Companion Guidelines on Replication & Reproducibility in Education Research

A Supplement to the *Common Guidelines for Education Research and Development*

A Report from
The National Science Foundation
and
The Institute of Education Sciences,
U.S. Department of Education

Nov 28, 2018

INSTITUTE OF EDUCATION SCIENCES

Introduction

The Institute of Education Sciences (IES) and the National Science Foundation (NSF) jointly issued the *Common Guidelines for Education Research and Development* in 2013 to describe "shared understandings of the roles of various types of 'genres' of research in generating evidence about strategies and interventions for increasing student learning" (IES and NSF, 2013: 7). In the intervening period, the education research community and federal policymakers have been increasingly attentive to the role of, and factors that promote and inhibit, replication and reproducibility of research.

In order to build a coherent body of work to inform evidence-based decision making, there is a need to increase the visibility and value of reproducibility and replication studies among education research stakeholders. The purpose of this companion to the *Common Guidelines* is to highlight the importance of these studies and provide crossagency guidance on the steps investigators are encouraged to take to promote corroboration, ensure the integrity of education research, and extend the evidence base. The companion begins with a brief overview of the central role of replication in the advancement of science, including definitions of key terminology for the purpose of establishing a common understanding of the concepts. The companion also addresses the challenges and implications of planning and conducting reproducibility and replication studies within education.

Background and terminology

Efforts to reproduce and replicate research findings are central to the accumulation of scientific knowledge that helps inform evidence-based decision making and policies. Purposeful replications of previous research that corroborate or disconfirm prior results are essential to building a strong, scientific evidence base (Makel and Plucker, 2014). From a policy perspective, replication studies provide critical information about the veracity and robustness of research findings, and can help researchers, practitioners, and policy makers gain a better understanding of what interventions improve (or do not improve) education outcomes, for whom, and under what conditions.

The *Common Guidelines* describe six genres of research: foundational, early-stage or exploratory, design and development, efficacy, effectiveness, and scale-up. The literature around replicability of research has primarily focused on causal impact studies (i.e., the efficacy, effectiveness, and scale-up genres). However, issues of replication are salient in other genres as well. For example, reproducibility and replication are critical for validating and extending early-stage or exploratory work. As the science develops, we may learn more about how issues of reproducibility and replication pertain to other genres of research discussed in the *Common Guidelines* and supported by IES and NSF (e.g., design and development).

Reproducibility refers to the ability to achieve the same findings as another investigator using extant data from a prior study. It has been described as "a minimum necessary condition for a finding to be believable and informative," (Subcommittee on Replicability and Science, 2015: 4).[1] Some reproducibility studies re-analyze data using the same analytic procedures to verify study results or identify errors in the dataset or analytic procedures. Others use different statistical models to see if changes in methods or assumptions lead to similar or different conclusions than the original study.

Multiple types of replications have been identified, and terminology to describe them proposed (e.g., Schmidt, 2009). In general, ***replication*** studies involve collecting and analyzing data to determine if the new studies (in whole or in part) yield the same findings as a previous study. As such, replication sets a somewhat higher bar than

[1] The Subcommittee uses somewhat different terminology in discussing the related issues of replicability and generalizability than employed here.

reproducibility and has been described as “the ultimate standard by which scientific claims are judged” (Peng, 2011: 1226).

> ***Direct replication*** studies seek to replicate findings from a previous study using the same, or as similar as possible, research methods and procedures as a previous study. The goal of direct replication studies is to test whether the results found in the previous study were due to error or chance. This is done by collecting data with a new, but similar, sample and holding all the research methods and procedures constant.
>
> ***Conceptual replication*** studies seek to determine whether similar results are found when certain aspects of a previous study’s method and/or procedures are systematically varied. Aspects of a previous study that may be varied include but are not limited to the population (of students, teachers, and/or schools); the components of an intervention (e.g., adding supportive components, varying emphases among the components, changing the ordering of the components); the implementation of an intervention (e.g., changing the level or type of implementation support, implementing under routine/typical as opposed to ideal conditions); the outcome measures; and the analytic approach.
>
> In efficacy, effectiveness, and scale-up research, the general goal of conceptual replications is to build on prior evidence to better understand for whom and under what conditions an education policy, program, or practice may or may not be effective. The research questions for a conceptual replication study would determine which aspects of the previous study are systematically varied. For instance, if the goal is to determine the generalizability of an intervention’s impacts for a particular group of students, the intervention would be tested with a different population of students, while holding all other aspects of the study the same. In comparison, for early-stage or exploratory research, the goal of a conceptual replication study would be to gather additional information regarding relationships among constructs in education and learning. For example, if the goal were to determine whether findings hold when different assessment tools are employed, data would be collected using different instruments from a prior study but keeping the construct or outcome (and all other methods and procedures) constant.

