The Problem of Miscitation in Psychological Science: Righting the Ship

Cory L. Cobb[1], Brianna Crumly[1], Pablo Montero-Zamora[2], Seth J. Schwartz[2], and Charles R. Martínez Jr.[3]

[1] Department of Human Development and Family Sciences, College of Human Sciences, Auburn University
[2] Department of Kinesiology and Health Education, College of Education, University of Texas at Austin
[3] College of Education, University of Texas at Austin

Scholarly citation represents one of the most common and essential elements of psychological science, from publishing research, to writing grant proposals, to presenting research at academic conferences. However, when authors mischaracterize prior research findings in their studies, such instances of miscitation call into question the reliability and credibility of scholarship within psychological science and can harm theory development, evidence-based practices, knowledge growth, and public trust in psychology as a legitimate science. Despite these implications, almost no research has considered the prevalence of miscitation in the psychological literature. In the largest study to date, we compared the accuracy of 3,347 citing claims to original findings across 89 articles in eight of top psychology journals. Results indicated that, although most (81.2%) citations were accurate, roughly 19% of citing claims either failed to include important nuances of results (9.3%) or completely mischaracterized findings from prior research altogether (9.5%). Moreover, the degree of miscitation did not depend on the number of authors on an article or the seniority of the first authors. Overall, results indicate that approximately one in every 10 citations completely mischaracterizes prior research in leading psychology journals. We offer five recommendations to help authors ensure that they cite prior research accurately.

> **Public Significance Statement**
> This article suggests that approximately one in every 10 citations across leading psychology journals is inaccurate. Such instances of miscitation may call into question the reliability and credibility of scholarship within psychological science. Scholars in psychology should be careful to ensure that they cite and characterize findings from prior research accurately in their studies.

Keywords: psychological science, citation, miscitation, scientific writing, recommendations

Supplemental materials: https://doi.org/10.1037/amp0001138.supp

Since its conception, the field of psychology has undergone considerable criticism in its efforts to establish itself as a serious scientific discipline (Coyne, 2016; Lilienfeld, 2010, 2012; Lykken, 1991; Meehl, 1967; Miller, 2004; Popper, 1959; Schmidt & Oh, 2016; Skinner, 1987). Concerns around psychology as a science have ranged from an overreliance on null hypothesis significance testing to the failure of independent research teams to replicate key findings from previous

Cory L. Cobb

experiments (e.g., Cumming, 2008; Krueger, 2001; Tackett et al., 2019). Many of these concerns in psychological science have been linked to the use of questionable research practices (QRPs), which refer to certain design, analytic, and reporting practices that increase the likelihood of finding evidence in support of a theory or hypothesis (John et al., 2012). Examples of QRPs include reporting only statistically significant findings from a study, failing to disclose experimental conditions to avoid presenting null findings, presenting post hoc hypotheses as if they had been specified a priori, and cherry-picking studies to cite that support an author's preferred hypothesis (Banks et al., 2016). Researchers may engage in QRPs purposefully or inadvertently, but such practices fall outside the scope of normative assumptions around well-conducted research. QRPs can be detrimental to the scientific progress of psychology as they may harm theory development, evidence-based practices, knowledge growth, and public trust in psychology as a legitimate science. QRPs can also mislead the field and can create the impression that there is more (or less) consensus around a research area or question than is the case (Ioannidis, 2005; Rudmin, 2007).

One QRP that has received relatively little attention in psychology, but that has important implications for psychological science, is the problem of miscitation. Although miscitation may entail several practices (e.g., incorrect reporting of bibliographic information in citations and references), we define miscitation as failure to adhere to the practice of providing a complete and correct account of the cited content of a study (Lazonder & Janssen, 2022).[1] Thus, miscitation represents the extent to which a citing claim corresponds to the theoretical claims or empirical results of the perspective or study that is being cited. Citation fidelity may vary in degrees ranging from *accurate* (i.e., the citing claim exactly reflects findings from the original article), to

somewhat accurate (i.e., the citing claim largely and correctly corresponds to the results from the original article but lacks qualifying nuance), to *inaccurate* (i.e., the citing claim does not correspond at all to the results of the original article). As will be discussed later below, miscitation has the potential to proliferate into erroneous consensus around a topic, harm theory-building efforts, and even influence evidence-based practice. Because trust in the reliability of published research is vital to the credibility of a scholarly journal and of a scientific discipline (Rivkin, 2020), and as the field of psychology strives toward becoming a stronger scientific discipline, it is important that citations of empirical work are accurate and that psychology researchers correctly characterize findings from prior work.

The problem of miscitation has been a growing concern for decades and has resulted in a large body of work, mostly in medical research (e.g., Armstrong et al., 2018; Hicks, 2021; Todd et al., 2007). Most of this work has examined the prevalence of citation inaccuracies across studies, with emphasis on how such inaccuracies can undermine the reliability of scientific findings. However, despite the important implications of citation accuracy for scientific research, little to no work has considered the prevalence of miscitation within psychological science. Yet, a recent (Rivkin, 2020) survey showed that upwards of 50% of psychologists reported having been miscited at some point in their career, and many reported taking shortcuts when citing research articles (e.g., relying on secondary citations; Klitzing et al., 2018). Thus, in the present study, we document the extent to which scholars (in)accurately cite empirical findings from published research studies across eight leading journals in psychology. We also provide recommendations for addressing miscitation in psychological science.

The Importance of Citation Accuracy Within Psychological Science

Scholarly publications serve as the catalysts through which scientific knowledge advances. The psychological community depends on peer-reviewed research articles to communicate current knowledge and to update the field when new developments occur. Most scientific publications are derivative in nature, such that authors build upon empirical findings from prior studies to support their own research. Indeed, one key component of publishing a research article involves citing findings from earlier work to develop a scholarly argument that substantiates important assumptions and claims that the author is advancing. Such reliance on prior work to validate one's research is not limited to publishing but also extends to other scholarly activities in psychology, ranging from writing grant proposals to obtain extramural research funding to presenting

---

[1] In educational and medical disciplines, this is generally referred to as quotation accuracy. However, because the term quotation in psychology means replicating the exact words of an author, we used the term miscitation instead.

Brianna Crumly

new research findings at scientific conferences. Thus, scholarly citation constitutes an essential practice that is central to many domains of psychological science.

According to American Psychological Association's *Ethical Principles of Psychologists and Code of Conduct* (American Psychological Association [APA], 2017), psychologists have an ethical responsibility to the Principle of Integrity such that they "seek to promote accuracy, honesty, and truthfulness in the science, teaching, and practice of psychology." In addition, the *Publication Manual of the American Psychological Association* (APA, 2020) suggests that citation of prior research assumes that citing authors have read and are familiar with the original article that is cited. The Office of Research Integrity (ORI, 2022) states that authors should avoid citing resources that they do not thoroughly understand and describes such citation practices as deceptive. All these principles appeal to the responsibility and collegiality of psychologists to represent the work of their peers in the scientific community properly and accurately. Hence, the practice of accurate and honest citation represents a key pillar of ethical conduct for psychologists, and authors have a duty to understand and familiarize themselves with the research findings that they cite.

