Introducing a Framework for Open and Reproducible Research Training (FORRT)

FORRT*

* FORRT is an open contributorship project in which all contributions are welcomed and formally recognized. An updated list of all involved is at the end of this document.

Abstract

Current norms for the teaching and mentoring of higher education are rooted in obsolete practices of bygone eras. Improving the transparency and rigor of science is the responsibility of all who engage in it. Ongoing attempts to improve research credibility have, however, neglected an essential aspect of the academic cycle: the training of researchers and consumers of research. Principled teaching and mentoring involve imparting students with an understanding of research findings in light of epistemic uncertainty, and moreover, an appreciation of best practices in the production of knowledge. We introduce a Framework for Open and Reproducible Research Training (FORRT). Its main goal is to provide educators with a pathway towards the incremental adoption of principled teaching and mentoring practices, including open and reproducible research. FORRT will act as an initiative to support instructors, collating existing teaching pedagogies and materials to be reused and adapted for use within new and existing courses. Moreover, FORRT can be used as a tool to benchmark the current level of training students receive across six clusters of open and reproducible research practices: 'reproducibility and replicability knowledge', 'conceptual and statistical knowledge', 'reproducible analyses', 'preregistration', 'open data and materials', and 'replication research'. FORRT will strive to be an advocate for the establishment of principled teaching and mentorship as a fourth pillar of a true scientific utopia.

[working document here: https://tinyurl.com/FORRTworkingDOC]

Keywords: Science Education, Teaching, Transparency, Open Science, Open Scholarship

The teaching and mentoring of transparent, reproducible and open research practices is the clearest indicator of the degree to which institutions and/or departments embody principles of credible science (FORRT, 2019).

Advancing science requires open, transparent, reproducible and cumulative research practices. Perhaps the oldest support for this adage dates back to the 1600s, when the newly founded Royal Society, the world's oldest scientific institution, chose as its motto *nullius in verba* ("take nobody's word for it"). The skepticism enshrined in its motto anticipated that scientific progress is optimized when the conditions for dissent are maximized. Optimal conditions occur when researchers – or any interested party – have unrestricted access to study protocols, data, and analyses. This allows researchers to not only reproduce results and incentivize replications, but also advance science via investigating the conditions under which findings hold and put competing hypotheses to the test. These fundamental features of science (McNutt, 2014; Miguel et al. 2014), gathered under the umbrella term "Open Science" or "Open Scholarship" (Crüwell et. al, 2019; Hampson, 2019), increase the robustness of scientific claims allowing for reliable interpretation of results.

Unfortunately, open practices have been shown to be underutilized in practice (Ioannidis, Munafo et al., 2014; O'Boyle, Banks, & Ginzalez-Mulé, 2014). Worse, questionable research and measurement practices are widespread (Flake & Fried, 2019; Fiedler & Schwartz, 2016; Lowenstein & Prelec, 2012) and coincide with the misalignment of incentives in the current system of academic publishing (Bakker, van Dijk, & Wicherts, 2012; Higginson & Munafo, 2016). Unsurprisingly, the so-called "replication crisis" ensued when the first replication initiatives released its findings (Franco, Malhotra, & Simonvits, 2014; Ioannidis, 2005; Open Science Collaboration, 2015; Simmons, Nelson, & Simonsohn, 2011).

Almost 10 years later, the replicability rate is estimated to be around 47%—across six large-scale replication projects, only 90 out of 190 studies have generated results similar to that claimed by original authors (Camerer et al., 2016; Camerer et al., 2019; Ebersole et al. 2018; Klein et al., 2014; Klein et al., 2018; Open Science Collaboration, 2015). Note that we use replicate to indicate testing the same hypotheses with data from a different sample, and reproduce to indicate returning the same results with the same data, i.e. by repeating analyses. These replication efforts have ushered in an era in social sciences where globally distributed networks of laboratories (e.g., the Psychological Science Accelerator; Moshontz et al. 2018) and collaborative initiatives (e.g., Mannheim Open Social Science Conference - Crowdsourced Replication Initiative, Political Science Replication Initiative, and Collaborative Replications and Education Project (CREP)) are leading in replicating studies across disciplines and countries.

If *nullius in verba* is the earliest example of the union between science and open practices, the latest crowning achievement might be that of researchers working at the Center for Open Science (COS). They proposed three overarching guiding principles to steer social sciences towards academic utopia:

1. opening scientific communication (Nosek, & Bar-Anan, 2012),
2. restructuring incentives and practices (Nosek, Spies, & Motyl, 2012), and
3. crowdsourcing science (Uhlmann et al., 2018).

Further, they have partnered with DARPA (the United States' Defense Advanced Research Projects Agency) to implement ways to systematically and reliably evaluate the credibility of scientific findings via artificial intelligence. SCORE (Systematizing Confidence in Open Research and Evidence) is projected to last three years with an estimated cost of $7.6 million (Center for Open Science, 2019). Taken together, these steps reflect a widespread awareness of, and call for, improved practices ushering in the "credibility revolution" (Vazire, 2018). This revolution includes higher standards of evidence, preregistration, direct replication, transparency, and openness in the process of science making (e.g. Nosek et al., 2018; Nosek & Lakens, 2014; Zwaan et al., 2017; Wagenmakers et al., 2012).

In our view, a *scientific utopia* has a fourth pillar, whose principal goal is to familiarize students, who are future consumers of science and perhaps themselves knowledge producers, with the intricacies of the process of science. We believe the teaching and mentoring of reproducible and open research practices is the clearest indicator of the degree to which institutions and/or departments embody principles of credible science. This demonstration goes beyond paying lip service to best practices, and ensures that students are *trained* to engage in these practices wheresoever. In this article, we first justify these assertions and then present a didactic infrastructure designed to recognize and support the teaching and mentoring of open and reproducible research in *tandem* with prototypical subject matters in higher education.

The need for integration

Nothing in science has any value to society if it is not communicated.

(Anne Roe, 1953, pp. 17)

The noblest of crafts, teaching is seen as the royal road to learning. It affords opportunities to guide young minds toward greater understanding while deepening one's own. Teachers, in addition to edifying students' education and future, strive to kindle that primal spark igniting passion for knowledge. In higher education, this pivotal mission falls on faculty, whose job description predominantly consists of teaching and research (along with, to a lesser extent, administrative tasks and service). The view that scholars are better equipped to impart into students' knowledge gained through research is enshrined in the academic ethos, and embedded into universities' histories, objectives, and institutions. The underlying rationale behind teacher-researchers is that those on the cutting edge of their discipline provide students a sense of authenticity and excitement through active involvement and contributions to the field's advancements. Marsh and Hattie (2002) add teacher-researchers are in better measure to present a birds-eye view of disciplines and "are more effective at instilling an actively critical approach to understanding complex research findings rather a passive acceptance of facts" (pp. 604).

We expand this argument to faculty with solely teaching responsibilities, in that these scholars are also well placed to integrate criticality, and an understanding of the process of research, into their teaching. We recognize the balance of scholars' teaching and research responsibilities varies widely, and we aim to ensure the Framework for Open and Reproducible Research Training (FORRT) is relevant for all teachers. Suitability as educators stems from familiarity with the procedures of knowledge generation, not from knowledge of scientific facts alone. This competence is habitually neglected in higher education – to the detriment of its students – where it is commonplace to give deference to communicating scientific/technical information over teaching a way to knowledge. Marks (2009) raises a fundamental question about science education:

> If science is a process of knowledge production, then is science education best expressed as teaching students the process or as teaching them the knowledge itself? If we focus on teaching students the accumulated knowledge, the facts of science, then we are not actually teaching them science. Rather, we are teaching them science's products, and indeed we are misleading them by substituting the teaching of scientific facts, as if it were the teaching of science itself (pp. 22).

In higher education, it is not sufficient to teach about what scientists know – and it may be deceiving to do so. Science education inevitably entails learning about *how* scientists learn *what* they know. This education is an instance in which the whole is greater than the sum of its parts; that is, a collection of scientific evidence (or the unreflective and decontextualized transmission thereof) does not yield scientific literacy. Thus, should higher education institutions, and scholars within, want to fulfill its prescribed goals, the integration between the teaching of subject matters and the tenants of modern science is *condicio sine qua non.*

Current practices and why academics should care

"Being able to identify reproducible research and appreciate openness and rigor is an essential skill to be a good consumer of research" (Crüwell et al., 2018).

Core to science, any science, is a system of methods of inquiry and knowledge production. While these vary widely from one field to another, contingent on studied phenomena (i.e., formal, natural, social and applied sciences) and on the use of methods (i.e., qualitative and/or quantitative), a common thread unites the different flavors of scientific enterprise: transparent, reproducible and cumulative research practices. Yet, it is still very common that graduates and undergraduates are deprived of ever hearing about these tenets through their education. As an example, while there have been several efforts to gauge academics' awareness and opinions of best research practices (Baker, 2016; Fanelli, 2018; Harris et al., 2018; Iqbal, et al., 2016; Raheel & Kujan, 2016), so far, only one study estimated its prevalence among students. Results are concerning. Chopik and colleagues (2018) reported that only 31% of the students they surveyed had heard of the reproducibility crisis. This finding is not ideal because hearing about these terms (e.g., reproducibility, credibility revolution, pre-registration, registered reports, HARKing, etc.) is a reasonable indicator of the quality of science education students are exposed to. Insofar higher education institutions aim to prepare students for an increasingly technological workforce whose advances affect everyone's lives in ever shorter time spans, its training ought to provide compulsory knowledge for practical adjudication of scientific output.

Furthermore, the idea that science education is only desirable for future academics is unbecoming, as good consumers of research should be able to identify reproducible research and understand the value of openness and rigor in scientific research (Crüwell et al, 2018). Since the majority of higher education students will not pursue academia, and thus will not receive further scientific training, higher education presents perhaps the last, best opportunity to formally teach prospective consumers of research how to conduct scientific research. In doing so, scholars not only increase the chances of leaving a lasting impression influencing students' beliefs about science, but also how they envision and perceive science, which ultimately dictates how they will interact with it in their future.

Consider the rate by which higher education is becoming the norm. In the US, master's degrees are as common now as bachelor's degrees were in the 1960s, and in Europe, countries have quintupled, on average, the population with tertiary degrees in the same period (World Bank, 2010). In sheer numbers, 4.7 million students graduated from tertiary education establishments in the EU-28 in 2016 and 4 times as many were still studying (EUROSTAT, 2016). Analogously, in the US, between 35-40% of the population indicate some sort of higher education degree (CPS, 2014). Today's bachelors and master's students are likely to be tomorrow's decision makers and opinion leaders. Even after leaving higher education, they will have numerous opportunities to impact science, positively or negatively, as discerning consumers of scientific research, or as patrons of science via voting and philanthropy.