Reproducing and replicating research in education science

In order to increase the visibility and value of reproducibility and replication studies, several challenges need to be addressed, including disincentives for conducting replications, difficulties implementing such studies, and complexities of interpreting study results. The following are some examples of these challenges.

Disincentives

Despite the importance of replications, there are a number of barriers and challenges to conducting and disseminating replication research, including a real or perceived bias by funding agencies, grant reviewers, and journal editors toward research that is novel, innovative, and groundbreaking (Travers, Cook, Therrien, and Coyne, 2016). In education, as in other research fields, a wide range of factors (e.g., publication bias; reputation and career advancement norms; emphases on novel, potentially transformative lines of inquiry) may dis-incentivize reproducibility and replication studies—or, as Coyne, Cook, and Therrien (2016: 246) suggest, tempt investigators to ‘mask’ or reposition conceptual replications, making it difficult to “systematically accumulate evidence about our interventions.”

Implementation challenges

As an investigator, one of the greatest challenges for replicating education research is the variability inherent in learning contexts (e.g., school-based settings). Indeed, given this variability, it has been argued that direct replications may be exceedingly difficult to conduct in education and the social sciences more generally (e.g., Coyne et al., 2016). Although direct replications may be challenging in education research, they may still be possible depending on the nature of the research questions and the context (e.g., the length of time between the previous study and the replication). Closely aligned conceptual replications (i.e., studies that are not direct replications but are as similar as possible to the original study) can serve a similar purpose and offer a more feasible alternative to direct replications (Coyne et al., 2016).

Interpreting findings

In theory, the ability to reproduce study findings should increase confidence in their veracity. However, reproducibility may mask repeated accidental or systematic errors. Re-analyses that yield identical findings may reflect identical flaws in the execution of the data analysis or other study procedure. On the other hand, when the results of an apparently well-designed and carefully executed study cannot be reproduced, there is a tendency to assume that the initial investigation was somehow flawed, calling into question the credibility of the findings. While this may be the case, scientists working in multiple disciplinary domains have documented a range of factors (e.g., differences in data processing, application of statistical tools, accidental errors by an investigator) that, intentionally or unintentionally, may limit the likelihood that findings will be duplicated when the research is repeated by the same, or separate, researchers (see, e.g., Earp and Trafimow, 2015; McNutt, 2014; Subcommittee on Replicability in Science, 2015). There are also complexities regarding the design and interpretation of replication studies. For instance, although there are various approaches or metrics for judging replication (e.g., requiring that effects are identical, requiring similar effect sizes) there is no consensus on the criteria that should be used to determine whether replication has occurred (Hedges and Schauer, 2018; Subcommittee on Replicability in Science, 2015). There is also the related issue of statistical power for replications and specifically the need for a large number of studies to obtain strong empirical test for replication (Hedges and Schauer, 2018). These challenges underscore that care must be taken in drawing conclusions from re-analyses and replication studies.

Guidelines for the education research community

Given the central role of replication research in the progress of science, it is important that the education field promotes the conduct and dissemination of reproducibility and replication studies. IES and NSF have long-standing commitments to supporting the reproducibility and replication of scientific work. For example, since 2004, IES has included a specific call for grant applications proposing replication studies under its Requests for Applications (e.g., Chhin, Taylor, and Wei, 2018). In addition, IES and NSF support the principles of open science (e.g., preregistration, data sharing, open access to publications) critical to replication and reproducibility. We offer the following guidelines to education stakeholders for thinking about and promoting reproducibility and replication in education research. These guidelines are consistent with, and in some cases, draw heavily from guidelines provided by scientific and professional organizations, advisory committees, and input provided in consultation with the field (see e.g., Cook, Lloyd, Mellor, Nosek, and Therrien, 2018; Coyne et al., 2016; Dettmer, Taylor, and Chhin, 2017; Nosek et al., 2015; Subcommittee on Replicability and Science, 2015). We also highlight the opportunities our agencies provide to support efforts to reproduce and replicate prior investigations and methodological research to inform the conduct of and interpretation of findings from replication studies.[2]

Guidelines for replication studies

Investigators are encouraged to submit proposals to conduct reproducibility and replication studies in response to relevant solicitations, announcements, and requests for applications from IES and NSF. Building on the original

[2] For more detailed information on current funding opportunities, see https://ies.ed.gov/funding/ and https://www.nsf.gov/funding/pgm_list.jsp?org=EHR

(2013) *Common Guidelines*, the following overarching principles for reproducibility and replication research are offered. For more detailed information about how to design, conduct, and interpret reproducibility and replication research see, for example, Coyne et al. (2016), Hedges and Schauer (2018), and Schmidt (2009).

