Routledge
Taylor & Francis Group

Educational Research and Evaluation

An International Journal on Theory and Practice

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/nere20

A decade of replication study in education? A mapping review (2011–2020)

Thomas Perry, Rebecca Morris & Rosanna Lea

To cite this article: Thomas Perry, Rebecca Morris & Rosanna Lea (2022): A decade of replication study in education? A mapping review (2011–2020), Educational Research and Evaluation, DOI: 10.1080/13803611.2021.2022315

To link to this article: https://doi.org/10.1080/13803611.2021.2022315

Published online: 31 Jan 2022.

EDUCATIONAL RESEARCH AND EVALUATION
https://doi.org/10.1080/13803611.2021.2022315

Routledge
Taylor & Francis Group

A decade of replication study in education? A mapping review (2011–2020)

Thomas Perry [a], Rebecca Morris [a] and Rosanna Lea [b]

[a]Department of Education Studies, University of Warwick, Coventry, UK; [b]Department of Education, University of Oxford, Oxford, UK

ABSTRACT

Replication studies in education are relatively rare. Of the few which are conducted, many are conceptual rather than direct replications. With so few replication studies, and many of those that are attempted producing null results, the scientific status of the evidence base for educational policy and practice is in question. Replicating Makel and Plucker's review of the education replication literature, conducted in 2014, this paper presents a mapping review looking at rates of replication in education research from 2011 to 2020. We provide an overview of the number of replication studies by replication type, year, outcome, authorship, and journal. Our results are consistent with those of Makel and Plucker, revealing very low but gradually increasing rates of replication study in education. We discuss the role of replication in producing a robust and trustworthy evidence base for policy and practice, and some of the challenges in operationalising definitions of replication we encountered.

KEYWORDS

Replication; mapping review; education; education science

Replications and education science

Researchers championing a more rigorous, scientific approach to education research report "dramatic" improvements in the last 2 decades, stemming from increased capacity in the education research community and demand from government-funded organisations focused on developing a rigorous policy- and practice-relevant evidence base (Hedges, 2018). Such accounts of education science often identify education as an area that has had particularly weak methodological standards relative to other disciplines and fields within social science and beyond (e.g., medical science), and difficulties in establishing a trustworthy knowledge base (Hedges, 2018). In this context, debates about scientific methodology and practice are highly pertinent for those concerned with the application of scientific approaches in education. Replication is at

CONTACT Thomas Perry tom.perry@warwick.ac.uk

the heart of these debates and, in particular, with regard to questions about the status of the knowledge that science produces in terms of its trustworthiness, applicability, rigour, and potential to be cumulative. In short, scientific knowledge ought to be verifiable, and replication study is the most direct way to test and ascertain whether that is the case. It is therefore of concern that replication study is relatively rare in education research (Makel & Plucker, 2014), and indications from other disciplines and fields (as discussed below) suggest that – if key results *were* tested – many would not replicate.

Perhaps more than any other issue, the so-called "replication crisis" has surfaced fundamental questions about science and the status of the evidence base (Marsden et al., 2018). High-profile papers from Ioannidis (2005a, 2005b) identified serious challenges in replicating studies in biomedical research, calling into question not just specific results, but current scientific practice more generally. Concerns about the unreliability of findings and calls for greater focus on replication have spread across disciplines and fields (Makel et al., 2019) including psychology (Open Science Collaboration, 2015; Pashler & Harris, 2012), political science (Key, 2016), and economics (Duvendack et al., 2017). These debates are also continuing across all areas within and connected to the field of education, for example, in educational psychology (Plucker & Makel, 2021), mathematics education (Aguilar, 2020), special education (Cook, 2014; Travers et al., 2016), engineering education (Benson & Borrego, 2015), chemistry education (Cooper, 2018), and second-language acquisition (Marsden et al., 2018).

Common to all of these discussions is the idea that replication is a vital part of the scientific process and for the development of a trustworthy evidence-base. We might expect that a healthy scientific field would be able to replicate important findings, both in principle and in practice, that is, with an appreciable proportion of studies being replications, a culture conducive to replication study, and with key results having been successfully replicated. In addition to verifying or identifying suspect substantive results in the field, replication can be utilised to advance understanding of methodological practice. Recent work, for example, has identified characteristics of studies that are more likely to successfully replicate (Altmejd et al., 2019) and that peers (Camerer et al., 2018), and even laypeople (Hoogeveen et al., 2020), can predict which studies will and will not successfully replicate. Replication is also valuable for identifying "hidden moderators", with replication failure often identifying new hypotheses and systematic barriers, as well as supporting researchers to understand "for whom, and in what contexts, evidence-based interventions work best" (Kim, 2019, p. 605).

