

Easing Into Open Science: A Tutorial for Graduate Students

Ummul-Kiram Kathawalla*¹

Priya Silverstein^{2,3}

Moin Syed¹

¹University of Minnesota, ²University of Surrey, ³Lancaster University

*corresponding author

Version Date: May 8, 2020

Submitted for review

Please address correspondence to:

Ummul-Kiram Kathawalla

kath0033@umn.edu

Abstract

This article provides a roadmap to assist graduate students to engage in open science practices. We suggest eight open science practices that novice graduate students could begin adopting today. The topics we cover include journal clubs, project workflow, preprints, reproducible code, data sharing, transparent writing, preregistration, and registered reports. To address concerns about not knowing how to engage in open science practices, we provide a difficulty rating of each behavior (easy, medium, difficult), present them in order of suggested adoption, and follow the format of *what*, *why*, *how*, and *worries*. We give graduate students ideas on how to approach conversation with their advisors/collaborators, ideas on how to integrate open science practices within the graduate school framework, and specific resources on how to engage with each behavior. We emphasize that engaging in open science behaviors need not be an all or nothing approach, but rather graduate students can engage with any number of the behaviors outlined.

Keywords: Open Science, Graduate Students, Tutorial, Reproducibility, Preregistration

Open science is best described as “an umbrella term used to refer to the concepts of openness, transparency, rigor, reproducibility, replicability, and accumulation of knowledge, which are considered fundamental features of science” (Crüwell et al., 2019, p. 3). The Open Science Movement developed in response to the culture of secrecy and skepticism that has been pervasive throughout scientific research. As doubt was cast upon foundational empirical work, there emerged a desire to better understand the conceptual, methodological, and analytic choices made throughout the research cycle, as doing so enhances knowledge among the scientific community and permits more informed assessments of credibility (Vazire, 2017).

Topics related to open science in psychology have received major attention in the last decade, including new terminology, new methodological and statistical procedures, new journals, and even whole new sub-fields (e.g., Meta-science, <https://metascience.com/>; see also Nelson et al., 2018; Spellman et al., 2018). This explosion, along with the fact that part of open science involves more rapid dissemination than the traditional scientific model, has resulted in a barrage of new findings, methods, and practices. All of this can be quite overwhelming to any researcher trying to get a handle on best practices, but especially for graduate students who are new to the field and are quickly trying to learn both methodological and substantive content.

The purpose of this tutorial is to provide a roadmap for how graduate students can wade through this confusion and begin to engage with open science practices. A sense of paralysis associated with not knowing where to begin with open science is a commonly expressed sentiment. Moreover, some may feel like they need to immediately adopt all open science practices if they are “truly doing open science.” We reject this “all or nothing” view and join with others who advocate for a selective approach to open science, with the accumulation of practices over time (Bergmann, 2018; Corker, 2018; Nuijten, 2019; Syed, 2019). Whereas there

are other excellent articles on how to get started with open science (e.g., Lewis, 2019; Nuijten, 2019), including Allen and Mehler's (2019) article on benefits to early career researchers, we see a great need for a tutorial that is student-focused and includes concrete suggestions. Of course, the relevance of our recommendations is not limited to graduate students—anyone who is new to open science should find them useful—but we prioritized the graduate student perspective. We hope that this tutorial can be the gateway for graduate students to easily begin using open scientific practices.

Easing into Open Science

In this tutorial we suggest eight open science practices that novice graduate students can begin adopting today. To address concerns about not knowing how to engage in open science practices, we provide a difficulty rating of each behavior (easy, medium, difficult) and present them in order of suggested adoption. We do not claim that this is the only order to follow, but rather is a sensible one given the ease of entry for each and how they build upon one another. That said, do not let any hesitance in adopting one of the behaviors stop you from attempting “later” ones. In discussing each practice, we follow the format of *what*, *why*, *how*, and *worries*. We begin by briefly applying this rhetorical form to open science in general, and then move into the specific behaviors.

What is Open Science?

Open science is a broad term that refers to a variety of principles and behaviors pertaining to transparency, credibility, reproducibility, and accessibility. There are numerous articles describing and debating open science in the literature, and it is beyond our scope to review all of the core issues. Thus, we direct interested readers to Cruwell and colleagues (2019), who provide an annotated open science reading list that covers major works across different open

science domains, as well as Yarkoni (2019), who provides a more conceptual perspective. In short, the foundational idea behind open science research practices is to be as transparent and open as possible—across all phases of the research cycle—so that readers can fully and appropriately evaluate the work.

Why Do Open Science?

Open science is *just good science* (Tennant, 2018). There are many unselfish reasons to do open science: disseminating reliable information/increasing faith in research, not being gatekeepers of knowledge, saving resources by enabling others to build on work that has already been done rather than reinventing the wheel, ensuring that readers can properly calibrate their inferences based on the quality of the work, and many many more. Allen and Mehler (2019) highlight the benefits and challenges involved with engaging in open science practices early in one's career, and specifically emphasize benefits for the future careers of graduate students, whether they remain in academia or not. Open science helps with project organization, and therefore efficiency in the long term, and facilitates collaborations that can beget new beneficial opportunities. All in all, there are many benefits—and few drawbacks—to engaging in open science practices.

How to do Open Science?