Social Justice and Inclusion

Perhaps the most overlooked aspect of the importance of science education, and the tenets of modern scholarship like transparency and open practices, is that of social justice. The current model of scientific production and educational practices reproduces global inequalities. Unfortunately, academia is still a privilege few can afford. Science is built on the same foundations of society itself, which means that science has inherited many of society's systematic barriers hindering the success of traditionally marginalized groups – whether based on gender, race, ethnicity, origin or social class. In science, wealthy and elite-educated individuals are over-represented, as are males, whites, and citizens of western industrialized rich countries.

Current academic structures around science communication, prestige-based funding, and teaching and mentoring practices perpetuate the power imbalances and exploitation of privilege by failing to address key issues salient to marginalized social and geographical contexts (Siler et al., 2018) and increasing the inequality around the diffusion of scientific ideas (Morgan et al., 2018). Known as the Matthew effect (Merton, 1968), it describes how, contrary to popular belief, the key factors predicting recognition of one's scientific work are factors like seniority, location, and institutional reputation – unfortunately, not quality. This then trickles down to funding, visibility and citations of one's scientific work (Merton, 1988). For example, Bol, de Vaan, and van de Rijt (2018) have shown that fund/grant winners slightly over the funding threshold accumulated more than twice the funding during the subsequent eight years as non-winners, even with near-identical review scores.

Similarly, junior researchers who co-author with prominent scientists have an ongoing competitive advantage, including continued co-authorships with top-cited scientists and a higher likelihood of becoming one themselves (Li et al., 2019). Differences in recognition, reputation and scientific productivity may be explained by cumulative advantage and search costs minimization by journal editors. The parallels with other forms of inequality which are widespread across societies – e.g., economic inequality wherein economic attainment is often a function of one's circumstances rather than merit – are not accidental. Rather, it is a reflection of societal arrangements over which Academe was built.

To mitigate the detrimental effects on the access to, learning about, and production of scientific content, it is paramount to create conditions for knowledge to become a public good - accessible to all members of society (Steltenpohl, Anderson, & Daniels, 2019). One way to achieve this goal is to integrate the teaching and mentoring of subject matters with open and reproducible research practices. Higher education institutions, and scholars therein, can then maximize every student's likelihood of present and future engagement with resources, facilitating the acquisition of knowledge and bolster opportunities that would otherwise be inaccessible to disadvantaged individuals. In leveling the playing field, abiding scholars and institutions would contribute towards greater diversity and democratization of science, and improved scientific literacy, consumption, and participation. The burden of implementing these efforts can be reduced if allied with new and existing initiatives such as open educational resources (OER), massive open online courses (MOOCs), and websites providing free access to millions of (otherwise paywalled) research papers and books. There is also a plethora of general and subject-specific infrastructures to achieve open education through new pedagogies such as Open Core Courseware, Git-Hub, and freely available statistical suites.

In addition to the lack of open materials and software, high-tech hardware can often impede the learning and practice of science. Across a variety of fields, hardware is a vital part of the scientific process, and advances in instrumentation have been central to scientific revolutions (e.g., telescopes, microscopes). By adopting principled research and education practices scientists have gone all-out to provide affordable, reliable, and easy to use instrumentation to enable as many people as possible to contribute to research (GOSH, 2019). With hopes to make open-source hardware ubiquitous by 2025, the Open Science Hardware community has emerged as a powerful movement to reduce barriers obstructing access to high-tech hardware necessary for the pursuit of scientific knowledge. From 3-D printed micropipettes labware meeting ISO standards (Brennan et al. 2017) to the sophisticated equipment used at Large Hadron Collider at CERN, a rich variety of free and open hardware, often exceeding the quality of its paid counterparts, can be found at the Open Hardware Repository. When providing non-scholars with similar access to, and knowledge about, deeply intricate scientific content, new scientific breakthroughs become possible. More than ever, members of the general public are able to participate in research projects, often in collaboration with, or under the direction of, professional scientists and their institutions. For example, a study surveying 388 scientific projects on biodiversity found that between 1.36 and 2.28 million citizen scientists have worked and meaningfully contributed to these projects generating between $667 million to $2.5 billion annually (Theobald et. al., 2015).

These developments are based on the idea that science is a community endeavor, wherein open inquiry and public participation is at the heart of the scientific enterprise. Here, the integration of teaching subject-matters along with open and reproducible scholarship sits at the very foundation of these advances as it gives students, citizen scientists, and aspiring scholars the necessary tools to engage with science-making. Furthermore, open teaching initiatives allow for a wider variety of societal actors – including researchers, citizen scientists, policy makers, businesses, and third sector organizations – to collaborate and contribute their expertise towards a common goal in science production. Ultimately, this means better aligning the scientific process and its outcomes with principled values, societal expectations and its needs.

In what follows, we advance our ideals about principled teaching and principled mentoring, which are foundational to the edification of FORRT as framework for open and reproducible research training.

Principled teaching

The objectives and the methods for the teaching of the responsible conduct of research should be nothing less than those used for the teaching of other skills and abilities valued within a discipline. (National Research Council, 2002)

There is something to be said about the ethics of current teaching practices. An increasing number of scholars have incorporated reproducible, transparent, and open principles to their research pipeline. However, the associated social and ethical responsibilities towards the teaching of these principles is yet to be realized. Moreover, when institutions and academics make efforts to assimilate open science into their educational programs to foster awareness about the intricacies of the scientific production, it usually is through the "research integrity" lens. Thus, it is framed more *for researchers* in the process of making research, rather than for imparting these lessons *to students*.

The view that a gap in information is justifiable regarding the intricacies of scientific production between researchers and non-researchers is not restricted to academics but also includes scientific institutions, open-science initiatives, and labs adopting best research practices (e.g. Fischer & Zigmond, 1996; Gifford, 1994; Hensel, 1991). But, there are exceptions to the rule. In 2002, the National Research Council makes several mentions on the importance of teaching 'students' (unfortunately, unspecified) about integrity in scientific research, whereupon they argue that “useful insight into the best practice for education in the responsible conduct of research comes by analogy to the education of students in the critical analysis of the research literature in their fields,” and, “education in the responsible conduct of research should be no less integral to the education of a researcher” (p. 85).

What is more, out of more than 200 actors in the Open-Science Grassroots Community Networks, only four mention ‘teach-’, mostly to indicate resources’ availability. Similarly, the literature review we conducted (described later and presented in table 2) to develop the framework aspect of FORRT, a surprisingly low frequency of mentions of ‘teach’, ‘teacher’, and ‘teaching’ were found in the surveyed readings. Taken together, and while certainly suggestive evidence, these are indicative that the norm in higher education is to (a) separate *students* from *aspiring researchers*, and (b) omit educating *students* about the replication crisis, credibility revolution, and open-scholarship. This oversight is not surprising given that social science disciplines have historically focused on single studies while relatively little attention is given to the issue of its replicability (*cf*., Asendorpf et al., 2013).

It is still common practice to teach subject matters, at undergraduate and graduate levels, for quantitative and qualitative science, without resolutely emphasizing that scientific claims should be taken in light of epistemic uncertainty – i.e., the intrinsic incertitude about the validity of truth claims (including scientific). Perhaps the most conspicuous case is the pervasive belief that statistical analyses give a ‘seal of objectivity’ to the research conclusions, whereas this preconception may be ill-advised (Freese & Peterson, 2018). Berger and Barry (1998) argue that acknowledging data interpretation’s inherent subjectivity allows us to recognize the role statistical analysis plays in integrating new evidence with existing knowledge. It is not difficult to find heated discussions in the literature on fine details of model specifications, but harder to find discussions on measurement and research design. The late astronaut Neil Armstrong illustrated this point: “i*f you are off by an inch on the landing, no big deal. If you are off by an inch on the takeoff, you miss the moon by a million miles*.” Inasmuch as the verity of quantitative scientific findings hinges on probabilistic uncertainty, research design, measurements and instruments, sampling methods and representativeness, we should communicate the facts of science relative to the process by which it was acquired. This idea is enshrined at the core of the open science movement, the credibility revolution (Vazire, 2018), open access (Nosek, & Bar-Anan, 2012; Tennent et al., 2016), and better measurement practices (Flake & Fried, 2019).

These notions are not restricted to quantitative sciences. The same tenets are applicable to the production of knowledge in general. Observational and qualitative research, for example, are contingent upon methods utilized, analytical choices, and theoretical framework. This is also true of literary and language studies; Queneau (1947), for example, has shown that it is possible to tell the same story in 99 different ways, each in a different style. Four centuries before, Erasmus of Rotterdam (1523), in *Copia: Foundations of the Abundant Style*, demonstrated 195 variations on the sentence "Your letter delighted me greatly." The garden of forking paths is an inescapable metaphor of non-deterministic objective epistemic systems (*cf.* Tannert, Elvers, & Jandrig, 2007). Merton (1957) asserted that given this path dependency, scientific progress can only occur through “organized skepticism” – i.e., continuous and iterative debates on the role of random and systematic errors, competing explanations, or interpretations (Peters & Dunwoody, 2016). Similar issues have been raised for qualitative research and it has been proposed that preregistration can be used to foster analytic transparency (Haven & Van Grootel, 2019)

Typically, academics are familiarized with, and utilize, heuristics in line with epistemic uncertainty, especially when it comes to judging the output of each other’s work. Indeed, epistemic uncertainty is one key aspect motivating the peer-review system. Unfortunately, however, science education – i.e., teaching and mentoring – has yet to take heed from the credibility revolution and realize that the lack of integration between field content and epistemic uncertainty undermines the paragon of higher education. In addition, it has the unfortunate consequence of teaching students a false sense of certainty about 'the facts' of science, which can be thought of as misleading, if not unethical (Marks, 2009).

These assertions find normative support in several social sciences' codes of conduct – we use psychology as an example, but do not intend to limit focus on any particular field. For example, the Ethical Principles of Psychologists and Code of Conduct of the American Psychological Association (APA) stipulates the following guidelines:

- When engaged in teaching or training, psychologists present psychological information accurately (Section 7.xx);
- Psychologists undertake ongoing efforts to develop and maintain their competence (Section 2.03); and
- Psychologists take reasonable steps to ensure that course syllabi are accurate regarding the subject matter to be covered, bases for evaluating progress, and the nature of course experiences (Section 7.03).

Considering the reproducibility crisis, we believe a probable interpretation of these guidelines involves being supportive of teaching subject-matters while communicating the possible caveats and uncertainty associated with any scientific work, and to focus on continuous self-improvement and education as a mean to teach Psychological Science accurately. Indeed, the ideas instilled on the APA's Code of Conduct portray the scientific enterprise as a continuous learning process, whose application in science education not only contributes to a more realistic portrayal of the subject matter but also helps shatter the academic glass ceiling - which ultimately could promote the curiosity and interest of students about the academic profession.