These debates sit within broader discussions of scientific methodology (including meta-scientific research) and research practice. Of particular note is the role of replication within wider accounts of scientific best practice and "Open Science", which include study preregistration and registration, open

materials and data, and transparency (e.g., van Dijk et al., 2021). Such reforms, according to Makel et al. (2019), "offer education research a path toward increased credibility and social impact" (p. 1). There are methodological, organisational, and cultural factors at play within these conceptions of effective science which stress the importance of effective funding bodies (Hedges, 2018) and the value and organisation of collaboration (Makel et al., 2019) as much as methodology. A credible prevailing description of effective open science is that it requires "the combination of large-scale collaboration with open and transparent research practices [to offer] education researchers opportunity to test theories, verify what is known about a topic, resolve disagreements, and explore new questions" (Makel et al., 2019, p. 1). This all suggests that verification and replication are key components of current understanding of the social scientific method. This perspective has been closely tied to quantitative and experimental approaches to education research. Replication, however, can be framed more generally as a principle for sound research, with authors arguing that the focus should fall on whether methods are transparent and replicable, and findings can be reproduced, irrespective of the research tradition (Leppink, 2017; Plucker & Makel, 2021). With a broad conception of replication that includes conceptual and approximate replication (see below), replication in some form arguably becomes applicable and of potential value across all areas of social research and for all knowledge claims.

In this vein, this paper presents a mapping review of the frequency of replications in education research; moreover, this study is an approximate replication of Makel and Plucker's study (2014), which examined the extent of replication in education research up to and including 2012. Our study examines a period of 1 decade, from 2011 to 2020, therefore overlapping with Makel and Plucker's data by 2 years, and updating their findings with the latest data. Before presenting this review, we outline definitions of replication in the literature that inform our methods and discussion of replication rates, and consider other reviews in this area.

Definitions of replication

In this study we use the categories of "direct", "approximate", and "conceptual" replication. These are defined and discussed here, and then used within the coding scheme for the mapping review. Common definitions distinguish replication in terms of how closely the replication study adhered to the focus, procedure, and context of the original study. With the greatest extent of adherence to the original study is the "direct" replication (sometimes synonymously referred to as "exact" replication elsewhere). It is possible to interpret this in a strict sense, rendering direct replication as something only possible in highly controlled conditions and "virtually unfeasible" outside of that (Aguilar, 2020, p. 40). Replication tends to be described as "direct" when it has

successfully used the same procedure and instruments, and a highly similar participant sample (Crandall & Sherman, 2016).

Many replication studies can be described as "close", or "approximate". While diverse, and with considerable overlap with "conceptual" replications (see below), these studies largely replicate the original study, but modify one or more substantive aspects, such as sample size or characteristics. We will refer to these as "approximate" replications. Johnson and Nicodemus (2016), for example, employed identical materials and procedures to Johnson et al. (2012), but used a more varied sample, and dropped one of the tasks that had not been feasible due to time constraints in the earlier study.

Approximate replications are similar to, but can be distinguished from, our third category of replication: "conceptual replication". While direct and approximate replication is focused on similarity of methods and procedure, conceptual replication is more focused on the original study's findings and theory:

> A conceptual replication is an attempt to test the same fundamental idea or hypothesis behind the original study, but the operationalizations of the phenomenon, the independent and dependent variables, the type and design of the study, and the participant population may all differ substantially. (Crandall & Sherman, 2016, p. 93)

As explored in (Morrison, in press, in the present Special Issue), supporters of conceptual replication not only hold that such variation is acceptable but, moreover, that planned variation in study samples can be a powerful approach to understanding whether a finding is generalisable and the conditions in which this is so. Conceptual variation can identify whether the original finding holds fast across variations in research design, methods, context, settings, sample, population, instrumentation, measures, elements of the intervention, procedures, duration, timing, and so forth (Morrison, in press). One interesting example from our own mapping review database (as reported below) is provided by Root et al. (2020), who describe how a series of 16 connected research studies had replicated, developed, and refined mathematics interventions for learners with special educational needs. They describe a sequence of replication studies that varied participants, settings, materials, intervention personnel, and task difficulty; they discuss how these had developed understanding of how, for whom, and in what conditions the interventions were successful.