This is why you are reading this article! There is an overwhelming amount of choice when it comes to how to engage with any individual open science behavior, and this abundance of choice can often leave people too confused to try anything. In this article we focus on different behaviors researchers can engage in and less on the tools they might use to enact those behaviors. We mostly focus on using the Open Science Framework (OSF; <http://osf.io>), which is a free, open source web application that supports various aspects of the research process (see

Foster & Deardorff, 2017; Nosek, 2014). This is not to say that OSF is the “best” tool for every behavior that we discuss, but one of its selling points is that it is a central location where researchers can carry out many open science behaviors, keeping everything together in one place and minimizing the need to learn multiple interfaces. But again, there are many different tools that can be used to engage with open science behaviors. We created an OSF project page (<https://osf.io/w5mbp/wiki/home/>) as a companion to this tutorial, which contains links to video tutorials, resources, and step-by-step guides for each of the behaviors listed.

Worries about doing Open Science

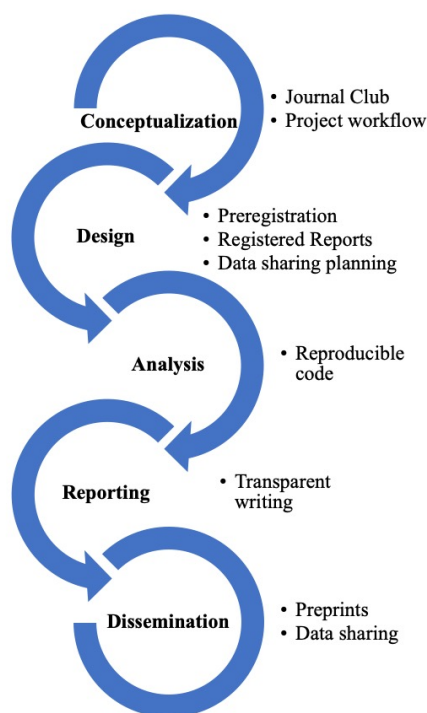
Open science seeks to upend the status quo, shifting norms about acceptable scientific practices and behaviors. For more senior researchers in the field, this could feel like open science advocates are arguing that everything they had learned and practiced was wrong, and thus may not be particularly favorable towards open science. This could create a challenging dynamic for graduate students who work with advisors or in departments who are not receptive to the open science movement (Koyama & Page-Gould, 2020; but see Christensen et al., 2020). Students may wonder, “*how do I sell this to my advisor?*” For many this is a very real concern, and thus in each section we include some common worries and how to address them with your advisor. However, we understand that talking to your advisor might not be possible for various reasons; in that case, we provide other ideas for creating your own community at your home institution (see Journal Club section). However, if that is not possible or seems out of reach, there are other open science organizations and people that you can connect with via the Society of the Improvement of Psychological Science (SIPS) or joining Academic Twitter to find other friendly community members. Additionally, by the end of this tutorial we hope you can identify some open science behaviors that you can practice in your own workflow (e.g., creating reproducible

code), even if your advisor is not supportive of engaging in other practices (e.g., preregistration). Other students may be worried, “*Won’t it make it harder to publish my research?*” It is absolutely the case that some practices associated with open science, such as larger sample sizes and more robust modeling, could result in some delays in publishing. However, some open science practices (e.g., posting preprints, Registered Reports) can speed up the process and make it easier to publish. Finally, a common concern is, “*What if I get it wrong?*” We hope this tutorial will help to ease that anxiety; there is no right way to do open science, and engaging in one open science practice is better than none—we are all welcome to feast as we choose from the “buffet of open science” (Bergmann, 2018).

Eight Open Science Practices Graduate Students Can Begin Right Now

In the remainder of this paper we review eight open science practices that have low barriers to entry but can have a sizable impact on research practice. These are by no means the only eight practices students new to open science should consider, and a different team might very well select a different set. We chose this set because they span different aspects of the research cycle (conceptualization, design, analysis, reporting, dissemination; Figure 1) and involve a range of difficulty levels for newcomers.

Figure 1. Open Science research practices across the research cycle



Open Science Journal Club (Level: Easy)

What? Organize a journal club with other students and staff to discuss issues surrounding reproducibility and open science. Usually, these take the format where one person leads the discussion each session after everyone has read the selected paper. They can range in how formal they are, from a presentation with slides followed by discussion to a completely open-ended discussion (with or without a moderator). It may even be possible that one already exists in your department that you could join!

Why? Before you can engage in open science practices you need to understand the lay of the land by becoming familiar with major works and issues. Journal clubs are a great way to do this, wherein you can learn about a new topic and critically engage with your colleagues. It can get lonely reading and working alone, so meeting in this way can create an environment in which you can socialize while also learning with others and building a community around open science. In addition to discussing the papers themselves, each article can serve as a conversation starter

about open science and reproducibility more generally. Seeing who attends the journal club can help locate others who are interested in or knowledgeable about open science, effectively establishing a network for collaborators or support. They can help you think about how you can approach conversations with your advisor or other faculty members in your department who may not be as interested in engaging with open science practices.

How? Contact colleagues in your department to inquire if anyone would be interested in participating. You only need one other person to get started! Once set up, the journal club can expand beyond your department to the wider university, facilitating interdisciplinary exchange. In addition to club attendees presenting, you can also invite external speakers to present either in person or remotely. It is also possible to organize these clubs completely remotely if you are unable to meet in person. One example of a very successful framework for an open science journal club is the ReproducibiliTea initiative: (<https://reproducibilitea.org/>), which has spread to over 50 institutions all over the world and provides a great starter pack including a list of papers to potentially discuss (<https://osf.io/3qrj6/>). For additional resources visit the accompanied OSF project page (<https://osf.io/w5mbp/wiki/home/>).

