The Open Science Movement and Clinical Psychology Training: Rigorous Science is
Transparent Science

Donald R. Lynam[1], David A. Sbarra[2], Jennifer Tackett[3], Priscilla Lui[4], Riley McDanal[5],
Katherine Schaumberg[6], Aidan G.C Wright[7], Yexinyu Yang[8], & Joshua D. Miller[9]

1 Department of Psychological Sciences, Purdue University

2. Department of Psychology, University of Arizona

3 Department of Psychology, Northwestern University

4 Department of Psychology, University of Washington

5 Department of Psychology, Stony Brook University

6 Department of Psychology, University of Texas at Austin

7 Department of Psychology, Department of Psychiatry, Eisenberg Family Depression Center,
University of Michigan

8 Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill

9 Department of Psychology, University of Georgia

*Author order 4 through 8 was determined alphabetically.

Corresponding author: Donald R. Lynam; dlynam@purdue.edu

Abstract

Clinical psychology, like other disciplines, is facing a replication/credibility crisis that undermines the evidentiary basis of our science. The Open Science Movement (OSM) offers solutions through practices such as preregistration, Registered Reports, and sharing of data, materials, and code. Clinical psychology has been slow to adopt these reforms, leaving trainees underprepared for emerging norms. This article reviews the factors that contributed to the crisis, outlines the necessary role of transparency in distinguishing rigorous from flawed research, and documents the limited uptake of open science practices in clinical journals and graduate training. We argue that transparency is an ethical as well as methodological imperative and propose a competency-based model for embedding open science principles into doctoral education. We conclude by calling on programs, journals, and accrediting bodies to make transparency a core requirement, essential for restoring credibility and advancing a cumulative, trustworthy clinical science.

Clinical psychology, like other areas of psychology, is facing a replication or credibility crisis[1]—a recognition that the empirical literature is on shakier ground than previously thought. Hengartner (2018) reviewed the state of psychotherapy research and argued that many of its published efficacy trials are riddled with biases that systematically inflate treatment effects. He concluded that the replication crisis is "no less pernicious and prevalent in clinical psychology" than in social psychology, underscoring how fragile much of the clinical evidence base may be. These concerns about the replicability of findings and the credibility of the field gave rise to the Open Science Movement (OSM)--a broad shift toward greater transparency, rigor, and reproducibility. The OSM emphasizes structural reforms—like preregistration, Registered Reports, provision of open data, materials, and code (whenever possible) and reproducibility checks—to build more transparent, trustworthy, self-correcting, and evaluable research practices. Unfortunately, clinical psychologists have been slow to join the OSM (Korbmacher et al., 2023; Tackett et al., 2019). This reluctance by the field of clinical psychology sets the next generation of clinical psychological scientists at a clear disadvantage.

In what follows, we explore why clinical psychology lags behind other sub-disciplines in joining the OSM, and what we can do about it. To move forward, clinical scientists must first understand the factors and practices that led to the crisis. We then outline a set of proposed solutions and discuss what they can and cannot be expected to do. Next, we propose changes for clinical psychology with a focus on doctoral training. At the heart of this proposal is this: graduate students in clinical psychology are eager for deeper training and more experience in open science (Van Til et al., 2025). Because transparency, which open science practices help to

---

[1] We use the terms the replication crisis and the credibility crisis somewhat interchangeably throughout. Both refer to concerns about the trustworthiness (e.g., robustness, believability) of the published literature and of the people who produce that literature (Vazire, 2018).

achieve, is a necessary condition for high-quality, ethical research, training programs that do nott

address these issues, are not meeting students' needs. Finally, we consider barriers to

implementing this training and outline ways these barriers might be addressed.

The Replication Crisis

We are approaching the 15-year anniversary of a confluence of events that catalyzed

what is now termed the "replication crisis" in psychological science. The history of this crisis is

well documented (Nelson et al., 2018) but we revisit key moments to highlight the *need* for the

OSM. In 2011, Bem published a series of studies seeming to indicate that precognition (i.e., the

ability to predict the future) existed in one of psychology's premier journals. The fact that he

could demonstrate statistically significant findings in eight of nine studies for a phenomenon

most empiricists do not believe exists raised grave concerns about his use of common but

problematic practices (e.g., selective reporting of studies and outcomes; outcome swapping, p-

hacking, hypothesizing after results are known [HARKing]; See Table 1 for a glossary of terms

used throughout this paper). Also in 2011, it was discovered that well-known social psychologist

Diederik Stapel had committed extensive fraud across his publications (Callaway, 2011). Finally,

in that same year, Simmons and colleagues (2011) demonstrated the impact of a set of field-wide

standard operating procedures (e.g., p-hacking) on false-positive rates. Specifically, they showed

that false positive rates were strongly influenced by the researcher degrees of freedom available

in most studies (e.g., decisions about when to stop data collection, whether to include covariates;

selectively choosing which results to report). In 2015, the Open Science Collaboration (2015)

published a landmark report revealing alarmingly low replication rates in cognitive and social

psychology. This collaborative group attempted to replicate 100 studies in which 97 had

demonstrated statistically significant results in the original study. Despite generally having

higher statistical power, this group found only 37 replications were statistically significant; of note, the effect sizes in the replications were roughly half that reported in the original studies. These reports sent shockwaves through the field and forced attention on an entire suite of standard research practices and the incentive structures within the field, calling for scientific reforms to safeguard the credibility of the field.

The Four Horsemen

Bishop (2019) identified four key contributors to the crisis—collectively referred to as the "four horsemen of irreproducibility": low statistical power, publication bias, p-hacking, and HARKing. Although we will discuss them separately, p-hacking and HARKing fall under a broader umbrella of questionable research practices (QRPs), described in a recent paper by Nagy et al. (2025), as "… ways of producing, maintaining, sharing, analyzing, or interpreting data that are likely to produce misleading conclusions, typically in the interest of the researcher" (p. 1). Up to 50% of researchers across scientific disciplines report engaging in QRPs, highlighting the continued urgency to address problematic research practices (Nagy et al., 2025). All four horsemen undermine the credibility of scientific conclusions and the integrity of the research process.

Statistical Power

Statistical power, the probability of correctly rejecting the null hypothesis (H_0) when it is false in the population, is well below the recommended 80% minimum in most psychological research studies for plausible effect sizes in the field (Bakker et al., 2012; Swiatkowski & Dompnier, 2017). The problem with underpowered studies is not just the increased risk of Type II errors (i.e., failing to reject the null when it is false in the population); underpowered studies also distort the broader research record. Although low power does not increase the nominal Type

I error rate for any given test (alpha, typically .05), it reduces the likelihood of detecting true effects. As a result, the relative proportion of false positives in the published literature increases (Button et al., 2013; Ioannidis, 2005). Equally important, null results from underpowered studies are uninformative—they do not provide strong evidence against the presence of an effect and cannot resolve scientific questions (Maxwell, 2004). Thus, it is difficult to publish underpowered studies that yield predominantly null findings. This often prompts additional testing and engagement in QRPs (e.g., changing outcomes, adding variables to analyses, exploring subgroups, trying alternative operationalizations) in the search for statistical significance, increasing the overall number of statistical tests being conducted. As more tests accumulate— within a single study, across a program of research, or across the field, especially under conditions of widespread low power and publication bias—the proportion of published findings that are actually false positives rises sharply (Bakker et al., 2012; Ioannidis, 2005; Simmons et al., 2011). Finally, underpowered studies also contribute to overestimated effect sizes entering the literature, as only the inflated and statistically significant results are likely to be published (Ioannidis, 2008).

Bakker et al. (2012) used simulations to demonstrate how low statistical power and QRPs—such as selectively reporting significant results or flexibility in data analysis—can dramatically inflate the actual Type I error rate in psychological research. Although researchers typically assume an alpha of 0.05, their simulations showed that when multiple QRPs are used in studies with small sample sizes, the true false positive rate can easily exceed 60%. Although initial reports of power in clinical work suggest decent sample size (Reardon et al., 2019), a focus on only sample size in the context of power is incomplete as different analytic approaches have vastly different sample size requirements (e.g., main effects vs. interactive effects). In fact,

the mean sample size reported by Reardon and colleagues in two premier clinical journals (Ns = 175-182) is still below the level at which basic correlations stabilize (Schonbrodt & Perugini, 2013) and far below the point where more complicated statistics like interactions stabilize (Castillo et al., under review).

P-hacking

P-hacking involves conducting analyses on a dataset until a desired, typically statistically significant, effect is found or selectively reporting subsets of analyses that yielded desired outcomes and suppressing others that did not (Simmons et al., 2011). These practices inflate Type I error rates. In their original paper, Simmons et al. examined the impact of specific types of p-hacking—running analyses on two different dependent variables, adding observations in the face of non-significant results, controlling for gender or allowing gender to interact with treatment, and dropping (or not) one of three experimental conditions. Although each practice alone slightly inflated the probability of a Type I error above the nominal 5% level (e.g., 9.5%, 7.7%, 11.7%, and 12.6%), their combined influence led to a Type I error rate of 60.7%.

Hypothesizing After the Results are Known (HARKing)

HARKing (Kerr, 1998) is another QRP, which occurs when a researcher present post-hoc hypotheses (i.e., a hypotheses generated after engaging with the data) as though they were *a priori* hypotheses. HARKing comes in many flavors including creating a hypothesis consistent with the known results, presenting one or more hypotheses known post-hoc that contradict the data (giving the false appearance of hypothesis testing), or suppressing *a priori* hypotheses that were not statistically significant (Kerr, 1998). All flavors allow exploratory work to masquerade as confirmatory work so as to make the work appear consistent with the hypothetico-deductive method. Unfortunately, HARKing undermines the scientific method; if hypotheses are generated

from the data, rather than tested against the data, no genuine test of a theory has occurred.

HARKing also leads to Type I errors being published and translated into theory, causing the loss

of valuable information related to one's original hypothesis and the presentation of biased,

inaccurate models of science (Murphy & Aguinis, 2019).

Publication Bias

Publication bias refers to a general bias in journals and among researchers for significant

and sexy (e.g., novel; counter-intuitive) results and against null results (Greenwald, 1975).

Reluctance to publish null results from underpowered studies is understandable, but there is no

reason to prevent publication of null results from high-powered studies. These biases push

researchers to search for statistically significant results in their data—setting the stage for QRPs,

which increase the prevalence of Type I errors in the literature. This bias also leads to the "file-

drawer problem," in which researchers suppress null results themselves because they believe

they will not be published, or they conflict with previously published findings (Rosenthal, 1979).

Publication bias influences the outcomes of meta-analytic work because published results may be

inflated by selection bias for statistically significant findings (Coursol & Wagner, 1986). These

biases can lead to the expenditure of time and resources on phenomena with a façade of

empirical support because null results have been suppressed (e.g., ego depletion).

The practices described above have been decried for decades. Tukey wrote about the

problems with multiple comparisons 70 years ago. Cohen was writing about the problems of low

statistical power in the 1960s. Greenwald wrote about bias against the null hypothesis in 1975.

