"Like Upgrading From a Typewriter to a Computer": Registered Reports in Education Research

Justin Reich

Massachusetts Institute of Technology

Hunter Gehlbach

Johns Hopkins University

Casper J. Albers

University of Groningen

Registered reports engender a review process in which reviewers evaluate the merits of the research question rather than the magnitude or direction of the results. In this process, authors first submit introduction, methods, and prespecified analyses sections of an article. Next, an initial peer-review process assesses the merits of the research question, in the absence of knowing the results. Worthy articles are granted an "in-principle acceptance." The authors, then conduct their study, analyze their data, write up the remainder of their article, and resubmit it. The in-principle acceptance means that if the authors adhere to the research design as originally submitted, then the article will be accepted. Building from seven of the earliest registered reports in education research, our introduction describes the benefits, opportunities, and challenges of registered reports as a means to increasing transparency in research and reducing the number of illusory results.

Keywords: *evaluation, open science, preregistration, registered reports, research methodology, research utilization, standard setting*

RECENT observers of the scientific process have called attention to a set of related issues that threaten the integrity of the scientific literature (Ioannidis, 2005; John et al., 2012; Van der Zee & Reich, 2018). These issues range from a lack of replication studies to "*p*-hacking" to a lack of transparency in analytic decision making (Gelman & Loken, 2014; Makel & Plucker, 2014; Simmons et al., 2011). In an ideal world, research that tackles compelling scientific questions and urgent public problems with the highest quality methodological approaches would be the most "publishable" research. Instead, scientific research has disproportionately been judged on the direction and magnitude of findings, or novelty of results, rather than on methodological rigor (Makel & Plucker, 2014). As a result, in many social science fields, illusory results buttressed by elaborate stories have saturated the research literature; null results are underrepresented. Gehlbach and Robinson (2018) liken the current state of science to the steroids eras in baseball. Performance-enhancing techniques in both have caused crises of confidence and credibility. One of the most promising innovations in addressing these issues is the focus of this *AERA Open* special topic: Registered Reports.

Registered reports are a new format of journal article where authors first submit a "Phase I" manuscript that includes the introduction, background and context, and methods (including prespecified analyses) sections of an article (Chambers, 2013; Chambers et al., 2015). Peer reviewers evaluate the importance of the research questions and the appropriateness of the data, method, and analytic approach, *in the absence of knowing what the results might be.* At this point editors can then grant the article "in-principle acceptance," request revisions, or even suggest modifications of the study design. An in-principle acceptance means that if the authors adhere to the research design as described in the Phase I manuscript, then the article will be accepted for publication regardless of the direction or magnitude of findings. After an in-principle acceptance, the authors then conduct their study or analyze their data, finish writing up their article, and submit a final draft that is then reviewed by the editor (and reviewers as needed).

Two notable benefits arise from this approach to the publication process. First, by shifting peer review earlier in the research process, authors benefit from the expertise of peer reviewers while there is still opportunity to make substantial methodological changes. Second, by granting in-principle acceptance before findings are known, the scientific publishing process removes several major incentives that drive researchers toward questionable research practices, such as mining for "significant" results. Over time, this practice should reduce the proportion of illusory results in the literature.

In this special topic, we are pleased to publish together seven of the first registered reports in education research. These articles address diverse topics: instructor-student relationships, norms for identifying gifted students, absenteeism, teacher perceptions of researchers, mischievous survey responders, text message–based interventions to improve college enrollment, and handwriting practice in Chinese language instruction. Furthermore, the authors showcase a range of methodologies: from lab studies to field experiments to secondary data analysis on both bespoke and also on widely used data sets. What these studies have in common is adherence to an authoring process designed to increase transparency and to a peer-review process designed to emphasize the quality of methods over the particular happenstance of findings. All the authors systematically described their methodological plans before data analysis (and in some cases, before data collection), negotiated those plans with peer reviewers and editors, and published both the analyses they preregistered along with relevant post hoc exploratory work.

Because registered reports are a relatively novel approach to the publication process in education research, we use this introduction to the Special Topic to illustrate the range of approaches used by authors in publishing registered reports. In the hope that this article might spark ideas for authors and editors interested in registered reports, we combine our own reflections with those of the authors' on the advantages, challenges, and lingering tensions involved in this publication process.