When authors miscite research, erroneous claims can proliferate and accumulate into false beliefs that are subsequently adopted by scholars in the field (Bareket et al., 2020). For example, many scholars often rely on secondary rather than primary sources for citations (Klitzing et al., 2018). If an author miscites findings from prior work in a published article, such miscitation can "domino effect" as other scholars subsequently cite and build upon the author's mischaracterization of prior research (ORI, 2022; Rudmin, 2007). This point is particularly salient given the digital age in which we live, where information can spread at uncontrollable rates.

Accurate citation of research is also critical to evidence-based decision-making. Applied researchers, educators, and policymakers often rely on summaries of evidence provided in scientific reports to make informed decisions (Tipton et al., 2021). When authors miscite research, these entities may make decisions that they believe are evidence-based, when in fact, they are based on misrepresentations of empirical findings. Miscitation of research studies may also challenge the credibility of psychological science in general. When psychological research is misrepresented, misinformation may propagate and diminish the integrity of psychology as a reliable science (Ecker et al., 2014). Even when corrected, the erroneous beliefs that derive from misinformation often persist (Johnson & Seifert, 1994; Nyhan & Reifler, 2010).

Citation accuracy carries several other important implications for psychological science. For example, a core tenet of theory building is that theories are constructed to explain phenomena that are presented within empirical data (Borsboom et al., 2021). However, when the findings used in constructing a theory are mischaracterized, that theory will likely rest upon faulty assumptions and premises that compromise the theory's validity and explanatory power. This problem becomes even more salient in applied work when interventions are developed based on specific theories of change. Miscitation of findings may also play a role in the current replication crisis (Open Science Collaboration, 2012). For instance, when attempting to replicate a study, an author must review and summarize prior research findings to develop hypotheses about what findings, if they were to emerge, would constitute a successful replication. However, if the author miscites prior literature, or relies on secondary sources that miscited the literature, it becomes difficult to ascertain whether a study was (or was not) replicated—because the hypotheses were formulated based on mischaracterizations of prior research findings.

Prevalence of Miscitation in the Scientific Literature

Most research on the prevalence of miscitation comes from the medical literature. Findings from a large body of work indicate that citation errors in medical research are relatively common and most prevalence rates of miscitation, depending on the journal and methodology, have ranged from 11% to as high as 41% (Armstrong et al., 2018; Fenton et al., 2000; Goldberg et al., 1993; Hicks, 2021; Jergas & Baethge, 2015; Lukić et al., 2004; Mogull, 2017; Pavlovic et al., 2021; Todd et al., 2007). Several other disciplines have found similar prevalence rates of miscitation in their respective literature including 11.1% in ecology (Todd et al., 2007), 10.6% in marine biology (Todd et al., 2010), 19% in physical geography (Haussmann et al., 2013), 25% in general science journals (Smith & Cumberledge, 2020), 26% in an education meta-analysis (Martella et al., 2021), and 7.6% in 25 highly cited education articles (Lazonder & Janssen, 2022).

Pablo Montero-Zamora

Although there has been little to no work on citation accuracy in the psychological literature, there is reason to suspect that miscitation occurs. In one of the few studies that has considered miscitation in the psychological literature, Rudmin (2007) noted that researchers in top journals systematically miscited a 1987 review article on the relationship between immigrant acculturation strategies and stress. Rudmin found that researchers subsequently cited the article as indicating that integration strategies were much less stressful than assimilation strategies for immigrants adapting to a new homeland. However, the original review article reported no such findings in that correlations for both assimilation and integration were nonsignificant and negatively related to stress. In a more recent study on citation behaviors in social psychology, experimental psychology, and educational sciences, Klitzing et al. (2018), found that one third of researchers in their sample reported relying on secondary rather than primary sources for citations, and approximately half of respondents reported having been miscited in their careers—57% reported never having received any formal citation-related training (Klitzing et al., 2018).

Given the implications of miscitation for science and the credibility of psychology at large, it is critical for the field to recognize this problem and take steps to prevent it. In the present study, we aim to document the extent to which miscitation may occur in top psychological science journals and to offer recommendations to right the proverbial ship.

Methodology

Journal Selection

To examine miscitation in psychological science, we assessed whether citing claims in published psychology journal articles corresponded to the findings of the original

articles cited.[2] For journal selection and to adequately represent the diverse set of subdisciplines found in psychological science, we chose eight high-impact journals, one per discipline, that are often considered the flagship journals of their respective disciplines. As displayed in Table 1, most psychological disciplines were represented in the study, including social/personality, cognitive, school, developmental, cultural, clinical/consulting, general, and counseling psychology.

Article Selection and Focus

After identifying target journals, we reviewed all articles from the most recent issue of each journal at the time the present study began. This decision was motivated by our desire to represent the state of the science by reviewing the most recent research possible. Across the eight journals, a total of 89 articles were included for analysis in the present study. Within each article, we elected to focus on citations only in the Introduction sections because (a) the Introduction section is where authors spend the most time reviewing and summarizing relevant literature and (b) the Introduction is where the rationale of a study is provided and serves as the foundation on which a study's primary hypotheses are based and substantiated.

Inclusion and Exclusion Criteria for Citations Reviewed

We developed inclusion and exclusion criteria regarding which citations to include in our analysis. The primary inclusion criterion was that citing claims must be based on an empirical study. Citations to nonempirical articles were excluded from the present analysis. Such nonempirical articles included books, literature reviews, theoretical articles, conceptual overviews, and articles focusing on general guidelines. We elected to focus on empirical citations because we can directly compare the degree to which a citing claim corresponds to the actual findings of a cited article. It is likely very difficult and substantially more subjective to determine whether, for example, a book or review article that covers large bodies of research aligns with a citing claim. Across the 89 Introduction sections we reviewed, a total of 4,953 citations were included. After omitting 1,606 citations based on our exclusion criteria, the final number of citations reviewed was 3,347 (see Table 1).

Coding Procedures

Because miscitation refers to the extent to which one has provided a complete and correct account of the cited content of a study (Lazonder & Janssen, 2022), miscitation varies by degree. However, because no validated scale exists for indexing the degree of miscitation in the literature, and because it is possible for two scientists to look at the same results and

---

[2] Data used for coding procedures and analysis are publicly available in the Open Science Framework at https://osf.io/meetings/apa and https://osf.io/fuae7/.

Seth J. Schwartz

interpret them differently, we decided to code citations according to three *verifiable* degrees of accuracy centering on the extent to which the results cited by an author (or author group) corresponded to the specific findings presented in the original article. Accordingly, we operationalized degree of citation accuracy, or miscitation, according to three categories: accurate (0), somewhat accurate (1), and inaccurate (2). To receive a code of *accurate*, a citing claim must correspond exactly to the findings of the original article and not omit key qualifiers. For example, if results from the original article indicated a significant association for boys but not for girls, then the citing claim explicitly noted that the significant association emerged for one sex but not the other.

Receiving a code of *somewhat accurate* indicates that a citing claim is mostly accurate, such that it is broadly related to or is in line with findings from the original article, but that key qualifiers or caveats were omitted despite being presented in the results of the original article. For instance, if a significant association emerged for boys but not for girls in the original article, and the citing authors make a claim about a significant finding for youth in general, then this example would receive a code of "somewhat accurate" because it omitted a key aspect of the original finding that effects were only significant for boys.

A code of *inaccurate* indicates that the citing claim (a) was not examined or reported in the original article or (b) directly contradicted the findings of the original article. For example, if a citing claim stated that there is a relationship between coping styles and substance use among Latinos, but coping styles were never assessed, this would be inaccurate because the citing claim was not examined or reported in the original article. Similarly, if the citing claim stated a positive relationship between authoritarian parenting and school performance, but the relationship was negative or null in the original article,

this claim would be inaccurate because the claim contradicts the finding from the original article (see Table 2, for coding categories and citation examples). Several additional examples can be found in the Supplemental Materials.