And Kerr wrote about HARKing in 1998. What changed after 2011 was the growing recognition

of the joint, cumulative impact of these practices which systematically distort the published

literature, give rise to inflated effects, false positives, and a misleading sense of credibility.

Not only have these practices harmed the field, many of these practices (e.g., p-hacking, HARKing) are inconsistent with APA's (and others; e.g., American Statistical Association) code of ethics and should be considered unethical not just questionable (Miller et al., 2025). Miller et al. (2025) argue that practices like HARKing and p-hacking represent clear violations of the APA Ethics Code, particularly its standards concerning integrity, scientific responsibility, and honesty. They emphasize that presenting post hoc hypotheses as if they were a priori misleads readers and reviewers, directly contravening Standard 8.10 on accurate reporting of results. Similarly, manipulating analyses to achieve statistical significance without full disclosure violates principles of transparency and openness embedded in the general principle of Integrity. These practices undermine trust in science, damage the cumulative knowledge base, and erode the professional obligation psychologists have to represent their work truthfully. Much as clinical psychologists work hard to avoid ethical violations in their clinical work (e.g., practicing outside one's competence; breaching confidentiality), they must work as assiduously to avoid ethical violations in their research practices. Importantly, clinical training programs must teach graduate students that such practices are unethical with the same level of attention and vigor as they instruct students about ethical practices in clinical applications.

The Open Science Movement (OSM)

This realization that multiple practices contributed to a less robust, empirically-grounded literature in psychology catalyzed the emergence of the OSM, which advocates a set of research practices designed to enhance the transparency, reproducibility, and credibility of psychological science. These practices include preregistration (e.g., Benning et al., 2019), the use of Registered Reports (e.g., Chambers & Tzavella, 2022), and the sharing of analytic code, statistical software output, study materials, and datasets.

Open Science Practices

Preregistration

Preregistration is among the workhorses of open science. It involves publicly

documenting a study's hypotheses, methods, and analysis plans, ideally *before* collecting or

accessing data.[2] This process helps distinguish between confirmatory and exploratory research

by creating a time-stamped, publicly accessible (or privately held) plan that can later be

compared to what was actually done. Typically submitted to platforms like the Open Science

Framework (OSF), a preregistration includes details such as the research question, a sample size

justification, exclusion criteria, variables to be measured, statistical analyses to be used, and

inferential criteria. The primary goal is to increase transparency and reduce QRPs like p-hacking

and HARKing. Importantly, preregistration *does not* prevent exploratory work – it simply

clarifies which findings were predicted in advance and which were discovered post hoc. Benning

and colleagues (2019) documented what they called the "registration continuum," noting that one

can and should write registrations even while data are being collected or even after data

collection is complete (but prior to analysis). Although preregistration prior to data collection is

ideal, it is not always possible given the common use of archival data in clinical psychology, and

there is still value to the registration of one's hypotheses and analytic plans after data have been

collected. Such registrations increase transparency. Templates exist for registering studies where

the data have already been collected that can help scholars navigate this process (Van den Akker

et al., 2021).

Registered Reports

---

[2] Although recent work recommends greater precision and a finer-grained distinction among registrations, study
protocols, analysis plans, and related outputs (Mayo-Wilson, Grant, Corker, & Moher, 2025), we use the term
preregistration broadly to encompass advance documentation of study protocols, hypotheses, analytic plans, and
statistical inference criteria (at least prior to data analysis).

Registered Reports are a publishing format designed to enhance transparency and reduce publication bias by conducting peer review *before* data collection, or, in the case of secondary data analysis, before analyses. Researchers submit a Stage 1 manuscript that includes the study rationale, hypotheses, detailed methodology, and proposed analyses.[3] This is reviewed by journal editors and peer reviewers, who evaluate the rigor and significance of the planned research—not the results. If accepted, the study receives in-principle acceptance (IPA), guaranteeing publication regardless of outcome, provided the researchers follow the approved protocol. After data collection and analysis, the completed manuscript (Stage 2) is submitted for final review to confirm adherence to the original plan and evaluate the interpretation of results. This format— often referred to as a "results free" peer review— reduces publication bias and incentivizes sound research design over flashy findings. It also removes incentives for p-hacking in that authors need not worry about finding statistically significant results.

Open Materials, Data, and Code

A final set of open science practices involves sharing the core components of a study—its raw data, analysis scripts, and study materials (e.g., surveys, stimuli, protocols)[4]—so others can evaluate, reproduce, and build upon the research. These materials are typically hosted on public repositories like the OSF, GitHub, or institutional archives, with appropriate documentation to ensure usability. Sharing data allows others to verify results and conduct secondary analyses; sharing code and raw output files ensures transparency in how results were derived; and sharing

---

[3] To be clear, not all journals in which clinical psychologists publish accept Registered Reports at this time. Although several premier journals do, such as *Clinical Psychological Science*, *Journal of Clinical and Consulting Psychology,* and *Journal of Psychopathology and Clinical Science*, others do not (e.g., *Psychological Assessment*). We believe and hope the adoption rate of this powerful tool will increase with time and exposure, as well as lobbying by authors, editorial board members, and even publishers.

[4] We do not discuss Open Access publishing here. Although it is often conceived of as part of a larger Open Science movement designed to make findings more accessible to scientists and the lay public, it is not a specific response to the replication/credibility crisis and it does not serve to make research more credible.

materials facilitates replication and adaptation in future studies. Open sharing promotes scientific credibility, accelerates progress, and aligns with growing norms and expectations in many fields and journals. Ethical considerations—such as participant privacy and intellectual property—must be addressed, but these are often manageable with de-identification, licenses, and embargoes (see Meyer, 2018). The importance of open sharing of all materials (i.e., data, code, and materials) is reflected by a recent change at the journal *Psychological Science*, where transparent practices, including the expectation that authors will make all primary data, original research material, and analysis scripts available in a third-party repository, are now the default (Hardwicke & Vazire, 2024). These changes also involve efforts by the journal to conduct reproducibility checks, ensuring all findings reported in a paper can be independently verified. To do this well, journals will have to invest more resources to support these activities (e.g., funds for consulting editors or other positions specifically tasked with checking correspondence between preregistrations and submissions, checking reproducibility, etc.). Given the profit margins for some academic publishers, greater investment in transparency and OS practices seems entirely feasible.

What Open Science Will (and Will Not) Do

Misconceptions as to what open science practices, particularly preregistration, will and will not do abound. It is important to be clear about what the field can expect if and, hopefully, when these practices become normative. Open science practices will not necessarily improve scientific rigor or replicability; that is, they will not necessarily improve research design, increase statistical power, or prevent researchers from capitalizing on researcher degrees of freedom (although specific, detailed preregistrations can curtail this to some degree).

What Open Science Will Not Do

Preregistration forces attention to several design issues—a priori hypotheses, power and

statistical precision, decisions regarding data screening (e.g., how outliers will be treated; whether data will be screened for response validity), covariate inclusion, multiple comparisons/analyses, etcetera. Attention to these issues may improve these aspects of the design. Bogdan (2025) recently reported that across the last 10 years, there has been an increase in the reporting of stronger p-values (i.e., p values < .01) in every subdiscipline of psychology. This presumably reflects an increase in statistical power and a decrease in p-hacking (or p-hacking to a more stringent alpha cut-off). But attention to these issues will not necessarily solve them. At best, preregistration might prevent HARKing, but this does not, by itself, lead to rigorous science. A researcher can always preregister an underpowered, poorly designed study with vague hypotheses, invalid and unreliable measures, inappropriate analyses, and non-specific inferential criteria (Rubin, 2020; Szollosi & Newell, 2020). Preregistration is not a sufficient condition for good science.

Registered Reports, on the other hand, have a greater capacity to improve rigor (Chamber & Tzavella, 2022). Within the Registered Reports (RRs) framework, a study's introduction, methods, and planned analyses are reviewed and a decision is reached about the manuscript. It is through the course of the revision that scientific rigor might be enhanced as the approach turns peer reviewers into informal collaborators who can guide the study authors towards more rigorous tests. Registered Reports are also the only open science practice that directly combats publication bias, as results have no bearing on acceptance. There is correlational evidence that Registered Reports are related to increased rigor (Soderberg et al., 2021) and a greater prevalence of null findings--56% in RRs vs. 4% in traditional report (Scheel et al., 2021).

As with preregistration, making data, code, and materials available will not necessarily improve the rigor of a study. Making data and code available allows other researchers to check

the code to see which analyses were actually run and to judge how appropriate these analyses were for the research questions asked. This is necessary because models are often ambiguously described and occasionally incorrect. Providing the data allows for a check on the reproducibility of the findings. Neither of these guarantees rigorous science, but they will make errors more apparent and safeguard against fraud.

Importantly, open science practices, particularly preregistration, do not prevent exploratory analyses or other methods reforms like robustness checks or multiverse analyses. In a recent critique of preregistration mandates that extends to preregistration itself, Klonsky (2024; cf, Vize et al., 2024) suggests that preregistration might discourage or prevent exploratory analyses, multiverse analyses, robustness checks, sensitivity analyses, and so on. Klonsky writes that preregistration mandates will "establish a norm that researchers should choose one or a subset of reasonable analyses and not others" (p. 14). He further suggests that preregistration offers "a blanket reward for a selective rather than comprehensive approach to understanding data, even though the latter approach is optimal" (p. 15). Finally, he argues that if preregistration becomes the norm, it will "devalue analyses conceived after engaging with the data" (p. 15).

We believe these concerns to be overblown and that any downsides to preregistration are likely to be offset by the benefits. Moreover, we are unaware of any clinical psychology journal currently mandating preregistration, nor any proposals from the field to do so. Importantly, a number of articles have discussed how preregistration can accommodate multiverse analyses or various other robustness checks (e.g., Hardwicke & Wagenmakers, 2023; Chambers & Tzavella, 2022; Simmons et al., 2021). In fact, these practices are *strengthened* by being preregistered due to the increased transparency. Exploratory analyses and analyses conceived after engaging with the data are valid analytic approaches within the preregistration framework; they simply need to

be labeled as deviations from the preregistered analytic plan[5] (see Phillips et al., 2025 for clearly

labelled deviations from a preregistration or O'Hara et al., 2020 for clearly labelled exploratory

analyses that were not part of the a priori analytic plan). The only thing that preregistration does

not allow is claiming that analyses conceived and carried out *after engaging with the data* were

planned. There are published guidelines for how best to report deviations from preregistrations

(Lakens, 2024; Willroth & Atherton, 2024). These critiques underscore the importance of

examining the principle and philosophy of open science practices rather than unnecessary

reification of the practices themselves.

What Open Science Will Do

The goal of preregistration and other open science practices is to increase the

transparency of research, allowing for evaluation of the severity of the test to which a given

claim has been subjected (Vize et al., 2024). The problem in the field is not just that there is

underpowered, p-hacked, and HARKed research populating our literature; the problem is that it

is difficult, if not impossible, to sort the good (i.e., rigorous) from the bad research in our

literature. This problem exists because we do not know how each result was arrived at.