Principled mentoring

"*If I have seen further it is by standing on the shoulders of giants*"

(Sir Isaac Newton)

Many academic achievements would not have been possible if it were not for exceptional mentoring and provision of resources. Central to the academic ethos, the relationship between mentor and mentee is one of scientific genesis, incremental contributions, and personal and professional development. For mentors, it is a chance to not only guide young minds towards greater understanding, but also to deepen one's own. For students, this stage can be liberating and empowering as, for many, it is only here that they leave the strictly formal, and often hierarchical, pupil-educator relationship to experience a more collaborative, levelled rapport. Undeniably, the typical mentorship involves a great deal of idiosyncratic bureaucratic procedures, but the imprimatur generated on students by coordinating in qualified capacity with professional academics should not to be underestimated. Exchanging constructive ideas and mutual feedback provides a unique, and often first, opportunity to be heard in an academic setting, to feel one can aggregate scientifically. Whether it is on an original project, existing, or as a part of a coordinated effort of a larger project, they all aim at producing knowledge. This exceedingly legitimizing academic experience is optimized when intrinsic and extrinsic motivations align, that is, when students feel they have something with which to contribute, and the product of it would be valued. One way to achieve this is to expose students to field-specific (and general) knowledge about the intricacies and procedures of science making, its flaws and merits. Apropos, open and reproducible tenets serve as an accessory to mentorship pedagogy.

Principled mentoring – mentorships integrating subject matters and open and reproducible science tenets — has three advantages: first, it endows students with the ability to discover on their own the myriad of ways their skillset allows them to do better than previous attempts. Second, it enables students to see their work and efforts serve a (higher) purpose – to aggregate and positively impact the scientific community. Third, it contributes to students' overall scientific and technological edification, yielding a better trained workforce and consumers of science.

While FORRT's ideals operate at the higher-education level, it has a role to play in the instruction of prospective scholars.

Training Future Academics

"Tout enseignant est enseigné, tout enseigné est enseignant"

(Inscription, Sciences Po, Paris)

As it goes, scholars' professional success depends on publishing, which, allied with journal editors' preferences for novel and positive contributions, effectively inflate the rate of published false effects. As scientists seek to uncover *truth* – conditional upon available evidence, scientific method and vigorous and open dissension – these dysfunctional incentives ultimately lead to a biased and unchallenged body of literature. In other words, the current structure contributes negatively to the process of science-making and knowledge accumulation (e.g. Smaldino & McElreath, 2016).

This notion, while counter-intuitive to scholars themselves, is particularly surprising to those outside academia, at least in part, because higher education students are largely left on the dark about the matters of science. The next generation of researchers is virtually unaware of Academia's defective structure of incentives, the way it manifests itself (i.e., garden of forking paths, researchers' degrees of freedom, selection bias, HARKing and *p*-hacking), and its consequences (i.e., inflated false positive, asymmetric distribution of p-values in publications, reproducibility crisis). And as these topics are not imbued at the onset of their professional careers, and as higher-ed students only know what they are taught or exposed to, it is naive to expect the next generation of researchers will understand, value, and maintain the better institutional constraints and open research procedures implemented at the dawning of the reproducibility crisis. For example, if future scholars are not taught about the perils of *p*-hacking (a large collection of practices intended to yield a 'significant' *p*-value), why would they care about pre-registering studies? At best, the lack of proper training delays the onset of best research practices. The fact that it is still common practice to teach subject matters without emphasizing that scientific claims should be taken in light of probabilistic uncertainty, research design, samples and measurements used undermines the beneficial change achieved in recent years. At the very least, by upholding the status quo in teaching practices scientists risk missing out on generational leaps of scientific progress - if not a recurrence of the closed and irreplicable practices that gave rise to the scientific crisis in the first place.

By no means is this discussion representative of the breadth of challenges facing aspiring scholars. In addition to the filter of inequality (economic, racial/ethnic and geographical), which fixes both quality and access of opportunities and educational resources: future scholars are only exposed to field-specific content. Too often graduate program lack adequate training in research methods, data acquisition, data management, and statistical coding. And often, graduate students are left on their own in learning about project and time management, publishing strategies and journal's idiosyncrasies, teaching skills, peer-review, managing research assistants, conferencing, applying for funding and granting institutions/foundations, navigating mental health issues, and dealing with micro- and macro-aggressions while minority. It is no surprise that, in 2016, a survey conducted by Nature, researchers identified better mentoring and supervision as one of the most likely strategies to improve science (Baker, 2016).

It would be easy to blame this unfortunate state of affairs solely on researchers, but here too, dysfunctional structures are at play. The context is that academic faculty positions are scarce, and number of applicants vastly outnumbers the available offers. Aspiring academics also must grapple with the reality that factors beyond the quality of their work affect the chances of their promotion substantially (e.g., gender, race, rank and prestige of university, journal impact factor; cf. Morgan et al., 2018; Van Dijk et al., 2014). Also worrisome is that the current incentive structure seems to favor quantity over quality in terms of published works (Allen & Mehler, 2019; Flier, 2017; Moher et al., 2018; Naudet et al., 2018), which is to say scholars are not incentivized to pursue better research practices such open and reproducible science because it puts them at a disadvantage compared to everyone else. It is under this arrangement that investigators are told that their job description is to teach and conduct research - as productively and as efficiently as possible - at the same time (Vicens & Bourne, 2009). But the truth of the matter is that academics are neither incentivized nor trained in the art of teaching, which in part, is a reality arises because academia is ultimately a market of prestige, in which administrative decisions such as hiring and promotion are almost exclusively grounded on research output and its impact (Fox, 1992; Garvin, 1980). So, researchers' commitment to teaching are often sensitive to the competing demands of research, administrative bureaucracy, mentoring, academic management and teaching.

Science is a community enterprise, which relies on a set of social norms, principles, and expectations defining a scientist's ideal conduct. Despite counterproductive structures and misaligned incentives, scientists persevere in their efforts to aspire increasingly improved goals. In the recent decade, academics made strides in steering the production of science towards *scientific utopia,* an ideal driving the adoption of transparent and reproducible practices (Nosek, & Bar-Anan, 2012; Nosek, Spies, & Motyl, 2012; Uhlmann et al., 2018; but see DORA and TOP guidelines at the incentives level). While there is a long way to go, these developments are already bearing fruit and substantive improvements were achieved with respect to the adoption of open and reproducible science practices (TAOS/COS). For example, the Tracking Adoption of Open Science (TAOS) project aims to provide insight on the adoption and spread of open science using an easily accessible method of assessment via the use of the OS tool at www.osf.io. Preliminary results indicate a rate of adherence of 35%, which is higher among the youngest scholars, and which has been strongest in Australian, the Netherlands and the UK – in comparison with the US, Germany and Canada (full report here: https://cos.io/blog/rise-open-science-psychology-preliminary-report/). This movement is heartening, and evidence that academia can change for the better.

Taken together, we hope to have depicted a revealing portrait of the current demands placed on teaching practices, its consequences and challenges. We also hope to have reasoned convincingly that the integration between teaching and mentoring of subject matters and open and reproducible science is not only desirable, but necessary. And we hope to have established with our review of implemented initiatives that there is not yet a way for willing scholars to easily transition from the traditional educational set-up to an integrated, open, and transparent teaching and mentoring pedagogy. Our project, a Framework for Open and Reproducible Research Training (FORRT), aims to rectify this.

FORRT: from Vision to Reality

The provision of instruction in the responsible conduct of research derives from a premise fundamental to doing science: the responsible conduct of research is not distinct from research; on the contrary, competency in research entails responsible conduct and the capacity for ethical decision-making (NIH, 2012, pp. 84).

In an effort to support this movement's sustainable development, we propose FORRT, a Framework for Open and Reproducible Research Training. As the tentative fourth pillar of *scientific utopia*, its main goal is to provide educators a pathway towards the incremental adoption of principled teaching and mentoring practices. That is, gradually integrating discipline content with open and reproducible science tenets, and opening ensuing materials. FORRT supports this endeavor through a three-pronged approach:

1. modifying the academic incentive structure regarding teaching and mentoring through assessment and commendations of principled teaching and mentoring, which in turn are documentable and relevant to the researchers' visibility, prestige, tenure and promotion reviews;
2. providing a comprehensive, straightforward, and accessible didactic framework to qualify and quantify current adherence to principled teaching and mentoring and providing a clear perspective going forward; and
3. equipping scholars with the necessary tools to acquire familiarity on open and reproducible research practices, its accompanying teaching and mentoring resources, and with successful and implemented pedagogies to help scholars realize principled education with ease.

Indeed, FORRT does not intend to further burden scholars' packed schedules, but rather facilitate the crystallization of open and reproducible tenets across researchers' professional activities in a feedback loop between cultivating better research practices, mentoring, and teaching. Accompanying these rather scientifically-oriented recommendations, FORRT also hopes to start a conversation on, and encourage debates about, both the ethics and social impact of teaching substantive topics without regard for scientific openness, epistemic uncertainty, and the credibility revolution (e.g., educating the consumers of science and potential mitigation of downstream consequences of current anti-science, anti-climate change and anti-vaccination movements; for examples for the former and latter, respectively). Lastly, FORRT also advocates for the progressive opening of teaching and mentoring materials as a means to facilitate access, discovery, and learning from extant teaching and mentoring pedagogies and materials to those who otherwise would be educationally disenfranchised. In sum, with FORRT's consolidation of the *what, why, and how* of improving teaching and mentoring practices, it hopes to contribute to the sustainability of the grassroots movement for the improvement of science and help foster social justice through the democratization of scientific educational resources and pedagogies.

In the previous sections, we hope to have persuasively discussed the *what,* and the *why* of FORRT. The following sections will focus on the *how*.

The framework: educational nexus

We conceptualize FORRT as an *educational nexus*, integrating its diverse components into an educational infrastructure serving those wishing to learn, adopt, and disseminate open and principled teaching. FORRT's e-learning platform is a hub for community-driven initiatives and resources. The focus lies not on aggregating resources but on making sense of existing initiatives and materials, and filling in the gap where none exist. Teachers' and researchers' time constraints are substantial, and any time or effort spent on education and learning are likely to pose a challenge to other commitments – sometimes deemed worthier, given the lack of incentives and recognition for quality teaching and mentoring. This challenge motivated FORRT to develop strategies and propose solutions to mitigate the effects of competing academic interests and help scholars implement open and principled education in their workflows.

Thematic clusters and sub-clusters. Based on an informal literature review, FORRT identified six clusters pervading open and reproducible literature: (1) reproducibility and replicability knowledge, (2) conceptual and statistical knowledge, (3) reproducible analyses, (4) preregistration, (5) open data and materials, and (6) replication research. FORRT further specifies its comprising components as a means to draw fuzzy boundaries between clusters while allowing for diversification and heterogeneity in how each chooses to integrate these sub-clusters with field content. Under Cluster 1, for example, six sub-clusters were identified: (1) 'reproducibility crisis and credibility revolution'; (2) 'exploratory and confirmatory analyses'; (3) 'questionable research practices, its theory, and prevalence'; (4) 'proposed initiatives for the improvement of science: on statistics, measurement, teaching, data sharing, code sharing, pre-registration, replication'; (5) 'ongoing debates' (e.g. incentives for and against open science); and (6) 'ethical considerations for improved practices'. The breakdown of each cluster into sub-categories provides scholars with useful information on the extant of open science scholarship, and how they are connected to one another.