1. Proposals should clarify how the given reproducibility or replication study would build on prior studies and contribute to the development of fundamental knowledge of ways to improve learning and other education outcomes. For example:
    a. For early-stage or exploratory research, proposals should explain how the reproducibility or replication study would contribute to the accumulation of knowledge regarding relationships among important constructs in education and learning and/or establish logical connections that might form the basis for future interventions or strategies to improve those outcomes.
    b. If conducting a replication of an impact study (e.g., efficacy, effectiveness, scale-up), proposals should establish the replication's potential to enhance understanding of the impact of a strategy or intervention under the same (direct replication) or under somewhat changed (conceptual replication) circumstances.

2. Proposals to conduct a conceptual replication should clearly specify the proposed variations from the prior study, along with a rationale for the proposed systematic variations.

3. Proposals for reproducibility or replication studies should ensure objectivity. If the original investigator is involved in the proposed reproducibility or replication study, safeguards need to be included to ensure the objectivity of the findings. At other times (e.g., in re-analysis studies), objectivity may be best accomplished by conducting a separate, independent investigation.

Designing studies with reproducibility and replicability in mind: Transparency and open science

Open science initiatives provide support for investigators seeking to reproduce or replicate a previous study and increase the likelihood that results from replications contribute to the development of theory and the building of a robust evidence base. With increased movement at the federal level toward making scientific research, including data and products, more accessible (e.g., requiring grantees to share data), the education research community should continue to support these efforts in ways that allow analyses and results of studies to be reproduced and replicated. Replication and reproducibility studies are predicated on access to detailed information about another's work (e.g., study designs, sampling plans, instrumentation, analytic methods) and, in the case of reproducibility, another's data. These guidelines are important for researchers performing initial studies as well as those performing replication and reproducibility studies, as a replication study could also serve as an initial study for another researcher.

4. Transparency is a necessary precondition when designing scientifically valid research. For all evaluations (initial and all replications) that test the impact of an intervention (i.e., efficacy, effectiveness, and scale-up), a pre-registration of the proposed research design and methods can help ensure the integrity and transparency of the proposed research.

5. Education research should continue to strive toward open data access policies, the development of commonly agreed upon data sharing guidelines, and the use of publicly available repositories to store data and other materials. In education research, the term data should continue to be defined in the broadest possible terms to include measures, data dictionaries and codebooks, social network analyses, user generated data, outcome data, and analytic models.

6. Analyses should be described in sufficient detail as to allow other researchers to reproduce the results using the same dataset.

7. Researchers should document the features (e.g., population, context, fidelity of implementation) of their study that would be salient to future replications.

8. Researchers should budget resources necessary to engage in the documentation, curation, and sharing activities necessary to facilitate efforts to reproduce and replicate their work.
9. To the extent possible, consent forms and Institutional Review Board (IRB) approvals should reference future public sharing of data and stipulate the conditions that will be put in place to protect the privacy of participants.

10. Researchers should be aware of data management policies across agencies including the Data Management for NSF EHR Directorate Proposals and Awards and the Policy Statement on Public Access to Data Resulting from IES Funded Grants along with the Frequently Asked Questions about Providing Public Access to Data document.

Reporting of research findings

Recognizing that the dissemination and publication stage of research is critically important to the overall goals of replication and reproducibility, the following guidelines are offered.

11. Data used to support claims in publications should be made available in public repositories along with data processing and cleaning methods, relevant statistical analyses, codebooks as well as analytic code.

12. Researchers should analyze and report how the results from their reproducibility or replication study compare to previous studies.

13. Researchers should clearly describe criteria used for exclusion of data or subjects, include results that were omitted for any reason (especially if the results do not support the main findings and/or hypotheses), and describe outcomes or conditions that were measured or used and are for some reason not included in the report.