Preregistration does not directly increase credibility; rather, it enhances transparency, which

enables evaluations of credibility—specifically, whether a claim has passed a severe test. That is,

transparency allows the field to see whether claims have passed severe tests, but it cannot by

itself make those tests severe. It is a necessary but not sufficient condition for rigorous science.

A test is severe if it has a high probability of revealing a claim to be false when it actually

is false (Lakens, 2019; Mayo, 2018). More broadly, the principle of severity addresses the

---

[5] Entirely exploratory studies should also be preregistered as exploratory. The preregistration of such studies
safeguards against later HARKing. Syed (2024) has discussed the myth (along with two others) that preregistration
is only relevant for certain kinds of studies.

question: When do data provide strong evidence for a claim or hypothesis? (Mayo & Spanos, 2011). The short answer: more severe tests yield stronger evidence (Vize et al., 2024). As Mayo writes:

> If data x agree with a claim C but the method used is practically guaranteed to find such agreement, and had little or no capability of finding flaws with C even if they exist, then we have bad evidence, no test (BENT)" (p. 5).

Underpowered research and practices like QRPs and HARKing undermine the severity of tests—they do not allow falsification. They lead to BENT studies where positive findings offer little to no evidential value (Mayo, 2018). Transparency allows such tests to be identified and their conclusions weighted accordingly.

The ability to transparently evaluate the severity of a test is essential, particularly because psychological theories often fail to constrain key analytic decisions (e.g., which measurement instrument to use, how to handle outliers, or which covariates to include; Lakens, 2019). Several scholars have argued that psychology's various crises—the replication crisis, the generalizability crisis, and the theory crisis—share a common root: empirical claims are frequently not based on severe tests (Claesen et al., 2022). The field might look different if researchers reported how studies were conducted, allowing others to assess the severity of the evidence provided.

We would be less inclined to give weight to papers that fished through data for significant results and retroactively wrote introductions to fit them (e.g., Bem, 2002). We would discount papers that selectively reported the only significant finding out of many tests conducted or selectively reported studies in multi-study manuscripts. We would question studies that made outlier and covariate decisions based on their influence on results. When such practices are hidden, it becomes impossible to distinguish severe tests from non-severe ones. Transparency is

thus a prerequisite for progress—this is the foundation of the OSM.

The Slow Progress of Clinical Psychology

For a variety of reasons, clinical psychology has been slower to embrace the OSM and to adopt widespread open science practices. Tackett and Miller (2019) noted that "the lack of engagement by the clinical psychology community…remains a glaring and disconcerting omission on the broader replicability landscape" (p. 487). The slow and non-urgent response may be due, in part, to the fact that some of the publications (e.g., Bem, 2011) or broader debates about evidentiary value fell in other subdisciplines of psychology (e.g., ego-depletion). Although there are some features of clinical psychology that reduce some concerns (e.g., Tackett et al., 2017; Tackett et al., 2019), the aforementioned issues are present and relevant to clinical science as well (e.g., bias against the null; low power; QRPs).

In an audit of 60 leading clinical psychology journals (first and second quartiles by impact factor), Nutu and colleagues (2019) examined editorial policies across five domains—preregistration, data sharing, preprints, reporting guidelines, and conflict-of-interest (COI) disclosure. They also assessed compliance in a sample of ~200 articles from journals with supportive policies in at least four domains. While 52 journals mandated COI disclosure, only 15 cited preregistration, only a single journal mandated data sharing, and fewer than half endorsed reporting guidelines. Engagement rates were terrible: 3% of articles were preregistered, 2% shared data, only one preprint was identified, just 19% followed reporting guidelines (mostly marked non-applicable), and only about half included a COI disclosure.

In a recent editorial on policies and engagement in Open Science at *Clinical Psychological Science*, Howard et al. (2024) reported on the percentages of articles receiving badges for preregistration, open data, and open materials. Importantly, this journal has been

awarding badges for 7 years and instituted Open Science as an editorial policy three years ago.

Following an initial uptake in badges across the first three years (from 2017 to 2019), rates have

remained relatively flat for open data and materials and risen slightly for preregistration.

Importantly, this plateau occurred at the start of a new editorial term during which formal badge

adjudication occurred for the first time in a clinical psychology outlet. This indicates that badges

alone (without adjudication) may be invalid indicators of adherence to OS practices—motivation

and interest is not enough, precision and accuracy of open science practices is also needed to

ensure they are promoting the intended principles. From 2020 to 2023, approximately 26% of

articles included open data while 20% included open materials. Across this same period, the

percentage of articles receiving badges for preregistration increased from 12% in 2020 to a high

of 27% in 2022 but fell in 2023 to 16%. From 2020 to 2023, approximately 35% of articles

received at least one badge. The field is not doing well when fewer than 40% of articles are

receiving at least one badge at a journal that established open science practices and principles as

a primary priority area for submissions.

Vize and Lynam (in press) compared the use of OS practices from 2021-2023 in two

personality disorder journals (*Journal of Personality Disorders* and *Personality Disorders:*

Theory, Research, and Treatment) to these practices in three personality journals (*Journal of*

Personality, *Journal of Research in Personality*, and *European Journal of Personality*).

Although the use of Registered Reports was comparably low across journal types, personality

journals showed much higher engagement in OS practices than did personality disorder journals.

The average percentage of articles sharing data in personality journals was 58.22% compared to

13.04% in personality disorder journals. A similar difference was observed in sharing analytic

code; on average, 65.10% of articles in personality journals shared code compared to 13.04% in

personality disorder journals. Although rates of preregistration were lower than rates of sharing data and code, the difference between journal types remained; in personality journals, 25.32% of studies were preregistered compared to 9.70% of studies in personality disorder journals.

This lack of engagement is reflected in the training that aspiring clinical psychologists receive as graduate students. In a recent Registered Report, Van Til et al. (2025) examined the state of graduate training in the replication crisis and OS practices in clinical psychology doctoral programs in the United States. This study included two sources of information—a review of research methods syllabi used in clinical psychology doctoral programs and a clinical psychology graduate student survey. The authors received syllabi from 112 clinical psychology programs and were able to code 104 of these; readings were coded into 23 different categories dealing with the replication crisis or open science methods. For the graduate student survey, Van Til and colleagues attempted to collect data from 100 doctoral students in each of three bins – (1) member of an APA accredited program, (2) member of an APA accredited program and a member of the Council of University Directors of Clinical Psychology (CUDCP) programs, and (3) member of the first two groups *and* the scientifically oriented Academy for Psychological Clinical Science (APCS).

Coverage in the research methods syllabi was generally thin across programs. The number of different readings on syllabi that were about open science or the replication/credibility crisis ranged from 0-17, with a median of 2 and an interquartile range of 0 to 5 (mean = 3.32; SD = 3.93). Fully 26% of programs did not have any assigned readings that fell into any of the 23 coded categories. Across all programs, the average percentage of readings dedicated to at least one of the included categories was 7.27. Disappointingly, readings criticizing open science were more frequent than readings dealing with HARKing, Registered Reports, measurement issues,

fraud, incorrect reporting of results, detecting "bad science" and open access.

The accompanying graduate student survey also revealed less than optimal training. Sixty-five percent of students had not completed a preregistration. Less than 9% had been involved in a Registered Report (RR); almost 30% reported they were "not sure" what an RR was. Students also rated their exposure to various open science principles and practices using a 1 (not at all) to 5 (a great amount) scale with 3 (a moderate amount) as the midpoint, and where those exposures occurred. Regarding the latter question, students endorsed greater exposure to all investigated open science-related principles and practices in their research methods class compared to the aggregate of the other queried sources (i.e., other classes, colloquia, discussions with advisor, social media, other), indicating that the limited exposure in methods classes was not being compensated for in other places. In terms of amount of exposure to open science topics, participants' exposure ranged, on average, from "none at all" to "a moderate amount." Exposure was highest for basic methodological concepts that are not new or specific to open science such as statistical significance (mean = 3.25) and reliability/measurement (mean = 3.05) and lowest for specific OS practices such as RRs (mean = 1.44) and preregistration (mean = 2.00). Notably, most of the comparisons across the three training categories (APA membership only; APA and CUDCP membership; APA, CUDCP and APCS membership) were not significant, suggesting that more scientifically oriented training programs are not doing more of this OS training than more clinically oriented programs. The most encouraging results from the survey were in relation to attitudes towards learning about the replication crisis and open science, 77.81% and 72.18% of respondents indicated they "liked" or "strongly liked" learning about each. We are doing our future colleagues a disservice if we are not preparing them for the most rigorous and exacting standards and norms of future psychological science.

What Has to Change

We believe that there are a variety of changes required to build a more robust and trustworthy clinical science literature, which include the adoption of various open science practices. Change is needed at all levels: student, faculty, programs, and institutions. Most of these changes are not specific to clinical psychology but apply to the entire field; however, if modifications to these approaches are needed for clinical psychology, they will only happen if clinical scientists broadly engage in these conversations and efforts.

Student Training: Toward Assessing Open Science Competencies

As evidenced by the results of Van Til et al. (2025), exposure to the factors contributing to the replication crisis and to OS principles and techniques in graduate training is inadequate. To avoid the practices that contribute to replication failures (i.e., low power, p-hacking, HARKing, and publication bias), one must understand and be able and willing to recognize them in one's own work. One of the more insidious features of these factors is that they are often committed without malevolent intent through subtle, psychologically motivated processes (Nagy et al., 2025). Practices like HARKing or p-hacking frequently feel justifiable to researchers because of motivated reasoning (Lund et al., 2018), the garden of forking paths (Gelman & Loken, 2014), moral blind spots (Ellemers, 2018), and/or practical considerations such as null findings in an underpowered study that preclude publication (and thus creates incentives to "find" a statistically significant result to enables publication). As a result, researchers may remain unaware that they are engaging in practices that compromise the integrity of their science (Fiedler & Schwarz, 2016) and may be ethical violations (Miller et al., 2025).

The fallacy of expertise/training/seniority should be familiar to clinical psychologists; much of the work on statistical versus actuarial judgment underscores the importance of

maintaining a "beginner's mind" (e.g., Grove et al., 2000). Related research finds that published academics across many fields (including psychology) routinely make fundamental errors in their interpretation of *p*-values; whereas, undergraduate students with no formal statistical training do not make such errors (McShane & Gal, 2016). These dynamics underscore the need for explicit training in research design, inference, and OS norms (Munafò et al., 2017; Vazire, 2017). Nagy et al. (2025) recently published a bestiary of 40 consensually defined questionable research practices; familiarizing students with this taxonomy, along with a discussion of their ethical status and scientific consequences, might mitigate against these practices.