Our inspection of the replication methods literature (e.g., Schmidt, 2009) and the present mapping review analysis suggest that use of the terms described above is not yet entirely consistent, both in terms of the specific terms that are preferred and how these are operationalised and understood within any given study. In addition to those defined above, there are a range of terms associated with replication in use including: "literal", "operational", "constructive", "partial", "concrete", "systematic", and "close". Conceptual replication can be challenging to define, and there is large variation in which and how many aspects of the original study are altered (Chhin et al., 2018; Morrison, in

press). Furthermore, there is the issue that many studies might be considered to be a replication, but are not described by the authors as being so (Chhin et al., 2018; Coyne et al., 2016). This presents a serious problem and limitation for studies that seek to identify the extent of replication study in education research, and do so – as in both the present study and Makel and Plucker (2014) – using a search strategy that assumes that replication studies will include "replicat*" in the text (see discussion below of Cook et al., 2016). Within the studies that were located, applying the distinctions above (between direct, approximate, and conceptual) also proved challenging, with many studies that did not clearly self-identify as a given type of replication, nor were easy to categorise due to a lack of clear-cut definitional categories. In our Methods section, we discuss the boundary lines between our definitions and the main issues with operationalising these. We also note that similar issues are evident in previous reviews in this area, to which we turn below. Moreover, in the Discussion section we return to the question of whether replication is more prevalent than suggested by our data due to studies not using the term, or preferring alternatives such as "follow-on" (Schmidt, 2009), or including replication elements within an extension study or as a scale-up study.

Previous reviews in this area

We start by briefly describing Makel and Plucker (2014), the study we set out to (approximately) replicate. We provide greater detail about Makel and Plucker's methods in our own Methods section, and concentrate here on the findings and significance of the study. Makel and Plucker analysed the complete publication history of the top-100 education journals (as ranked by the 5-year impact factor in January 2013) to identify replication studies in education. They found very low rates of replication, at 0.13% of published articles, but a small increase in the proportion of replications over a period of several recent decades. They also found that success rates appeared higher in education than those previously reported in medicine. The success rates of education were more closely aligned to psychology, where the majority of studies were successfully replicated. They also drew attention to the relationship between authorship and success: with replications being considerably less likely to have found the same result when done by unique authors (54%) compared to the original authors (70.6% success for studies with at least one author the same in a new publication and up to 88.7% success for the same authors in the same publication). Alongside their results, Makel and Plucker discussed the value and challenges of doing replication study and presented an argument in favour of replication in education science, along the lines of the discussion presented earlier in this paper.

In later work, Makel et al. (2016) conducted a similar survey of the prevalence and success rate of replication specifically in the area of special education. Using similar methods, they estimated that 0.52% of all articles self-reported as being

a replication of previously published findings, and that over 80% of these replications were successful, meaning that they obtained the same substantive result as the original study. Again, they found that, when the original study and the replication study shared at least one author in common, the rate of successful replication was substantially higher. Remarkably, Lemons et al. (2016) conducted an "inadvertent concurrent replication" of Makel et al. (2016) for the same special edition of the *Remedial and Special Education* journal (focused on replication research). That is to say that both Lemons et al. and Makel et al. (2016) independently, concurrently, and unbeknownst to each other, conducted a review of replication rates in special education journals. Both are published in that special edition and provide a useful comparison. Lemons et al. estimated the overall replication rate at 0.41% of all published articles, a figure higher than the 0.52% from Makel et al. (2016). They also found the same relationship between authorship and replication success. However, there were also marked differences between the papers that became apparent through Lemons et al. obtaining and comparing the two datasets. Relative to Makel et al. (2016), Lemons and colleagues were considerably more likely to consider a study a conceptual replication (90.0% vs. 51.5%), and to describe the outcome a mixed success (27.1% vs. 12.7%). Agreement on the author overlap was 95.8%; agreement on the type of replication between the two studies was 58.3%, and agreement on the success rate category was 66.7%. More starkly, the rate of agreement on article inclusion was only 15.2%; that is, Lemons et al. identified 70 replication studies, Makel et al. (2016) identified 109 studies, and the overlap between them was only 24 studies. There were several reasons behind this: First, the differences in how journals were indexed in different databases meant that different studies were obtained in searches; second, the inclusion or not of studies where the replication occurred within the same manuscript (included by Makel et al., 2016, but not by Lemons et al., 2016); and, third, the exclusion of studies for more-to-less vague references to which previous findings or studies were being replicated. Of the 24 replications included in both reviews, all had the term "replicat*" located in the article abstract.