Table 1. FORRT's Cluster and Sub-clusters.

I. Reproducibility Crisis and Credibility Revolution	**II. Conceptual and Statistical Knowledge**	**III. Reproducible analyses**	**IV. Open data and materials**	**V. Preregistration**	**VI. Replication research**
Attainment of a grounding in the motivations and theoretical underpinnings of reproducible and open research. Integration with field specific content (i.e., or grounded in the history of replicability)	*Enacting this principle indicates that students attain a grounding in fundamental statistics, measurement, and its implications.*	*Reproducible analyses allow the checking of analytic pipelines and facilitate error correction. Enacting this principle requires students to move towards transparent and scripted analysis practices*	*Enacting this principle indicates that students have attained a grounding in open data and materials in both; using and sharing*	*Preregistration entails laying out a complete methodology and analysis before a study has been undertaken. This facilitates transparency and removes several potential QRPs.*	*Replication research takes a variety of forms, each with a different purpose and contribution. Reproducible science requires replication research.*
i. Reproducibility crisis and credibility revolution.	*i.* The logic of null hypothesis testing, p-values, Type I and II errors (and when and why they might happen).	*i.* Strengths of reproducible pipelines.	*i.* Knowledge of traditional publication models. Open access publishing, preprints	*i.* Purpose of preregistration - distinguishing exploratory and confirmatory analyses, transparency measures	*i.* Purposes of replication attempts - what is a 'failed' replication?
ii. Exploratory and confirmatory analyses.	*ii.* Limitations and benefits of NHST, Bayesian and Likelihood approaches.	*ii.* Scripted analyses compared with GUI.	*ii.* Reasons to share; for science, and for one's own practices	*ii.* Preregistration and registered reports - strengths and differences	*ii.* Large scale replication attempts
iii. Questionable research practices (its 'theory'), and prevalence.	*iii.* Effect sizes, Statistical power, Confidence Intervals.	*iii.* Data wrangling	*iii.* Repositories; e.g. OSF, FigShare, GitHub	*iii.* When can you preregister? Can you pre-register secondary data?	*iii.* Distinguishing conceptual and direct replications
iv. Proposed improvement science initiatives on statistics, measurement, teaching, data sharing, code sharing, pre-registration, replication.	*iv.* Research Design, Sample Methods, and its implications for inferences.	*iv.* Programming reproducible data analyses.	*iv.* Accessing/sharing others data, code, and materials	*iv.* Writing a preregistration.	*iv.* Conducting replication studies; challenges, limitations, and comparisons with the original study
v. Ongoing debates, (e.g. incentives for and against open science).	*v.* Questionable research (QRPs) & measurement practices (QMPs).	*v.* Open source and free software.	*v.* Ethical considerations	*v.* Comparing a preregistration to a final study manuscript.	*v.* Registered Replication Reports
vi. Ethical considerations for improved practices.	*vi.* Understand the relationship between all of the above.	*vi.* Tools to check yourself and others; statcheck, GRIM, and SPRITE	*vi.* Examples and consequences of accessing un/open data	*vi.* Conducting a preregistered study.	*vi.* The politics of replicating famous studies

Curated readings. The first is FORRT's entry-level selection of open and reproducible science literature aiming to convey its basic tenets as succinctly as possible. This provides unfamiliar researchers a starting point – a guided tour, if you will – of open and principled education principles. FORRT's entry-level curated list is designed such that small-time commitments can yield meaningful improvements in reproducible and open-science teaching. As shown in *Table 4* below, it provides two essential sources for open education, principled teaching, principled mentoring, and two for each FORRT cluster. These curated lists incentivize incremental improvements while decreasing informational and time encumberment.

Crowdsourced Resources Nexus. FORRT recognizes the importance to democratize the access and development of teaching and mentoring materials for open and reproducible science. For this, FORRT aims to serve as a centralized educational platform (i.e., *nexus*) for a rich, diversified, and up-to-date collection of open and reproducible science resources for teaching and mentoring. These live educational resources collectively provide a comprehensive and plural library of online teaching and mentoring materials. Currently, it lists approximately 600 resources between publications, books, blogs, syllabi, podcasts, educational videos, OS courses, OSF projects, software, a variety of websites and other collections of resources.

FORRT verifies all resources therein and adds relevant tags, description, origin, and appropriate clusters. This massive undertaking is done as a mean to maximize usability and fulfill its potential. In addition, this database is open to the community's input. That is, scholars add educational resources via a Google form that feeds into the live database, which increases its impact and visibility. The form is straightforward and can be easily filled in with tags so that the entered resources are appropriately categorized for easier retrieval. FORRT is currently devoting resources to transform this database into a ‘back-end’ source for a more functional, scalable and visualizable educational asset.

Implemented pedagogies. FORRT aims to collect and catalog instances in which principled education are implemented. FORRT has created a centralized platform aiming to provide detailed examples of successful pedagogies across social sciences – in teaching, mentoring, and openness of these materials – in the real world. We hope that interested parties – i.e., scholars, instructors, educational institutions aiming to develop or reform degree programs and courses – can use FORRT’s pedagogies as an initial template towards the creation of their personalized pedagogies. FORRT intends to facilitate the initial costs of integrating field specific education with open and reproducible science tenets.

For example, it has been suggested that one potential solution to the replicability crisis is to make replication an explicit part of higher education pedagogy (Everett & Earp, 2015; Frank & Saxe, 2012; Grahe et al., 2012; Hawkins et al., 2018; King et al., 2016; LeBel, 2015; Standing, 2016). So, in addition to formal teaching of courses on replications (Janz, 2015), mentorship offers an important avenue into the dissemination of open and reproducible research principles. One strong example of principled mentoring is a consortia approach (Button et al., 2018) in which students collaborate on an overarching project (while still ensuring sufficient independence). Importantly, students are mentored through a research process that maps more effectively to best practices; studies are typically pre-registered, they are sufficiently statistically powered to address the core research questions, and open research practices are followed throughout. Similarly, individual courses can be leveraged to mentor students through open and reproducible practices. For instance, the Collaborative Replications and Education Project (CREP; Grahe et al. 2018) offers students with training opportunities to conduct large-scale collaborative replication research projects, and training and support for teachers wishing to guide their students through these projects. Thus, principled education can leverage the aforementioned ideals to teach and train students (open to all students, not only those pursuing a career in academia) the procedures of science and its intricacies along with hands-on-approach to relevant concepts (such as reproducibility, replication, pre-registration, analytical flexibility, etc.), involving tasks addressing the broad range of skills that they will need as professionals, such as understand and communicate data insights (collect it or perform experiments), precise and complete record keeping, accurate reporting of results, and a commitment to oral and written presentations and publication.

The framework assessment: thematic clusters, sub-clusters, breadth and depth

We set out to design a pedagogic assessment framework enabling a nuanced perspective on the degree of open and reproducible research teaching and mentoring. Ideally, such a framework should be – at the same time – comprehensive but uncomplicated, informative and expedient. For this, FORRT conducted a literature review on open and reproducible practices, which is presented in *Table 2*.

Pedagogical Assessment. In the first instance, we aim that teachers use the FORRT assessment to benchmark current teaching and mentoring practices. We have developed a survey to collect this information. There are no pre-requisites to take the survey. The survey was designed to be self-explanatory, but both during the survey and in FORRT's website (https://forrt.netlify.com) there are instructional materials and videos (via platforms such as YouTube and Vimeo) with which scholars can further inform themselves. The *working* version can be found in the supplemental materials (see appendix I). The assessment process is straightforward. The survey starts with a debriefing along with a consent form, after which they are redirected to a page explaining FORRT's *what* and the *why*. Here, respondents are provided with links to FORRT's framework (website, manuscript, resources, social media). What follows is a simple questionnaire on personal information such professional affiliation and teaching and mentoring experience. Respondents are then given the opportunity to choose to evaluate their teaching, mentoring or both. The questions format and content are identical – except that, for teaching, the questions allow for more granular assessment. This delineation is included so to facilitate comparisons between these two essential educational aspects.

Table 2. FORRT's associated literature, descriptions, organized by Cluster and Sub-clusters.

Cluster	Description		Sub-cluster	Papers and Initiatives
I. Reproducibility Crisis and Credibility Revolution	*Attainment of a grounding in the motivations and theoretical underpinnings of reproducible and open research. Integration with field specific content (i.e., or grounded in the history of replicability)*	*i.*	Reproducibility crisis and credibility revolution.	*Munafo et al. (2017); Vazire, S. (2018); Chambers (2017); Crüwell et al. (2019); Merton (1968; 1988); Baker (2016); Edwards & Roy (2017).*
		ii.	Exploratory and confirmatory analyses.	*Wagenmakers et al. (2012); Lin & Green (2016); Chambers (2017); Wagenmakers, Dutilh, & Sarafoglou (2018).*
		iii.	Questionable research practices (its 'theory'), and prevalence.	*Simmons, Nelson, & Simonsohn (2011); John, Loewenstein, & Prelec (2012); Wicherts et al. (2016); Gelman & Loken (2013); Smaldino & McElreath (2016).*
		iv.	Proposed improvement science initiatives on statistics, measurement, teaching, data sharing, code sharing, pre-registration, replication.	*Klein et al. (2018); Crüwell et al. (2019); Peng (2015); Munafo et al. (2017).*
		v.	Ongoing debates, (e.g. incentives for and against open science).	*Pashler & Harris (2012); Fanelli (2018); Bahlai et al. (2019); Chen et al. (2019); Drummond (2018); Fell (2019).*
		vi.	Ethical considerations for improved practices.	*Chopik et al. (2016); Bol et al. (2017); Edwards & Roy (2017); Jones (2007); Fell (2019).*
II. Conceptual and Statistical Knowledge	*Enacting this principle indicates that students attain a grounding in fundamental statistics, measurement, and its implications.*	*i.*	The logic of null hypothesis testing, p-values, Type I and II errors (and when and why they might happen).	*Lakens' MOOC; Gelman & Carlin (2014); Banerjee et al. (2007).*
		ii.	Limitations and benefits of NHST, Bayesian and Likelihood approaches.	*Cumming (2014); Greenland et al. (2016); Nuzzo (2014); Etz et al. (2018); Wagenmakers et al. (2018).*
		iii.	Effect sizes, Statistical power, Confidence Intervals.	*Button, et al. (2013).; Perugini et al. (2014); Greenland et al. (2016); Pek & Flora (2017); Brysbaert et al. (2018); Lakens (2013)*
		iv.	Research Design, Sample Methods, and its implications for inferences.	*Perugini et al. (2014); Wicherts et al. (2016); Gervais et al (2015).*
		v.	Questionable research (QRPs) & measurement practices (QMPs).	*Flake & Fried (2018; 2019); Hussey & Hughes (2019); Flake, Pek, & Hehmen (2017); Rodebaugh et al. (2016).*
		vi.	Understand the relationship between all of the above.	

Table 2. FORRT's associated literature, descriptions, organized by Cluster and Sub-clusters.