To determine which codes to assign to citations, we first documented the citing claim verbatim in an excel file and then reviewed the Method and Results sections of each cited article to assess the degree of correspondence. Two independent coders, a PhD student (second author) working alongside the first author of the present study and a postdoctoral research fellow at a separate institution (third author), compared each citing claim against results from cited articles and provided codes accordingly. For each citation code, and using the coding criteria outlined above, coders focused specifically on whether the empirical claim made by the citing article reflected the results as presented in the original article. When a code was assigned to a citation, each coder provided a brief rationale for selecting that code. To minimize interrater bias in the codes assigned, the two coders had no contact throughout the entirety of the 8-month coding process and thus were completely independent. Prior to coding, the first author of the present study provided identical training to both coders, including discussion of coding schemes, providing textual examples of each code, and working through several practice examples. Moreover, the first author was readily available to address any questions or issues that may have occurred to coders during the coding process.

To provide readers with an explicit breakdown of why citations were classified into *somewhat accurate* and *inaccurate* categories, we grouped the rationale provided by raters for each code into general themes.[3] As noted above, when a code was assigned to a citation, each coder provided a rationale for providing that code. These rationales served as the basis for the groupings. Explicitly highlighting the themes from these two categories will provide researchers with a greater understanding of the reasons why citations were classified into each category.

Regarding the *somewhat accurate* category ($n = 311$), we identified four primary reasons for classifying citations into this category: overgeneralization, overspecification, double-barreled claims, and miscellaneous. *Overgeneralization* ($n = 165$) refers to instances where a citing claim generalized a finding for a specific group or process to a larger group or process that was not directly assessed in the original article (e.g., findings for Chinese immigrants generalized to Asian populations in general). *Overspecification* ($n = 56$) refers to instances where the original study's findings were based on a general population or process, whereas the citing claim applied that finding to a specific population or process that was not directly assessed in the original study (e.g., findings for a diverse sample of Asians applied to South Koreans

---

[3] For transparency, these categories were analyzed and added as part of the peer review process.

Charles R. Martínez Jr.

specifically). *Double-Barreled Claims* ($n = 54$) regarded those that entailed two claims within a single statement, whereas the original study only supported one of those claims (e.g., citing claims referred to links among racism, depression, and substance use, whereas only racism and depression were assessed in the original study). *Miscellaneous* ($n = 36$) refers to citing claims that were largely accurate but that entailed minor errors in the reporting of results from the original study (e.g., a citing claim correctly reports the directional nature of a relationship but misrepresents the magnitude of the statistic—stating a correlation of .26 is moderate in strength when it is small).

Regarding the *Inaccurate* category ($n = 318$), two general themes emerged. Specifically, all citing claims in the Inaccurate category were there because (a) they did not appear in the cited article in that the results stated in the citing claim were never assessed or reported in the original study ($n = 195$) or (b) they directly contradicted the findings from the original study, for example, if a citing claim stated that there was a positive relationship between two variables when the original study found a negative or null relationship ($n = 123$). See Supplemental Materials, for several additional examples of these groupings.

Interrater Reliability

Once final codes were complete, and because our coding data were on an ordinal scale, we computed weighted kappa as an index of interrater reliability. Weighted kappa corrects for agreement that may occur by chance and considers how different kinds of disagreement should be differentially weighted—on an ordinal scale, the distance between two codes of 0 and 1 carries less weight than the distance between codes of 0 and 2 (Cohen, 1968). Acceptable levels of kappa range from .61 to .80 (substantial agreement) to .81–1.00

(almost perfect agreement; Landis & Koch, 1977). For the present study, the reliability coefficient was .77 ($p < .001$, CI [.74−.79]), again indicating strong interrater reliability. As supplemental indices of interrater reliability, we also computed raters' percentage of agreement (ratio of agreement to the number of total observations) as well as the intraclass correlation coefficient (degree of agreement within classes of raters). Generally, acceptable levels for percentage of agreement and intraclass correlations are >.80 for both indices (Hallgren, 2012; Koo & Li, 2016). Results indicated strong interrater reliability for both percentage of agreement (90.2%) and the intraclass correlation (.91). Regarding codes for which there was disagreement between coders, most discrepancies occurred between codes of 0 and 1 or between 1 and 2. These discrepancies, along with any of the rationales provided by raters for each code, were addressed in subsequent meetings among the first author and the two coders after all independent coding had been completed. During these meetings, each discrepancy was reviewed and discussed while revisiting results from the originally cited article, and the group reached consensus on each code. When discrepancies were difficult to resolve (e.g., coders could not clearly determine whether a citation should receive a 1 or 0), we gave authors the benefit of the doubt and assigned the lesser of the two codes (e.g., assigning 0 rather than 1). Thus, our final codes represent more conservative estimates in terms of miscitation rates.

Supplemental Analyses

The following analyses were conceived and specified a priori. To consider whether the degree of miscitation in each article was related to the number of authors for that article, we summed the citation codes (0, 1, 2) for each of the 89 articles (higher values represent greater degree of miscitation) and regressed these values onto the number of authors. This analysis was based on the possibility that having more authors on a article might reduce the degree of miscitation—the assumption being that more authors may be more likely to catch citation errors.

Furthermore, we considered the possibility that the degree of miscitation may depend on seniority level of the first author. For example, having more (or less) experience in scholarly publication might be linked with greater (or lower) proclivity toward miscitation. We conducted a chi-square analysis of independence between the summed index of miscitation described above and the citing first author's seniority level. Seniority was coded according to the following: student = 0, postdoctoral fellow = 1, assistant professor = 2, associate professor = 3, full professor = 4, and those in industry or private practice = 5. Of course, the specific code assigned is not important to the chi-square test, as it treats the categorical variable as a nominal scale. To identify each first author's rank, we examined their affiliation on the cover page

Table 1
Journal Information and Citation Counts

Journal name	Discipline	Volume/issue/year	Number of articles	Number of citations	Number of citations excluded	Total citations reviewed
1. *Journal of Personality and Social Psychology*	Personality/social psychology	120/6/2021	13	937	338	599
2. *Cognitive Psychology*	Cognitive psychology	128/NA/2021[a]	4	231	88	143
3. *Journal of School Psychology*	School psychology	87/NA/2021[a]	5	399	102	297
4. *Developmental Psychology*	Developmental psychology	57/6/2021	15	853	224	629
5. *Cultural Diversity and Ethnic Minority Psychology*	Cultural psychology	27/3/2021	24	1,261	427	834
6. *Psychological Science*	General psychology	32/8/2021	12	415	126	289
7. *Journal of Consulting and Clinical Psychology*	Consulting and clinical psychology	89/7/2021	6	250	83	167
8. *Journal of Counseling Psychology*	Counseling psychology	68/4/2021	10	607	218	389
Totals	—	—	89	4,953	1,606	3,347

Note. NA = not applicable.
[a] Only volume or issue number provided by journal but not both.

of each published article reviewed and visited their respective department websites to determine rank.