Other reviews in this area that we located include Marsden et al. (2018), who estimated the mean rate of replication in second language (L2) research as being 1 in 400 articles (0.25%) and found no direct replications in the field. Again, they also found that overlap of authorship was linked with higher rates of successful replication. Furthermore, the availability of the initial study materials was conducive to successful replication. Another example is that by Chhin et al. (2018), who examined the number of replications in efficacy and effectiveness grants funded by the US Institute of Education Sciences (IES). They found that, between 2006 and 2016, no direct replications were funded, but almost half of the grant applications funded "can be considered conceptual replications that vary one or more dimensions of a prior study" (p. 594), even

though the majority did not explicitly state their intention as conducting a replication. This is consistent with the point made above: Many studies are likely to be replications without being described as such by the authors. We located one study that used a range of synonyms for replicate in the assessment process: Cook et al. (2016) examined 83 intervention studies in special education in 2013 and 2014. Of the 83 intervention studies, 53% ($n = 44$) indicated that the purpose of the study was to "*replicate, expand, extend, add to, or use as their basis* one or more previous studies" (p. 228, italics added), and 73% ($n = 61$) included a comparison of the findings of previously conducted research. Cook et al. categorised 31% ($n = 26$) as replications, since they met both of these criteria. All of these 26 were considered conceptual replications; many of which had numerous "parent" studies. No direct replications were found. The final example of a replication review reported here is by Banerjee et al. (2018). They provide another example of replication being used to study replication, as their aim was to "replicate the methods used by Cook and colleagues and conduct a systematic review of the quantity, type, and findings of replication research in early childhood special education (ECSE)" (p. 246). Banerjee et al. located 78 intervention studies in three ECSE journals from 2012 to 2016. Using the same approach as Cook et al., they found that 56% ($n = 44$) were replications, a figure considerably higher than Cook et al.'s estimate of 31%. Studies such as those by Chhin et al., Cook et al., and Banerjee et al. suggest that calculating rates of replication using studies that self-identify as replications might be a gross under-estimate. On the other hand, many studies that self-identify as replication, and those identified by researchers actively looking for replications (using substantive criteria), may not be so if stricter definitions were used. This is especially problematic when it comes to the definitionally challenging and complex concept of the conceptual replication.

Overall, the literature reviewed above suggests that replication study is widely considered to have a significant role in scientific research approaches, and is advocated as a useful approach for verifying and improving the methodological practice and the evidence base in an area. Reviews suggest, however, that rates of replication study are low and the rate of direct replication is very low indeed. We identified a workable set of definitions, but also outlined inconsistency in their use and application, with both substantive and methodological studies identifying challenges both locating potential replications and reliably categorising them as such. The category of conceptual replication appears to be a particular definitional "grey area" situated between overt, purposeful attempts to reproduce a previous finding, and studies that fall within the topic area and aim to extend previous results but have little methodological overlap with original studies. We have these challenges in view as we progress to the methods of this approximate replication study, where, as well as detailing our methods, we discuss some of the boundary issues when identifying and categorising replication studies.

Methods

This study is an approximate replication of Makel and Plucker (2014). It is a mapping review in the sense that our aim is to provide an overview of an area of the research literature (replications in education) in terms of its size and key features (i.e., type, authorship, place of publication, result). The replication is "approximate" as we are deviating from their reported methods in several respects detailed below, but fundamentally adhering to the aims and approach and providing comparable results. Our replication was an "approximate", rather than a "direct" one, for several reasons. First, we chose to survey all education research, rather than the top-100 journals. Second, we additionally coded for "approximate replications", effectively separating conceptual and approximate replications, with the former deviating more substantially from the procedure (but not the concept) of the original study. Third, we did not calculate citation rates or gain ratios within our mapping of the replication literature. Fourth, we chose to include within-study replications (i.e., multi-study papers). Finally, and perhaps most significantly, we focused on a more recent time period. Details of these are provided below. These deviations resurface the definitional questions discussed earlier in the paper. In particular, one might argue that the use of a different time period renders our piece a conceptual replication, as it seeks to test out a finding in a new "population" (i.e., education research from 2011–2020) and ascertain if the original result generalises over time; it is also possible to consider this study an extension study rather than a replication study at all. Our view is that we are seeking to replicate the original results of Makel and Plucker, that is, that rates of replication in education are low, gradually increasing, and that the success of the replication is associated with author and publication location. We describe this as an approximate replication because of the deviations from the original study's methods, discussed above, but not a conceptual one, as our primary intention was to test whether the original results remain the case, as opposed to deepening our understanding of whether these hold true in a different context or set of circumstances.

