided the materials needed to replicate the research, the study was replicated some 90 or so times and subjected to a meta-analysis. Additionally, as was the case with Milgram and Asch, the article was published in a highly regarded peer-reviewed journal and was authored by an esteemed social psychologist working at a top-tier research university. Instead, in Bem's case, immediate calls for replication were issued. Recall that for the Open Science Collaboration and others who were concerned with the controversial findings of the Bem article, along with a number of other publications with unexpected findings, replication became the key tool to be used to address and assess the reliability and validity of these studies. Given the reasons discussed in the introduction of this article, this makes a great deal of sense. What makes less sense is the differential treatment of this article compared to others that have presented controversial findings before. Unlike Milgram's study and others with unexpected results that had not been replicated for a long time, Bem's study, which was immediately subjected to replication, is not cited regularly in IPTs, at least not in the 2017 and 2018 editions we reviewed. What was it, then, about Bem's study that kicked off the replication crisis and resulted in its omission from IPTs?

An overly easy answer would be that the findings of Bem's study are not only unexpected, but they are also unexplainable within the scientific worldview of mainstream psychology and the fundamental assumptions psychologists use to understand reality. The problem with this response is that Bem's now infamous *Feeling the Future*

study was in no way the first on psi phenomena, nor was it the first of its kind to embrace replication. By the time his study was published in 2011, 70 years of solid empirical research had already been conducted in this area, with replication being one of its chief cornerstones. Readers of this journal who are versed in the history of Parapsychology, a term coined by German philosopher Max Dessoir and taken up by J.B. Rhine, will recall that repeatability and replication formed for Rhine the crux of his entire philosophy, which launched the American Parapsychology movement (McVaugh & Mauskopf, 1976).

Rhine believed it was necessary to conduct huge numbers of repeated trials within a single study as well as across separate studies. For example, in one single study, he conducted 2800 trials. By 1940, he had accumulated close to a million trials across 33 precognitive experiments, 27 of which produced statistically significant results. Several other labs successfully attempted replications of Rhine's work. A meta-analysis by Honorton & Ferrari (1989) examined 309 independent precognitive experiments conducted by 62 experimenters from 1935 to 1987 and found statistically significant results as well.

Rhine's strength in numbers philosophy is reflected in the many meta-studies conducted in the past three decades within the field of parapsychology, including those looking at pre-sentiment experiments, which, in a related vein to Bem's studies, measure automatic physiological responses of participants prior to receiving a future stimulus. Tressoldi (2011) analyzed 37 of these experiments across six protocols, finding an overall effect size of a Cohen's d of 0.26. According to Rabin (2016), the odds against chance for these were "6.3 × 1017, i.e., 625,000,000,000,000,000 to 1, which is "almost identical to the average effect size reported in 25,000 experiments conducted over a century of social psychology research…it allows us to seriously entertain the idea that presentiment *does* in fact exist." (p.1). Mossbridge et al (2012) conducted another meta-analysis the following year of published and unpublished presentiment studies, finding similar results. Clearly, researchers studying psi phenomena conduct scientific experimentation, replication, and meta-analyses, and they publish the results of their research in peer-reviewed journals, so the claim that Bem's (2011) research was treated differently because psi phenomena fall outside of the purview of psychological science and reproducibility would need some hefty justification.

One application that follows from the implication of the differential treatment of the Bem (2011) study, and others like it, is for future research to examine the worldview assumptions that inform mainstream psychology, to investigate potential systemic biases against research on psi phenomena, and to examine the ways in which replication narratives might be exploited to justify this differential treatment. Such research would be very helpful in better understanding the context surrounding the replication crisis and could be a useful resource to instructors who teach critical and scientific thinking about replication. It is beyond the scope of this article to draw or even suggest any conclusions about the differential treatment Bem's (2011) research received compared to other studies with controversial findings, including studies that went unreplicated for a long time. We note here only that despite the substantial body of research on psi phenomena just reviewed, and despite the central focus on replication across that research, all of the textbooks we examined were devoid of any mention of Bem's research and of this history of precognitive studies and adherence by parapsychologists to scientific standards to achieve repetitive and reliable results. Yet, that research was placed front and center as evidence of the replication crisis.