14. Final reports to funding agencies should include details about how all data and relevant supporting documentation are being made available and can be accessed.

IES- and NSF-funded reproducibility and replication studies

The idea that knowledge advances through progressive iterations of prior work is central to the presentation of the six education research genres originally set out in the 2013 *Common Guidelines for Education Research and Development*. As described there, NSF's and IES's complementary missions are such that NSF focuses relatively more on the first three genres or research types (foundational research, early-stage or exploratory research, and design and development research), while IES "concentrates its investments on developing and testing the effectiveness of well-defined curricula, programs, and practices that could be implemented by schools" (p. 7). Exhibit 1 provides examples of IES and NSF awards with explicit reproducibility and/or replication goals.

Exhibit 1: Examples of IES- and NSF-supported studies with an emphasis on replication &/or reproducibility

A Randomized Control Trial of a Tier 2 Kindergarten Mathematics Intervention

Ben Clarke, Principal Investigator
https://ies.ed.gov/funding/grantsearch/details.asp?ID=1327

This study is an example of a *conceptual replication* that was built in to a larger efficacy project funded under IES's Special Education Research Grants program. The replication study was conducted by the same investigators as the original study. However, objectivity was ensured by using an external entity from the Boston area to collect data and an independent evaluator to conduct statistical analyses. The purpose of the replication study was to test whether the findings from the initial efficacy study (conducted one year prior) of a Tier 2 kindergarten math intervention, *ROOTS*, would replicate when researchers varied three key instructional and contextual elements. Similar to the initial efficacy study, researchers employed a randomized controlled trial where students were either assigned to receive the *ROOTS* Tier 2 program in addition to Tier 1 core math instruction (intervention) or to receive Tier 1 core instruction only (comparison condition). The intervention, population of students, outcome measures, and analyses were all the same as the initial investigation.

Researchers systematically varied the following aspects of the replication study: 1) the geographic region, 2) the timing of intervention onset, and 3) the instruction provided in the comparison condition. First, the original study took place in rural and suburban schools in Oregon whereas the replication took place in urban and suburban schools in Massachusetts. Researchers varied the setting to determine whether the effects held up for students in schools with different sociodemographic characteristics (e.g., more racial/ethnic diversity and a higher percentage of students from low-income backgrounds). Second, in the replication study, the intervention began approximately two months earlier in the year than it did in the initial efficacy study. Researchers varied the timing to determine whether earlier intervention onset led to stronger results for at-risk kindergarteners. Third, relative to the initial efficacy study, the comparison condition in the replication included math programs with stronger evidence for improving students' math achievement. As such, the replication provided a more stringent test of the efficacy of *ROOTS*.

Findings from the replication study showed significant positive effects of *ROOTS* on proximal and distal measures of math achievement. Effects on a researcher-developed measure of early numeracy skills, a standardized measure of whole number understanding (*Test of Early Mathematics Ability-Third Edition)*, and a curriculum-based measure of early numeracy proficiency were replicated in the conceptual replication study. Both the initial and replication studies found effects in the same direction and at similar levels of statistical significance and effect sizes fell within or exceeded the upper bound of those reported in the initial efficacy study. Unlike the initial efficacy study, the replication did not find statistically significant positive impacts of the intervention on a measure of oral counting. Yet, the replication study showed significant positive impacts two distal measures of math achievement (*Number Sense Brief Screen* and *Stanford Early School Achievement Test*), which were not observed in the initial efficacy study.

Selected Publications:

Clarke, B., Doabler, C. T., Smolkowski, K., Kurtz Nelson, E., Fien, H., Baker, S. K., & Kosty, D. (2016). Testing the immediate and long-term efficacy of a Tier 2 kindergarten mathematics intervention. *Journal of Research on Educational Effectiveness*, *9*(4), 607634.

Doabler, C. T., Clarke, B., Kosty, D. B., Kurtz-Nelson, E., Fien, H., Smolkowski, K., & Baker, S. K. (2016). Testing the efficacy of a tier 2 mathematics intervention: A conceptual replication study. *Exceptional Children, 83*(1), 92-110.

Scaling Up the Implementation of a Pre-Kindergarten Mathematics Curricula: Teaching for Understanding with Trajectories and Technologies

Douglas Clements, Principal Investigator
https://nsf.gov/awardsearch/showAward?AWD_ID=0228440

This study is an example of a *conceptual replication*. The investigators sought to replicate and scale-up a previously developed Pre-K mathematics intervention, *Building Blocks*, with additional supports for implementation.