Searching

Makel and Plucker (2014) used the top-100 education journals (using the 5-year impact factors on the Web of Knowledge Journal Citation Reports data) for their study database. In contrast, we used the Web of Science "Education" category. We are of the view that limiting the education research field to a specific number of journals, and the potential for changes in the "panel" membership between 2012 and 2019, meant that the use of the database category was more comprehensive, collecting data from the entire "population" of education research. Our database contains replication studies from 154 different journals. From this point, our study was much the same as that of Makel and Plucker.

In February 2021, we searched for all publications in the Web of Science's Science Citation Index Expanded (SCI-EXPANDED) and Social Sciences Citation Index (SSCI), for the entire years of 2011 through to 2020 (10 years in total). From this general search, we located 226,044 results. We also recorded the number of articles published by year to use in our replication rate calculations. Then, we conducted a search for "replicat*" in all fields within the education category (ALL = replicat* AND WC = educ*). This search yielded 1,919 results, already less than 1% of total education research publications. These results were located in the following categories (with overlap): Education/Educational Research (894), Psychology Educational (591), Education Special (392), and Education Scientific Disciplines (298). We exported these records into our database, from which we removed a small number of duplicates and records with a missing year that, when we located this information, were found to be published in 2021 (we believe that the Web of Science records were for the preprints). Our final search database included 1,892 records.

Inclusion and coding

All 1,892 records including the term "replicat*" were analysed. A team of three researchers completed this work in two screening stages. In the first stage, studies were split between the three researchers who inspected and organised all allocated studies into one of four groups:

- ***Replication*** – studies identified at Stage 1 as being replications.
- ***Not replication*** – studies were only placed in the "not replications" group where the researcher was confident that this was the case;
- ***Second opinion*** ($n = 134$) – any study that was thought to be on the borderline of being a replication or not was placed for consideration in a separate group;
- ***Not education*** – several studies, despite their inclusion in the education category, were considered not to be education research, with only a slight or unclear education aspect;
- ***Literature review*** – any study that was not a replication, but was identified as being valuable to the present study's literature review was placed in a separate folder (e.g., reviews of replication as reported above). To these we added 16 further studies known to the authors, bringing the total in this database group to 69 records.

With regard to what we defined as a replication, we followed the rule that even if a study self-defined as a replication, we only categorised it as such if it clearly identified the specific study they were attempting to replicate. Studies that aimed to replicate a general phenomenon, or results demonstrated from a set of studies without providing details on which aspects of the methodology

were being replicated (e.g., "We attempted to replicate the effect of x on y, as described by … "), were not coded as a replication. We did not include research instrument validation studies (e.g., replicating the psychometric properties of a psychological measurement instrument); we recognise the importance of this practice but preferred to focus on replications of substantive results in a topic area, and were concerned that many studies would self-describe as "validating" the instrument rather than "replicating" previous results. We *did* include instances of replication studies within the same manuscript. We did not have any criteria relating to methods, so in principle we could have located qualitative replication pieces (see discussion above). While it was evident that there were very few qualitative replication studies in our database, we did not code for methods and thus have not formally reported this observation in our results.

After this first round of screening, the team discussed the issues and questions arising from the records marked as needing a "second opinion". We discuss these below. Following this, all records in the second-opinion group were categorised based on this discussion by a single researcher (the lead author) as either being replications or not. Of the 134 second-opinion papers, 44 were categorised as not being replications, and 90 were categorised as replications and added to the main folder. During the second screening process and coding of the replication studies, six duplicates were identified, and 88 more papers were identified on further inspection as not being either in education or replications, and were therefore excluded. In total we identified 442 replication studies, which were all coded using the coding scheme below. An overview of the screening process and exclusion decisions for both screening stages is provided in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009) diagram shown in Figure 1.

Coding

This study coded all replication studies in the final database on the following variables:

- **Type of replication** – direct, approximate, conceptual, or mixed (see below);
- **Result** – success (i.e., original result supported); failure (i.e., original result not supported); mixed (i.e., more than one knowledge claim, where at least one was a success, and at least one was a failure); unclear;
- **Authorship** – same (i.e., overlap of at least one author between replication and original study) or different;
- **Journal** – same or different;
- **Same publication replication** – yes or no. This identified studies where there were several studies reported in a single manuscript, of which one or more later studies replicated one or more earlier studies in the same piece. They were by definition within the same author and journal.