Results

Results for the degree of citation accuracy, both across and within each journal, are presented in Table 3. Results across all journals indicate that 81.2% of citations were classified as accurate (range = 71.3%–84.1%), 9.3% as somewhat accurate (range = 5.9%–16.8%), and 9.5% as inaccurate (range = 6.2%–14.1%). Thus, nearly one in every 10 citations in the Introduction sections of leading psychology journal articles was inaccurate. In addition, approximately 9.3%, or one in every 11 citations, were somewhat accurate such that authors correctly matched the basic findings but omitted key qualifiers in their presentation of that study's results.

The proportions of miscitation are also presented for each journal separately. As noted in Table 3, *Psychological Science* had the highest proportion of citations in the accurate category (84.1%), whereas the *Journal of Consulting and Clinical Psychology* had the lowest proportion of citations in this category (71.3%). The *Journal of Consulting and Clinical Psychology* had the highest proportion of citations in the somewhat accurate category (16.8%), whereas *Developmental Psychology had* the lowest proportion of citations in this category (5.9%). *Developmental Psychology* had the highest proportion of citations in the inaccurate category (14.1%), whereas *Psychological Science* (6.2%) had the lowest proportion in this category.

Next, we assessed the relationship between miscitation frequency and number of authors on a given citing article. Number of authors ranged from 1 to 12 authors across citing articles. Results indicated that the number of authors on a research article was not significantly related to the degree of miscitation ($b = .56$, $SE = .43$, $p = .20$). Regarding the association between first author's seniority level and degree of miscitation, results from the chi-square test indicated no significant relationship between these variables, $\chi^2(130) = 142.06$, $p = .22$. The proportions of citing articles according to seniority were: student (18.4%), postdoctoral fellow (17.2%), assistant professor (26.4%), associate professor (10.3%), full professor (19.5%), and private practice/industry (8.0%). The seniority status for authors of two separate studies could not be identified and was thus excluded from analyses.

Discussion and Recommendations

In the present analyses, we assessed the degree to which authors who publish in leading scientific journals in psychology accurately cite and represent findings from prior work in their research studies. Results of our review of 3,347 citations across 89 articles in eight journals indicated that, on average, approximately 81.2% of citations accurately characterized findings from cited research articles. However, we also found that approximately the same proportions of citations either

Table 2
Coding Categories and Citation Examples

Category	Code	Criteria
Accurate	0	• Cited claim is explicitly verified/verifiable in the original document in terms of key findings such that the cited claims in the target article correspond precisely to the findings of the original article. • No missing key qualifiers such that the cited findings in the target article did not omit any details, nuances, or caveats from the findings of the original article. • Example: If the original article found a significant effect for boys but not girls, then the target article explicitly notes that the effect was only significant for boys rather than stating a significant effect more broadly. • Example: If the original article found a significant effect for Latino immigrants but not for U.S.-born Latinos, then the target article explicitly notes that the effect was only significant for immigrants rather than stating a significant effect for Latino populations more broadly.
Somewhat accurate	1	• Missing key qualifier such that the cited claim in the target article is largely accurate and is related to the findings of the original article, but omits important details, nuances, or caveats found in the original article (e.g., overgeneralized a specific finding to a larger group, overspecifies a general finding to a specific group, correctly reports the nature of a relationship but provided inaccurate statistics from the original article). • Example: The target article cites a finding about *adolescents* but findings from the original article were based on youth in *early childhood*. In this sense, although the claim in the target article regards youth in general, it is "somewhat accurate" because it fails to specify that findings from the original article were based on youth in a different developmental period. • Example: A target article that is focused on an HIV positive population cites findings from the original article that is based on a convenience student sample, many of which were not HIV positive. In this sense, the citation is only "somewhat accurate" because although related to HIV, the authors failed to specify the very different sample on which the original findings were based.
Inaccurate	2	• Cited claim reported in target article do not correspond to the actual claims found in the original article such that (a) findings in the target article were not assessed and thus nonexistent in the original article or (b) the result provided in the target article directly contradicted the result from the original article (e.g., opposite direction of relationship). • Example: The target article cites findings that a certain coping style among Latinos is protective against the adverse effects of discrimination. However, the original article from which findings are cited did not assess coping styles (nonexistent in the original article). • Example: The target article cites findings that there was significant variation found in ethnic/racial socialization among parents that are passed to their children, whereas the original article found no variation in ethnic/racial socialization (contradicted the finding of the original article).

Note. See Supplemental Materials, for additional examples.

(a) omit important aspects of an article's key findings or (b) completely misrepresent prior findings altogether. Perhaps the most alarming finding is that approximately one in every 10 citations is completely inaccurate. Our results thus yield important implications for the reliability of scholarly writing in psychology and the credibility of psychological science.

Findings from the present study parallel those in other disciplines that found similar rates of miscitation, from 11% to 41% in the medical literature (e.g., Fenton et al., 2000; Goldberg et al., 1993; Pavlovic et al., 2021), to 25% in general science journals (Smith & Cumberledge, 2020), to 12%–25% in education journals (Lazonder & Janssen, 2022; Martella et al., 2021). It is encouraging to see that the proportion of inaccurate citations in psychological science appears to be lower than those in many other recognized disciplines. However, for a field that has long strived toward becoming a serious scientific discipline (e.g., Coyne, 2016; Lilienfeld, 2010, 2012; Schmidt & Oh, 2016), it is critical for scholars in the field to recognize the problem of miscitation and work to ensure that characterizations of prior research findings are accurate.

As noted earlier, if unchecked, miscitation can carry potentially serious implications for psychological science. Apart from constituting an ethical imperative (APA, 2017, 2020;

Office of Research Integrity, 2022), consequences associated with miscitation may include, but are not limited to, a proliferation of misrepresented findings that build upon one another, the development of theories that are constructed based on dubious empirical claims, funding decisions reached by accepting citing claims at face value, and policy decisions that are thought to be grounded in evidence but that are in fact based on miscited evidence. Miscitation of prior work can also undermine public trust in psychological science because incidents of misrepresented research, particularly in high-impact journals, can cast doubt on the credibility of the field and of those scientific journals in which these misrepresentations appear. Moreover, the degree of miscitation that we found occurred in leading psychology journals. This raises the question that if we found considerable miscitation in leading journals, what might the degree of miscitation be in less prestigious outlets? Future research is needed to compare how rates of miscitation might vary between higher versus lower impact psychological science journals.

Supplemental analyses indicated that the degree of miscitation was not significantly related to either number of authors on an article or to the first author's seniority level. These findings are informative because they suggest that

Table 3
Proportion of Citation Accuracy by Journal and Overall

Journal/discipline	0 (Accurate)	1 (Somewhat accurate)	2 (Inaccurate)	% Accurate to overall	% Somewhat accurate to overall	% Inaccurate to overall	% Somewhat accurate and inaccurate to overall
Cognitive Psychology	118	15	10	82.5	10.5	7	17.5
Journal of Counseling Psychology	324	37	28	83.3	9.5	7.2	16.7
Journal of Consulting and Clinical Psychology	119	28	20	71.3	16.8	11.9	28.7
Developmental Psychology	503	37	89	80	5.9	14.1	20
Journal of Personality and Social Psychology	483	61	55	80.6	10.2	9.2	19.4
Journal of School Psychology	227	46	24	76.4	15.5	8.1	23.6
Psychological Science	243	28	18	84.1	9.7	6.2	15.9
Cultural Diversity and Ethnic Minority Psychology	701	59	74	84	7.1	8.9	15.9
Overall	2,718	311	318	81.2	9.3	9.5	18.8
Mean %				80.3	10.6	9.1	19.7

miscitation in psychological science appears to be more of a general problem, rather than one that is based on certain characteristics such as number of authors or author seniority. Thus, our findings indicate that researchers at all levels of experience are susceptible to misciting prior work and that having more authors on an article may not prevent or reduce the degree of miscitation that occurs.