Cluster	Description		Sub-cluster	Papers and Initiatives
III. Reproducible analyses	*Reproducible analyses allow the checking of analytic pipelines and facilitate error correction. Enacting this principle requires students to move towards transparent and scripted analysis practices*	*i.*	Strengths of reproducible pipelines.	*ProjectTIER; Gandrud (2016); Software Carpentry.*
		ii.	Scripted analyses compared with GUI.	*Gandrud (2016);*
		iii.	Data wrangling	*https://gupsych.github.io/data_skills/; (Fox) https://www.youtube.com/playlist?list=PLmvNihjFsoM5hpQdqoI7onL4oXDSQ0ym8*
		iv.	Programming reproducible data analyses.	*Wilson et al. (2017); Gandrud (2016); Software Carpentry; https://learningstatisticswithr.com/book/; https://sites.trinity.edu/osl/; (Fox https://www.youtube.com/playlist?list=PLmvNihjFsoM5hpQdqoI7onL4oXDSQ0ym8)*
		v.	Open source and free software.	*Chao (2009);*
		vi.	Tools to check yourself and others; statcheck, GRIM, and SPRITE	*Nuijten et al. (2017); Brow & Heathers (2016); van der Zee, Anaya & Brown (2017);*
IV. Open data and materials	*Enacting this principle indicates that students have attained a grounding in open data and materials in both; using and sharing*	*i.*	Knowledge of traditional publication models. Open access publishing, preprints	*Tennant (2016); Hardwicke et al. (2018); Klein et al. (2018); Siler et al. (2018); Rouder (2015)*
		ii.	Reasons to share; for science, and for one's own practices	*Klein et al. (2018); Tennant (2016); Piwowar et al. (2013); Rouder (2015); Levenstein & Lyle (2018); Stodden (2011)*
		iii.	Repositories; e.g. OSF, FigShare, GitHub	*Rouder (2015); Soderberg (2018); Gilmore et al. (2018); osf.io; figshare.com; github.com*
		iv.	Accessing/sharing others data, code, and materials	*Klein et al. (2018); Piwowar et al. (2013); Joel et al. (2018); Wicherts et al. (2006)*
		v.	Ethical considerations	*Siler et al. (2018); Walsh et al. (2018); Ross, Iguchi, & Panicker (2018); Hand (2018)*
		vi.	Examples and consequences of accessing un/open data	*Rouder (2015); Walsh et al. (2018); Houtkoop et al. (2018); Wicherts et al. (2006); Peng (2015)*

Table 2. FORRT's associated literature, descriptions, organized by Cluster and Sub-clusters.

Cluster	Description		Sub-cluster	Papers and Initiatives
V. Preregistration	*Preregistration entails laying out a complete methodology and analysis before a study has been undertaken. This facilitates transparency and removes several potential QRPs.*	*i.*	Purpose of preregistration - distinguishing exploratory and confirmatory analyses, transparency measures	*Nosek et al. (2018); Wicherts et al. (2016); Dal-Re et al. (2014); Lin & Green (2016)*
		ii.	Preregistration and registered reports - strengths and differences	*Chambers et al. (2013); Chambers, Feredoes, Muthukumaraswamy, & Etchells (2014);*
		iii.	When can you preregister? Can you pre-register secondary data?	*Chambers et al. (2015); Mertens & Krypotos (2019); Haven & Van Grootel (2019); Kirtley et al. (2019)*
		iv.	Writing a preregistration.	*https://cos.io/prereg/; https://cos.io/blog/10-preregistration-tips/;*
		v.	Comparing a preregistration to a final study manuscript.	*Wiley's Reviewing Registered Reports*
		vi.	Conducting a preregistered study.	
VI. Replication research	*Replication research takes a variety of forms, each with a different purpose and contribution. Reproducible science requires replication research.*	*i.*	Purposes of replication attempts - what is a 'failed' replication?	*Zwaan, Etz, Lucas, & Donnellan (2018); Fidler & Wilcox (2018); Frank & Saxe (2012); Van Bavel et al. (2016); García (2016);*
		ii.	Large scale replication attempts	*Open Science Collaboration. (2015); Klein et al. (2018); CREP refs count here too?; Van Bavel et al. (2016); https://manyprimates.github.io/;*
		iii.	Distinguishing conceptual and direct replications	*Van Bavel et al. (2016); Kunert (2016); Simons (2014); Zwaan, Etz, Lucas, & Donnellan (2018);*
		iv.	Conducting replication studies; challenges, limitations, and comparisons with the original study	*Grahe et al. (2012): Frank & Saxe (2012); Wagge et al. (); Lenne & Mann (2016); Wagge et al. (2019); Stanley & Spence (2014);*
		v.	Registered Replication Reports	*Simons, Holcombe, & Spellman (2014); Alogna et al. (2014); Eerland et al. (2014); Psychological Science ongoing Replications.*
		vi.	The politics of replicating famous studies	*Neuliep & Crandall (1990; 1993); Example on "conservative brain" and scathing personal attack after failed replication.*

Breadth and Depth. In addition to distinguishing thematic categories, FORRT created an assessment that (a) can gauge the least and most levels of adherence to principled education – and thereby able to incorporate newcomers and experts; (b) is applicable to both teaching and mentoring; (c) delineates a steadfast route frontward, per knowledge area; and (d) differentiates between information versus application. FORRT operationalized these competing requirements by combining the extent to which students are exposed to thematic content (i.e., not yet enacted, opportunities for some, course requirement) and the level to which students are required to engage with it (i.e., not yet enacted, knowledge, practice, application), within each thematic area. FORRT termed these as *breadth* and *depth*, respectively, as shown in *Tables 2* and *3*.

Assessing one's teaching materials (along with our devoted resources) can become a learning experience in itself, particularly for those whose knowledge about open and reproducible research practices are not yet solidified. To encourage this learning, the survey is designed in a way that allows respondents to save and continue at a later point, and to go back and forth inside the survey. These features are also aimed at galvanizing sharing of educational materials (e.g., syllabi, slides, lab manual, etc.) which may require some time to prepare. While non-mandatory, during the assessment, teachers and mentors will be able to upload supporting materials or provide links containing these materials as a mean to (a) foster open education; and (b) support the self-assessment, which will be verified. With this, FORRT hopes to fortify the normalization and institutionalization of open and reproducible scientific tenets, and foster open educational practices. On the last stage of the survey, there are two small modules about demographics and open and reproducible practices applied to respondents' research pipelines.

Table 3. Description of breadth and depth of teaching and mentoring.

Breadth describes how widely teaching is distributed and takes three levels:	
	Not yet enacted, minimal breadth, or no evidence: Although the principle may be discussed, it does not form part of a course / module component.
	Opportunities for some: The principle is taught, for example, as part of an optional course or elective or workshop.
	Course requirement for all: The principle is taught to all students, for example as part of the core course, compulsory module, or practical workshop.
Depth describes the degree to which students interact with the core principles and takes four levels:	
	Not yet enacted, minimal depth, or no evidence: the curricula does not formally include teaching on the principle in question. Although instructors might mention the principle, it is not covered in depth and/or does not feature as part of the syllabi.
	Knowledge: Students are taught about the principle and are required to demonstrate sufficient understanding.
	Practice: In addition to acquiring a knowledge base in the principle, students also put this understanding into practice in practical exercises, e.g. as part of a workshop or technical exercises.
	Application: In addition to acquiring a knowledge base in the principle and practical skills; students are required to apply this understanding and skill set in a research project.

FORRT personnel will then proceed to authenticate logged details and verify its accuracy in relation to shared educational materials, after which respondents will then be awarded a teaching or a mentoring badge. FORRT will provide in its website a list containing those abiding by principled and open educational practices. Ideally, FORRT adopters would be able to revise their statements as course content changes, with the lofty ideal of moving towards required application of open and reproducible science practices.

Community feedback and institutional grassroots advocacy. As the details of self-report teaching and mentoring practices are logged and verified, FORRT will have at its disposal a database that combines mentoring and teaching evaluations with its possible causes and correlates. Not only it provides the community with a birds' eye view into current practices, but it may also help elucidate the factors behind the emergence, development, and sustainability of scientific integrity, principled teaching and open education. For example, in coordination with institutions like the UK's Reproducibility Network, the German Open-Science Network (NOSI), and the Dutch *NWO* – all of which share a common mission: investigating factors that contribute to research integrity, robust research practices, promoting training activities, and disseminating best practices – FORRT's assessment may provide these stakeholders data with which to inform and substantiate policy proposals and advocacy. Indeed, FORRT is in a prime position to mediate the demands between empirical and normative components on a range of emerging and contemporary (politico-)scientific debates. FORRT hopes to work with national and regional 'open science and education' institutions – and agencies, organizations, groups and individuals, both in wider society and the local community – to ensure that FORRT's potential is useful and impactful beyond its main educational scope, and to all those seeking to engage the scientific community.

Taking stock. The presented assessment protocol – i.e., clusters, sub-clusters, width and breadth – translates educational ideals into a tangible and coherent framework. This operationalization aims at enabling FORRT to fit under its umbrella a wide array of courses, mentoring styles, and their respective pedagogies. This asssessment allows for a granular quantification and qualification of open and principled education into FORRT ideals. For example, one course might include a few lecture slides on the importance of replication research whereas another might require that students complete a replication study. Both are important contributions but by observing these qualitative differences FORRT communicates an ideal to strive towards. We acknowledge that it is possible for all courses to score the highest levels of adherence to the FORRT pedagogical assessment (i.e., application). To a certain extent, that is precisely the point: to guide and incentivize scholars towards better teaching, to the extent it is within their control, and afterwards use FORRT's resources to advocate for institutional change allowing for the full realization of principled education. As with Science, cumulative development in teaching is indispensable.

The upshot of FORRT's assessment framework is the production of high-quality data on open and principled education, which can be used to evaluate current practices and substantiate evidence-based policies and advocacy of scientific organizations/agencies.

Building the fourth pillar of scientific utopia: FORRT advocacy

We envisage a scientific utopia in which openness and reproducibility is considered the normative practice. In order to realise this true scientific utopia we need a fourth pillar - principled teaching and mentoring – to ensure these principles are taught to all prospective researchers. To achieve this, academia must do more to value, support, and incentivise teaching. Imagine a scientific culture in which open teaching materials are recognised for their cumulative and community value. Imagine a scientific culture in which academics share, applaud, and cite these resources as we would an empirical research article – a culture in which teaching contributions are as valued for hiring and promotion as research papers. This is the pinnacle of FORRT's vision for a scientific utopia.

Cultural shifts in academic practices require both top-down and grassroots initiatives. We can see both when taking stock from the credibility revolution. From the top-down, we see research-funding councils requiring data sharing and journals offering the registered report format. From the ground-up, we see individual researchers engaging in open and reproducible research, and many early career led initiatives leading the proliferation of reproducibility knowledge. All of these efforts have been strengthened – and in some cases made possible by –

advocacy for rigorous and transparent research practices. FORRT aims to provide a platform for this advocacy – towards building a scientific utopia supported by a fourth pillar of principled teaching and mentoring.