Our training must include much more attention to the issue of transparency and open science practices. These practices are often understood as a mere set of technical procedures—preregistration, data sharing, registered reports—akin to using regression or checking reliability. But this is an incomplete understanding. Open science is not a statistical technique or a specific method; it is not merely a set of procedures. It is a commitment to a meta-principle of transparency that underlies the entire scientific enterprise. Just as falsifiability is a foundational criterion for scientific claims, transparency is foundational for evaluating the credibility of those claims. Practices like preregistration, data sharing, and code availability are instantiations of this principle, and it is essential that they be done well. These practices do not belong to a specific methodological school or type of analysis—they apply to all scientific inquiry (Munafò et al., 2017; Nosek et al., 2015). Even methodologically sophisticated work can be subject to bias, rationalization, and error (Fiedler & Schwarz, 2016; Gelman & Loken, 2014), underscoring the need for transparency in all cases. Thus, training in open science is not a niche skill—it is core scientific training. From this perspective, we find it deeply problematic that some investigators state, simply, that they "do not do open science" (in their laboratories) or they object to

transparency; this position stands in direct contrast to a belief that one is engaging in high-quality science. There can be no rigorous science without transparency. These issues—both the factors contributing to the replication crisis and the principles and tools of open science—must be mastered. We suggest this will require far more than two readings in our research methods classes (Van Til et al., 2025).[6]

One way to think about the training needs in this area is to consider student competencies and the explicit scientific skills that are needed to practice open science. This end-point training goal has the benefit of moving beyond knowledge alone and moving from intention to application. From this perspective, the floor of open science graduate training is a knowledge-based curriculum whereas the ceiling is determined through training-related activities and requirements that compel students to engage in increased transparency. As noted above, simply preregistering a study does not ensure its rigor. These activities will need to be completed in a mentored fashion, and all doctoral mentors have experience supervising dissertation proposals and critiquing the depth of these documents. The same degree of supervision should be applied to students' open science competencies, which also requires a commitment from faculty mentors to learn and adopt these practices. Table 2 lists a set of core competencies that we believe are necessary and may be sufficient for training students in better open science practices. As illustrated in the table, we pair each of these competencies with a set of readings and propose program-related activities that could be used to promote and evaluate students' proficiency in the specific behavioral skills.

At the student level, implementation can be incentivized by requiring students to preregister at least one required project during their graduate training; several of the current

---

[6] A list of potential readings covering all topics coded by Van Til et al. (2025) is provided in the supplement.

authors have been able to implement this requirement within their areas to some modest effect.

End of semester grants/papers assignments in core courses (e.g., Research Methods;

Psychopathology) can be replaced with end of semester preregistrations; two of the current

authors implement this in their research methods classes.[7] Student awards can be established;

these are already established at several programs that might serve to further motivate meaningful

student engagement in open science practices (Supplemental Table 1 provides potential criteria

for such an award).

Training in open science will also prepare students to navigate the recently revised

Transparency and Openness Promotion (TOP) Guidelines—a framework for journals to design

and implement publication standards that reflect the values of transparency, rigor, and

reproducibility. The 2025 revision of the TOP Guidelines (Grant et al., 2025) organizes standards

into three overarching categories: research practices, verification practices, and verification

studies. The research practices category, which remains the centerpiece of the framework,

includes seven specific standards: citation, data transparency, code transparency, materials

transparency, design and analysis transparency, study registration, and analysis plan registration.

These components reflect the key domains in which transparent reporting is expected and

provide clear guidance about what openness looks like in practice. Each standard includes three

levels: Level 1 requires disclosure (e.g., stating whether data are shared), Level 2 requires the

behavior (e.g., actually sharing the data), and Level 3 requires verification (e.g., confirming that

the data permit analytic reproducibility). Together, these levels create a progressive model that

distinguishes between mere intention, execution, and verifiability—an approach that can be

---

[7] For students conducting research that might be perceived as being controversial in the current political environment, preregistrations might be kept private until the associated manuscript is submitted. In particularly high stakes work, preregistrations may not be made public except to share them with peer reviewers during the review process until the manuscript is set to be published online.

directly mapped onto graduate training programs committed to fostering open science behaviors. Our experience is that the TOP Guidelines can be used as instructive training tools for students to "reverse engineer" what high-fidelity open science looks like in practice; in much the same way we may use treatment fidelity rating scales can be used to enhance psychotherapy training, the TOP Guidelines can be used for pedagogical purposes.

Implementation

Beyond exposure and training in courses, students need to implement these practices in their own work. These efforts will likely be of limited utility, however, if graduate students return to labs where open science is not actively supported or prioritized—and worse, opposed-- by faculty. The most effective way for faculty to support open science practices among their students is to model it in their own research. At a minimum, faculty can create space for students to engage in these practices independently. Many mentors already engage with their students in behaviors that are preregistration adjacent through dissertation and thesis proposals; posting those documents to the OSF is a natural next step. Likewise, incorporating consent form language that allows for data sharing, or uploading final analysis code, are easy entry points that provide structure and practice without much cost. Kathawalla et al. (2021) offer a clear, concrete guide for how students and mentors can begin incorporating these practices, with strategies tailored for low-resource labs and resistant environments. Their focus on small, feasible steps makes the process approachable—even for faculty who are just getting started or who remain unsure about how to proceed.

At the departmental and institutional levels, meaningful change will require a reorientation of incentive structures to explicitly value transparency. Currently, faculty hiring, merit evaluations, and promotion decisions tend to reward grant funding, media visibility, and

publications in premier journals—often with little consideration of how the research was produced (Nosek et al., 2012; Ioannidis, 2005). Said differently, in terms of variance explained, institutional incentive structures that reward these activities are arguably the strongest drivers undermining scientific credibility (e.g., Smaldino & McElreath, 2016). This dynamic sends a clear message: flashier or grant funded work is better work, regardless of its reproducibility or rigor. Transparency itself must be treated as a core academic contribution and, as mentioned above, not a niche area or specialty pursuit. Practices such as preregistration, sharing open data, code, and materials, and Registered Reports should be directly recognized in hiring and promotion decisions (Munafò et al., 2017; Chambers, 2019). Faculty applicants should be asked to talk about these issues in job talks— whether findings come from studies that were preregistered and followed other OS practices. Applicants who are not yet engaged in such practices can be asked about future plans in this vein.

Departments should revise evaluation rubrics to include transparent practices as evidence of research quality; with these revisions, transparency is not merely personal choice, it is a scientific job requirement. Institutions should also provide infrastructure and training to support these practices. Unless transparency is institutionally incentivized—on par with publications and funding—it will remain an individual choice rather than a collective norm. An excellent example is the University of Maryland's Toolkit for Aligning Incentives 2.0 (Dougherty et al., 2024), which provides a variety of suggestions for reforming evaluation policies in line with values like openness, equity, and rigor. The toolkit offers factsheets and concrete guidance for departments seeking to revise hiring, tenure, and promotion practices to reward behaviors that reflect the core academic mission of contributing to the public good. Among its recommendations are the inclusion of transparent, reproducible, and publicly accessible research outputs in evaluation

rubrics, as well as recognition of community-engaged scholarship and value-aligned teaching. The toolkit provides a values-to-behavior worksheet that maps institutional principles—such as accessibility, rigor, and inclusivity—onto specific, observable faculty contributions. This kind of structure helps departments operationalize values in concrete and evaluable ways. Without such reforms, faculty remain subject to limited metrics (e.g., publication counts, impact factors) that do not necessarily reflect research quality and may even actively disincentivize transparency. Aligning incentives with core scholarly values is not just about fairness—it is essential for restoring credibility in psychological science and ensuring its contributions are both trustworthy and socially meaningful. Another relevant resource departments and universities may consider is the San Francisco Declaration on Research Assessment (see: https://sfdora.org/), which provides guidelines for holistic assessments of research performance.

Barriers to Implementation

Presently, none of the above suggestions are implemented widely. In general, students are not being trained in the replication crisis or open science practices (Van Til et al., 2025). Most clinical psychology faculty are not engaging in these practices in a routine or meaningful manner. We are aware of very few programs that have incorporated open science into their hiring, merit, or promotion rubrics. Obviously, there are barriers that must be overcome.

Cultural Resistance

One huge barrier to the broader adoption of open science practices is cultural resistance within the academic community, particularly among more senior scholars. These researchers often built careers under different norms—ones that rewarded (or at least allowed) flexibility in data analysis, prioritized novel findings, and rarely required transparency. For some, open science is experienced not as an invitation to improve how science is conducted, but as a rebuke

of earlier practices that were once fairly normative. This resistance occasionally surfaces in

public commentary, such as Fiske's (2016) editorial decrying "methodological terrorists" and

characterizing early OS advocates as overly hostile critics. Senior scholars may be hesitant to

question the practices and assumptions that supported their earlier success (Lakens, 2019;

Vazire, 2018). Additionally, researchers embedded in high-prestige institutions may feel less

pressure to adopt open science practices, either because their success insulates them from

changing norms or because they do not see clear incentives to do so (Bakker et al., 2020; Nosek

& Bar-Anan, 2012). Until these attitudes shift or institutional structures (e.g., journals, funders,

hiring and promotion committees) create clearer expectations—reform is likely to remain partial

and uneven.[8]

Underpowered Research

A major barrier to the adoption of open science practices lies in research contexts where

large samples are difficult to obtain. Studies involving rare populations—such as individuals

with low-base-rate disorders—or costly methodologies (e.g., neuroimaging; biological assays)—

often operate with constrained sample sizes. When combined with the typically small effect sizes

found in psychology (Gignac & Szodorai, 2016; Schäfer & Schwarz, 2019), these constraints

result in chronically underpowered studies. As noted earlier, low statistical power not only

increases the likelihood of Type II errors but also inflates effect size estimates and raises the

false positive rate among significant findings (Button et al., 2013; Ioannidis, 2005). Recent large-

scale analyses have underscored the severity of this issue: Marek et al. (2022), using data from

---

[8] Much has been written about perverse incentives and their effect on science and scientific reform. We ascribe to Yarkoni's (https://talyarkoni.org/blog/2018/10/02/no-its-not-the-incentives-its-you/) perspective on this matter. He writes, "If you find yourself unable to do your job without regularly engaging in practices that clearly devalue the very science you claim to care about, and this doesn't bother you deeply, then maybe the problem is not actually The Incentives—or at least, not The Incentives alone. Maybe the problem is You."

the Adolescent Brain Cognitive Development (ABCD) study, demonstrated that brain-behavior

correlations in neuroimaging are often much smaller than previously assumed, and that

thousands of participants are typically required to achieve adequate power in typical between

subjects studies. This makes clear that most published MRI studies—often conducted in the

context of single labs via R01 grant mechanisms that support samples of fewer than 100

participants—are drastically underpowered. In such contexts, researchers may be reluctant to

preregister analyses, commit to sharing data, or submit to the rigor of Registered Reports—

fearing that transparent methods will expose a fragility in their results. In this context, open

science practices may be experienced as existential threats. It is these factors that appear to

underlie the results of Bogdan (2025) who found less increase in stronger p-values at highly

ranked research universities; the authors reported that two-thirds of the association could be

explained by highly ranked universities preferring laborious, expensive, and subtle research

topics where effect sizes are likely to be exceedingly small. Addressing this barrier requires

structural solutions—such as multi-site collaborations, shared data repositories, and alternative

evaluation criteria—that make transparent science feasible even under difficult constraints.