Worries. Students have different relationships with their advisors with regards to how much permission they would need to set up a journal club. However, in most cases journal clubs can absolutely be student-generated and student-run. It may be worth telling your advisor and inviting them to attend, yet, making it clear they have no obligation to attend. Also, you may feel that learning about open science practices is taking you away from time that could be spent on your own research. This is a common worry when engaging in any “extracurricular activities.” Although it does take time, it can be a relatively small time investment for how much you can learn, and it will help you get through the stack of papers you have on your desk that you need to

read (or browser tabs you keep open). You can choose how often to hold the journal clubs and when to start and stop holding them, so they could be anything from weekly to termly, whatever works well for you and your colleagues.

Project Workflow (Level: Easy)

What? Project workflow refers to how you organize projects and move through the various stages of the research cycle. This includes your file folder structure, document naming conventions, version control, cloud storage, and other details. It also includes the choice of who has access to the project (e.g., collaborators, the public) and when in the process they have access (e.g., at all times, upon publication).

Why? Having a dedicated project workflow system helps keep your research organized, enhancing reproducibility, minimizing mistakes, and facilitating collaborations with others and future-you. Making your project open to your advisors and any other collaborators (even if not open to the public) through working in shared folders can ensure everyone has access to everything in an organized fashion and saves the hassle of emailing infinite versions of documents. If you do choose to make the project public at any point, everything will be almost complete, and you will not have to create a system from scratch.

How? One option is to set up a private or public project page on the OSF. Soderberg (2018) provides instructions on how to: (1) create an OSF account, (2) create a project, (3) add collaborators, (4) upload files, and (5) review additional capabilities of an OSF project. What you decide to include on your project page is up to you. Some people find it helpful to include all study materials and anonymized data on the project page as they go, others only use the project page for final documents. You can alternatively organize a project using a variety of cloud-based storage providers that link directly to your computer (e.g., Google Drive, Dropbox, Box). Many

cloud-based storage providers also offer the option of linking directly to an OSF project so that you can have the benefits of both. Many research teams will not want to make all of their work public from the get-go, but even just *imagining* that the project will be public can encourage taking an outside perspective that will lead to improvements in organization. When joining a new lab as a graduate student or beginning a new collaboration, ask the project leaders about their project workflow. It is entirely possible that they do not have a formally specified workflow, and just asking about it could initiate new ways of doing things. If they are not as receptive as you would like, find a compromise between what you would find most useful and what they are used to, and if you are a new student, then it might be useful to set some review dates for when you will discuss if the current approach is working well. For additional resources visit the accompanied OSF project page (<https://osf.io/w5mbp/wiki/home/>).

Worries. You may feel some apprehension at the idea of having your workflow process public. You definitely do not have to make your project page public right away. You can wait until the project is complete, if you choose, and can clean up your project page if/when you eventually make it public. Relatedly, you may be unclear on what you are allowed to share. We encourage you to consult with the General Data Protection Regulation (GDPR; <https://gdpr.eu/>) and/or your institution's ethics board. For the most part, as long as you are working with de-identified data, you do not need to worry. Lastly, you may be concerned that if other people use your materials (e.g., survey design, code) that this is detrimental, as someone else is "profiting" from your hard work, but actually you can gain credit yourself in the form of citations (you can create a citable DOI for your OSF project).

Preprints (Level: Easy)

What? The term *preprint* originally referred to a version of a manuscript that was publicly available prior to being submitted for peer-review. Although that still remains true, preprints now also include manuscripts that are under review, or author-formatted versions of accepted articles prior to publication or even after publication (sometimes called postprints).

Why? Posting a manuscript before submitting to a journal allows for a wider range of feedback than what is afforded through peer review and can help improve a paper prior to submission by identifying any major flaws. Posting an article after submission, but before acceptance, gets a version of the paper out as soon as possible for sharing findings and interest, as well as keeping a record of what the paper looked like before it underwent the review process. Posting a manuscript after it has been accepted to a journal allows for the paper to be shared faster than it may be published and allows for an open access version of the paper to be shared. Preprints are also a great way to share work that does not continue to publication, providing greater access to the full body of available literature.

How? There are many different available preprint hosts with varying levels of moderation (e.g., arXiv, bioRxiv, SSRN) and emphasis on specific disciplines (see <https://osf.io/preprints/>). Here we focus on PsyArXiv (<https://psyarxiv.com/>), a preprint server developed for psychology by OSF, that is free to submit to and uses post-submission moderation. To submit a preprint to PsyArXiv, you must create an OSF account (which you did when establishing your workflow) and then submit directly from the PsyArXiv homepage. When doing so, you have the option to link the preprint with an existing OSF project if you have one.

As mentioned, you can post a preprint at different points of the publication timeline. If you are posting an article that has already been accepted for publication at a journal, in most cases you are able to post an author-formatted version to a preprint server, but not the final

publisher-formatted version (see Worries below for details). See Syed (2020) for suggestions on how to manage a preprint across the publication timeline, including keeping your preprint updated, linking to publication DOI, and merging citations in Google Scholar. The most challenging decision may be picking which license to choose for your submission. We recommend using the CC-BY license, as it allows other researchers to build upon your work while still citing the original work and not using the words verbatim (see Moshontz, 2018). For additional resources visit the accompanied OSF project page (<https://osf.io/w5mbp/wiki/home/>).