Future research may want to investigate potential systemic biases within mainstream psychology that target research that not only presents controversial or anomalous findings, but that is open to theories and concepts that are inconsistent with the worldview assumptions of normative psychology and is, as a result targeted and/or excluded from the mainstream of the discipline, often on the basis of claims of a scientific nature that do not in fact apply. One of the authors of this study has conducted exactly this kind of research regarding another excluded conceptualization (theism) in which the claims of mainstream psychology that such a conceptualization has no place in an empirical psychology have been clearly refuted and shown to reflect disciplinary worldview biases (e.g., Reber et al., 2012; Slife & Reber, 2009; Slife & Reber, 2021).

## CONCLUSION

The importance of replication to psychological science has been well established since the birth of the discipline and has been reaffirmed through the recent replication crisis (Ebersole et al., 2019; Klein, Cook, et al., 2019; Klein, Vianello, et al., 2018; Makel et al., 2012; Open Science Collaboration, 2015). Though psychologists may disagree about the label "crisis," they have certainly become more

vigilant about replication and the issues attending it, and they have strengthened their resolve to demonstrate the importance of replication within scientific psychology (Shrout & Rodgers, 2018). Several IPT authors have updated the content of their texts to reflect this reaffirmation and increased accountability, but given ever tighter word and page limits, there is only so much detail they can go into on this or any topic.

This study examined the content of the most popular IPTs for common themes and trends in the coverage of replication, the replication crisis, and the replication status of the studies cited and described in these textbooks. The results of that examination suggest several strategies instructors could take to expand upon the content of the IPT in teaching their students to think scientifically and critically about replication in psychology. As instructors create activities, assignments, and experiential learning opportunities that challenge students to think more carefully and deeply about the principles of replication, they help to ensure that this pillar of science remains strong and front of mind, not only among psychology majors who are the future stakeholders of the discipline but also among the general public, many of whom will have been students in an Introduction to Psychology course on their way to other majors and careers.

## REFERENCES


Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. P*sychological Monographs*, *47*(1), i–171. https://doi.org/10.1037/h0093360

American Psychological Association. (2023). *APA guidelines for the undergraduate psychology major, version 3.0: Empowering people to make a difference in their lives and communities*. American Psychological Association. https://www.apa.org/about/policy/undergraduate-psychology-major.pdf

Anglin, S. M., & Edlund, J. E. (2019). Perceived need for reform in field-wide methods and the teaching of replication, interpretation, and transparency. *Psychology Learning & Teaching*, *19*(1), 60–76. https://doi.org/10.1177/1475725719859453

Archer, J. (2004). Sex differences in aggression in real-world settings: A meta-analytic review. *Review of General Psychology*, *8*(4), 291–322. https://doi.org/10.1037/1089-2680.8.4.291

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. https://doi.org/10.1037/h0093718

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*, 452–454. https://doi.org/10.1038/533452a

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice Hall. https://doi.org/10.4135/9781446221129

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press. https://doi.org/10.1017/cbo9780511759185

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407–425. https://doi.org/10.1037/a0021524

Bem, D. J., Tressoldi, P. E., Rabeyron, T., & Duggan, M. (2016). Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of random future events [version 2; peer review: 2 approved]. *F1000Research*, *4*, 1188. https://doi.org/10.12688/f1000research.7177.2

Block, J., & Kuckertz, A. (2018). Seven principles of effective replication studies: Strengthening the evidence base of management research. *Management Review Quarterly*, *68*, 355–359. https://doi.org/10.1007/s11301-018-0149-3

Burger, J. M. (1999). The foot-in-the-door compliance procedure: A multiple process analysis review. *Personality and Social Psychology Review*, *3*(4), 303–325. https://doi.org/10.1207/s15327957pspr0304_2

Burger, J. M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist*, *64*(1), 1–11. https://doi.org/10.1037/a0010932

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

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., Taylor, A., & Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, *297*(5582), 581–584. https://doi.org/10.1126/science.1072290

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Chandler, M. (2019, June 28). AAP calls for closer regulation of Amazon and Google. *The Bookseller*. https://www.thebookseller.com/news/aap-calls-closer-regulation-amazon-and-google-1027576

Cherry, K. (2023). What is replication in psychology research? *Verywell Mind*. https://www.verywellmind.com/what-is-replication-2795802

Ciccarelli, S. K., & White, J. (2017). *Psychology: An exploration* (5th ed.). Pearson.