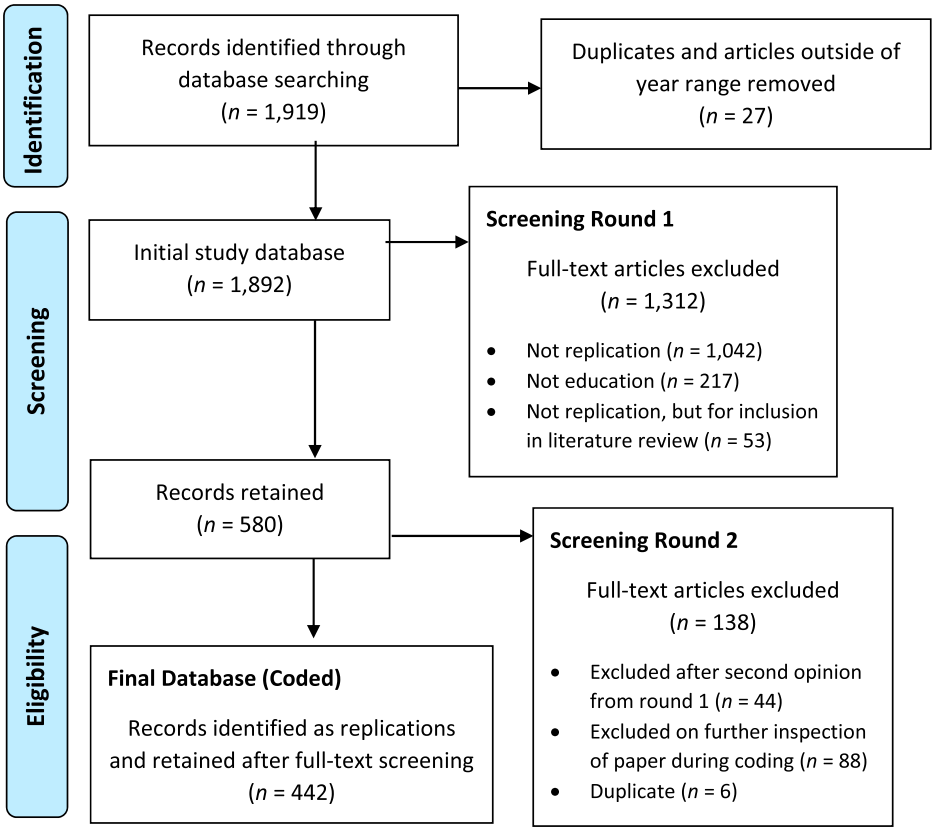

Figure 1. PRISMA flow diagram of eligibility screening process.

We based the determination of "success" or "failure" of the replication on author self-reports. This was partially a practical decision, as our time constraints did not allow analysis of study results based on our own criteria; this also reflected our perspective that what counts as a success will be topic- and outcome-specific and cannot be reduced to general, objective criteria. The small number of studies in the "mixed" category were multi-part studies that conducted more than one type of replication, usually describing their aims as "replicating and extending" a previous study. To determine whether studies were approximate, conceptual, or direct replications, we applied the definitions set out in the initial literature review above. As was discussed earlier, there are not common and precise definitions, and there are potential issues with reliability – particularly in relation to identifying approximate and conceptual replications. To explore this further, we conducted reliability testing on a sample of the database. We selected a random sample of 10% of the overall database (190/1,892) using a random number generator. This subset was then split equally three-ways between the researchers; the papers were then screened and coded for a second time using the same procedure as described above. The original and second set of coding results were then brought back

together to assess the level of agreement. The level of agreement when identifying replications was fairly high, with 94.7% of the sample of 190 having the same result. Thirty-five studies were identified as being a replication study. For these, there was a moderate 77.8% level of agreement for the type and a high level of agreement of 91.1% for the result, 95.6% for the author, 100% for the journal, and 93.3% for the study being an in-study replication.

Unlike Makel and Plucker (2014), we did not calculate citation rates for replication studies. One other deviation between our study and that of Makel and Plucker was that we did not calculate gain ratios to communicate changes in replication rates. These did not feature prominently in the paper and, in our view, made the findings less, rather than more, accessible to readers. Finally, we conducted a brief review of the papers that were flagged as needing a second opinion, and discussed challenges of coding replication types and other items within the team. We recorded key challenges and discussed what this revealed in the short section below. We will make available on request our database for any researchers wishing to examine these decisions, inspect the database, or conduct any other further analysis.

Definitional boundaries

As described above, the research team identified studies that were at the borderline of our categories, both for replication type and with regard to whether or not the study was deemed to be a replication. In total, 134 studies were not clear-cut and required further consideration. Ninety of these were ultimately deemed to be replications, and 44 not. There were also challenges when it came to the second step of categorising the replications as direct, approximate, conceptual, and mixed, which we discussed as a team during the coding process.