To be sure, the presence of miscitation in psychological research does not necessarily suggest that the results from a given empirical study are invalid. It is quite possible that an author can mischaracterize findings from prior work but still pose valid hypotheses that are corroborated by subsequent empirical analysis. However, the use of miscitation to construct a scientific argument can cast doubt on the reliability of the remainder of one's study. The research design and results of a study can be valid even as the scientific premise and foundation of the study are questionable because of miscited claims in the Introduction section of the article and in the justification for the study itself. Nevertheless, even isolated incidents of miscitation, especially when discovered, may cause one to question the quality of psychological research and whether the larger discipline (or journal) is credible. Moreover, it is difficult to ascertain the degree to which the miscitation damages the scientific progress of psychological science. Research has yet to identify the relationship between inaccurate characterizations of prior published research and the scientific advancement of a discipline. These caveats notwithstanding, miscitation still has the potential to harm the field and undermine longstanding efforts for psychology to establish itself as a credible scientific discipline.

To right the proverbial ship, we offer several recommendations for the field that may help to reduce miscitation in psychological science. These recommendations are not "fix all" solutions to the problem of miscitation. Rather, they represent certain research practices that, if followed, can help scholars to be more accurate in their reporting of prior research as well as to protect against additional criticisms that could further harm the scientific status of the field.

Recommendation 1: Recognize the Importance of Citation Accuracy for Psychological Science

To date, there has been little to no research on miscitation in psychological science. Instead, most attention has been on other QRPs that can undermine the scientific integrity of the field. These QRPs include heavy reliance on null hypothesis significance testing, reporting only significant findings, and cherry-picking evidence to support one's argument, to name just a few (Banks et al., 2016; Cumming, 2008; John et al., 2012; Krueger, 2001). This emphasis on other QRP's notwithstanding, proper citation is one of the most foundational assumptions across products generated by psychological science (e.g., journal articles, books, grant proposals,

conference presentations), and unchecked miscitation is a problem in psychology that can exert far-reaching consequences. The first step to address the problem of miscitation is to recognize its prevalence and understand its potential implications. Although some may consider accurate reporting of scientific findings as common knowledge, the fact that nearly one in 10 citations across leading journals in the present study were completely inaccurate suggests that many scholars in the field continue to miscite prior research.

Recommendation 2: Cross-Check Citing Claims to Ensure They Are Accurate

According to governing bodies in the field (APA, 2017), psychologists have an ethical obligation to promote accuracy, honesty, and truthfulness in the science and practice of psychology. This principle indicates that psychologists have a responsibility to accurately represent their colleagues' research findings and to ensure that the findings they cite clearly align with the original results of the cited studies. For authors, this responsibility entails reviewing the articles that authors cite and understanding the findings that they incorporate into their articles. As noted within the *Publication Manual of the American Psychological Association* (APA, 2020), citing prior research assumes that the authors understand the work they are citing. For editors and reviewers (e.g., for journal articles, grant proposals, conference abstracts), however, it would be unfeasible to cross-check every citation with the original article. In these cases, we recommend explicitly noting the potential issue of miscitation and encouraging authors to ensure that the research they have cited is accurate. In this sense, the editor or reviewer may raise this point to authors that miscitation represents an underappreciated potential threat to the validity of psychological science and request that they check to ensure that all citations are correct.

Recommendation 3: Cite Primary Sources Rather Than Relying on Secondary Sources

A commonly suspected source of miscitation involves heavy reliance on secondary rather than primary sources for citation. A recent survey among psychologists and education scientists (Klitzing et al., 2018) indicated that approximately one third reported relying on secondary rather than primary sources of citations. When authors rely on secondary sources to support a claim, they take the citing authors' representation of prior findings at face value. That is, authors assume that an author's characterization of prior research is accurate, perhaps because it is published in the scientific literature. However, when possible, authors should always consult the primary source to ensure that they are reporting the findings accurately. Because published work is typically derivative—present work builds on prior work—inaccurate secondary citations can result in a proliferation of false

beliefs that are subsequently adopted and carried forward by other scholars in the field (Bareket et al., 2020). Although consulting primary sources requires a greater time investment, such a practice is necessary to promote accurate reporting of scientific findings in the literature. Relatedly, authors may consider avoiding citing large-scale review articles rather than citing the primary sources themselves. Citing review articles, rather than the primary source, takes some of the credit away from authors of the original studies and assumes that the authors of the review article accurately characterized the original findings.

Recommendation 4: Consult the Method and Results Sections of Cited Articles, Rather Than Rely on Summaries in an Abstract Section

In the digital age, it is easy to search for a topic area on *Google Scholar* or *PsycINFO*, identify an article, glance at the abstract of an article, and cite the source. However, such reflexive reliance on information presented in an abstract can be problematic for several reasons. Perhaps the most concerning problem is that authors do not always include complete information vis-à-vis the primary results of a study, which can lead the citing researcher to exclude important information (e.g., qualifiers, effect sizes, design considerations). Many journals, for example, have restrictive word limits (e.g., 100–150 words) that do not allow researchers to provide complete information in the Abstract. Although Klitzing et al. (2018) found that several authors do consider the Method and Results sections of the articles they cite, others reported using the abstract as their primary or exclusive source of information. We recommend that, when authors are not very familiar with a research article, they should examine the Method and Results sections to ensure that they do not misrepresent a study's findings. Of course, we are not suggesting that abstracts are necessarily unreliable sources of information, per se. Rather, we are suggesting that abstracts are often incomplete and do not—and generally cannot—provide a full account of the study's findings.

Recommendation 5: Be Clear and Precise When Reporting the Original Findings of a Study

A common issue that we encountered in our analysis of miscitation was for authors to characterize findings from prior work in imprecise terms. For example, in several instances, authors described findings from prior work vaguely, such that the claim was related to the topic of the cited study but did not necessarily align with the study's findings. Many authors overgeneralized findings to larger populations when results from the cited study represent specific results (e.g., effects found for young children but generalized to adolescents). This means that miscitation can occur due to failing to include qualifying nuances from the

results of cited articles (e.g., effects emerging for boys but not for girls) as well as failure to note the boundary conditions of a study's results. Another example is citing meta-analyses to substantiate a relationship between two variables. Although meta-analyses may note the size of a relationship among variables, they often also include many studies that did not find significant relationships. Indeed, a key purpose of meta-analytic research is to identify important moderators of findings across studies and to note how the nature of a study's results, including the effect sizes that emerge from those studies, may vary as a function of those moderators. Yet, researchers often cite a meta-analysis as evidence of a general relationship, when in fact that relationship might not exist across some studies and for some populations. Eagly (2011) provides a good example of how meta-analyses have been frequently miscited to skew, and even contradict, findings from the original studies. Thus, it is important for authors to use clear and precise language when summarizing or citing prior research findings. Precision of language avoids vague descriptions and provides readers with an informed understanding of a cited study's findings.