Worries. Most of the worries around posting preprints pertain to what is allowable and how doing so will impact the peer-review process. Many authors worry that posting a preprint will be treated as “published” and thus they will not be able to submit their manuscript for publication in a journal; however, this is not true in most cases. Before posting a preprint, authors should consult Sherpa Romeo (<http://sherpa.ac.uk/romeo>; <https://v2.sherpa.ac.uk/romeo/>), which tracks the restrictions and rules for most journals, indicating whether journals allow posting of preprints pre-submission (usually yes), posting of author-formatted accepted articles (usually yes), and posting of publisher-formatted accepted articles (usually no). There may also be concern that posting a preprint will decrease the number of times an article is cited; however, this is not the case as preprints have actually been found to increase the number of citations (Fraser et al., 2019) and you are able to merge citations between a preprint and published article in Google Scholar (Syed, 2020). Another concern with posting a preprint prior to submission is that someone else could “scoop” you, stealing your idea and running the study themselves before you are able to publish your work. Although this is possible, it is also very unlikely. Moreover, all articles are posted with a date/time stamp and therefore there is a clear temporal record.

Finally, students may have concerns that their advisors or other collaborators will not be open to posting preprints. Of course, you should always consult with coauthors prior to posting preprints. There should be little concern with posting author-formatted versions of accepted articles, but some may be more skittish about posting prior to submission to a journal. There may be worries about posting a version of paper that will later be changed; however, you are able to post as many updated revisions to your preprint as needed. Overall, we recommend engaging in a conversation to determine the source of the concern and then providing the preceding explanations for the commonly expressed concerns.

Reproducible Code (Level: Medium)

What? Reproducible code for data analysis and visualizations (e.g. tables, figures) refers to a detailed, written version of your code that would allow someone else (or your future self) to generate the same output reported in your manuscript.

Why? Reproducible code plays an important role in documenting the analysis pipeline, allowing for detection of errors, ease of modification in the analysis or visualization, and facilitation of sharing and collaboration. Using analysis scripts makes it easy to make small changes to repeat similar steps within the same project, or scripts from one project can be used as a starting point for another, therefore saving a lot of time in the long run. Finally, as noted, reproducible code allows for other people to transparently reproduce your analyses or visualizations.

How? Using R/RStudio (<https://www.r-project.org/>, <https://rstudio.com/>) is a popular choice for creating reproducible code. There are many helpful resources available online to learn R, and although it is hard work at the beginning, it does pay off in the long run. However, contrary to what seems to be popular belief, you do not need to use R to create reproducible

code. Although there are some potential drawbacks that we will not discuss here, windows-based programs where the user points and clicks options for analysis (e.g. SPSS Statistics; <https://www.ibm.com/products/spss-statistics>; JASP; <https://jasp-stats.org/>) can also be used in a reproducible way. In SPSS Statistics, a good starting point for beginners is to select the analysis options in the windows, then press the “paste” button rather than “OK.” Doing so will paste the analysis script into a new “syntax” file that can be modified, executed, annotated, and saved for future use. Similarly, options selected via point-and-click in JASP can be exported to a reproducible script. When preparing analysis code, be sure to add clear and extensive comments (including headers and explanations) to guide your future self and other people who may look at your code. For additional resources visit the accompanied OSF project page (<https://osf.io/w5mbp/wiki/home/>).

Worries. Many researchers have worries about learning something new because of the heavy time commitment and steep learning curve. This is completely understandable—coding can take a lot of time to learn from scratch. While learning does take time, there are many helpful tutorials and online resources to help the learning process (see <https://osf.io/w5mbp/wiki/home/>). Perhaps start with some of the introductory resources for getting started in R that we have suggested and see how it goes. If learning a script-based program is too daunting, start by pasting/exporting code via window selection to get accustomed to working with analysis code. If you do this for all analyses, add helpful comments, and save this file, you will greatly increase the quality of your workflow.

Another worry is that others will see mistakes in your code (especially when you are starting out) or you may conduct some analyses in a less-than-elegant way. However, this concern can actually be reframed as a positive, as error detection allows for awareness of the

problem and correcting the error as early as possible. Additionally, it is obviously the case that you can also make mistakes using point-and-click methods, it is just less likely that anyone would notice. To ensure this happens earlier rather than later, you can have a “coding buddy,” where you check each other’s code, or join a local coding club. Finally, choosing an analysis program is not an all-or-nothing decision—some people like to use R for all their data cleaning and manipulation but do statistical analyses using another program, or some may only use R for data visualization. Whatever you choose to use, just be sure that you are keeping documentation of all that you do.

Sharing Data (Level: Medium)

What? Sharing data pertains to making the de-identified dataset used for a project available to other researchers. Importantly, this means posting the data on OSF or another data repository for researchers to download and use, or establishing a formal system through which others can access the data (useful for sensitive data). There is enough evidence to indicate that stating “data available on request” is not sufficient to constitute engaging in the process of data sharing (e.g., Wicherts et al., 2006). Sharing data often also includes sharing analysis scripts (see section on reproducible code), especially the scripts related to data cleaning and labeling.