Situating the Seven Registered Reports in This Special Topic

Merk and Rosman (2019) conducted a pair of experiments on students enrolled in a teacher education program to evaluate how novice teachers evaluate sources of education research. The first exploratory study was conducted before preregistration, and it showed that novice teachers see expert education researchers as less benevolent but with greater expertise than typical teachers, which the authors describe as "smart but evil." Merk and Rosman conducted a second study as a registered report with preregistered hypotheses to confirm (or disconfirm) their initial findings. They found similar biases but with moderated effect sizes. The article

demonstrates how an exploratory finding and a more robust confirmatory design can be published in the same registered report, adding a replication to the scientific literature.

Lu et al. (2019) used the ASSISTments online homework submission system to test the importance of Chinese handwriting practice in Chinese language learning. In a pilot study, the authors found that students who spent time on handwriting practice scored lower on word recognition, and they argue that learning time is best spent on other tasks. In this registered report, they find similar results again. Apart from a short description of the preregistration process, this registered report is similar to many published experimental trials in online learning. This example shows that a registered report that goes as planned looks a lot like a traditional article. The key difference is that readers can assess the article knowing that the analytic plan was preregistered and then adhered to, unless otherwise noted in the article.

By contrast, the article from Robinson et al. (2019) shows how registered reports provide a more transparent process for publishing studies that don't unfold as intended. In pilot work in high schools, researchers found that interventions to improve instructor-student relationships lead to higher grades among treated students (Gehlbach et al., 2016). Robinson et al. tested a similar intervention in a higher education context and found null results. Nevertheless, exploratory work on pretest surveys demonstrated theoretically interesting correlations between instructor-student relationships and student performance that had not been documented previously at the collegiate level. In a traditional publication, the authors might have felt pressure to minimize the broader experimental context and publish only the "significant" survey findings. In a registered report, the field benefits from the publication of a well-designed experimental study with null results, and the exploratory survey findings are presented in a fuller context of the larger initiative. Some critics of open science have raised concerns that preregistering hypotheses, as is required in the registered report process, might eliminate the possibility of discovering important findings through additional data analysis. This article shows that the researchers retained the flexibility to pursue these additional interesting lines of inquiry as the vagaries of their research agenda unfolded, but they published their results in a way that clearly delineates results that come from a priori versus post hoc analyses.

Kramer (2020) presents a large-scale field experiment testing the efficacy of different framings for text messages designed to support the transition from high school graduation to college enrollment and attendance. The study with implemented through partners, and data collection was already underway when the call for this Special Topic was posted. While it might have been ideal to have her registered report peer reviewed before the experimental design, Kramer submitted her Phase I manuscript before she conducted any data analysis. This illustrates that advances in open science

are not binary or one-size-fits-all, but rather adaptable to the circumstances of particular studies. Through the peer review process, reviewers and editors guided Kramer to be more specific in a priori planning for her subgroup analyses, and while Kramer couldn't change aspects of the experimental design, she could change the hypotheses for her analytic design before examining her data.

Kirksey (2019) gathered attendance and grade data from one large California school district to estimate the magnitude of harms from absenteeism. In a reflection on the registered report process, Kirksey described how the preregistration process forced him, in his words:

> to be more intentional in my conversations with the school district providing the data. Knowing that I would be expected to follow through using the empirical models I specified *a priori*, I thought more deeply about what data I would have access to and what coding, construction of variables, and inclusion of covariates would be most ideal for someone who was unfamiliar with the data (e.g., reviewers).

Peters et al. (2019) as well as Cimpian and Timmer (2019) used widely available data sources for secondary analysis. Peters et al. use data from 10 third-grade cohorts of students taking the Northwest Evaluation Association Measures of Academic Progress test, and they evaluate different criteria for assigning students to receive gifted education services. Policymakers and researchers have devised numerous systems for identifying potential recipients of gifted education services. The preregistration process allowed the researchers to make timestamped predictions about those different approaches in advance of testing various combinations.