Caveats and Considerations

The present study represents a massive undertaking in that we analyzed the accuracy of 3,347 empirical citations from 89 articles across eight psychology disciplines. Because ours was the largest study of its kind to date, there are several caveats to consider regarding the interpretation and the social import of our results. We offer these caveats to provide transparency regarding the limitations and boundary conditions of our findings.

First, all coding procedures in our study consisted of ratings from two separate coders. Any time two or more individuals are involved in a coding process, there exists the possibility that they (or other raters) may view and interpret results from studies differently. Thus, our findings may be indicative, in part, of how our research team operationalized and assessed miscitation—and scholars should keep this in mind when interpreting our results. This limitation notwithstanding, we were rigorous in our approach and followed established procedures for our coding process and analysis. Specifically, our process occurred over a period of 8 months and entailed two independent coders who had no contact with one another during the coding process. Moreover, we employed established statistical procedures to assess the degree of reliability between coders, including weighted kappa, which accounts for citations that may have been agreed upon by chance. Across all reliability indices, codes evidenced high degrees of reliability. In addition, at least in our estimation, the issue of interrater subjectivity in coding is more problematic in qualitative studies that analyze larger bodies of textual information (e.g., long excerpts of text). In the present study, however, we were focused specifically on whether a cited empirical claim

aligned with the results reported in the original study. In this sense, we believe that interrater subjectivity poses less of a threat to the validity of our findings than what might be the case for purely qualitative studies analyzing larger bodies of text.

Another consideration vis-à-vis the present study is that raters were not experts in several of the topic areas in which citations were reviewed. Thus, it is possible that having less expertise in a content area may have influenced how raters evaluated citation accuracy. It is also possible, however, that having substantial expertise in a topic area could also influence how one might evaluate the accuracy of a citation—strong expertise in a content area may lead to a preference for a certain finding or theoretical approach. Perhaps a consideration in this regard is that raters in the present study did not evaluate empirical support for a theory or content area but rather assessed whether citing claims matched findings from the original articles. In this sense, we believe that being an expert in a topic area was likely less of an issue given our focus on assessing empirical (statistical) claims. Nevertheless, because the degree to which expertise impacts the evaluation of citation accuracy is unknown, we cannot rule out this possibility.

Furthermore, analysis from the present study relied on coding criteria developed by our research team. The rationale behind this decision is that, to date, there is no validated measure with which to assess miscitation. Accordingly, it was necessary to develop verifiable criteria to assess the degree to which citing claims aligned with results from original articles. Although the coding categories in the present study were developed by the research team, we erred on the side of parsimony to identify the least number of variable categories possible. For this reason, we noted that citing claims may either be accurate, inaccurate, or fall somewhere in between (somewhat accurate). To err on the side of caution, we labeled the first two categories accurate and somewhat accurate, and the final category Inaccurate. This strategy permitted the research team more easily determine whether a citing claim is either (a) accurate to some degree or (b) completely inaccurate. With this strategy, we believe that most reasonable researchers could compare a cited empirical claim to the findings from the original article and determine whether it is completely inaccurate or whether it falls into the accurate or somewhat accurate categories.

Finally, we developed several recommendations to guide researchers in providing more accurate citations. As with most recommendations in the social sciences, these recommendations are not perfect, and none of them will completely solve the problem of miscitation in psychological science. Thus, we do not claim to provide a "catch-all" solution to miscitation. However, we believe that these recommendations have the potential to improve scientific reporting and assist the field of psychology in becoming a stronger scientific discipline.

Conclusion

As the discipline of psychology strives toward becoming a more rigorous science, it is important for the field to note areas in need of work and to improve upon them. Although there are myriad QRPs that scholars have noted as problematic over the years, miscitation represents one questionable practice that has been highly neglected in the empirical literature. This dearth of research attention is present despite the potentially far-reaching implications of miscitation for the reliability, accuracy, and trustworthiness of scholarly writing in psychological science. Our study is one of the only to date to consider miscitation in psychological science and, to our knowledge, is the most extensive across any discipline (3,347 citations across 89 articles). For the most part, authors of published articles in psychology have accurately cited and characterized findings from prior research. However, the finding that nearly 20% of citations either left out important nuances or mischaracterized prior work altogether is concerning. We have offered several concrete recommendations to assist the field in providing more accurate citations, in terms of citation practices to adopt as well as to avoid. We believe that accurate representation of scientific work is critical to strengthening psychology's position as a rigorous science, and we hope that our study takes an important step in that regard. As noted by Gupta et al. (2005, p. 140), "To be useful, references need to be cited and quoted correctly. References are akin to mortar, which not only binds the bricks together in a wall but also lends it the most vital things, (i.e., strength and durability)."

References

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct*. Retrieved August 9, 2022, from https://www.apa.org/ethics/code#:~:text=Principle%20C%3A%20Integrity,or%20intentional%20misrepresentation%20of%20fact

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). https://doi.org/10.1037/0000165-000

Armstrong, M. F., Conduff, J. H., III, Fenton, J. E., & Coelho, D. H. (2018). Reference errors in otolaryngology–head and neck surgery literature. *Otolaryngology—Head and Neck Surgery*, *159*(2), 249–253. https://doi.org/10.1177/0194599818772521

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

Bareket, R., Schonberg, M. A., & Schonmann, Y. (2020). Make quotations great again: A proposal to reduce false-knowledge. *BMJ Evidence-Based Medicine*, *25*(1), 12–14. https://doi.org/10.1136/bmjebm-2018-111109

Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, *16*(4), 756–766. https://doi.org/10.1177/1745691620969647

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213–220. https://doi.org/10.1037/h0026256

Coyne, J. C. (2016). Replication initiatives will not salvage the trustworthiness of psychology. *BMC Psychology*, *4*(1), Article 28. https://doi.org/10.1186/s40359-016-0134-3

Cumming, G. (2008). Replication and p intervals: *P* values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, *3*(4), 286–300. https://doi.org/10.1111/j.1745-6924.2008.00079.x

Eagly, A. H. (2011). A mis-citation classic. In R. M. Arkin (Ed.), *Most underappreciated: 50 prominent social psychologists describe their most unloved work* (pp. 250–253). Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199778188.003.0050

Ecker, U. K. H., Swire, B., & Lewandowsky, S. (2014). Correcting misinformation—A challenge for education and cognitive science. In D. N. Rapp & J. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 13–38). MIT Press.