Funding

      With an entire institute devoted to funding mental health research (NIMH) and two others

devoted to funding research on specific types of mental health problems (NIDA and NIAAA) the

grant culture exerts a strong influence on the field, and not always for the better. Lilienfeld

(2017) suggested that the grant culture is a primary contributor to the replication crisis. He was

hopeful that the replication crisis might serve as a necessary reckoning rather than a disaster,

marking a shift toward greater humility and methodological rigor in psychological science. But

he argued that the deeper threat lies elsewhere—in the institutional pressures created by the grant

culture. The relentless focus on securing funding distorts priorities, rewarding fundraising over meaningful contributions and pushing researchers toward questionable practices, narrow lines of inquiry, and safe, fundable topics.[9] It fosters confirmation bias, intellectual hyper-specialization, and overpromising, all while leaving little time for the kind of deep, reflective thinking that fuels real insight. He felt that unless psychology grapples seriously with these structural distortions, the field's self-correction will remain superficial.

Statistical Training

Applied (e.g., clinical) researchers may be weaker quantitatively and may not fully appreciate how the horsemen work to undermine the field (Aiken et al. 1990, 2008). To fully appreciate the corrosive effect the horsemen have on the literature requires an understanding of sampling variability, the probabilistic nature of our statistics, statistical power, the problems caused by multiple testing (within a study, across a program of research, or within the field), and the influence of selection (i.e., publication bias). These are not simple ideas, and most people will require multiple exposures (e.g., multiple methods/statistics classes) to fully grasp them. Scholars at all levels of expertise and training must commit to an ongoing dialogue, in the spirit of scientific humility and self-improvement. In this way, senior scientists not only catch up so as to provide students and junior scientists the level of scientific training they require, but further serve as role models for the practice of continuous self-improvement and self-awareness.

Sensitive Data

Another persistent barrier to open science is the challenge of sharing sensitive data,

---

[9] A recent news story noted that Duke University School of Medicine will begin actively reducing salaries of faculty who are not sufficiently successful at securing external funding, despite the incentives this creates for the use of problematic research practices and a political landscape where external funding is increasingly limited, especially for certain topics. https://www.dukechronicle.com/article/2025/07/duke-university-school-of-medicine-implements-faculty-productivity-guidelines-external-grant-funding-requirements-salary-reductions .

particularly in clinical, neuropsychological, and developmental research. These datasets often contain personally identifiable or protected health information, raising legitimate concerns about participant confidentiality and regulatory compliance (e.g., HIPAA, GDPR). Even when data are de-identified, the risk of re-identification—especially in small or richly detailed samples—can be non-trivial. As a result, many researchers working with clinical populations perceive data sharing as either infeasible or ethically fraught. This has created a structural disincentive: those working in high-risk or high-stakes domains are less able to meet open data expectations and may feel unfairly penalized in a reform movement that sometimes assumes a one-size-fits-all standard. But there are solutions. Controlled-access repositories, data use agreements, and synthetic datasets offer pathways to balance transparency with privacy (Gilmore et al., 2018). To lower barriers, institutions and journals must provide clear, realistic guidance on responsible data sharing, including flexible models tailored to sensitive contexts. Otherwise, the push for transparency risks privileging low-risk, convenience-sample research—ironically reinforcing some of the weaknesses Open Science was meant to fix.

A Vision for the Future

The field needs to recognize this truth: to be credible, research must be transparent. Other researchers need to have access to the decisions and processes that led to the conclusions. As noted earlier, transparency does not guarantee important or rigorous science. One can write a Registered Report on a trivial topic. One can preregister a horribly flawed study with inappropriately small samples, invalid measures, and fundamentally mismatched analytic approaches. But one cannot have credible science without transparency because to evaluate rigor—the severity of a test— one requires access to the choices the scientist makes along the way. Credible science is important in all domains of psychology but *particularly so in clinical—*

all stakeholders must know the truth as to the reliability and validity of our findings for assessments, therapy modalities, studies of etiology, and so on. These are important issues with practical ramifications for actual people. Ethical research practices call for transparency.

We look forward to increased training in the replication crisis and open science practices. Although it takes some training and practice to reach full competencies in such approaches, beginning the process is easy. Resources (e.g., readings, lectures, templates, examples) abound. Preregistration does not require much additional time. Students are already completing variants of Registered Reports vis-a-vis thesis and dissertation proposals. Many faculty members are already engaged in preregistration-adjacent work in the way of grant, thesis, and dissertation proposals. Data and code already exist on our own computers and servers. What is required is a commitment to transparency because it is important to the field and the right thing to do, and because the next generation of clinical psychological scientists will only be trained and equipped to meet the demands of modern science if we do so.

We look forward to more journals following the lead of Vazire and her team at *Psychological Science* (Hardwicke & Vazire, 2023) who are making transparency the default and not merely an option. Top journals should not just encourage transparent practices—they should require them. The same is true for reviewers at these journals; one needs to know how results were arrived at in order to fully evaluate the rigor of a study. If authors refuse to engage in practices that increase transparency, that should preclude publication in these journals. Publication in the top journals should be available only to research that is transparent; after all, transparency is a necessary condition for good science. There may be some issues specific to clinical psychology that make some OS practices less feasible (e.g., the sharing of sensitive data), but other OS practices are still available (i.e., preregistration, Registered Reports, sharing of code). Minimally, all journals

can move to higher levels of the TOP guidelines.

We hope that psychology departments in the US in general and clinical psychology programs more specifically will pivot towards valuing transparency among faculty and graduate students. This will involve (a) encouraging evaluations of open science practices for all job candidates across the field, and (b) building OS evaluation criteria into promotion and tenure evaluations. Clinical graduate training programs should bake these practices into their training in the same way they require competencies in clinical work, research, interpersonal and professional skills—these should be evaluated regularly along with other competencies. However, this requires a faculty body that values and incorporates these approaches into their own work and models this for early career researchers both formally and informally. As with many things, the younger generation is well ahead of us—the survey results from Van Til (2025) indicate that our graduate students think these topics are important for science and critical to their training. Our call for change in clinical psychology is not only a plea to improve methodological rigor, but also an ethical imperative (Miller et al., 2025) and a topic we have failed to address for far too long.

Finally, we call on accrediting bodies—the American Psychological Association (APA) and the Psychological Clinical Science Accreditation System (PCSAS)—to give these issues the attention they deserve. These organizations are among the most influential levers for change in the field. When accrediting bodies introduce new requirements, programs respond. By incorporating open science and transparency standards into accreditation criteria, they could accelerate the adoption of practices that strengthen the credibility and integrity of clinical psychology.

References

Aiken, L. S., West, S. G., Sechrest, L., & Reno, R. R. (1990). Graduate training in statistics, methodology, and measurement in psychology: A survey of PhD programs in North America. *American Psychologist, 45(6)*, 721–734. https://doi.org/10.1037/0003-066X.45.6.721

Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology: Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. *American Psychologist, 63(1),* 32–50.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, 7(6), 543–554.

Bakker M, Veldkamp CLS, van Assen MALM, Crompvoets EAV, Ong HH, Nosek BA, et al. (2020) Ensuring the quality and specificity of preregistrations. *PLoS Biology 18(12):* e3000937. https://doi.org/10.1371/journal.pbio.3000937

Bem. D.J. (2002). Writing the Empirical Journal Article. In Darley, J. M., Zanna, M. P., & Roediger III, H. L. (Eds). *The Compleat Academic: A Career Guide*. Washington, DC: American Psychological Association.

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

Benning, S. D., Bachrach, R. L., Smith, E. A., Freeman, A. J., & Wright, A. G.C. (2019). The registration continuum in clinical science: A guide toward transparent practices. *Journal of Abnormal Psychology, 128(6),* 528.

Benning, S. D., & Smith, E. A. (2023). The registration continuum in personality disorder

      studies: Theory, rationale, and template. *Personality Disorders: Theory, Research, and*

      Treatment, 14(1), 5-18.

Bishop, D. (2019). Rein in the four horsemen of irreproducibility. *Nature, 568(7753)*, 435. doi:

      10.1038/d41586-019-01307-2.

Bogdan, P.C. (2025). One decade into the replication crisis, how have psychological results

      changed? *Advances in Methods and Practices in Psychological Science*, 8(2).

      doi:10.1177/25152459251323480

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò,

      M. R. (2013). Power failure: why small sample size undermines the reliability of

      neuroscience. *Nature reviews. Neuroscience*, *14*(5), 365–376.

      https://doi.org/10.1038/nrn3475

Callaway, E. Report finds massive fraud at Dutch universities. *Nature 479*, 15 (2011).

      https://doi.org/10.1038/479015a

Castillo, A., Baranger, D., Miller, J.D., Vize, C., & Lynam, D.R. (2025). *When do*

      interaction/moderation effects stabilize in linear regression? Manuscript submitted for

      publication.

Chambers, C. D. (2019). *The seven deadly sins of psychology: A manifesto for reforming the*

      culture of scientific practice. Princeton University Press.

Chambers, C.D., & Tzavella, L. (2022). The past, present and future of Registered Reports.

      Nature Human Behavior, 6, 29-42.

Claesen, A., Lakens, D., Vanpaemel, W., & van Dongen, N. (2022). Severity and crises in

      science: Are we getting it right when we're right and wrong when we're wrong?

PsyArXiv Preprints. https://doi.org/10.31234/osf.io/ekhc8

Coursol, A., & Wagner, E. E. (1986). Effect of positive findings on submission and acceptance

rates: A note on meta-analysis bias. *Professional Psychology: Research and Practice,*

17(2), 136–137. https://doi.org/10.1037/0735-7028.17.2.136

Devezer, B., Nardin, L. G., Baumgaertner, B., & Buzbas, E. O. (2019). Scientific discovery in a

model-centric framework: Reproducibility, innovation, and epistemic diversity. *PLOS*

ONE, 14(5), e0216125. https://doi.org/10.1371/journal.pone.0216125

Dougherty, M. R., Carter, C., McKiernan, E., Tananbaum, G., Parker, Z., & Garcia, C. (2024,

February 21). Toolkit for Aligning Incentives 2.0. https://doi.org/10.31219/osf.io/8x4e9

Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social*

Psychological and Personality Science, 7(1), 45-52.

https://doi.org/10.1177/1948550615612150

Fiske, S. T. (2016). *A call to change science's culture: Supplanting intellectual bullying with*

collaboration. *APS Observer, 29(9*). https://www.psychologicalscience.org/observer/a-

call-to-change-sciences-cultureGelman, A., & Loken, E. (2014). The statistical crisis in

science: Data-dependent analysis—a "garden of forking paths"—explains why many

statistically significant comparisons don't hold up. *American Scientist, 102*(6), 460–466.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences

researchers. *Personality and Individual Differences, 102*, 74–78.

https://doi.org/10.1016/j.paid.2016.06.069

Gilmore R.O., Lorenzo Kennedy, J., & Adolph, K.E.. (2018). Practical solutions for sharing data

and materials from psychological research. *Advances in Methods and Practices in*

Psychological Science, 1, 121-130.

Grant, S., Corker, K. S., Mellor, D. T., Stewart, S. L. K., Cashin, A. G., Lagisz, M., … Nosek, B.