Cimpian et al. (2018) had previously examined the Youth Risk Behavior Survey and found that an unlikely percentage of boys reporting to be homosexual also reported a variety of unusual behaviors, for example, extreme heroin use. This study argued that that these unlikely behaviors were due to "mischievous responders," and this data integrity issue can lead to overstated risks to LGBTQ (lesbian, gay, bisexual, transgender, and queer or questioning) youth. In this special topic, Cimpian and Timmer (2019) publish a preregistered, confirmatory replication of their previous study using the new 2017 release of the same Youth Risk Behavior Survey. As before, they demonstrate that mischievous responders are threatening the integrity of inferences from these large national surveys.

Mischievous responders pose a special risk to research integrity because our current publication system is tilted toward the publication of provocative and extreme results. In their reflection, Cimpian and Timmer describe some of the misaligned incentives in their subfield, and how registered reports might help realign those incentives:

> The current incentives in academia can perversely lead researchers to ignore threats due to mischievous responders. For example, ignoring mischievous responders can save researchers a lot of difficult work trying to identify potential mischievous responders, while giving researchers smaller *p*-values that are more likely to get published, lead to grant funding, and get media coverage. With registered reports, however, researchers could specify in advance how they would identify and remove suspected mischievous responders, and the research design could be accepted for in-principle publication regardless of whether the removal of suspected mischievous responders took a *p*-value from $p < .001$ to $p = .72$. Registered reports could incentivize researchers to conduct the difficult work of trying to improve data validity by not penalizing researchers for "losing" significant findings.

In publishing a rigorous replication of prior findings on mischievous responders, Cimpian and Timmer demonstrate how registered reports can give readers improved confidence in the inferences from a single study while paving the way for an improved system of education research.

Author and Editor Perspectives on Registered Reports

As might be expected from early adopters, the editors and authors involved in this Special Topic were enthusiastic about the design, writing, and review process. One author team noted, "Going from the old way of doing research, in which reviews are received after all the work has been done, to registered reports is like upgrading from a typewriter to a computer." Registered reports lead to better study design, more streamlined work, and clearer distinctions for readers about what are preplanned, confirmatory analyses and what are post hoc, exploratory analyses.

Both authors and editors agreed that the peer review process for this special topic was a friendlier, more productive set of exchanges than might be typically found in education research publishing. Editors and authors both noted that the process felt more like formative assessment—helping authors conduct the best possible study, rather than merely performing a gatekeeping function. Authors described the feedback as "a much more collaborative and friendly process" and less adversarial, "really about working together with the reviewers." Since this feedback happens earlier in the project, reviewers can suggest major methodological changes, and authors in most cases have a genuine opportunity to respond and adapt their study designs. From our perspective as editors, reviewers appeared less critical when they didn't know the results of the study, and the criticisms that were shared were more constructive and generative.

Several authors argued that the process of designing a study for a registered report was simultaneously easier and more rigorous. The preregistration process requires researchers to think through the details of design prior to implementation and data collection. Of course that's best practice for most quantitative studies, but the registered report process provides additional incentives for thoughtful preplanning. Several authors argued that the process either took no extra time— merely shifted it—or actually saved time compared with traditional publishing. Since reviewers and authors agree on

analyses up front, the iteration happens at the ideation phase, not executing new analyses with data. One team observed, "We saved so much time on analyses by not having to re-do ten different analyses to address reviewer criticisms."

Editors and authors agreed that one of the best benefits of the format was that authors felt freer to be candid in the reporting and discussion of results. In the current publishing system, when researchers anticipate that findings will be evaluated on their magnitude and novelty, they may feel pressure to have their results conform to a neat, tidy, positive story. One author team wrote,

> Writing the discussion section after Phase I acceptance was a new experience to us: Without fearing that the paper might be rejected, we felt that the discussion became more honest and vivid as we could, for example, give sharper opinions.

After in-principle acceptance, researchers can report their findings and be straightforward about null effects, possible confounds, and additional possible interpretations.

Finally, we note with enthusiasm that three of our studies have graduate students as first authors. We believe that the upcoming generation of education researchers will be champions of new approaches to open science.

Lingering Challenges and Tensions for Future Editors of Registered Reports

Many authors noted in their reflections, and we as editors agree, that not all studies are appropriate for a registered report. Exploratory work may not need early rounds of review or preplanning. Many of the studies published in this special topic are the culmination of lines of inquiry that carry over multiple studies. Registered reports may be more appropriate for this kind of confirmatory research. All the studies published here use quantitative methodologies, and the benefits of registered reports for qualitative or design-based work is less clear.