Clancy, S. A. (2005). *Abducted: How people come to believe they were kidnapped by aliens*. Harvard University Press. https://doi.org/10.4159/9780674029576

Collisson, B., Kellogg, J., & Rusbasan, D. (2015). Perceptions of psychology as a science: The effect of citation within introductory textbooks. *The Free Library*. https://www.thefreelibrary.com/Perceptions+of+-psychology+as+a+science%3a+the+effect+of+citations...-a0403448489

Coon, D., Mitterer, J. O., & Martini, T. S. (2022). *Introduction to psychology*: Gateways to mind and behavior (16th ed.). Cengage Learning.

Diener, E., & Biswas-Diener, R. (2024). The replication crisis in psychology. In R. Biswas-Diener & E. Diener (Eds.), *Noba textbook series: Psychology*. DEF Publishers. http://noba.to/q4cvydeh

Dobyns, Y. H., Dunne, B. J., Jahn, R. G., & Nelson, R. D. (2004). The MegaREG experiment: Replication and interpretation. *Journal of Scientific Exploration*, *18*(3), 369–397.

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, *16*(12), 939–944. https://doi.org/10.1111/j.1467-9280.2005.01641.x

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*(1), 109–128. https://doi.org/10.1037/0033-2909.110.1.109

Ebersole, C. R., Mathur, M. B., Baranski, E., Bart-Plange, D., Buttrick, N., Chartier, C. R., Corker, K. S., Corley, M., Hartshorne, J., Ijzerman, H., Lazarevic, L., Rabagliati, H., Ropovik, I., Aczel, B., Aeschbach, L. F., Andrighetto, L., Arnal, J., Arrow, H., Babincak, P., ... Nosek, B. A. (2019, October 3). Many Labs 5: Testing pre-data collection peer review as an intervention to increase replicability. *Advances in Methods and Practices in Psychological Science*, *3*(3), 309–331. https://doi.org/10.1177/2515245920958687

Feest, U. (2019). Why replication is overrated. *Philosophy of Science*, *86*(5), 895–905. https://doi.org/10.1086/705451

Feldman, R. S. (2018). *Understanding psychology* (14th ed.). McGraw-Hill Education.

Ferguson, C. J. (2015). Do angry birds make for angry children? A meta-analysis of video game influences on children's and adolescents' aggression, mental health, prosocial behavior, and academic performance. *Perspectives on Psychological Science*, *10*(5), 646–666. https://doi.org/10.1177/1745691615592234

Fidler, F., & Wilcox, J. (2021). Reproducibility of scientific results. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2021 ed.). Stanford University Press. https://plato.stanford.edu/archives/sum2021/entries/scientific-reproducibility/

Frank, M. C., & Saxe, R. (2012). Teaching replication. *Perspectives on Psychological Science*, *7*(6), 600–604. https://doi.org/10.1177/1745691612460686

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, *4*(2), 195–202. https://doi.org/10.1037/h0023552

Gernert, D. (2010). How to reject any scientific manuscript. *Journal of Scientific Exploration*, *22*(2), 233–243.

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, *8*(341), Article 341ps12. https://doi.org/10.1126/scitranslmed.aaf5027

Gottesman, I. I. (1991). *Schizophrenia genesis: The origins of madness*. W. H. Freeman.

Grahe, J., Brandt, M., Ijzerman, H., & Cohoo, J. (2014). Replication education. *Association for Psychological Science*. https://www.psychologicalscience.org/observer/replication-education

Gray, K., Arnott-Hill, E., & Benson, O. (2021). Module 4: The science of psychology: Tension and conflict in a dynamic discipline. In *Introduction to psychology*. College of DuPage Digital Press. https://cod.pressbooks.pub/introductiontopsychology/chapter/the-science-of-psychology-tension-and-conflict-in-a-dynamic-discipline/

Griggs, R. A. (2017). *Psychology: A concise introduction* (5th ed.). Worth Publishers.