Why? There are several compelling reasons for sharing your data. First, data sharing allows others to reproduce the analyses reported in a paper, providing checks on quality and accuracy, and to expand on the analyses through fitting alternative models and conducting robustness tests. Second, most datasets have use beyond what is reported in a paper. This includes secondary data analysis that addresses different questions altogether and inclusion in meta-analyses, where researchers having access to the raw data is a major benefit. Third, sharing data may be required by the funding source for the project or the journal the article is published

in (see <https://www.topfactor.org/>). Sharing your data upon submission to a journal can be looked upon favorably by reviewers, even if they choose not to do anything with it, as it indicates a commitment to transparency.

How? The *how* of data sharing is why we conceptualize it as *medium* difficulty. There are a lot of complexities associated with data sharing, and so we direct all readers to Meyer's (2018) excellent article, "Practical Tips for Ethical Data Sharing," rather than covering all of those complexities here. Graduate students should read that article and share it with their supervisors and research team, as it addresses many of the commonly expressed concerns. Because the possibility of sharing must be indicated in the consent forms provided for participants, it can sometimes be difficult to publicly share data after the fact. Thus, a good time to initiate discussions about sharing data is during the project design phase. At that point, you can be sure to include appropriate clauses in the consent forms and ethics protocols that describe both your intent and your plan for sharing the raw data in an ethical way. Importantly, including these details does not require you to share your data, but rather allows for the option—if your supervisor is uncertain about sharing the data, you can revisit this with them once you have finished the project. Whether sharing your data or not—but especially if you are—it is critical to provide a data codebook that includes information on the structure of the dataset (e.g., what variable names correspond to, measurement levels); data sharing is only useful if it is understandable to an outsider. We have found Arslan's (2019) *codebook* R package is an excellent way to use reproducible code to create a codebook during the data cleaning phase. Finally, there are loads of different sites where you can share your data, so it can be a bit overwhelming to choose. For simplicity's sake, we suggest sharing data on OSF (this keeps the

data centralized with the rest of the project materials). For additional resources visit the accompanied OSF project page (<https://osf.io/w5mbp/wiki/home/>).

Worries. There are three major worries with data sharing. First, as mentioned above, are ethical concerns, for which we again direct readers to Meyer (2018). Second, many datasets yield more than one article, and therefore you may be worried that other researchers will conduct the analyses you had planned if you share your data. Two ways to mitigate this problem are 1) to delay sharing the full dataset until you have answered all planned questions, sharing only the variables/cases relevant for the initial article and 2) preregistering (see below) all planned analyses from the get-go and indicating these plans in the data documentation. Finally, there is sometimes concern about other researchers benefiting from all the hard work you put in to collecting the data, and that you will not get credit for subsequent use. However, by applying at least a CC-BY license to the data set, anyone who uses the data is obligated to attribute the data to you (e.g. through citing the associated paper or through using a DOI for the OSF project).

Transparent manuscript writing (Level: Medium)

What? To write a transparent, clear, and reproducible manuscript, it is helpful to follow manuscript writing guidelines or standards. Guidelines or standards include suggestions on the level of detail that should be included within each section of an article. This includes being transparent (e.g., about hypotheses and participant characteristics), as well as stating and justifying decisions (e.g., about the stopping rule and analyses).

Why? Writing a transparent and clear manuscript helps other researchers better understand how the research was conducted and allows researchers to better calibrate the implications of the findings. It may even speed up the peer review and publication process, as reviewers will not have to ask for information that was not immediately clear from the

manuscript. Writing transparently also facilitates replication attempts because other researchers can more easily follow the method and analysis, as well as a more cumulative science, as subsequent studies can more accurately build upon each other.

How? There are a few helpful guidelines and standards that are easy to follow. Whereas Bem (2004) was once considered a gold standard for writing journal articles, it has since fallen out of favor due to many of his recommendations not supporting transparent and honest writing. Rather, Gernsbacher (2018) provides excellent recommendations that are current and consistent with best practice in open science. Additionally, the American Psychological Association has published Journal Articles Reporting Standards for quantitative studies (Appelbaum et al., 2018) and qualitative/mixed method studies (Levitt et al., 2018). Importantly, transparent writing does not actually begin with the writing phase. It is helpful to maintain a log that documents all decisions made in the project across all phases. Often decisions will be made in meetings or through email and keeping these all in one place can really help when it comes to writing.

Worries. A common worry is that writing transparent, detailed manuscripts will take more time and slow down the research process. Although following standards may involve including more detail than one would normally, it may save time in the revision process. Additionally, it can feel embarrassing to report, for example, that experimenter error resulted in missing data for some measures, or that funding ran out, so the sample is smaller than intended. Further it may feel vulnerable to report that, although you had a certain analysis in mind, after reading more you realized another method was better. However, being open about these things will normalize the humanness of research and help others designing similar studies, and good reviewers will appreciate this.

Following these transparent recommendations may result in the paper being too long, either for a journal's length requirement or just for your own taste. If you are worried about this, some additional information can be uploaded as supplementary materials (e.g., a detailed description of the procedure including materials; results from robustness tests, etc.). As long as this is signposted clearly in the paper, and openly available (e.g. on the OSF project), then this is perfectly fine, and certainly preferable to the alternative of not providing the information at all.

Preregistration (Level: Medium)

What? In the most basic form, preregistration refers to posting a timestamped outline of the research questions, hypotheses, method, and analysis plan for a specific project prior to data collection and/or analysis (see Nosek et al., 2018).