The original study was conducted with 68 preschool children and initial results indicated that the combined strategies of the *Building Blocks* curriculum resulted in significant mathematical learning gains in favor of the experimental group (effect size = .85). The replication involved implementing the program in 25 Head Start and State Preschool classrooms in diverse locations of California and New York. This replication included support for teachers, technical and pedagogical coaching during implementation, and materials and active roles for parents and administrators. The researchers systematically varied the student population being served and the geographic location of the study. The researchers were interested to learn if and how *Building Blocks* was effective for a diverse group of students most at risk for poor performance in mathematics and when the program was implemented on a larger scale.

In the scaling-up replication study, the team conducted a randomized field trial design and implemented *Building Blocks* along with enhanced supports and tools for implementation. The replication design involved classrooms serving children at risk for later school failure and the team examined the impact of the program on mathematics learning across two domains: number and geometry (*Building Blocks Assessment of Early Mathematics*). The study also included measures of fidelity and classroom observations. Implementing the program with high levels of fidelity in the intervention settings resulted in significantly higher mean scores compared to control and substantially greater gains in children's mathematics achievement in the intervention group compared to the control (effect size = .62). Given the similarity in the observed effect sizes and the statistical significance in favor of treatment across the two studies, the results from this conceptual replication supported findings from the initial study.

Selected Publications:

Sarama, J., & Clements, D. H. (2004). Building blocks for early childhood mathematics. *Early Childhood Research Quarterly*, *19*(1), 181-189.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks project. *Journal for Research in Mathematics Education, 38*(2), 136-163.

Sarama, J., Clements, D. H., Starkey, P., Klein, A., & Wakeley, A. (2008). Scaling up the implementation of a pre-kindergarten mathematics curriculum: Teaching for understanding with trajectories and technologies. *Journal of Research on Educational Effectiveness*, *1*(2), 89-119.

Project Early Reading Intervention

Deborah Simmons, Principal Investigator
https://ies.ed.gov/funding/grantsearch/details.asp?ID=370

This study is an example of a *conceptual replication* that was built in to a larger efficacy project funded under IES's Special Education Research Grants program. The purpose of the replication study was to evaluate whether the findings from the initial efficacy study (conducted one year prior) of the *Early Reading Intervention* (*ERI*), a supplemental kindergarten reading program, would generalize to a different geographical location and under different instructional conditions. Similar to the initial efficacy study, researchers employed a randomized controlled trial where students were either assigned to receive *ERI* (intervention condition) or to receive the school's core reading instruction (comparison condition). The design, measures, methods, and procedures utilized in the replication study were similar to those employed in the initial efficacy study. One potential limitation was that there was overlap in the investigators who conducted the initial study and the replication study.

The replication differed from the initial efficacy study in terms of the geographic region (the replication was conducted in Florida and the initial study in Connecticut and Texas) and the instructional context. More specifically, the original study took place in school districts where the core reading instruction was less coordinated and as such, varied within and across classrooms and schools. For instance, most schools used a combination of commercial reading programs and less structured reading instruction and did not provide supplemental reading intervention to kindergarteners. Because the goal of the replication was to determine if intervention impacts would replicate in schools with a different instructional context, the replication was conducted in a school district in Florida characterized by more coordinated and consistent policies and practices around core reading instruction and intervention (e.g., teachers routinely received professional development related to evidence-based reading strategies, students at-risk for reading difficulties received supplemental reading intervention). Thus, the replication provided a more stringent test of the efficacy of ERI than the original trial.

Unlike the findings from the initial efficacy study, results from the replication study showed no statistically significant impacts of ERI compared to core reading instruction on any of the reading outcome measures. Results of the initial efficacy trial showed that students who received ERI significantly outperformed those who received core reading instruction on foundational alphabetic, phonemic, and untimed decoding skills. Additional analyses indicated that intervention students in the replication study responded similarly to the intervention relative to intervention students in the original study, but that there were statistically significant differences in reading outcomes among students in the comparison condition in the replication study versus the original study. Although both groups of comparison students showed similar levels of achievement on reading measures at pre-test, comparison students in the replication study significantly outperformed comparison students in the initial study on a variety of reading measures (i.e., phonemic awareness, letter sound knowledge, nonsense word fluency, and word identification) at post-test. Thus, researchers concluded that the differences in findings across the initial and replication studies were largely due to the differences in the reading instruction provided in the comparison condition and students' response to that instruction.