    A. (2025, February 3). TOP 2025: An Update to the Transparency and Openness

    Promotion Guidelines. https://doi.org/10.31222/osf.io/nmfs6_v2

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological*

    Bulletin, 82(1), 1–20. https://doi.org/10.1037/h0076157

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus

    mechanical prediction: a meta-analysis. *Psychological Assessment, 12(1)*, 19-30.

Hardwicke, T. E., & Vazire, S. (2024). Transparency is now the default at Psychological

    Science. *Psychological Science, 35(7)*, 708-11.

    https://doi.org/10.1177/09567976231221573

Hardwicke, T. E., & Wagenmakers, E. J. (2023). Reducing bias, increasing transparency and

    calibrating confidence with preregistration. *Nature Human Behaviour*, *7*(1), 15–26.

    https://doi.org/10.1038/s41562-022-01497-2

Hengartner, M. P. (2018). Raising awareness for the replication crisis in clinical psychology by

    focusing on inconsistencies in psychotherapy research: How much can we rely on

    published findings from efficacy trials? *Frontiers in Psychology, 9,* Article

    256. https://doi.org/10.3389/fpsyg.2018.00256

Howard, A. L., Neal, T. M. S., Kirtley, O. J., Urry, H. L., & Tackett, J. L. (2024). Open science

    at Clinical Psychological Science: Reflections on progress, lessons learned, and

    suggestions for continued improvement. *Clinical Psychological Science, 13(1),* 195-204.

    https://doi.org/10.1177/21677026241255882)

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2(8),*

    e124.

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology,*

   19(5), 640–648. https://doi.org/10.1097/EDE.0b 013e31818131e7

Kathawalla, U.K., Silverstein, P., & Syed, M. (2021). Easing into open science: A guide for

   graduate students and their advisors. *Collbara: Psychology, 7*, 18684.

   https://doi.org/10.1525/collabra.18684

Kerr N. L. (1998). HARKing: hypothesizing after the results are known. *Personality and Social*

   Psychology Reeview, *2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203

Klonsky E. D. (2025). Campbell's law explains the replication crisis: Pre-registration badges are

   history repeating. *Assessment*, *32*(2), 224–234.

   https://doi.org/10.1177/10731911241253430

Korbmacher, M., Azevedo, F., Pennington, C.R. et al. (2023). The replication crisis has led to

   positive structural, procedural, and community changes. *Communications Psychology, 1*,

   3. https://doi.org/10.1038/s44271-023-00003-2

Lakens, D. (2019). The value of preregistration for psychological science: A conceptual

   analysis. *Japanese Psychological Review, 62*, 221–230

Lakens, D. (2024). When and how to deviate from a preregistration. *Collabra, 10 (1): 117094*

Lilienfeld, S. O. (2017). Psychology's replication crisis and the grant culture: Righting the ship.

   Perspectives on Psychological Science, *12*(4), 660-664.

   https://doi.org/10.1177/1745691616687745

Lind, T., Erlandsson, A., Västfjäll, D., & Tinghög, G. (2022). Motivated reasoning when

   assessing the effects of refugee intake. *Behavioural Public Policy, 6(2),* 213–236.

   https://doi.org/10.1017/bpp.2018.41

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., …

Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require

thousands of individuals. *Nature*, *603*(7902), 654–660. https://doi.org/10.1038/s41586-022-04492-9

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research:

Causes, consequences, and remedies. *Psychological Methods, 9(2)*, 147–163.

Mayo, D. G. (2018). *Statistical inference as severe testing: How to get beyond the statistics

wars*. Cambridge: Cambridge University Press.

Mayo, D. G. and Spanos, A. (2011) Error Statistics in *Philosophy of Statistics , Handbook of

Philosophy of Science Volume 7*, (General editors: Dov M. Gabbay, Paul Thagard and

John Woods; Volume eds. Prasanta S. Bandyopadhyay and Malcolm R. Forster.)

Elsevier: 1-46.

Mayo-Wilson, E., Grant, S., Corker, K. S., & Moher, D. (2025). *Consistent and precise

description of research outputs could improve implementation of open science*.

MetaArXiv. https://doi.org/10.31222/osf.io/q6xwp_v4

McShane, B. B., & Gal, D. (2016). Blinding us to the obvious? The effect of statistical training

on the evaluation of evidence. *Management Science, 62(6),* 1707-1718.

Meyer, M.N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices

in Psychological Science, 1*, 131-144. doi:10.1177/2515245917747656

Miller, J.D., Phillips, N.L., & Lynam, D.R. (2025). Questionable research practices violate the

APA Code of Ethics. *Journal of Psychopathology and Clinical Science,* 113-114.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., et al. (2017). A manifesto for reproducible

science. *Nature Human Behaviour, 1*, 0021. https://doi.org/10.1038/s41562-016-0021

Murphy, K. R., & Aguinis, H. (2019). HARKing: How badly can cherry-picking and question

trolling produce bias in published results? *Journal of Business and Psychology, 34*(1), 1–

17. https://doi.org/10.1007/s10869-017-9524-7

Nagy, T., Hergert, J., Elsherif, M. M., Wallrich, L., Schmidt, K., Waltzer, T., ... & Rubínová, E.

(2025). Bestiary of questionable research practices in psychology. *Advances in Methods

and Practices in Psychological Science*, *8*(3), 25152459251348431.

doi:10.1177/25152459251348431

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review

of Psychology, 69,* 511–534. https://doi.org/10.1146/annurev-psych-122216-011836

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., …

Yarkoni, T. (2015). Scientific standards. Promoting an open research

culture. *Science*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication.

Psychological Inquiry, 23(3), 217-243. https://doi.org/10.1080/1047840X.2012.692215

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration

revolution. *PNAS, 115(11),* 2600–2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia II: Restructuring incentives and

practices to promote truth over publishability. *Perspectives on Psychological Science,

7*(6), 615–631. https://doi.org/10.1177/1745691612459058

Nutu, D., Gentili, C., Naudet, F., & Cristea, I. A. (2019). Open science practices in clinical

psychology journals: An audit study. *Journal of Abnormal Psychology, 128(6)*, 510–516.

https://doi.org/10.1037/abn0000414

O'Hara, K.L., Grinberg, A.M., Tackman, A.M., Mehl, M.R., & Sbarra, D.A. (2020). Contct with

an ex-partner is associated with psychological distress after marital separation. *Clinical*

Psychological Science, 8, 450-463.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science.

Science, 349, aac4716. DOI:10.1126/science.aac4716

Phillips, N. L., Carter, N. T., King, K. M., Owens, M. M., Hyatt, C. S., Lynam, D. R., & Miller,

J. D. (2025). Additive and interactive relations of personality and cognition with

externalizing behaviors. *Clinical Psychological Science, 13,* 951-977.

Pittarello, A., Leib, M., Gordon-Hecker, T., & Shalvi, S. (2015). Justifications shape ethical

blind spots. *Psychological Science*, *26*(6), 794–804.

https://doi.org/10.1177/0956797615571018

Reardon, K. W., Smack, A. J., Herzhoff, K., & Tackett, J. L. (2019). An N-pact factor for

clinical psychological research. *Journal of Abnormal Psychology*, *128*(6), 493–499.

https://doi.org/10.1037/abn0000435

Retraction Watch. (n.d.). *Diederik Stapel*. Retraction Watch.

https://retractionwatch.com/category/diederik-stapel/

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological*

Bulletin, 86(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rubin, M. (2020). Does preregistration improve the credibility of research findings? *Quantitative*

Methods for Psychology, 16(4), 376-390. doi: 10.20982/tqmp.16.4.

Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible

computational research. *PLOS Computational Biology, 9(10),* e1003285.

https://doi.org/10.1371/journal.pcbi.1003285

Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological

research: Differences between sub-disciplines and the impact of potential biases.

Frontiers in Psychology, 10, 813. https://doi.org/10.3389/fpsyg.2019.00813

Scheel, A.M., Schijen, M.R.M.J., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with Registered Reports. *Advances in Methods and Practices in Psychological Science,* 4(2), 1-12. doi:10.1177/25152459211007467

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

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

Smaldino, P.E., & McElreath, R.(2016). The natural selection of bad science. *Royal Society Open Science, 3*, 160384. http://dx.doi.org/10.1098/rsos.160384

Soderberg, C. K. (2018). Using OSF to share data: A step-by-step guide. *Advances in Methods and Practices in Psychological Science, 1(1)*, 115–120. https://doi.org/10.1177/2515245918757689

Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J., Thorn, F. S., Vazire, S., Esterling, K. M., & Nosek, B. A. (2021). Initial evidence of research quality of registered reports compared with the standard publishing model. *Nature Human Behaviour*, *5*(8), 990–997. https://doi.org/10.1038/s41562-021-01142-4

Świątkowski, W., & Dompnier, B. (2017). Replicability crisis in social psychology: Looking at the past to find new pathways for the future. *International Review of Social Psychology, 30*(1), Article 111-124. https://doi.org/10.5334/irsp.66

Syed, M. (2024). Three persistent myths about open science. *Journal of Trial & Error*, *4*(2).

Szollosi, A., Kellen, D., Navarro, D. J., Shiffrin, R., van Rooij, I., Van Zandt, T., & Donkin, C.

(2020). Is preregistration worthwhile? *Trends in Cognitive Sciences, 24(2),* 94–95.

https://doi.org/10.1016/j.tics.2019.11.009

Szollosi, A., & Newell, B. R. (2020). People as intuitive scientists: Reconsidering statistical

explanations of decision making. *Trends in Cognitive Sciences, 24*(12), 1008–

1018. https://doi.org/10.1016/j.tics.2020.09.005

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*, 579-

604. doi: 10.1146/annurev-clinpsy-050718-095710.

Tackett, J. L., Brandes, C. M., & Reardon, K. W. (2019). Leveraging the Open Science

Framework in clinical psychological assessment research. *Psychological*

Assessment, *31*(12), 1386.

Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D.,

Oltmanns, T. F., & Shrout, P. E. (2017). It's time to broaden the replicability

conversation: Thoughts for and from clinical psychological science. *Perspectives on*

Psychological Science, *12*(5), 742-756. https://doi.org/10.1177/1745691617690042

Tackett, J. L., & Miller, J. D. (2019). Introduction to the special section on increasing

replicability, transparency, and openness in clinical psychology. *Journal of Abnormal*

Psychology, 128(6), 487–492. https://doi.org/10.1037/abn0000455

van den Akker, O. R., Weston, S. J., Campbell, L., Chopik, W. J., Damian, R. I., Davis-Kean, P.

E., Hall, A. N., Kosie, J. E., Kruse, E., Olsen, J., Ritchie, S. J., Valentine, K. D., van 't

Veer, A. E., & Bakker, M. (2021). Preregistration of secondary data analysis: A template

and tutorial. *Meta-Psychology, 5*, Article 2625. https://doi.org/10.15626/MP.2020.2625

van Ravenzwaaij, D., de Jong, M., Hoekstra, R., Scheibe, S., Span, M. M., Wessel, I., &

Heininga, V. E. (2025). De-Identification when making data sets findable, accessible, interoperable, and reusable (FAIR): Two worked examples from the behavioral and social sciences. *Advances in Methods and Practices in Psychological Science, 8(2),* 25152459251336130.