Fenton, J. E., Brazier, H., De Souza, A., Hughes, J. P., & McShane, D. P. (2000). The accuracy of citation and quotation in otolaryngology/head and neck surgery journals. *Clinical Otolaryngology and Allied Sciences*, *25*(1), 40–44. https://doi.org/10.1046/j.1365-2273.2000.00322.x

Goldberg, R., Newton, E., Cameron, J., Jacobson, R., Chan, L., Bukata, W. R., & Rakab, A. (1993). Reference accuracy in the emergency medicine literature. *Annals of Emergency Medicine*, *22*(9), 1450–1454. https://doi.org/10.1016/S0196-0644(05)81995-X

Gupta, P., Yadav, M., Mohta, A., & Choudhury, P. (2005). References in Indian Pediatrics: Authors need to be accurate. *Indian Pediatrics*, *42*(2), 140–145.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Haussmann, N. S., McIntyre, T., Bumby, A. J., & Loubser, M. J. (2013). Referencing practices in physical geography: How well do we cite what we write? *Progress in Physical Geography*, *37*(4), 543–549. https://doi.org/10.1177/0309133313482135

Hicks, R. W. (2021). How accurate are your citations? *Journal of the American Association of Nurse Practitioners*, *33*(9), 667–669. https://doi.org/10.1097/JXX.0000000000000645

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, *2*(8), Article e124. https://doi.org/10.1371/journal.pmed.0020124

Jergas, H., & Baethge, C. (2015). Quotation accuracy in medical journal articles-a systematic review and meta-analysis. *PeerJ*, *3*, Article e1364. https://doi.org/10.7717/peerj.1364

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

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1420–1436. https://doi.org/10.1037/0278-7393.20.6.1420

Klitzing, N., Hoekstra, R., & Strijbos, J. W. (2018). Literature practices: Processes leading up to a citation. *Journal of Documentation*, *75*(1), 62–77. https://doi.org/10.1108/JD-03-2018-0047

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Krueger, J. (2001). Null hypothesis significance testing. On the survival of a flawed method. *American Psychologist*, *56*(1), 16–26. https://doi.org/10.1037/0003-066X.56.1.16

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lazonder, A. W., & Janssen, N. (2022). Quotation accuracy in educational research articles. *Educational Research Review*, *35*(1), Article 100430. https://doi.org/10.1016/j.edurev.2021.100430

Lilienfeld, S. O. (2010). Can psychology become a science? *Personality and Individual Differences*, *49*(4), 281–288. https://doi.org/10.1016/j.paid.2010.01.024

Lilienfeld, S. O. (2012). Public skepticism of psychology: Why many people perceive the study of human behavior as unscientific. *American Psychologist*, *67*(2), 111–129. https://doi.org/10.1037/a0023963

Lukić, I. K., Lukić, A., Gluncić, V., Katavić, V., Vučenik, V., & Marusić, A. (2004). Citation and quotation accuracy in three anatomy journals. *Clinical Anatomy*, *17*(7), 534–539. https://doi.org/10.1002/ca.10255

Lykken, D. T. (1991). What's wrong with psychology anyway? In D. Cicchetti & W. M. Grove (Eds.), Thinking clearly about psychology: Vol. 1. *Matters of public interest* (pp. 2–39). University of Minnesota Press.

Martella, A. M., Yatcilla, J. K., Martella, R. C., Marchand-Martella, N. E., Ozen, Z., Karatas, T., Park, H. H., Simpson, A., & Karpicke, J. D. (2021). Quotation accuracy matters: An examination of how an influential meta-analysis on active learning has been cited. *Review of Educational Research*, *91*(2), 272–308. https://doi.org/10.3102/0034654321991228

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*(2), 103–115. https://doi.org/10.1086/288135

Miller, G. A. (2004). Another quasi-30 years of slow progress. *Applied & Preventive Psychology*, *11*(1), 61–64. https://doi.org/10.1016/j.appsy.2004.02.010

Mogull, S. A. (2017). Accuracy of cited "facts" in medical research articles: A review of study methodology and recalculation of quotation error rate. *PLOS ONE*, *12*(9), Article e0184727. https://doi.org/10.1371/journal.pone.0184727

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

Office of Research Integrity. (2022). *Citing sources that were not read or thoroughly understood*. Retrieved on August 9th, 2022, from https://ori.hhs.gov/citing-sources-were-not-read-or-thoroughly-understood

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, *7*(6), 657–660. https://doi.org/10.1177/1745691612462588

Pavlovic, V., Weissgerber, T., Stanisavljevic, D., Pekmezovic, T., Milicevic, O., Lazovic, J. M., Cirkovic, A., Savic, M., Rajovic, N., Piperac, P., Djuric, N., Madzarevic, P., Dimitrijevic, A., Randjelovic, S., Nestorovic, E., Akinyombo, R., Pavlovic, A., Ghamrawi, R., Garovic, V., & Milic, N. (2021). How accurate are citations of frequently cited papers in biomedical literature? *Clinical Science*, *135*(5), 671–681. https://doi.org/10.1042/CS20201573

Popper, K. R. (1959). *The logic of scientific discovery*. Basic Books.

Rivkin, A. (2020). Manuscript referencing errors and their impact on shaping current evidence. *American Journal of Pharmaceutical Education*, *84*(7), Article ajpe7846. https://doi.org/10.5688/ajpe7846

Rudmin, F. W. (2007). Acculturation alchemy: How miscitations make biculturalism appear beneficial. *Psykologisk Tidsskrift*, *11*, 4–10.

Schmidt, F. L., & Oh, I. S. (2016). The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else? *Archives of Scientific Psychology*, *4*(1), 32–37. https://doi.org/10.1037/arc0000029

Skinner, B. F. (1987). Whatever happened to psychology as the science of behavior? *American Psychologist*, *42*(8), 780–786. https://doi.org/10.1037/0003-066X.42.8.780

Smith, N., Jr., & Cumberledge, A. (2020). Quotation errors in general science journals. *Proceedings of the Royal Society Series A: Mathematical Physical and Engineering Sciences*, *476*(2242), Article 20200538. https://doi.org/10.1098/rspa.2020.0538

Tackett, J. L., Brandes, C. M., King, K. M., & Markon, K. E. (2019). Psychology's replication crisis and clinical psychological science. *Annual Review of Clinical Psychology*, *15*(1), 579–604. https://doi.org/10.1146/annurev-clinpsy-050718-095710

Tipton, E., Spybrook, J., Fitzgerald, K. G., Wang, Q., & Davidson, C. (2021). Toward a system of evidence for all: Current practices and future opportunities in 37 randomized trials. *Educational Researcher*, *50*(3), 145–156. https://doi.org/10.3102/0013189X20960686

Todd, P. A., Guest, J. R., Lu, J., & Chou, L. M. (2010). One in four citations in marine biology papers is inappropriate. *Marine Ecology Progress Series*, *408*, 299–303. https://doi.org/10.3354/meps08587

Todd, P. A., Yeo, D. C., Li, D., & Ladle, R. J. (2007). Citing practices in ecology: Can we believe our own words? *Oikos*, *116*(9), 1599–1601. https://doi.org/10.1111/j.0030-1299.2007.15992.x

Supplemental Materials

Additional Examples of Citation Coding Procedures Taken from Target Articles

Accurate

Example 1:
Claim: Munley (1974) published an early content analysis of JCP. He coded articles published over 19 years (1954–1972) into 17 content areas. The most prominent content area he observed was research on counseling process and counseling outcome, which together made up 25% of the publications: outcome research accounted for 15%, process research accounted for 9%, and combined process and outcome research accounted for 1%. This category was followed by research on vocational behavior (18% of publications) and research on the development and validation of measures (17% of the publications). Support: Classified articles and brief reports appearing in the 1st 19 volumes of the Journal of Counseling Psychology into 17 content categories and recorded certain methodological features from each research article.; Considering related content categories together, counseling outcome and process research account for the largest percentage of empirical research published (25%) , with outcome research leading process research (15% versus 9%). Qnly a small of the research reviewed studied outcome - and process variables together (1%). Research on vocational behavior and research on tests and measurements account for the second and third largest percentages, 18% and 17%, respectively. Code: 0 Rationale: The cited findings in the target article correspond directly and precisely to the findings of the original article.