How FORRT will advocate

For teachers seeking support. We do not expect teachers to shoulder additional burdens, nor do we advocate all teaching materials be adapted to adhere to FORRT principles. The educational NEXUS will be a platform for teachers to Find, Access, Interoperate, and Reuse (FAIR; Wilkinson et al., 2016) teaching materials. FORRT will act as a pedagogical resource, including information about, for example, context and prior knowledge, class assignments and assessments, class delivery, and student expectations and outcomes. Our intention is that resources are maximally useful to teachers wishing to use and adapt them. FORRT will advocate for the sharing of these valuable resources that can reduce some of the burden on teachers looking to incorporate open and reproducible research principles into their teaching.

For teachers sharing materials. FORRT's educational NEXUS aims to adopt a format not unlike data or software articles – in which data or code are attached to a paper that describes them in detail. The key implication FORRT will advocate for is that these resources are not only used, but cited and recognised. Future teachers should be able to build on the shoulders of giants, alleviating burden and improving the quality of teaching. FORRT will support and recognise these giants; in the current academic landscape, citing teaching resources would help begin the transition to the scientific teaching utopia we have described.

For diversity and equality in academic contributions. FORRTs advocacy is not only for the merging of subject teaching with teaching about the process of science (including transparency, reproducibility, and rigour); our advocacy efforts will also aim to elevate the standing of principled teaching in academia and nurture a culture of sharing, developing, and valuing open teaching materials. FORRT will advocate that these teaching contributions – particularly open teaching materials, such as the excellent pedagogies included in the NEXUS – be given a similar standing to research contributions. This inequality does no justice to the teaching of open science. FORRT will advocate for bridging this divide.

To achieve this goal, we will collect data from FORRT assessments as described previously. The aim is to curate a large representative dataset capturing the current state of open and reproducible teaching practices. We hope that this data will provide an indication of what best practices in principled teaching are out there, and what areas are best placed for improvements and curation of further teaching materials. Our aim is that FORRT, alongside national and international open science organisations, will use this data to advocate for further integration of openness and reproducibility into teaching courses – and to raise the profile of principled teaching and mentoring across academia.

Final thoughts

FORRT was inspired, in part, from existing realizations of the credibility revolution, such as the Transparency and Openness Promotion Guidelines (TOP; Nosek et al. 2015) so that journals are able to evaluate and promote the eight TOP standards in research practice (including: preregistration, data transparency, design and analysis transparency, and replication). TOP's continued aim is to align these critical features of a mature scientific discipline with practices required by journals and funding bodies through incremental levels of implementation of each standard: disclose, require, and verify. This enables journals to select which level of each standard is most appropriate to their needs before implementation. The bar for entry is relatively low, and the levels offer guidance on how the journal may become more transparent and open in the future.

At the time of writing, over 5,000 journals and organizations have become TOP guidelines signatories, of which over 850 have implemented TOP (see https://cos.io/our-services/top-guidelines/ for a list of current signatories). We hope to achieve a similar goal with FORRT in providing a tool to incentivize and evaluate the teaching and mentoring open and reproducible science alongside the necessary resources. While TOP targets journals and funders, other efforts, such as the San Francisco Declaration on Research Assessment (DORA; https://sfdora.org/) target institutions and individual researchers. DORA advocates that academic institutions support improvements in research practices in research assessment. DORA proposes that institutions consider the scientific quality of all research outputs (e.g. datasets and software, in addition to journal articles) over metric-based assessments, such as the number of publications, impact factors, or H-indices.

Among other initiatives, TOP and DORA are important steps to incentivize and support researchers to engage in open and reproducible research practices. These guidelines highlight that improvements in research practices can be evaluated and incentivized. The onus is shared between journals, funders, and academic institutions to promote improvements in research practices. It is then the responsibility of individual researchers to remain up-to-date with these advancements and to meet these improved standards of open, transparent, and reproducible research. We hope that FORRT will meet the pressing need to evaluate and incentivize the teaching and mentoring of these practices that has not yet been the focal point of other initiatives.

When achieved, principled teaching and mentoring benefit students, teachers, science and society. For students, it exposes them to a more honest educational process, especially with respect to epistemic uncertainty and its consequences to the facts of science; it breeds an atmosphere of constructive curiosity and problem-solving; it bequeaths first-hand practical experience in probing the production of knowledge which in turn trains students to be good consumers of science; and it helps shatter the academic glass ceiling and promote engagement with the academic profession.

For teachers, FORRT offers a didactic entry-point for both newer and older generations to better research practices; it foments researchers' continued development and gradual implementation of open and reproducible teaching practices; it hopes to connect scholars, their general and field-specific insights to facilitate the transition into principled teaching and mentoring and facilitate the discovery of content, tools and implemented pedagogies; it cuts time and streamlines the education pipeline by simplifying access to (a) a curated and vetted list of open and reproducible science resources and (b) providing a platform for the centralization of educational, research and teaching materials and its implemented pedagogies.

For science, FORRT has the potential to positively influence the reproducibility and overall quality of researchers' works via a feedback loop between class content/discussions and practice; it galvanizes and accelerates the dissemination of reliable, rigorous, replicable, open, reproducible science to a previously uninformed population: undergraduate and graduate students.

For society, FORRT creates the conditions for knowledge production and its process to become a public good, accessible to all its members, either through their formal tertiary education, or through internet connectivity; it is one more safeguard against the harmful effects of unsubstantiated beliefs by non-experts in today's public discourse impacting public opinion and policy on heavily researched and consensus issues like vaccination and climate change; it helps mitigate structural and geographical inequalities in the access of cutting-edge scientific, teaching and mentoring materials, which ultimately fosters the democratization of science and bolster students' (and interested parties') opportunities that would otherwise be inaccessible to disadvantaged individuals.

With FORRT, scholars now have an unprecedented opportunity to shape the minds and future of the consumers of science and the next generation of academics. Numerous scholars consider contributing to the next generations' edification and professional development a significant part of their responsibilities. Scholars would also agree that if the time they spend on teaching and mentoring meaningfully influenced a student's life – i.e., broadened their horizons, imparted expertise into the acquisition of knowledge independently, fostered the conditions to discover connections of seemingly unrelated phenomena, helped develop the self-confidence to challenge outdated dogma – all would be worth it. In this sense, At its best, FORRT should be the answer to the question, ‘‘What are the best conceivable educational practices in higher education and how to achieve it?”

Open Practices Statement

The materials and data for these studies are available online, and can be found at https://osf.io/g29eu/

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

Allen, C., & Mehler, D. M. (2019). Open science challenges, benefits and tips in early career and beyond. PLoS Biology, 17(5), e3000246.

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., ... & Perugini, M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*(2), 108-119.

Bahlai et al. (2019). Open science isn't always open to all scientists. *American Scientist, 107*(2), 78. DOI: 10.1511/2019.107.2.78

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature News, 533(7604), 452.

Baker, M. (2016). Is there a reproducibility crisis? Nature, 533(7604), 3–5.

Banerjee, A., Chitnis, UB., Jadhav, SL., Bhawalkar, JS., Chaudhury, S. (2009). Hypothesis testing, type I and type II errors. Industrial Psychiatry Journal, 18(2), 127-131. https://doi.org/10.1111/j.1740-9713.2015.00827.x

Bol, T., de Vaan, M., & van de Rijt, A. (2018). The Matthew effect in science funding. Proceedings of the National Academy of Sciences, 115(19), 4887-4890.

Brown, N. J., & Heathers, J. A. (2016). The GRIM test: A simple technique detects numerous anomalies in the reporting of results in psychology. Social Psychological and Personality Science, 1948550616673876. http://journals.sagepub.com/doi/pdf/10.1177/1948550616673876

Brysbaert, M. and Stevens, M. (2018) Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. Journal of Cognition, 1(1): 9, pp. 1–20, DOI: https://doi.org/10.5334/joc.10

Button. K. S. (2018). Reboot undergraduate courses for reproducibility. *Nature, 561*, 287. doi: 10.1038/d41586-018-06692-8

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.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., et al. (2016). Evaluating replicability of laboratory experiments in economics. Science, 351, 1433–1436. DOI: 10.1126/science.aaf0918

Camerer CF, Dreber A, Holzmeister F, Ho T-H, Huber J, Johannesson M, et al. (2019). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nature Human Behaviour, 2(9), 637–44. https://doi.org/10.1038/s41562-018-0399-z

Center for Open Science (2019). Can machines determine the credibility of research claims? The Center for Open Science joins a new DARPA program to find out. *Center for Open Science.* https://cos.io/about/news/can-machines-determine-credibility-research-claims-center-open-science-joins-new-darpa-program-find-out/

Chambers, C. (2017). The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice. Princeton University Press. http://dx.doi.org/10.1515/9781400884940

Chambers, C. D. (2013). Registered reports: a new publishing initiative at Cortex. Cortex, 49(3), 609-610.

Chambers, C. D., Feredoes, E., Muthukumaraswamy, S. D., & Etchells, P. (2014). Instead of “playing the game” it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. AIMS Neuroscience, 1(1), 4–17.

Chambers, C.D., Dienes, Z., McIntosh, R.D., Rotshtein, P., & Willmes, K. (2015). Registered Reports: Realigning incentives in scientific publishing. Cortex, 66, A1-2. DOI: 10.1016/j.cortex.2015.03.022

Chao, Lee (2009) Utilizing open source tools for online teaching and learning Information Science (Hershey PA & London).

Chen et al. (2019). Open is not enough. Nature: physics, 15, 113-119. https://doi.org/10.1038/s41567-018-0342-2

Chopik, W. J., Bremner, R. H., Defever, A. M., & Keller, V. N. (2018). How (and Whether) to Teach Undergraduates About the Replication Crisis in Psychological Science. Teaching of Psychology, 45(2), 158–163. https://doi.org/10.1177/0098628318762900

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J., … Schulte-Mecklenbeck, M. (2018, November 16). 8 Easy Steps to Open Science: An Annotated Reading List. https://doi.org/10.31234/osf.io/cfzyx

Cumming, G. (2014). The new statistics: Why and how. Psychological Science, 25(1), 7-29. https://doi.org/10.1177/0956797613504966

Current Population Survey, 2014. U.S. Census Bureau, Annual Social and Economic Supplement. Retrieved from: https://www.census.gov/data/tables/2014/demo/educational-attainment/cps-detailed-tables.html

Dal-Ré, R., Ioannidis, J. P., Bracken, M. B., Buffler, P. A., Chan, A.-W., Franco, E. L., La Vecchia, C., Weiderpass, E. (2014). Making prospective registration of observational research a reality. Sci. Transl. Med. 6(224). DOI: 10.1126/scitranslmed.3007513

Drummond, C. (2018); Reproducible research: a minority opinion. Journal of Experimental & Theoretical Artificial Intelligence, 30(1), 1-11. https://doi.org/10.1080/0952813X.2017.1413140

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., … Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. Journal of Experimental Social Psychology, 67, 68–82. https://doi.org/10.1016/j.jesp.2015.10.012

Edwards, M. A., & Roy, S. (2016). Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. Environmental Engineering Science. DOI: 10.1089/ees.2016.0223

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

EUROSTAT. (2016). Tertiary education statistics (ISSN 2443-8219). Retrieved June 24, 2019, from https://ec.europa.eu/eurostat/statistics-explained/index.php/Tertiary_education_statistics

Everett, J. A. C., & Earp, B. D. (2015). A tragedy of the (academic) commons: interpreting the replication crisis in psychology as a social dilemma for early-career researchers. Frontiers in psychology, 6, 1152.