The main challenges with this coding were as follows: first, the question of preplanning and specification. There were authors that had used the abstract or discussion section to comment that their results "replicated" previous named studies or findings without the intention being made explicit in the methods (Sakkal & Martin, 2019). Where studies are not pre-registered, or the intention to replicate is not evident in the methods, replication findings are open to issues with post hoc analysis, sometimes referred to as HARKing, or hypothesising after the results are known (Kerr, 1998). Similarly, there were numerous studies where the results or studies being replicated were not clear, where authors discussed numerous and/or general findings from across the literature and then stated that they would seek to replicate these. In many cases, authors referred to numerous studies, coupled with the aim of replicating the findings, or a known effect from "previous research" in "the literature" (e.g., Flores et al., 2014, pp. 77, 75, respectively) or "in previous works" (e.g., Liew & Tan, 2016, p. 104); there were also cases within these where the present authors felt that the finding to be tested was specified too generally to constitute a productive

and challenging test, for example, seeking to replicate a "positive" result or general relationship. There was a tendency for authors to state that findings replicated the original study, so long as their effect was also statistically significant, or if a relationship between variables was in the same direction. While the value of different levels of granularity in findings is field- and question-specific, we had expected to see greater consideration of the similarity of effect sizes.

Another definitional issue related to sub-analyses, especially in the context of replications within a single paper. There were instances of multiple analyses of the same data, or subgroup analysis of these being described as replication (Amemiya & Wang, 2018; Froiland & Worrell, 2016); we viewed use of a separate dataset to test the same result to be a replication, but not the use of the same dataset with sub-analysis. Similarly, within-publication replications were not always clear-cut in terms of whether the purportedly separate studies were replications or merely arms of a single study. There was one study using an ABAB design, that described the second two conditions as replicating the first two (Collins et al., 2016).

Results

Replication rates in education research (2011–2020)

In total, we identified 442 replication studies in education from 2011 to 2020. We now compare the proportions of replications in different coding categories to the equivalent figures from Makel and Plucker (2014), which latter are given as the second figure in each set of parentheses. In our database of 442 replication studies, 145 (32.8% vs. 28.5%) were direct replications, 200 (45.3% vs. 69.3%) were conceptual replications, 82 (18.6%) were approximate replications (this code was not used in Makel & Plucker, 2014, but was closest to conceptual replication, see above), and 15 (3.4% vs. 2.3%) were mixtures of these types. Most were considered a success (i.e., the replication produced the same result as the original study), with 299 (67.7% vs. 67.4%) judged as successful; 41 (9.3% vs. 13.1%) did not provide the same substantive result (replication "failures"), 95 (21.5% vs. 19.4%) had a mixed result, and for seven (1.6%) studies the result was unclear (not coded by Makel & Plucker, 2014). We found that 60.4% of studies had the "same author" (i.e., an overlap of at least one member of the author team for both the original and replication studies), and 36.7% were published in the same journal (not including within-publication replications); 26.0% of the replication studies were "within-publication" replications, where more than one study was reported in a single paper. In Table 1 we provide figures for the number of replications and number of direct replications (a subset of all replications) by year. The figures immediately above are not comparable with Makel and Plucker's, due to our decision to include within-publication replications.

Table 1 shows an increasing number of total publications in education research across the period of time used. There is also a gradual increase in the number of replications and in the number of direct replications, although the trend is less clear in the latter. Makel and Plucker (2014) estimated an overall rate of replication, across entire histories of journals of 0.13%; with a gradual increasing trend, this figure was in excess of 0.2% in 2010–2012 (specific figures are not provided by year). This means that our overall replication rate is broadly consistent but lower than that in Makel and Plucker. The replication rate is very low, at around 0.20%, or 1 in every 500 publications overall. It has gradually increased in the last decade, from 0.13% in 2011 to 0.27% in 2020. Rates of direct replication are lower still, at around a third of this level, with slightly more volatility. These results are visualised in Figure 2.

Replication characteristics and success

Tables 2 and 3 provide the success rate of replication studies against, first, the replication type and, second, its authorship and publication location, respectively. Table 2 shows that success rates are between 59.8% and 73.5% across replication types, with reasonable consistency with the overall figure of 68.1% and Makel and Plucker's (2014) figure of 67.6%. There were no other marked differences by replication type, perhaps other than a suggestion that approximate and mixed replications are more likely to fail. Table 3 reveals more notable differences and, in particular, a very low rate (2.9%) of replication failure for within-publication replications (i.e., multiple studies within a single paper) compared to replications by different authors in different journals (18.5% failure rate).