Education research is multifaceted and complex, and there are tensions between the demands of field implementations and the constraints of the registered reports process. One author submitted their Phase I manuscript after field implementation, and after a third party had conducted data collection, but before analysis. One limitation of this timing is that if reviewers had recommended different data collection approaches to the author, they would not be able to adapt them. We chose to proceed with reviewing the article, because we view open science practices as a continuum rather than as a binary. Being in perfect compliance with ideal registered report conditions is good, but preregistering studies before data analysis (even if not before data collection) is still preferable to no preregistration at all.

Preregistration is also best viewed along a continuum. Open Science Framework and other groups have put together templates and guidelines for rigorous and precise preregistration of exactly what equations are to be tested. However, particularly in education research, data do not always arrive in perfect form. Eliminating satisficers (Barge & Gehlbach, 2012) or mischievous responders, transforming nonnormal data, handling missing responses, and so on, may be hard to prespecify (Gehlbach & Robinson, 2018). Thus, the extent to which registered reports require preregistration on independent platforms, the level of precision of the prespecification, and what happens when the data come back with unexpected challenges are all issues that need to be worked out over time as more education research is published in the form of registered reports.

Manuscript length emerged as an additional topic of reflection. We asked authors to submit detailed methodological plans in the Phase I manuscript, and when findings and additional exploratory analysis were included in the main articles, they could become quite long. In future efforts, we would recommend that authors continue to be very thorough in Phase I, and then editors could encourage authors to move many of the Phase I methodological details into supplementary materials. In a future editorial process, we'd recommend editors provide more explicit coaching around how to publish short, focused articles aimed at a wide readership that are accompanied by extensive additional detail for specialists in online supplementary materials.

Given that there are multiple rounds of review, as editors we questioned when we wanted to include peer reviewers in the evaluation of Phase 2 manuscripts. Our bias was toward treating decisions around Phase 2 manuscripts as editorial rather than peer review decisions. If we thought researchers faithfully executed on their Phase I plans without complication, we accepted their articles as editors. Sometimes, rather than sending an article back to peer reviewers for a full review, we asked a single reviewer to check on some specific detail. The field may develop norms and standards to guide editors as to when peer review should extend into Phase 2.

To the extent that editors want the review process to continue to serve a gatekeeping function (perhaps a questionable goal in a world of unlimited server space), new screening criteria will need to be generated to determine what Phase I manuscripts are publishable. We will need new norms around which types of hypotheses are sufficiently interesting to carry a whole paper. We ultimately accepted seven of eight Phase I manuscripts. After the peer review process, we believed all of these studies represented well-designed, well-written contributions to the research literature.

On Volunteering to Advance Open Science

A final note on our role and position as editors: We proclaim no special expertise in producing or editing registered

reports. As authors, we have published preregistered studies in the past, but none of us had published, reviewed, or edited a registered report before this special topic. We had no invitation from *AERA Open* or *AERA*, we simply responded to an open call to propose special topics. If education research is to benefit from advances in open science, those advances will be brought about, subfield by subfield, by groups of inexperienced, well-intentioned volunteers who raise their hands to try to make education research a little bit more trustworthy and a little bit more useful to practitioners and policymakers.

As you read the seven articles in this special topic, if you are excited about the approaches presented here, we encourage you to raise your hand to advance open science in your part of education research. Write a proposal to your favorite journal proposing a special issue on registered reports. Host a symposium of registered reports in your division, SIG, or conference. Plan a registered study and ask your favorite journal to review it as a registered report. Encourage a journal editor to differentiate articles (e.g., through badging or different sections) with prespecified analyses versus exploratory analyses. Talk to conference program committee chairs, *AERA* division heads, NSF (National Science Foundation) program officers, and others about what you can do to advance registered reports and open science.

Finally, you can submit your next article here, to *AERA Open*, as a registered report. Seeing the results of this special topic, the publication committee at *AERA* and the editors of *AERA Open* have agreed to make registered reports a new submission type for the journal. Our hope is that some of the most rigorous, transparent, candid research in the decades ahead will be published here, in this new format.