Why? The primary purpose of preregistration is to more clearly distinguish between the decisions and hypotheses made before the study began ("confirmatory" research) and decisions made after seeing the data or in the absence of specific hypotheses ("exploratory" research). By outlining the research questions, hypotheses, method, and analysis plan prior to data collection and analysis, your intentions and predictions are clear and time stamped. It can often be useful to have "proof" of what was decided before the project in case you or your advisors/collaborators forget. More generally, preregistration is helpful for making sure you have thought through the project fully before starting data collection or analysis.

How? Preregistration is best conceptualized as a detailed, written version of your study design. That said, preregistration plans can be prepared at various levels of specificity. They can be brief, only describing the research questions, hypotheses, and how you will test them, or they can include more detail about the handling of missing data, planned analyses, code that will be used, a decision tree based upon data, and other predetermined decisions. In general, however,

the more detail you can provide the better. The variety of preregistration templates available on OSF (<https://osf.io/zab38/>) correspond to these variations in specificity and represent different types of research contexts (e.g., experimental, secondary data analysis, replications).

Preregistration plans in OSF can be kept as drafts and are editable until officially registered, at which point they are locked. You can make the plan public immediately or embargo it for a specific amount of time, up to four years (e.g. until you expect data collection will be finished).

Talk to your advisors/collaborators about preregistration, why you would like to do it, and how involved they would like to be in the preregistration process. Again, one way to think about preregistration plans is that they are written versions of the study design.

Advisors/collaborators should at least read and approve the content of the preregistration plan before it is registered, even if you did not go through the sections together; however, going through the plan together can also be a great way to discuss intricacies of the project and sort out any issues before data collection begins. If your advisors are not on board with preregistration, a step in the right direction is to use a preregistration template to record all design decisions, even if you do not formally register it. For additional resources visit the accompanied OSF project page (<https://osf.io/w5mbp/wiki/home/>).

Worries. Perhaps the most common worry about preregistration is that it stifles creativity and discovery by barring exploratory analyses. This is false. Exploratory analyses are perfectly acceptable in preregistered studies, you just have to clearly specify that the analyses are exploratory, and you cannot reframe exploratory analyses as though they had always been confirmatory (e.g., HARKing; Kerr, 1998). Another common concern is that you will make mistakes by preregistering the wrong thing or realizing a better approach after the fact. The fact is, you *will* make mistakes (Rouder et al., 2019), but fortunately preregistration is not a binding

contract. You can always create a new registration of changes, depending on when in the research cycle they occur, or, alternatively, just be transparent about these changes in your write-up—the key is to be transparent about what you did. Remember, you do not need to cover every detail of your study in your preregistration and you can decide how detailed you would like to be. For those just starting out with preregistration, we suggest creating a simple plan focused on the research questions, hypotheses, and how you will test them.

Registered Report (Level: Difficult)

What? Most journals in psychology, as well as other sciences, make decisions about whether to accept articles based on the pattern of results. In contrast, Registered Reports seek to shift the selection process to make acceptance based on sound conceptualization and design (Chambers & Tzavella, 2020). Registered Reports involve a two-part submission process, where authors first submit a *Stage 1 proposal*, which includes the introduction, method, and analysis plan—all before data collection and/or analysis has been done. This Stage 1 proposal is what is submitted for peer review, with the ultimate outcome being an “in-principle acceptance” (IPA), which means that the journal guarantees to publish the article, regardless of the results, so long as the accepted plan is followed and executed with sufficient quality. Upon completion of the project, the results and discussion sections are added to the Stage 1 manuscript and re-submitted to the journal as a *Stage 2 manuscript*, which is then reviewed to ensure adherence and quality. Registered Reports are related to preregistration in that they involve making plans prior to beginning the project, but Registered Reports are even more detailed and undergo the review process before data collection and/or analysis.

Why? Publication bias, wherein articles are disproportionately accepted for publication based on meeting a specific threshold (e.g., $p < .05$) or because they fit with prevailing wisdom,

is a major problem in psychology (Scheel et al., 2020). Registered Reports help to reduce publication bias of only significant results by guaranteeing publication regardless of the substantive outcome. They are intended to center studies that are informative regardless of the outcome through strong conceptualization and design. Additionally, the Registered Report process is almost certain to lead to a stronger project, as the design and analyses are peer reviewed at a stage when changes can actually be made and not just relegated to the limitations section, preventing “fatal flaws” that could doom the manuscript from publication altogether.

How? We strongly encourage readers to consult <https://www.cos.io/rr>, which contains a wealth of resources, including an extremely informative FAQ and Kiyonaga and Scimeca (2019), who have created a helpful practical guide. In terms of graduate students submitting Registered Reports, there may be different considerations for different types of programs. Worldwide, there is variation in how research is conducted during a Ph.D. program. In many countries the thesis/dissertation consists of a series of studies that are pursued across the course of study. Because some of the studies may need to be published prior to defense, time can be a major concern in such programs (but see below). In contrast, Ph.D. dissertations in the U.S. typically consist of a single well-designed study, or series of studies, undertaken towards the end of the program. Interestingly, most such programs split the dissertation process into two parts. The first is a proposal, where the student submits the introduction, method, and planned analysis section to a committee for review and approval. The second part is the defense, where the student reports on the findings. This sequence is almost identical to the Stage 1/Stage 2 process of Registered Reports, and thus dissertations in the U.S. are very well-suited for the format.