Etz, A., Gronau, Q.F., Dablander, F. et al. Psychon Bull Rev (2018) 25: 219. https://doi.org/10.3758/s13423-017-1317-5

Fanelli, D. (2018). Opinion: Is science really facing a reproducibility crisis, and do we need it to? PNAS, 115(11), 2628-2631. https://doi.org/10.1073/pnas.1708272114

Fell, M.J. The Economic Impacts of Open Science: A Rapid Evidence Assessment. Preprints 2019, 2019050302 (doi: 10.20944/preprints201905.0302.v1).

Fidler, F, Wilcox, J. Reproducibility of scientific results. In: Zalta, EN , ed. The Stanford Encyclopedia of Philosophy. Winter 2018 edition. Stanford, CA: Metaphysics Research Lab. https://plato.stanford.edu/archives/win2018/entries/scientific-reproducibility. Published December, 3, 2018.

Flake, J. K., & Fried, E. I. (2019, January 17). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. https://doi.org/10.31234/osf.io/hs7wm

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct Validation in Social and Personality Research: Current Practice and Recommendations. Social Psychological and Personality Science, 8(4), 370–378

Flier, J. (2017). Faculty promotion must assess reproducibility. Nature News, 549(7671), 133.

Fox, Nick. (2018). Writing Reproducible Scientific Papers in R. https://www.youtube.com/playlist?list=PLmvNihjFsoM5hpQdqoI7onL4oXDSQ0ym8

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

Freese, J., & Peterson, D. (2018). The emergence of statistical objectivity: changing ideas of epistemic vice and virtue in science. Sociological theory, 36(3), 289-313.

Fried, E. I., & Flake, J. K. (2018, March). Measurement matters. Observer. Retrieved from https://www.psychologicalscience.org/observer/measurement-matters

Gandrud, C. (2016) Reproducible research with R and R studio. New York; CRC Press

García, FM. (2016). Replication and the Manufacture of Scientific Inferences: A Formal Approach. International Studies Perspectives, 17(4), 408–425. https://doi.org/10.1093/isp/ekv011

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing Type S (sign) and Type M (magnitude) errors. Perspectives on Psychological Science, 9(6), 641-651.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. Unpublished manuscript. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

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.

Gilmore, R. O., Kennedy, J. L., & Adolph, K. E. (2018). Practical Solutions for Sharing Data and Materials From Psychological Research. Advances in Methods and Practices in Psychological Science, 1(1), 121–130. https://doi.org/10.1177/2515245917746500

GOSH (2019). Gathering for Open Science Hardware. Retrieved 21 March 2019, from http://openhardware.science/gosh-manifesto/

Grahe, J. E., Reifman, A., Hermann, A. D., Walker, M., Oleson, K. C., Nario-Redmond, M., & Wiebe, R. P. (2012). Harnessing the Undiscovered Resource of Student Research Projects. Perspectives on Psychological Science, 7(6), 605–607. https://doi.org/10.1177/1745691612459057

Grahe, J. E., Brandt, M. J., Wagge, J. R., Legate, N., Wiggins, B. J., Christopherson, C. D., . . . LePine, S. (2018). Collaborative Replications and Education Project (CREP). Retrieved from https://osf.io/wfc6u/

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. European Journal of Epidemiology, 31(4), 337–50. http://doi.org/10.1007/s10654-016-0149-3

Hampson, G. (2019). Issue Brief 4: Plan S & the Quest for Global Open. Open Scholarship Initiative Proceedings, 2019-. https://doi.org/10.13021/osi2019.2400

Hand DJ. Aspects of Data Ethics in a Changing World: Where Are We Now?. Big Data. 2018;6(3):176–190. doi:10.1089/big.2018.0083

Hardwicke TE. et al. (2018) Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal Cognition. R. Soc. open sci. 5: 180448. http://dx.doi.org/10.1098/rsos.180448

Harris, J. K., Johnson, K. J., Carothers, B. J., Combs, T. B., Luke, D. A., & Wang, X. (2018). Use of reproducible research practices in public health: A survey of public health analysts. PloS one, 13(9), e0202447.

Haven, Tamarinde., L. & Van Grootel, Leonie. (2019). Preregistering qualitative research. *Accountability in Research, 26*(3), 229-244, DOI: 10.1080/08989621.2019.1580147

Hawkins, R. X., Smith, E. N., Au, C., Arias, J. M., Catapano, R., Hermann, E., ... & Salehi, S. (2018). Improving the replicability of psychological science through pedagogy. Advances in Methods and Practices in Psychological Science, 1(1), 7-18.

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V. M., Nichols, T. E., & Wagenmakers, E.-J. (2018). Data Sharing in Psychology: A Survey on Barriers and Preconditions. Advances in Methods and Practices in Psychological Science, 1(1), 70–85. https://doi.org/10.1177/2515245917751886

Hussey, I., & Hughes, S. (2018, November 19). Hidden invalidity among fifteen commonly used measures in social and personality psychology. https://doi.org/10.31234/osf.io/7rbfp

Iqbal, S. A., Wallach, J. D., Khoury, M. J., Schully, S. D., & Ioannidis, J. P. (2016). Reproducible research practices and transparency across the biomedical literature. PLoS biology, 14(1), e1002333.

Janz, N. (2016). Bringing the Gold Standard into the Classroom: Replication in University Teaching. International Studies Perspectives, 17(4), 392–407. https://doi.org/10.1111/insp.12104

Joel, S., Eastwick, P. W., & Finkel, E. J. (2018). Open Sharing of Data on Close Relationships and Other Sensitive Social Psychological Topics: Challenges, Tools, and Future Directions. Advances in Methods and Practices in Psychological Science, 1(1), 86–94. https://doi.org/10.1177/2515245917744281

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. Psychological science, 0956797611430953.

Jones, NL. (2007). A code of ethics for the life sciences. Sci Eng Ethics, 13, 25-43. DOI: 0.1007/s11948-006-0007-x

Kirtley, O. J., Lafit, G., Achterhof, R., Hiekkaranta, A. P., & Myin-Germeys, I. (2019, April 10). Making the black box transparent: A pre-registration template for studies using Experience Sampling Methods (ESM). https://doi.org/10.31234/osf.io/seyq7

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., et al. (2014). Investigating variation in replicability: A "many labs" replication project. Social Psychology, 45, 142–152. https://doi.org/10.1027/1864-9335/a000178

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., … Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. Advances in Methods and Practices in Psychological Science, 1(4), 443–490. https://doi.org/10.1177/2515245918810225

Kunert, R. (2016). Internal conceptual replications do not increase independent replication success. Psychonomic bulletin & review, 23(5), 1631-1638.

Kretser, A., Murphy, D., Bertuzzi, S., Abraham, T., Allison, D. B., Boor, K. J., ... & Jacobs-Young, C. (2019). Scientific Integrity Principles and Best Practices: Recommendations from a Scientific Integrity Consortium. Science and engineering ethics, 1-29.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Frontiers in Psychology, 4:863. 10.3389/fpsyg.2013.00863

Lakens, D. Improving your statistical inferences. Online course. https://www.coursera.org/learn/statistical-inferences

Lenne & Mann (2016). CREP project report. https://osf.io/sdj7e/

Levenstein, M. C., & Lyle, J. A. (2018). Data: Sharing Is Caring. Advances in Methods and Practices in Psychological Science, 1(1), 95–103. https://doi.org/10.1177/2515245918758319

Li, W., Aste, T., Caccioli, F., & Livan, G. (2019). Early coauthorship with top scientists predicts success in academic careers. *Nature Communications, 10*, 5170. doi:10.1038/s41467-019-13130-4

Lin, W., & Green, D. P. (2016). Standard operating procedures: A safety net for pre-analysis plans. PS: Political Science & Politics, 49(3), 495-500.

Mertens, G., & Krypotos, A. (2019, February 20). Preregistration of secondary analyses. https://doi.org/10.31234/osf.io/ph4q7

Merton R, K. (1957) Social Theory and Social Structure. Free Press.

Merton, R., K. (1968). The Matthew effect in science. Science, 159(3810), 56-63. 10.1126/science.159.3810.56

Merton, R., K. (1988). The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property. ISIS, 79, 606-623. 10.1086/354848

Moher, D., Naudet, F., Cristea, I. A., Miedema, F., Ioannidis, J. P., & Goodman, S. N. (2018). Assessing scientists for hiring, promotion, and tenure. PLoS biology, 16(3), e2004089.

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., ... & Castille, C. M. (2018). The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science, 1*(4), 501-515. https://doi.org/10.1177/2515245918797607

Morgan, A. C., Economou, D. J., Way, S. F., & Clauset, A. (2018). Prestige drives epistemic inequality in the diffusion of scientific ideas. EPJ Data Science, 7(1), 40.

Munafo, M. R., et al. (2017). A manifesto for reproducible science. Nature Human Behaviour, 1, 0021. DOI: 10.0138/s41562-016-0021.

National Research Council. (2002). Integrity in scientific research: Creating an environment that promotes responsible conduct. National Academies Press.

Naudet, F., Ioannidis, J., Miedema, F., Cristea, I. A., Goodman, S. N., & Moher, D. (2018). Six principles for assessing scientists for hiring, promotion, and tenure. Impact of Social Sciences Blog, 1-5.

Navarro, D. (2019). Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.6.1). https://learningstatisticswithr.com/book/

Neuliep, J. W., & Crandall, R. (1990). Editorial bias against replication research. Journal of Social Behavior & Personality, 5(4), 85-90.

Neuliep, J. W., & Crandall, R. (1993). Reviewer bias against replication research. Journal of Social Behavior & Personality, 8(6), 21-29.

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. NYU Press.

Nosek, B. A., Ebersole, C. R., DeHaven, A., & Mellor, D. (2017). The Preregistration Revolution. PNAS, 115(11), 2600-2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., Ofiesh, L., Grasty, F., Pfeiffer, N., Mellor, D. T., Brooks, R. E., III, … Baraniuk, R. (2019, June 7). Proposal to NSF 19-565 to Create a STEM Education Research Hub. https://doi.org/10.31222/osf.io/4mpuc

Nuijten, M. B., Van Assen, M. A. L. M., Hartgerink, C. H. J., Epskamp, S., & Wicherts, J. M. (2017). The validity of the tool “statcheck” in discovering statistical reporting inconsistencies. Preprint retrieved from https://psyarxiv.com/tcxaj/.