Van Til, K., Phillips, N. L., Du, T. V., Rose, L., Miller, J. D., & Lynam, D. R. (in press). Open science training in APA-accredited clinical psychology programs: A registered report. *Clinical Psychological Science.*

Vazire, S. (2017). Quality uncertainty erodes trust in science. *Collabra: Psychology* , *3 (1):* 1. doi: https://doi.org/10.1525/collabra.74

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. Perspectives on Psychological Science, 13(4), 411–417. https://doi.org/10.1177/1745691617751884

Vize, C.E., & Lynam, D.R. (2025). *Open science practices in personality disorder journals*. Manuscript submitted for publication.

Vize, C. E., Phillips, N. L., Miller, J. D., & Lynam, D. R. (2025). On the use and (mis)uses of preregistration: A reply to Klonsky (2024). *Assessment, 32*, 235-243.

Willroth, E. C., & Atherton, O. E. (2024). Best laid plans: A guide to reporting preregistration deviations. *Advances in Methods and Practices in Psychological Science, 7(1),* Article 25152459231213802. https://doi.org/10.1177/25152459231213802

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. *PLOS Computational Biology, 13(6)*, e1005510. https://doi.org/10.1371/journal.pcbi.1005510

Wingen, T., Berkessel, J. B., & Englich, B. (2022). No preprint is an island: How preprints enter

the journal system. *PLOS ONE, 17(8),* e0272808.

https://doi.org/10.1371/journal.pone.0272808

Yarkoni, T. (2018, October 2). No, it's not the incentives—it's you. *[citation needed].* Retrieved

August 12, 2025, from https://talyarkoni.org/blog/2018/10/02/no-its-not-the-incentives-

its-you/

Table 1.

Glossary of Terms and Impact

Term	Definition	Impact
Low statistical power	Power = probability of correctly rejecting H_0	Effect sizes in the literature are overestimates; increases Type II errors; increases proportion of false positives in the literature; prompts additional testing/QRPs
p-hacking	conducting multiple analyses on a dataset until a desired, typically statistically significant, effect is found or selectively reporting subsets of analyses that yielded desired outcomes while suppressing others that did not	Undermines scientific credibility by inflating effect sizes and increasing the likelihood of false positives.
HARKing	Presenting a hypothesis generated after engaging with the data as though it were an *a priori* hypothesis	Allows exploratory work to masquerade as confirmatory work which provides unwarranted credibility; fails to provide a test of the hypothesis; pushes Type I errors into literature; loss of information about original hypothesis
Publication bias	General bias among editors, reviewers, and authors for statistically significant and/or novel results and against null findings	Pushes researchers to find statistically significant results in their data which sets the stage for QRPs; pushes Type I errors into the literature; keeps disconfirming evidence out of the literature; undermines meta-analytic work
QRPs	Broad umbrella term for methods of collecting, analyzing, interpreting, or reporting data that increase the likelihood of misleading conclusions, often serving the researcher's interests. Includes specific p-hacking practices, HARKing, and other methods	Increased Type I error rates, overestimates of effect size; Allows exploratory work to masquerade as confirmatory work which provides unwarranted credibility
Transparency	The comprehensive documentation and sharing of research procedures, data, and analytic decisions, allowing others to trace how findings were produced.	Increases reproducibility; allows evaluation of the strength and credibility of the research
Preregistration	The process of publicly	Distinguishes exploratory work from

	documenting a study's research questions, hypotheses, design, and planned analyses before data collection or analysis begins.	confirmatory work; prevents HARKing; limits flexibility in data analysis (QRPs); increases evaluability
Registered Reports	A publishing format in which the study's rationale, design, and analysis plan are peer-reviewed and accepted before data collection. If accepted, the journal commits to publishing the results regardless of outcome, provided the authors follow the approved protocol.	Distinguishes exploratory work from confirmatory work; prevents HARKing; limits flexibility in data analysis (QRPs); may improve rigor (including statistical power); overcomes publication bias
Open materials	Sharing the stimuli, surveys, codebooks, or other materials used to conduct a study	evaluate, replicate, or adapt the research procedures
Open data	Making the raw or cleaned data from a study publicly available in a usable format, with sufficient documentation to allow others to understand and reanalyze it	facilitates reproducibility; allows identification of statistical errors
Open code	Providing access to the analysis code (e.g., R scripts, SPSS syntax) used to process and analyze the data	Facilitates reproducibility

Table 2

Open Science Student Competencies, Suggested Readings, Illustrative Program Activities, and Example Behavioral Evidence of Competency

Competency	Readings	Illustrative Program Activities	Behavioral Evidence
Knowledge of the history and emergence of the open science movement	Munafò et al., 2017; Open Science Collaboration, 2015; Korbmacher et al., 2023; Chambers, 2019; Hardwicke & Vazire, 2024	Short lecture + seminar mapping OS timeline in psychology/clinical; 'replication crisis' case map; journal policy scavenger hunt (identify transparency policies across 3 clinical journals)	1-page timeline explaining key events, actors, and policy shifts in the OSM movement; brief memo comparing two journals' transparency policies with examples.
Awareness of corrosive & unethical research practices (QRPs) and how to avoid them	Simmons et al., 2011; Kerr, 1998; Button et al., 2013; Ioannidis, 2005	Guided critique of a published paper to identify researcher dfs; power analysis workshops	Marked-up critique identifying researcher dfs and potential QRPs; documented power analysis justifying sample size; revised methods section that pre-commits decision rules.
Drafting, organizing, and finalizing (and posting) preregistrations; knowledge of Registered Reports	Nosek et al., 2018; Chambers & Tzavella, 2022; Tackett, Brandes, & Reardon, 2019; van den Akker et al., 2021; Willroth & Atherton, 2024	In-class OSF preregistration build; peer-review of Stage-1 proposals via the Peer Community In (PCI) Registered Reports; mock Stage-1 Registered Report submission.	Public OSF preregistration meeting a quality rubric; Stage-1 style introduction + methods + analysis plan; peer-review responses. Conduct a peer review of a Stage 1 PCI Registered Report
Knowledge of principles of replication and reproducibility	Open Science Collaboration, 2015; Munafò et al., 2017; Hardwicke & Wagenmakers, 2023	Compare/contrast replication outcomes to original studies (e.g., in large-scale Registered Replication Reports); consider design feature differences.	Replication report with effect-size comparison to original, prereg link, and reflection on reproducibility.
Skills in computational reproducibility & code transparency; experience publishing code	Wilson et al., 2017; Sandve et al., 2013; Hardwicke et al., 2018	Reproducible analysis pipeline in R/Python using a template repositiories; README, environment file, and scripted	Generate reproducible code that is successfully rerun by an independent peer or faculty.

		workflow; code review swap—in class activity to reproduce code with a published paper.	
Familiarity with FAIR data practices & repository use	Gilmore et al., 2018; Soderberg, 2018; van Ravenzwaaij et al. (2025)	Workshop to create a dataset 'FAIR checklist': metadata, codebook, licenses; deposit a de-identified demo dataset in OSF with a landing page.	Deposited dataset with README, codebook, variable dictionary, license, and persistent identifier; passes a FAIRness spot-check rubric.
Navigation of the preprint ecosystem	Kathawalla et al., 2021; Wingen et al., 2022; PsyArXiv overview	Draft & post a preprint with clear 'status' and versioning; add lay summary and limitations to preprint page.	Live preprint with DOI; version history; public-facing summary communicating peer-review status and caveats.
Knowledge of data-sharing best practices & experience sharing data	Meyer, 2018; Gilmore et al., 2018	Draft data-sharing plan for an IRB protocol; generate de-identification workflow; publish an anonymized teaching dataset with usage notes.	IRB-ready data-sharing section; published data package with de-identification rationale and reuse notes.
Critical evaluations of open-science claims	Rubin, 2020; Szollosi et al., 2020; Devezer et al., 2019	Structured debate: advantages, limitations, and scope conditions of preregistration /RRs/replication.	Active participation in open debate; paper summarizing pros and cons.
Awareness of incentives & institutional policy structures	Nosek et al., 2012; TOP Guidelines 2015; TOP 2025 update; DORA	Policy audit: evaluate your program/department/journal against TOP/DORA; propose a feasible incentive (e.g., OS badges, RR adoption) and an implementation plan.	Brief policy memo benchmarking against TOP/DORA and recommending a specific change with timeline and stakeholders; student-led self-study of program promotion criteria in light of DOR guidelines.

Supplemental Materials

Supplemental Table 1.

Transparency and Openness Graduate Student Paper Award Suggested Criteria

1. Is this a preregistration or a Registered Report?
Neither (0)/Preregistration (.5)/Registered Report (1)

2. Were your hypotheses preregistered with a time stamp on osf.io, aspredicted.org, or similar (If no hypotheses were registered, was the study preregistered as exploratory)?
No (0)/Yes (.5)

3. Was the design of your study and/or the variables preregistered with a time stamp on osf.io, aspredicted.org, or similar?
No (0)/Yes (.5)

4. The target sample size and its rationale along with criteria for ending data collection were determined in advance and preregistered with a time stamp on osf.io, aspredicted.org, or similar?
No (0)/Yes (.5)

5. There were no deviations from the planned sample size/termination criteria during data collection or the deviations were made transparent in the work.
No (there were unnoted deviations) (0)
Yes (there were no deviations or deviations were noted) (.5)

6. Analysis plan was preregistered with a time stamp on osf.io, aspredicted.org, or similar?
(0 = No, .25 = Minimal detail, .5 = Substantial detail)

7. The final analysis did not reveal any deviations from the pre-registered analysis plan or the deviations were made transparent in the work.
No (0)/Yes (.5)

8. Analysis code was preregistered with timestamp on osf.io, aspredicted.org, or similar?
No (0)/Yes (1)

9. The raw data for this study have been published (e.g., on osf.io, PURR [Purdue University Research Repository], or similar).
No (0)/Yes (1)

10. Materials (instructions, stimuli, questionnaires, description of measurement procedures, etc.) have been made available online (if possible, i.e., not copyright protected), together with all notes necessary for replication of the methodology by an independent investigator?
No (0)/Yes (1)

11. Has the work been made publicly available (e.g., psyarxiv.com, ZORA)?

No (0)/Yes (1)

Supplemental Table 2

Potential Readings in Each Coding Category Identified by Van Til et al. (in press) generated by members of the Open Science Committee of the Academy of Psychological Clinical Science

Factors Contributing to the Replication Crisis

Documenting the replication crisis

- Bakker, M., Van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*(6), 543-554.
- Bem, D. J. (2011). Feeling the future: Experimental evidence for abnormal retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100*, 407-425.
- Bishop, D. (2019). Rein in the four horsemen of irreproducibility. *Nature*, *568*(7753).
- Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science*, *7*(6), 562-571.
- Miller, J. D., Phillips, N. L., & Lynam, D. R. (2025). Questionable research practices violate the American Psychological Association's Code of Ethics. *Journal of Psychopathology and Clinical Science.*
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.
- Reuters Staff. (2011, November 1). Dutch psychologist admits he made up research data. Reuters.