Regardless of the program structure, to submit a Registered Report as part of your Ph.D., it will be important to plan ahead. You should talk to your advisors as early as possible in your

graduate program, and if they are new to Registered Reports, be sure to review the resource links beforehand to know some answers to worries/questions they are likely to have. For additional resources visit the accompanied OSF project page (<https://osf.io/w5mbp/wiki/home/>).

Worries. Preparing a Registered Report may be complicated if your advisor/s, committee members, or collaborators are not supportive of the idea. They will need to be committed to the idea, as it is not something most students can do on their own. What students can do is make sure they are as informed as possible about the benefits of Registered Reports and responses to common worries and try to respond to whatever *their* specific worries are. For those students in the U.S. or other countries that use a similar model, reminding them that the process is the same as how they approach dissertations could be useful to help them see the value.

As noted, time can be a major concern. Preparing a Registered Report means needing to write a manuscript-length document and submit it to an appropriate journal, then waiting and working through the review-revision cycle until you get an IPA and can conduct the study. This can certainly take some time, however a substantial amount of time is saved on the back-end: the analyses are already planned out, and you know that the manuscript will be accepted and, thus, you do not need to submit to multiple journals, waiting through the review process of each until the paper “finds a home.” Moreover, while the Stage 1 proposal is under review, students can finish preparing everything needed to get started with the project upon receipt of the IPA. Nevertheless, it is important to keep in mind whether the Registered Report timeline makes sense for when you would be able to collect data and complete the study.

Conclusion: This Is Just the Beginning

In the preceding sections, we have detailed eight specific open science behaviors that students can engage in *right now*. The behaviors vary from easy to difficult, but they are all

doable. Indeed, our definition of “difficult” is in the context of a novice open science practitioner. In the context of the full open science movement, there are many more behaviors to engage in across the research cycle, some of which are quite advanced, including using RMarkdown to write reproducible manuscripts (Mair, 2016), pushing analysis code to Github (<https://github.com/>) or some other repository, and developing preregistration plans that include full analysis scripts based on simulated data (for additional examples see <https://www.osc.uni-muenchen.de/toolbox/index.html>). The purpose of this tutorial is *not* to overwhelm you with possible practices, but rather to present a few practices that can help ease you into open science. Once you do, there is no turning back!

Author Contributions:

UK & MS conceptualized the paper; UK, PS, & MS wrote the paper.

Conflicts of Interest:

The author(s) declare that there were no conflicts of interest with respect to the authorship or the publication of this article.

Acknowledgments:

We thank Linh Nguyen and Aisha Udochi for their helpful comments on this manuscript.

Funding:

The first author's work is supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. (00039202).

Disclaimer:

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Allen, C. & Mehler, DMA. (2019). Open science challenges, benefits and tips in early career and beyond. *PLoS Biol* 17(5): e3000246. <https://doi.org/10.1371/journal.pbio.3000246>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3-25. <http://dx.doi.org/10.1037/amp0000191>
- Arslan, R. C. (2019). How to Automatically Document Data With the codebook Package to Facilitate Data Reuse. *Advances in Methods and Practices in Psychological Science*, 2515245919838783. <https://doi.org/10.1177/2515245919838783>
- Bem, D. J. (2004). Writing the empirical journal article. In J. M. Darley, M. P. Zanna, & H. L. Roediger III (Eds.), *The compleat academic: A practical guide for the beginning social scientist*. (p. 185–219). American Psychological Association.
- Bergmann, C. (2018). *How to integrate open science into language acquisition research*. Student workshop at the 43rd Boston University Conference on Language Development (BUCLD), Boston, USA.
<https://docs.google.com/presentation/d/1bdICPzPOFs7V5aOZA2OdQgSAvgoB6WQweI21kVpk9Gg/edit#slide=id.p>
- Center of Open Science. (n.d.) *Registered Reports*. <https://www.cos.io/rr>
- Center of Open Science. (n.d.) *Top Guidelines*. <https://www.topfactor.org/>
- Chambers, C. D., & Tzavella, L. (2020). Registered Reports: Past, Present and Future [Preprint]. *MetaArXiv*. <https://doi.org/10.31222/osf.io/43298>

- Christensen, G., Wang, Z., Levy Paluck, E., Swanson, N., Birke, D., Miguel, E., & Littman, R. (2020). Open Science Practices are on the Rise: The State of Social Science (3S) Survey. *UC Berkeley: Center for Effective Global Action*.
<https://escholarship.org/uc/item/0hx0207r>
- Corker, K. S. (2018, September 12). Open Science is a Behavior. *Center of Open Science*.
<https://cos.io/blog/open-science-is-a-behavior>
- Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., ... Schulte-Mecklenbeck, M. (2018, November 16). 7 Easy Steps to Open Science: An Annotated Reading List. <https://doi.org/10.31234/osf.io/cfzyx>
- Fetzer Franklin Fund. (n.d.) *Metascience: The Emerging Field of Research on the Scientific Process*. <https://metascience.com/>
- Foster, E. D., & Deardorff, A. (2017). Open science framework (OSF). *Journal of the Medical Library Association: JMLA*, 105(2), 203. doi: 10.5195/jmla.2017.88
- Fraser, N., Momeni, F., Mayr, P., & Peters, I. (2019). The effect of bioRxiv preprints on citations and altmetrics. *bioRxiv*, 673665. <https://doi.org/10.1101/673665>
- General Data Protection Regulation. (n.d.) *Complete guide to GDPR compliance*. <https://gdpr.eu/>
- Gernsbacher, M. A. (2018). Writing empirical articles: Transparency, reproducibility, clarity, and memorability. *Advances in methods and practices in psychological science*, 1(3), 403-414. <https://doi.org/10.1177/2515245918754485>
- Github. (n.d.) *Github*. <https://github.com/>
- JASP. (n.d.) *JASP*. <https://jasp-stats.org/>
- Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, 2(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