Selected Publications:

Coyne, M. D., Little, M. E., Rawlinson, D. M., Simmons, D. C., Kwok, O., Kim, M., Simmons, L. E., Hagan-Burke, S., & Civetelli, C. (2013). Replicating the impact of a supplemental beginning reading intervention: The role of instructional context. *Journal of Research on Educational Effectiveness, 6*(1), 1-23.

Simmons, D. C., Coyne, M. D., Hagan-Burke, S., Kwok, O., Simmons, L., Johnson, C., Zou, Y., Taylor, A. B., McAlenney, A. L., Ruby, M., & Crevecoeur, Y. C. (2011). Effects of supplemental reading interventions in authentic contexts: A comparison of kindergarteners' response. *Exceptional Children*, *77*(2), 207-228.

References

Chhin, Christina S., Taylor, Katherine A., and Wei, Wendy S. (2018). Supporting a culture of replication: An examination of education and special education research grants funded by the Institute of Education Sciences. *Educational Researcher*. Advanced online publication. doi: 10.3102/0013189X18788047

Cook, Bryan G., Lloyd, John W., Mellor, David, Nosek, Brian A., and Therrien, William J. (2018). Promoting open science to increase the trustworthiness of evidence in special education. *Exceptional Children*. Advance online publication. doi: 10.1177/0014402918793138

Coyne, Michael D, Cook, Brian G., and Therrien, William J. (2016). Recommendations for replication research in special education: A framework of systematic, conceptual replications. *Remedial and Special Education, 37*(4): 244-253.

Dettmer, Amanda M., Taylor, Katherine A., and Chhin, Christina S. (2017). Responses to the Request for Public Comment on IES-Funded Efficacy, Replication, and Effectiveness Studies. Retrieved from https://ies.ed.gov/whatsnew/pdf/Response_to_Public_Comment_on_Goals_3_and_4.pdf.

Earp, Brian D., and Trafimow, David. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, *6*: 1-11.

Hedges, Larry V., and Schauer, Jacob M. (2018). Statistical analyses for studying replication: Meta-analytic perspectives. *Psychological Methods*. Advanced online publication. doi: 10.1037/met0000189

Institute of Education Sciences (IES), U.S. Department of Education. Resources for Researchers: Frequently Asked Questions About Providing Public Access To Data. Retrieved from https://ies.ed.gov/funding/datasharing_faq.asp.

Institute of Education Sciences (IES), U.S. Department of Education. Resources for Researchers: Policy Statement on Public Access to Data Resulting from IES Funded Grants. Retrieved from https://ies.ed.gov/funding/datasharing_policy.asp.

Institute of Education Sciences (IES), U.S. Department of Education and National Science Foundation (NSF). (2013). *Common Guidelines for Education Research and Development*. Retrieved from IES website: https://ies.ed.gov/pdf/CommonGuidelines.pdf. Also available as NSF 13-126 from NSF website: https://www.nsf.gov/pubs/2013/nsf13126/nsf13126.pdf.

Makel, Matthew C., and Plucker, Jonathan A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, *43*(6): 304-316.

McNutt, Marcia. (2014). Reproducibility. *Science*, *343*(6168): 229. Retrieved December 2017 from http://science.sciencemag.org/content/sci/343/6168/229.full.pdf.

National Science Foundation. (2017a). Data Management for NSF EHR Directorate Proposals and Awards. Alexandria, VA: National Science Foundation. Retrieved October 2017 from https://www.nsf.gov/bfa/dias/policy/dmpdocs/ehr.pdf.

Nosek, Brian A., Alter, George, Banks, George C., Borsboom, Denny, Bowman, Sara D., Breckler, Steven J., ... and Yarkoni, Tal (2015). Promoting an open research culture. *Science*, *348*(6242), 1422-1425.

Peng, Roger D. (2011). Reproducible research in computational science. *Science, 334*(6060): 1226-1227.

Schmidt, Stefan. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2): 90-100.

Subcommittee on Replicability in Science. Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic sciences. (2015). *Social Behavioral and Economic Science Perspectives on Robust and Reliable Science*. Arlington, Virginia: National Science Foundation. Retrieved December 2017 from https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf.

Travers, Jason C., Cook, Bryan G., Therrien, William J., and Coyne, Michael D. (2016). Replication research and special education. *Remedial and Special Education*, *37*(4): 195-204.