Statistical power

- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience*, *14*(5), 365-376.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.
- Gervais et al. (2015). A powerful nudge? Presenting calculable consequences of underpowered research shifts incentives towards adequately powered designs. *Social Psychological and Personality Science, 6*, 847-854
- Higginson, A. D., & Munafo, M. R. (2016). Current incentives for scientists lead to underpowered studies with erroneous conclusions. *PLoS Biology, 14*(11), e2000995.
- Reardon, K. W., Smack, A. J., Herzhoff, K., & Tackett, J. L. (2019). An N-pact factor for clinical psychological research. *Journal of Abnormal Psychology*, *128*(6), 493.

HARKing

- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality*

and social psychology review, *2*(3), 196-217.

QRPs / Researcher *df*

- Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist*, *102*(6), 460–465. https://www.americanscientist.org/article/the-statistical-crisis-in-science
- Gopalakrishna, G., ter Riet, G., Vink, G., Stoop, I., Wicherts, J. M., & Bouter, L. M. (2022). Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands. *PLoS ONE, 17*(2), e0263023.
- 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.
- Nagy, T., Hergert, J., Elsherif, M. M., Wallrich, L., Schmidt, K., Waltzer, T., ... & Rubínová, E. (2025). Bestiary of questionable research practices in psychology. *Advances in Methods and Practices in Psychological Science*, *8*(3), 25152459251348431.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, *22*(11), 1359-1366.
- Stefan, A. M., & Schönbrodt, F. D. (2023). Big little lies: A compendium and simulation of p-hacking strategies. *Royal Society Open Science*, *10*(2), 220346.

Publication bias

- Bartoš, F., Maier, M., Wagenmakers, E.-J., Nippold, F., Doucouliagos, H., Ioannidis, J. P. A., Otte, W. M., Sladekova, M., Deressa, T. K., Bruns, S. B., Fanelli, D., & Stanley, T. D. (2024). Footprint of publication selection bias on meta-analyses in medicine, environmental sciences, psychology, and economics. *Research Synthesis Methods, 15*(3), 500–511.
- Ferguson C. J., Heene M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, 7, 555–561.
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science, 345*(6203), 1502–1505.
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological bulletin*, *82*(1), 1-20.

Incentive structures

- Edwards, M. A., & Roy, S. (2017). Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environmental Engineering Science, 34*(1), 51–61.

- Gervais, W. M., Jewell, J. A., Najle, M. B., & Ng, B. K. (2015). A powerful nudge? Presenting calculable consequences of underpowered research shifts incentives toward adequately powered designs. *Social Psychological and Personality Science, 6*(7), 847-854.
- Higginson, A. D., & Munafo, M. R. (2016). Current incentives for scientists lead to underpowered studies with erroneous conclusions. *PLoS Biology, 14*(11), e2000995.
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7*(6), 615–631.
- Yarkoni, T. (2018, October 2). No, it's not the incentives. It's you [Blog post]. Talyarkoni.

Fraud

- Lewis-Kraus, G. (2024, November 19). The Business-School Scandal That Just Keeps Getting Bigger. The Atlantic.
- McKie, R. (2012, September 13). False positives: fraud and misconduct are threatening scientific research. The Guardian.

Measurement issues / Reliability

- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in methods and practices in psychological science*, *3*(4), 456-465.

Practices that work against error detection

- Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE, 6*(11), e26828.

Statistical significance testing

- Lakens, D., Scheel, A. M., & Isager, P. M. (2022). Justify Your Alpha: A Primer on Two Practical Approaches. *Advances in Methods and Practices in Psychological Science, 5*(2), 25152459221080396
- McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon Statistical Significance. *The American Statistician, 73*(sup1), 235–245. https://doi.org/10.1080/00031305.2018.1527253

Incorrect reporting of results

- Nuijten, M. B., Hartgerink, C. H. J., Van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013).

Behavior Research Methods, 48(4), 1205–1226. https://doi.org/10.3758/s13428-015-0664-2

Open Science

General open science readings

- Cook, B. G., Fleming, J. I., Hart, S. A., Lane, K. L., Therrien, W. J., Van Dijk, W., & Wilson, S. E. (2022). A how-to guide for open-science practices in special education research. *Remedial and Special Education*, *43*(4), 270-280. https://pmc-ncbi-nlm-nih-gov.libproxy.lib.unc.edu/articles/PMC9426775/
- Kathawalla, U. K., Silverstein, P., & Syed, M. (2021). Easing into open science: A guide for graduate students and their advisors. *Collabra: Psychology, 7*(1), 18684. https://psycnet.apa.org/record/2021-39459-001
- Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., ... & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature human behaviour*, *1*(1), 0021. https://pmc-ncbi-nlm-nih-gov.libproxy.lib.unc.edu/articles/PMC7610724/
- Nosek, B. A., Alter, G., Banks, G. C., et al. (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425.
- Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry, 23*(3), 217–243. https://doi.org/10.1080/1047840X.2012.692215
- Van Til, K., Phillips, N. L., Du, T. V., Rose, L., Miller, J. D., & Lynam, D. R. (2025). Open-Science Training in American Psychological Association–Accredited Clinical-Psychology Programs: A Registered Report. *Clinical Psychological Science*, 21677026251331550.
  - Registration: https://osf.io/w92cs

Pre-registration

- Benning, S. D., Bachrach, R. L., Smith, E. A., Freeman, A. J., & Wright, A. G. (2019). The registration continuum in clinical science: A guide toward transparent practices. *Journal of Abnormal Psychology*, *128*(6), 528-540
- Benning, S. D., & Smith, E. A. (2023). The registration continuum in personality disorder studies: Theory, rationale, and template. *Personality Disorders: Theory, Research, and Treatment*, *14*(1), 5-18.
- Lakens, D. (2019). The value of preregistration for psychological science: A conceptual analysis. Japanese Psychological Review, 62(3), 221–230.
- Lakens, D. (2024). When and how to deviate from a preregistration. Collabra: Psychology, 10(1), 117094.
- Moher, D., Liberati, A., Tezlaff, J., Altman, D.G., & The PRISMA group (2009).

Preferred reporting items for systematic reviews. *Journal of Clinical Epidemiology, 62*, 1006-1012.

- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. Proceedings of the National Academy of Sciences, 115(11), 2600–2606.
- Van den Akker, O., Weston, S., Campbell, L., Chopik, B., Damian, R., Davis-Kean, P., ... & Bakker, M. (2021). Preregistration of secondary data analysis: A template and tutorial. *Meta-psychology*, *5*, 2625.
- Vize, C. E., Phillips, N. L., Miller, J. D., & Lynam, D. R. (2025). On the use and misuses of preregistration: A reply to Klonsky (2024). *Assessment*, *32*(2), 235-243.

Registered reports

- Chambers, C. D. (2013). Registered reports. *A new publishing initiative at Cortex. Cortex*, *49*(3), 609-610.
- Chambers, C. D., & Tzavella, L. (2022). The past, present, and future of Registered Reports. *Nature Human Behavior, 6*(1), 29-42.
- Scheel, A. M., Schijen, M. R., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with registered reports. *Advances in Methods and Practices in Psychological Science*, *4*(2), 25152459211007467.

Open data

- Walsh, C. G., Xia, W., Li, M., Denny, J. C., Harris, P. A., & Malin, B. A. (2018). Enabling open-science initiatives in clinical psychology and psychiatry without sacrificing patients' privacy: Current practices and future challenges. *Advances in Methods and Practices in Psychological Science*, *1*(1), 104-114.
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. American Psychologist, 61(7), 726–728.

Open code

- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., ... & Yarkoni, T. (2016). How open science helps researchers succeed. eLife, 5, e16800. https://doi.org/10.7554/eLife.16800

Open materials

- Morey, R. D., Chambers, C. D., Etchells, P. J., Harris, C. R., Hoekstra, R., Lakens, D., ... & Wagenmakers, E.-J. (2016). The peer reviewers' openness initiative: Incentivizing open research practices through peer review. *Royal Society Open Science, 3*(1), 150547. https://doi.org/10.1098/rsos.150547

Open access

- Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., ... & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ, 6*, e4375. https://doi.org/10.7717/peerj.4375

Replication

- 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

Detecting "bad" science

- Bishop, D. V. M. (2020). The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice. *Quarterly Journal of Experimental Psychology, 73*(1), 1–12. https://doi.org/10.1177/1747021820968013

General reforms

- Lynam, D. R., Hyatt, C. S., Hopwood, C. J., Wright, A. G., & Miller, J. D. (2019). Should psychologists sign their reviews? Some thoughts and some data. *Journal of Abnormal Psychology*, *128*(6), 541. https://pubmed.ncbi.nlm.nih.gov/31368733/
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*(6), 615-631.
- Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, *69*(1), 487-510.
- 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.
- Tackett, J. L., Brandes, C. M., & Reardon, K. W. (2019). Leveraging the Open Science Framework in clinical psychological assessment research. *Psychological Assessment*, *31*(12), 1386. https://psycnet.apa.org/record/2019-13441-001?doi=1
- Tackett, J. L., & Miller, J. D. (2019). Introduction to the special section on increasing replicability, transparency, and openness in clinical psychology. *Journal of Abnormal Psychology*, *128*(6), 487. https://pubmed.ncbi.nlm.nih.gov/31368726/
- Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., ... & Shrout, P. E. (2017). It's time to broaden the replicability conversation: Thoughts for and from clinical psychological science. *Perspectives on Psychological Science*, *12*(5), 742-756.

High power

- Baranger, D.A.A., Finsaas, M.C., Goldstein, B.L., Vize, C.E., Lynam, D.R., Olino, T.M.

(2023). Tutorial: Power Analyses for Interaction Effects in Cross-Sectional Regressions. *Advances in Methods and Practices in Psychological Science, 6*, 1-13.

- Vize, C.E., Baranger, D.A.A., Finsaas, M. C., Goldstein, B. L., Olino, T., & Lynam, D. R. (2023). Moderation effects in personality disorder research. *Personality Disorders: Theory, Research, and Treatment, 14*, 118-126.

Critiques of open science movement

- Bahlai, C. A., Bartlett, L. J., Burgio, K. R., Fournier, A. M. V., Keiser, C. N., Poisot, T., & Whitney, K. S. (2019). Open Science Isn't Always Open to All Scientists. American Scientist, 107(2), 78–82. thttps://www.americanscientist.org/article/open-science-isnt-always-open-to-all-scientists
- Lui, P. P., Gobrial, S., Pham, S., Giadolor, W., Adams, N., & Rollock, D. (2022). Open science and multicultural research: Some data, considerations, and recommendations. *Cultural Diversity & Ethnic Minority Psychology*, *28*(4), 567. https://pubmed.ncbi.nlm.nih.gov/35404633/
- Ross-Hellauer, T. (2022). Open science, done wrong, will compound inequities. Nature, 603(7901), 363. https://doi.org/10.1038/d41586-022-00724-0