Griggs, R. A., & Whitehead, G. I. (2014). Coverage of the Stanford prison experiment in introductory social psychology textbooks. *Teaching of Psychology*, *41*(4), 318–324. https://doi.org/10.1177/0098628314549703

Griggs, R. A., & Whitehead, G. I. (2015). Coverage of Milgram's obedience experiments in social psychology textbooks: Where have all the criticisms gone? *Teaching of Psychology*, *42*(4), 315–322. https://doi.org/10.1177/0098628315603065

Gurung, R. A. R., Hackathorn, J., Enns, C., Frantz, S., Cacioppo, J. T., Loop, T., & Freeman, J. E. (2016). Strengthening introductory psychology: A new model for teaching the introductory course. *American Psychologist*, *71*(2), 112–124. https://doi.org/10.1037/a0040012

Hedges, L., Hullman, J., Rometty, G., Tackett, J., & Uzzi, B. (2024, February 28). 'An existential crisis' for science: What IPR scholars are doing to solve the

replication crisis. *Northwestern University Institute for Policy Research*. https://www.ipr.northwestern.edu/news/2024/an-existential-crisis-for-science.html

Hirst, W., Phelps, E. A., Meksin, R., Vaidya, C. J., Johnson, M. K., Mitchell, K. J., & Olsson, A. (2015). A ten-year follow-up of a study of memory for the attack of September 11, 2001: Flashbulb memories and memories for flashbulb events. *Journal of Experimental Psychology: General*, *144*(3), 604–623. https://doi.org/10.1037/xge0000055

Hockenbury, S. E., & Nolan, S. A. (2018). *Psychology* (8th ed.). Worth Publishers.

Holden, C. J., & Goodwin, G. (2018). The importance of replication. In V. Zeigler-Hill & T. Shackelford (Eds.), *Encyclopedia of personality and individual differences*. Springer. https://doi.org/10.1007/978-3-319-28099-8_1352-1

Honorton, C., & Ferrari, D. C. (1989). "Future telling": A meta-analysis of forced-choice precognition experiments, 1935–1987. *Journal of Parapsychology*, *53*(4), 281–308.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581–592. https://doi.org/10.1037/0003-066X.60.6.581

Iso-Ahola, S. E. (2020). Replication and the establishment of scientific truth. *Frontiers in Psychology*, *11*, 2183. https://doi.org/10.3389/fpsyg.2020.02183

Kalat, J. W. (2017). *Introduction to psychology* (11th ed.). Cengage Learning.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, *65*(4), 681–706. https://doi.org/10.1037/0022-3514.65.4.681

Kasschau, R. A. (2003). *Understanding psychology*. Glencoe/McGraw-Hill.

Katz, D. (2022). Don't judge a book by its cover: A case study and comparative analysis of popular vs. academic psychology books. *Journal of Scientific Exploration*, *36*(2), 233–250. https://doi.org/10.31275/20222463

Klein, R. A., Cook, C. L., Ebersole, C. R., Vitiello, C. A., Nosek, B. A., Chartier, C. R., Ahn, P. H., Brady, A. J., Christopherson, C. D., Clay, S., Collisson, B., Crawford, J. T., Cromar, R., Gardiner, G., Gosnell, G. L., Grahe, J., Hall, C., Howard, I., … Ratliff, K. A. (2019). Many Labs 4: Failure to replicate mortality salience effect with and without original author involvement. *PsyArXiv*. https://doi.org/10.31234/osf.io/vef2c

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

Korbmacher, M., Azevedo, F., & Pennington, C. R. (2023). The replication crisis has led to positive structural, procedural, and community changes. *Community Psychology*, *1*, 3. https://doi.org/10.1038/s44271-023-00003-2

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, *330*(6006), 932. https://doi.org/10.1126/science.1192439

Kirsch, I. (2010). *The emperor's new drugs: Exploding the antidepressant myth*. Basic Books.

Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, *37*(6), 822–832. https://doi.org/10.1037/0022-3514.37.6.822

Laumann, E. O., Gagnon, J. H., Michael, R. T., & Michaels, S. (1994). *The social organization of sexuality: Sexual practices in the United States*. University of Chicago Press.