Nuijten, M. B. et al. (2017). Journal Data Sharing Policies and Statistical Reporting Inconsistencies in Psychology. Collabra: Psychology, 3(1): 31, pp. 1–22. DOI: https://doi.org/10.1525/collabra.102*

Nuzzo, R. (2014). Statistical errors: P values, the ‘gold standard’ of statistical validity, are not as reliable as many scientists assume. Nature, 506(7487), 150-152.

Okune, A., Hillyer, R., Albornoz, D., Posada, A., & Chan, L. (2018, June). Whose Infrastructure? Towards Inclusive and Collaborative Knowledge Infrastructures in Open Science. In ELPUB 2018.

Open Science Collaboration (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. Perspectives on Psychological Science, 7, 657–660.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716.

Open Stats Lab. https://sites.trinity.edu/osl/

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. Perspectives on Psychological Science, 7, 531-536.

Pek, J., & Flora, D. B. (2017). Reporting effect sizes in original psychological research: A discussion and tutorial. Psychological Methods. http://doi.org/10.1037/met0000126

Peng, R. (2015). The reproducibility crisis in science: A statistical counterattack. Significance, 12(3). https://doi.org/10.1111/j.1740-9713.2015.00827.x

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. Perspectives on Psychological Science, 9, 319-332.

Peters, H. P., & Dunwoody, S. (2016). Scientific uncertainty in media content

Piwowar, Heather A., and Todd J. Vision. (2013). "Data reuse and the open data citation advantage", PeerJ, 1:e175. DOI 10.7717/peerj.175.

ProjectTIER. https://www.projecttier.org/

Queneau, R. (1947). Exercices de Style. Alma Books.

Raheel, S. A., & Kujan, O. B. (2016). The opinion and response of health professionals associated with academics about the research design and methods: A study. Journal of International Society of Preventive & Community Dentistry, 6(2), 154.

Rodebaugh, T. L., Scullin, R. B., Langer, J. K., Dixon, D. J., Huppert, J. D., Bernstein, A., . . . Lenze, E. J. (2016). Unreliability as a threat to understanding psychopathology: The cautionary tale of attentional bias. Journal of Abnormal Psychology, 125(6), 840-851. http://dx.doi.org/10.1037/abn0000184

Ross, M. W., Iguchi, M. Y., & Panicker, S. (2018). Ethical aspects of data sharing and research participant protections. American Psychologist, 73(2), 138-145. http://dx.doi.org/10.1037/amp0000240

Rouder, J.N. (2015). The what, why, and how of born-open data. Behavioral Research. DOI 10.3758/s13428-015-0630-z (https://link.springer.com/article/10.3758/s13428-015-0630-z)

Siler K, Haustein S, Smith E, Larivière V, Alperin JP. Authorial and institutional stratification in open access publishing: the case of global health research. PeerJ. 2018;6:e4269. Published 2018 Feb 19. doi:10.7717/peerj.4269

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.

Simons, D. J. (2014). The Value of Direct Replication. Perspectives on Psychological Science, 9(1), 76–80. https://doi.org/10.1177/1745691613514755

Simons, Holcombe, & Spellman (2014); Alogna et al. (2014); Eerland et al. (2014); https://www.psychologicalscience.org/publications/replication/ongoing-projects;

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. Royal Society open science, 3(9), 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/10.1177/2515245918757689osf.io; figshare.com; github.com

Steltenpohl, C.N., Anderson, A.J., & Daniels, K.M. (2019). Giving community psychology away: A case for open access publishing. *Global Journal of Community Psychology Practice*, 10(3), 1-14. https://www.gjcpp.org/en/article.php?issue=33&article=199

Software Carpentry. https://software-carpentry.org/

Stanley, D. J., & Spence, J. R. (2014). Expectations for replications: Are yours realistic?. Perspectives on Psychological Science, 9(3), 305-318.

Stodden, V. C. (2011). Trust your science? Open your data and code. Amstat News, 409, 21-22.

Tannert, C., Elvers, H. D., & Jandrig, B. (2007). The ethics of uncertainty: In the light of possible dangers, research becomes a moral duty. EMBO reports, 8(10), 892-896.

Tennant JP, Waldner F, Jacques DC, Masuzzo P, Collister LB, Hartgerink CH. The academic, economic and societal impacts of Open Access: an evidence-based review. F1000Res. 2016;5:632. Published 2016 Sep 21. doi:10.12688/f1000research.8460.3

Uhlmann, E. L., Chartier, C. R., Ebersole, C. R., Errington, T. M., Kidwell, M., Lai, C. K., … Nosek, B. A. (2018, August 13). Scientific Utopia: III. Crowdsourcing Science. https://doi.org/10.31234/osf.io/vg649

U.S. Census Bureau (2011). Statistical Abstract of the United States: Education. Retrieved from https://www.census.gov/prod/99pubs/99statab/sec04.pdf

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. Proceedings of the National Academy of Sciences, 113(23), 6454-6459.

van der Zee T, Anaya J, Brown NJL. 2017. Statistical heartburn: An attempt to digest four pizza publications from the Cornell Food and Brand Lab. PeerJ Preprints 5:e2748v1 https://doi.org/10.7287/peerj.preprints.2748v1

Van Dijk, D., Manor, O., & Carey, L. B. (2014). Publication metrics and success on the academic job market. Current Biology, 24(11), R516-R517.

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

Vicens, Q., & Bourne, P. E. (2009). Ten simple rules to combine teaching and research.

Wagenmakers , E.-J., Dutilh, G., & Sarafoglou, A. (2018). The Creativity-Verification Cycle in Psychological Science: New Methods to Combat Old Idols. Perspectives on Psychological Science, 13(4), 418–427. https://doi.org/10.1177/1745691618771357

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Mass, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. Perspectives on Psychological Science, 7(6), 632–638. doi:10.1177/1745691612463078Lin & Green (2016);

Wagenmakers, EJ., Love, J., Marsman, M. et al. Psychon Bull Rev (2018) 25: 58. https://doi.org/10.3758/s13423-017-1323-7

Wagge JR, Brandt MJ, Lazarevic LB, Legate N, Christopherson C, Wiggins B and Grahe JE (2019) Publishing Research With Undergraduate Students via Replication Work: The Collaborative Replications and Education Project. Front. Psychol. 10:247. doi: 10.3389/fpsyg.2019.00247

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. https://doi.org/10.1177/2515245917749652

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. American Psychologist, 61(7), 726.

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R. C., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. Frontiers in psychology, 7.

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. Frontiers in Psychology, 7, 1832.

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

Wittgenstein Centre for Demography and Global Human Capital (2015). Wittgenstein Centre Data Explorer Version 1.2. Retrieved from: http://dataexplorer.wittgensteincentre.org/wcde-v2/

World Bank (2010). Education Statistics. Retrieved from https://datacatalog.worldbank.org/dataset/education-statistics

Zwaan, R.A., Etz, A., Lucas, R.E, Donnellan, M.B. (2018). Making replication mainstream. Behavior and Brain Sciences, 41. https://doi.org/10.1017/S0140525X17001972

FORRT Contributors

FORRT is operated as an open contributorship – we welcome meaningful contributions from all, and all of these contributions will be recognized. Our current contributions taxonomy (adapted from CRediT; https://casrai.org/credit/) includes the following roles. Our only request is that credited contributions are meaningful. Specific contributions should include a note of the contribution, e.g. "Dissemination (Teaching Conference 2019)". If you are interested in contributing to the manuscript, a google doc version can be found here: https://tinyurl.com/FORRTworkingDOC

- **FORRT Ambassadors:** Coordination of FORRT activities, oversight and leadership responsibility,
- **Conceptualisation:** Ideas; formulation and/or evolution of overarching project goals and aims
- **Writing – Original:** Preparation, creation, and writing of the initial draft of the FORRT manuscript
- **Writing – Feedback:** Providing useful feedback for the improvement of the FORRT manuscript
- **Writing – Revisions:** Contributing to the editing and revising the FORRT manuscript
- **Website development:** Contributions to developing and maintaining the FORRT website in all capacities
- **Resource curation:** Assisting with the management, organization, and annotation of any of the FORRT teaching resource collections
- **Submitted teaching resources:** Submitted own teaching resources for curation
- **Completed FORRT self-assessment:** Submitted a FORRT pedagogical self-assessment of one's own course(s)
- **Dissemination:** Presenting FORRT at meetings/conferences; dissemination internal and external to one's own department/institution; social media engagement
- **Feedback on project / manuscript:** We want to recognize all feedback and critical input across the entirety of FORRT. Even if simply to act as a signatory in support of FORRT.
- **And many more other valuable contributions:** We will add to the contributor roles as necessary and as the project develops.

Contributors: Flavio Azevedo, Sam Parsons, Carl Michael Galang, Heather Urry, Kristen Lane, Lisa DeBruine, Benjamin Le, Donnald Tellinghuisen, Madeline Harms

Full contributions:

Flavio Azevedo
Roles: Conceptualisation (SIPS hackathon 2018), Writing - Original, Writing – Revisions, Writing - Feedback, Website development, Resource curation, Dissemination (ProjectTIER workshop, SIPS 2019), FORRT Ambassador
Email: flavio.azevedo@uni-jena.de
Affiliations: University of Jena
ORCID: 0000-0001-9000-8513

Sam Parsons
Roles: Conceptualisation (SIPS hackathon 2018), Writing - Revisions, Writing - Feedback, Dissemination (ProjectTIER workshop, SIPS 2019), FORRT Ambassador
Email: sam.parsons@psy.ox.ac.uk
Affiliations: University of Oxford
ORCID: 0000-0002-7048-4093

Carl Michael Galang
Roles: Conceptualisation (SIPS hackathon 2018), Writing - Feedback, Resource curation
Email: galangc@mcmaster.ca
Affiliations: McMaster University
ORCID: 0000-0002-4088-8476

Heather Urry
Roles: Conceptualisation (SIPS hackathon 2018)
Email:
Affiliations:
ORCID:

Kristen Lane
Roles: Conceptualisation (SIPS hackathon 2018)
Email:
Affiliations:
ORCID:

Lisa DeBruine
Roles: Conceptualisation (SIPS hackathon 2018)
Email:
Affiliations:
ORCID:

Benjamin Le

Roles: Conceptualisation (SIPS hackathon 2018)
Email:
Affiliations:
ORCID:

Donald Tellinghuisen
Roles: Conceptualisation (SIPS hackathon 2018)
Email:
Affiliations:
ORCID:

Madeline Harms
Roles: Conceptualisation (SIPS hackathon 2018)
Email:
Affiliations:
ORCID:

Erin Buchanan
Roles: Writing – Feedback,
Email:
Affiliations:
ORCID:

Crystal N. Steltenpohl
Roles: Writing – Revisions, Writing - Feedback
Email: cnsteltenp@usi.edu
Affiliations: University of Southern Indiana
ORCID: 0000-0001-5049-9354