Somewhat Accurate

Example 1: JCCP
Claim: "TJA is also a risk factor for the development of chronic opioid use" Support: "This study evaluated perioperative opioid use patterns in patients who underwent TKA from 2007 to 2013. Although prescription narcotic use tended to diminish by 6 months after surgery, over one-third of patients who used opioids preoperatively continued to use them chronically after 6 months. In contrast, a minor fraction of patients who were opioid-naïve before surgery continued opioid use chronically." Code: 1

Rationale: Missing qualifier that the risk factor for chronic opioid use, rather than for TJA specifically, was for people who were already using them preoperatively.

Inaccurate

Example 1: JCP Article 2

Claim:

"Research has consistently demonstrated that community connectedness is related to decreased depression and anxiety"

Support:

"To examine the unique effects of each support variable when other support variables were controlled for, three multivariable regressions were conducted in which all three social support variables were entered as simultaneous predictors of anxiety and depression symptoms and resilience (Supplemental Table 2). In these analyses age, sex assigned at birth, income, education, gender identity, and sexual identity were included as covariates. Results indicated that having more family support continued to predict less depression and anxiety symptoms and more resilience even when support from friends and community connectedness were controlled. Similarly, support from friends continued to predict less depression and anxiety symptoms when support from family and community connectedness were controlled for. However, community connectedness did not predict anxiety and depression symptoms or resilience in the multivariable models. None of the three-way interactions of the support variables were significant, indicating that the effects of the social support variables were consistent across levels of the other social support variables"

Code: 2

Rationale: Support article explicitly states that community connectedness was not predictive of outcome variables which is a direct contradiction to the claim made in original article.

Examples of Citation Themes for Somewhat Accurate and Inaccurate Categories Taken from Target Articles

Somewhat Accurate

Overgeneralization: The original study's findings were based on a specific population or process, whereas the citing claim generalizes this finding to a broader population or process that was not assessed in the original study
Example Claim: Other-oriented processes and their influence on support giving have been studied among people motivated by individualistic values Example Support: The current study employed a 14-day dyadic daily-experience paradigm to measure daily sacrifices in romantic relationships. We predicted that individuals higher in communal strength toward their romantic partners would experience more positive emotions when making daily sacrifices, would feel more appreciated by their partners for the sacrifices, and would experience greater relationship satisfaction. Furthermore, we predicted that the reason why individuals higher in communal strength experience more delight when sacrificing for their partner is because they feel that their sacrifices are an authentic reflection of their true selves... we found that the higher individuals were in communal strength toward their romantic partner, the more they experienced positive emotions during daily sacrifices, felt appreciated for their efforts, and reported high relationship satisfaction on days when they made sacrifices. Code: 1 Rationale: The original study was in regard to romantic relationships while the cited claim regards general support giving – not that of romantic relationships
Overspecification: The original study's findings were based on a general population or process, whereas the cited claim applies that finding to a specific population or process that was not directly assessed in the original study
Example Claim: A core characteristic of collectivism that is particularly salient among Chinese individuals is prioritizing the needs of in-group members and concern for one's own social standing Example Support: We achieved three aims by developing a new instrument that assesses collectivism among Asians and validating it in two independent samples. First, the BCQ appears to comprise a set of items that appropriately represent collectivism from an individual psychological standpoint. Second, our results yielded preliminary empirical support for a bifactor structure of collectivism that is consistent with

emerging research...... Specifically, people of Asian descent seem to differentiate being proactive in prosocial tendencies from being cordial, collaborative, and deferential to maintain harmony. Additionally, incremental to broadband collectivism, there seems to be an affective component in collectivism with regard to people's concerns for how others perceive oneself (both the individual and the extended self). Although honoring family has not been included consistently in measuring collectivism in Asian cultures, our findings show that this content is an important part of the broader construct. Collectivism also seems to be represented adequately by items that tap the concern for one's reputation in which the self is best understood by both the individual and the family.

Code: 1

Rationale:

The claim specifically indicates the association for Chinese individuals, however the cited study does not differentiate between Asian countries

Double Barreled: The cited claim makes two claims within a single statement, whereas the original study only supported one of those claims.

Example Claim:

Avowing a different identity label than an ascribed one represents a marginalized group defining and creating a new reality that empowers its members and generates a sense of pride

Example Support:

Among survey respondents who expressed a preference for one term or the other, preference for the label "black" or "African-American" was nearly equally divided. Still, several interesting cleavages emerged among respondents. Gender and level of education did not influence terminological preference, but age, city size, and region did, with younger residents of large cities outside the South expressing more enthusiasm for the label "African-American." Moreover, those whose grammar schools had been half black and half white, as opposed to all or mostly black or almost all white, and those for whom racial identity was more important were significantly more likely than others to prefer the term "African-American" to "black."

Code: 1

Rationale:

The cited study found that there are differences in preferred labels but did not expand on how the choice in labels effects an individual's sense of empowerment or pride. Pride was not assessed in this context.

Miscellaneous: The cited claim is largely accurate but there are minor errors in the statement or in the reporting or results from the original study.

Example Claim:

The family context, however, is often the most influential context in which youth learn about the heritage culture

Example Support:

Explored the role of the ethnic family background and ethnic socialization in the social cognitive development of ethnic identity in Mexican-American children… Individually administered scales assessed parental generation of migration; parental education; mothers' cultural orientation; mothers' teaching about Mexican culture, ethnic pride, and discrimination; Mexican objects in the home; and children's ethnic identity. As predicted, the socialization indices functioned as a mediator of the influence of ethnic family background on their children's ethnic identity.

Code: 1

Rationale:

The cited study found family context to be influential, however, this context was not compared to others so the claim that it is the most influential context could not be evaluated

Inaccurate

Nonexistent in Original Study: The cited claim was not assessed in any way in the original study

Example Claim:

Youth who experience more bullying/harassment offline (such as being verbally or physically attacked about some physical, mental, or social trait) are at heightened risk for worse mental, physical, and academic outcomes over time than youth who have experienced less victimization.

Example Support:

Compared to uninvolved adolescents, cybervictims and cyberbully-victims were significantly more likely to report symptoms of depression (cybervictims: odds ratio [OR] = 1.44, 95% confidence interval [CI] [1.00, 2.06]; cyberbully-victims: OR = 1.54, 95% CI [1.13, 2.09]) and social anxiety (cybervictims: OR = 1.52, 95% CI [1.11, 2.07]; cyberbully-victims: OR = 1.44, 95% CI [1.10, 1.89]) but not below average well-being (cybervictims: relative risk ratio = 1.28, 95% CI [.86, 1.91]; cyberbully-victims: relative risk ratio = 1.38, 95% CI [.95, 1.99]) at 1 year follow-up, after adjustment for confounding factors including baseline mental health.

Code: 2

Rationale:

The claim specifically regards offline bullying, the cited study does not provide support for this claim as it only measured cyber bullying

Contradiction: The cited directly contradicts the finding from the original study.

Example Claim:

Some research, including that conducted using adolescent samples supports an immigrant paradox, whereby risk behaviors are more prevalent among later-generation, compared to first generation immigrant Latinos/as

Example Support:

We used data from 1466 youth (ages 8–16) enrolled in the Hispanic Community Health Study of Latino Youth (SOL Youth), a probability sample of Hispanic/Latino youth living in Chicago (IL), Miami (FL), Bronx (NY), and San Diego (CA). We found no evidence of an immigrant paradox.

Code: 2

Rationale:

The cited claim states that evidence for an immigrant paradox was supported, whereas the original study explicitly states that there was no evidence for an immigrant paradox