References

Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, *53*(2), 182–200. https://doi.org/10.1007/s11162-011-9251-2

Chambers, C. D. (2013). Registered reports: A new publishing initiative at Cortex. *Cortex*, *49*(3), 609–610. https://doi.org/10.1016/j.cortex.2012.12.016

Chambers, C. D., Dienes, Z., McIntosh, R. D., Rotshtein, P., & Willmes, K. (2015). Registered reports: realigning incentives in scientific publishing. *Cortex*, *66*, A1–A2. https://doi.org/10.1016/j.cortex.2015.03.022

Cimpian, J. R., & Timmer, J. D. (2019). Large-scale estimates of LGBQ-heterosexual disparities in the presence of potentially mischievous responders: A preregistered replication and comparison of methods. *AERA Open*, *5*(4). https://doi.org/10.1177/2332858419888892

Cimpian, J. R., Timmer, J. D., Birkett, M. A., Marro, R. L., Turner, B. C., & Phillips, G. L. (2018). Bias from potentially mischievous responders on large-scale estimates of lesbian, gay, bisexual, or questioning (LGBQ)–heterosexual youth health disparities. *American Journal of Public Health*, *108*(Suppl. 4), S258–S265. https://doi.org/10.2105/AJPH.2018.304407

Gehlbach, H., Brinkworth, M. E., King, A. M., Hsu, L. M., McIntyre, J., & Rogers, T. (2016). Creating birds of similar feathers: Leveraging similarity to improve teacher–student relationships and academic achievement. *Journal of Educational Psychology*, *108*(3), 342–352. https://doi.org/10.1037/edu0000042

Gehlbach, H., & Robinson, C. D. (2018). Mitigating illusory results through preregistration in education. *Journal of Research on Educational Effectiveness*, *11*(2), 296–315. https://doi.org/10.1080/19345747.2017.1387950

Gelman, A., & Loken, E. (2014). The statistical crisis in science: Data-dependent analysis—a "garden of forking paths"—explains why many statistically significant comparisons don't hold up. *American Scientist*, *102*(6), 460–466. https://doi.org/10.1511/2014.111.460

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

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Kirksey, J. J. (2019). Academic harms of missing high school and the accuracy of current policy thresholds: Analysis of preregistered administrative data from a California school district. *AERA Open*, *5*(3). https://doi.org/10.1177/2332858419867692

Kramer, J. W. (2020). Experimental evidence on the effects (or lack thereof) of informational framing during the college transition. *AERA Open*, *6*(1). https://doi.org/10.1177/2332858420908536

Lu, X., Ostrow, K. S., & Heffernan, N. T. (2019). Save your strokes: Chinese handwriting practice makes for ineffective use of instructional time in second language classrooms. *AERA Open*, *5*(4). https://doi.org/10.1177/2332858419890326

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, *43*(6), 304–316. https://doi.org/10.3102/0013189X14545513

Merk, S., & Rosman, T. (2019). Smart but evil? student-teachers' perception of educational researchers' epistemic trustworthiness. *AERA Open*, *5*(3). https://doi.org/10.1177/2332858419868158

Peters, S. J., Rambo-Hernandez, K., Makel, M. C., Matthews, M. S., & Plucker, J. A. (2019). Effect of local norms on racial and ethnic representation in gifted education. *AERA Open*, *5*(2). https://doi.org/10.1177/2332858419848446

Robinson, C. D., Scott, W., & Gottfried, M. A. (2019). Taking it to the next level: A field experiment to improve instructor-student relationships in college. *AERA Open*, *5*(1). https://doi.org/10.1177/2332858419839707

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

Van der Zee, T., & Reich, J. (2018). Open education science. *AERA Open*, *4*(3). https://doi.org/10.1177/2332858418787466

Authors

JUSTIN REICH is an assistant professor of comparative media studies at MIT and the director of the MIT Teaching Systems Lab. As a learning scientist, he studies learning at scale and advocates for open science.

HUNTER GEHLBACH is a professor and vice dean at the Johns Hopkins School of Education as well as the director of research at Panorama Education. An educational psychologist by training and a social psychologist at heart, his interests lie in improving the social and motivational contexts of schools and improving approaches to teaching environmental education.

CASPER J. ALBERS is a professor in applied statistics and data visualisation at the Faculty of Behavioural and Social Sciences of the University of Groningen. He studies psychological dynamic models, and time series, and statistical communication.