- Kiyonaga, A., & Scimeca, J. M. (2019). Practical Considerations for Navigating Registered Reports. *Trends in neurosciences*, 42(9), 568-572. doi: 10.1016/j.tins.2019.07.003.
- Koyama, J. & Page-Gould, E. (2020, February). *The open science conversation* [Poster presented]. Society for Personality and Social Psychology, New Orleans, LA.
<https://osf.io/7mqud/>
- Levitt, H. M., Bamberg, M., Creswell, J. W., Frost, D. M., Josselson, R., & Suárez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 26-46.
<http://dx.doi.org/10.1037/amp0000151>
- Lewis Jr, N. A. (2019). Open Communication Science: A Primer on Why and Some Recommendations for How. *Communication Methods and Measures*, 1-12.
<https://doi.org/10.1080/19312458.2019.1685660>
- Ludwig-Maximilians-Universitat Munchen. (n.d.) *Open Science Center: Toolbox*.
<https://www.osc.uni-muenchen.de/toolbox/index.html>
- Mair, P. (2016). Thou shalt be reproducible! A technology perspective. *Frontiers in psychology*, 7, 1079. <https://doi.org/10.3389/fpsyg.2016.01079>
- Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science*, 1(1), 131-144. <https://doi.org/10.1177/2515245917747656>
- Moshontz, H. (2018, May 4). Licensing your work on PsyArXiv. *PsyArXiv Blog*.
<http://blog.psycharxiv.com/2018/05/14/licensing-work-psycharxiv/>
- Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual review of psychology*, 69, 511-534. <https://doi.org/10.1146/annurev-psych-122216-011836>

- Nosek, B. A. (2014, February 28). Improving my lab, my science with the open science framework. *APS Observer*, 27(3).
<https://www.psychologicalscience.org/observer/improving-my-lab-my-science-with-the-open-science-framework>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600-2606.
<https://doi.org/10.1073/pnas.1708274114>
- Nuijten, M. B. (2019). Practical tools and strategies for researchers to increase replicability. *Developmental Medicine & Child Neurology*, 61(5), 535–539.
<https://doi.org/10.1111/dmcn.14054>
- Open Science Framework. (n.d.) *Open Science Dashboard*. <http://osf.io>
- OSF Preprints. (n.d.) *OSF Preprints*. <https://osf.io/preprints/>
- PsyArXiv Preprints. (n.d.) *PsyArXiv*. <https://psyarxiv.com/>
- R. (n.d.) *The R Project for Statistical Computing*. <https://www.r-project.org/>
- ReproducibiliTea. (n.d.) *ReproducibiliTea*. <https://reproducibilitea.org/>
- Rouder, J. N., Haaf, J. M., & Snyder, H. K. (2019). Minimizing mistakes in psychological science. *Advances in Methods and Practices in Psychological Science*, 2(1), 3-11.
<https://doi.org/10.1177/2515245918801915>
- RStudio. (n.d.) *RStudio*. <https://rstudio.com/>
- Scheel, A. M., Schijen, M., & Lakens, D. (2020). An excess of positive results: Comparing the standard Psychology literature with Registered Reports. *PsyArXiv*,
<https://psyarxiv.com/p6e9c/>
- Sherpa Romeo (n.d.) *Sherpa Romeo*. <http://sherpa.ac.uk/romeo> or <https://v2.sherpa.ac.uk/romeo/>

Soderberg, C. K. (2018). Using OSF to share data: A step-by-step guide. *Advances in methods and practices in psychological science*, 1(1), 115-120.

<https://doi.org/10.1177/2515245918757689>

Spellman, B. A., Gilbert, E. A., & Corker, K. S. (2018). Open science. *Stevens' handbook of experimental psychology and cognitive neuroscience*, 5, 1-47.

<https://doi.org/10.1002/9781119170174.epcn519>

SPSS Statistics. (n.d.) *IBM SPSS Statistics*. <https://www.ibm.com/products/spss-statistics>

Syed, M. (2019). The Open Science Movement is For All of Us. *PsyArXiv*.

<https://psyarxiv.com/cteyb/>

Syed, M. (2020, May 7). Managing Preprints across the Publication Pipeline. *Get Syeducated*.

<https://getsyeducated.blogspot.com/2020/05/managing-preprints-across-publication.html>

Tennant, J [DARIAH-EU]. (2018, June 1). Open Science is just good science [Video]. Youtube.

https://www.youtube.com/watch?v=UEEcwRUgQu8&feature=emb_title. [Talk delivered at the Oxford University Seminar on Reproducibility and Open Research.

<http://users.ox.ac.uk/~phys1213/ReproAtASC.html>]

Vazire, S. (2017). Quality uncertainty erodes trust in science. *Collabra: Psychology*, 3(1). DOI:

<https://doi.org/10.1525/collabra.74>

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61(7), 726–728.

<https://doi.org/10.1037/0003-066X.61.7.726>

Yarkoni, A. T. (2019, July 13). I hate open science. *[Citation needed]*.

<https://www.talyarkoni.org/blog/2019/07/13/i-hate-open-science/>