Lehrer, J. (2010, December 5). The truth wears off. *The New Yorker*. https://www.newyorker.com/magazine/2010/12/13/the-truth-wears-off

Lewandowsky, S., & Oberauer, K. (2020). Low replicability can support robust and efficient science. *Nature Communications*, *11*, 358. https://doi.org/10.1038/s41467-019-14203-0

Lilienfeld, S. O., Lynn, S. J., & Namy, L. L. (2017). *Psychology: From inquiry to understanding* (4th ed.). Pearson.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*(6), 537–542. https://doi.org/10.1177/1745691612460688

McBurney, D. H. (1996). *How to think like a psychologist: Critical thinking in psychology*. Prentice Hall.

McGowan, V. (2020). Institution initiatives and support related to faculty development of open educational resources and alternative textbooks. *Open Learning: The Journal of Open, Distance and e-Learning*, *35*(1), 24–45. https://doi.org/10.1080/02680513.2018.156232

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

McVaugh, M., & Mauskopf, S. H. (1976). J. B. Rhine's *Extra-sensory perception* and its background in psychical research. *Isis*, *67*(2), 221–233. https://doi.org/10.1086/351583

Miller, B., & Gentile, B. F. (1998). Introductory course content and goals. *Teaching of Psychology*, *25*(2), 89–96. https://doi.org/10.1207/s15328023top2502_2

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67*(4), 371–378. https://doi.org/10.1037/h0040525

Mossbridge, J., Tressoldi, P., & Utts, J. (2012). Predictive physiological anticipation preceding seemingly unpredictable stimuli: A meta-analysis. *Frontiers in Psychology*, *3*, 390. https://doi.org/10.3389/fpsyg.2012.00390

Myers, D. G., & DeWall, C. N. (2018). *Exploring psychology in modules* (11th ed.). Worth Publishers.

Myers, D. G., & DeWall, C. N. (2020). *Exploring psychology in modules* (13th ed.). Worth Publishers.

National Academies of Sciences, Engineering, and Medicine. (2019). Chapter 7: Confidence in science. In *Reproducibility and replicability in science*. The National Academies Press. https://doi.org/10.17226/25303

National Research Council. (2002). Chapter 3: Guiding principles for scientific inquiry. In *Scientific research in education*. National Academies Press. https://doi.org/10.17226/10236

Nature. (2021, December 15). Replicating scientific results is tough – but essential. *Nature*. https://www.nature.com/articles/d41586-021-03736-4

Nosek, B. A., & Errington, T. M. (2017). Reproducibility in cancer biology: Making sense of replications. *eLife*, *6*, e23383. https://doi.org/10.7554/eLife.23383

Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLOS Biology*, *18*(3), e3000691. https://doi.org/10.1371/journal.pbio.3000691

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

Pavlov, I. P. (1927). *Conditioned reflexes*. Oxford University Press.

Peels, R., & Bouter, L. (2023). Replication and trustworthiness. *Accountability in Research*, *30*(2), 77–87. https://doi.org/10.1080/08989621.2021.1963708

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*(3), 193–198. https://doi.org/10.1037/h0049234

Phelps, E. A., Berkman, E. T., & Gazzaniga, M. S. (2022). *Psychological science* (7th ed.). W. W. Norton & Company.

Rabin, D. (2016). Presentiment. In *Psi encyclopedia*. Society for Psychical Research. https://psi-encyclopedia.spr.ac.uk/articles/presentiment

Rabeyron, T. (2020). Why most research findings about psi are false: The replicability crisis, the psi paradox and the myth of Sisyphus. *Frontiers in Psychology*, *11*, 562992. https://doi.org/10.3389/fpsyg.2020.562992

Rathus, S. A., Veenvliet, S. G., & Maheu, S. J. (2017). *Psych: Introduction to psychology*. Cengage Learning.

Reber, J. S., Slife, B. D., & Downs, S. D. (2012). A tale of two theistic studies: Illustrations and evaluation of a potential program of theistic psychological research. *Research in the Social Scientific Study of Religion*, *23*, 191–212. https://doi.org/10.1163/9789004242870_012

Rosenhan, D. L. (1973). On being sane in insane places. *Science*, *179*(4070), 250–258. https://doi.org/10.1126/science.179.4070.250

Ruscio, J. (2006). *Critical thinking in psychology: Separating sense from nonsense* (2nd ed.). Cengage Learning.

Savic, I., Berglund, H., & Lindstrom, P. (2005). Brain response to putative pheromones in homosexual men. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(20), 7356–7361. https://doi.org/10.1073/pnas.0407998102

Scargle, J. D. (2000). Publication bias: The "file-drawer" problem in scientific inference. *Journal of Scientific Exploration*, *14*(1), 91–106.

Schacter, D. L., Nock, M. K., & Wegner, D. M. (2018). *Psychology* (4th ed.). Worth Publishers.

Schacter, D. L., Gilbert, D. T., Nock, M. K., & Wegner, D. M. (2020). *Psychology* (5th ed.). Worth Publishers.

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90–100. https://doi.org/10.1037/a0015108

Sheldrake, R. (1998). Experimenter effects in scientific research: How widely are they neglected? *Journal of Scientific Exploration*, *12*(1), 73–78.

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, *69*, 487–510. https://doi.org/10.1146/annurev-psych-122216-011845

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*(1), 76–80. https://doi.org/10.1177/1745691613514755

Slife, B. D., & Reber, J. S. (2009). Is there a pervasive implicit bias against theism in psychology? *Journal of Theoretical and Philosophical Psychology*, *29*(2), 63–79. https://doi.org/10.1037/a0015151

Slife, B. D., & Reber, J. S. (2021). Against methodological confinement: Toward a pluralism of methods and interpretations. *Psychology of Religion and Spirituality*, *13*(1), 14–18. https://doi.org/10.1037/rel0000256

Slife, B. D., Reber, J. S., & Richardson, F. C. (Eds.). (2005). *Critical thinking about psychology: Hidden assumptions and plausible alternatives*. American Psychological Association. https://doi.org/10.1037/10885-000

Stanovich, K. E. (2007). *How to think straight about psychology*. Allyn & Bacon.

Sternberg, R. J., & Halpern, D. F. (Eds.). (2020). *Critical thinking in psychology* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781108684354

Stevenson, I. (1999). What are the irreducible elements of the scientific enterprise? *Journal of Scientific Exploration*, *13*(2), 257–270.

Sturrock, P. A. (1994). Applied scientific inference. *Journal of Scientific Exploration*, *8*(4), 491–508.

Tolman, E. C., & Honzik, C. H. (1930). Introduction and removal of reward and maze performance in rats. *University of California Publications in Psychology*, *4*, 257–275. https://psycnet.apa.org/record/1931-02280-001

Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin*, *140*(5), 1332–1360. https://doi.org/10.1037/a0037173

Tressoldi, P. E. (2011). Extraordinary claims require extraordinary evidence: The case of non-local perception, a classical and Bayesian review of evidences. *Frontiers in Psychology*, *2*, 117. https://doi.org/10.3389/fpsyg.2011.00117

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Wade, C., Taris, C., Somers, S., & Shin, L. (2017). *Invitation to psychology* (7th ed.). Pearson.

Wade, C., Taris, C., Somers, S., & Shin, L. (2020). *Invitation to psychology* (8th ed.). Pearson.

Wagge, J. R., Brandt, M. J., Lazarevic, L. B., Legate, N., Christopherson, C., Wiggins, B., & Grahe, J. E. (2019). Publishing research with undergraduate students via replication work: The Collaborative Replications and Education Project. *Frontiers in Psychology*, *10*, 247. https://doi.org/10.3389/fpsyg.2019.00247

Warne, R. T., Astle, M. C., & Hill, J. C. (2018). What do undergraduates learn about human intelligence? An analysis of introductory psychology textbooks. *Archives of Scientific Psychology*, *6*(1), 32–50. https://doi.org/10.1037/arc0000038

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, *3*(1), 1–14. https://doi.org/10.1037/h0069608

Weiten, W. (2017). *Psychology: Themes and variations* (10th ed.). Cengage Learning.

Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, *39*(4), 202–217. https://doi.org/10.1037/teo0000137

Youyou, W., Yang, Y., & Uzzi, B. (2023). A discipline-wide investigation of the replicability of psychology papers over the past two decades. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(6), e2208863120. https://doi.org/10.1073/pnas.2208863120

Zientek, L. R., Werner, J. M., Campuzano, M. V., & Nimon, K. (2018). Writer's forum: The use of Google Scholar for research and research dissemination. *New Horizons in Adult Education and Human Resource Development*, *30*(1), 39–46. https://doi.org/10.1002/nha3.20209