Replication rates by journal

Finally, we look at rates of replication by journal. As detailed above, we have included any journal publishing education research. In this paper we report

Table 1. Replication frequency and proportion in education research by year.

Publication Year	Total publications	Replication studies		Direct Replications	
	Freq.	Freq.	%	Freq.	%
2011	18,534	24	0.13	8	0.04
2012	19,814	27	0.14	8	0.04
2013	20,110	29	0.14	8	0.04
2014	20,009	39	0.19	13	0.06
2015	20,189	43	0.21	7	0.03
2016	21,870	43	0.20	13	0.06
2017	22,285	43	0.19	17	0.08
2018	23,542	52	0.22	23	0.10
2019	29,355	61	0.21	21	0.07
2020	30,336	81	0.27	27	0.09
Total	226,044	442	0.20	145	0.06

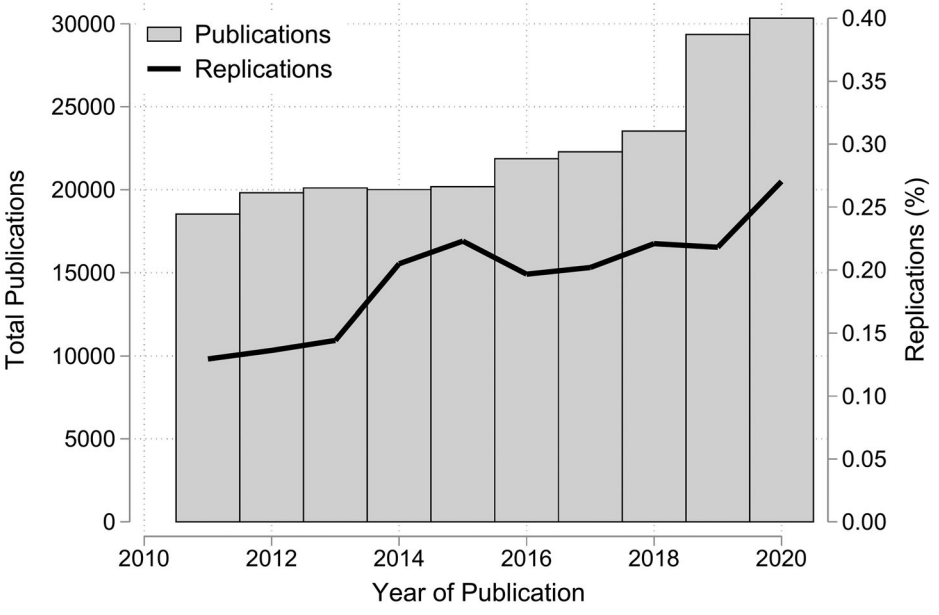

Figure 2. Replication frequency and proportion in education research by year.

all journals with two or more replication studies. In Table 4 we report all journals with two or more replication studies of any type, and with a rate of 0.6% or higher (i.e., replication as a proportion of total journal publications). In Appendix 1 we also indicate journals with two or more replications, but with a rate lower than 0.6%. In total, there are eight journals with a replication rate above 2% (i.e., 1 in 50 papers are replication studies). A further 23 have a rate above 1% and up to 2%.

Discussion

Summary of results

In this study we conducted an "approximate" replication of Makel and Plucker's (2014) paper "Facts Are More Important Than Novelty: Replication in the Education Sciences". Despite this study being only an "approximate" replication, we judge our study to be addressing the same substantive question

Table 2. Replication results by type (% to 1 decimal place).

	Replication Result									
Replication Type	*Failure*		*Mixed*		*Success*		*Unclear*		*Total*	
Direct	13	9.0%	35	24.1%	94	64.8%	3	2.1%	145	100.0%
Approximate	12	14.6%	18	22.0%	49	59.8%	3	3.7%	82	100.0%
Conceptual	14	7.0%	38	19.0%	147	73.5%	1	0.5%	200	100.0%
Mixed	2	13.3%	4	26.7%	9	60.0%	0	0.0%	15	100.0%
Total	41	9.3%	95	21.5%	299	67.6%	7	1.6%	442	100.0%

Table 3. Replication results by authorship and location (% to 1 decimal place).

	Replication Result									
Authorship and publication	Failure		Mixed		Success		Unclear		Total	
Within-publication replication	3	2.9%	16	15.2%	85	81.0%	1	1.0%	105	100.0%
Same journal and author	2	6.1%	6	18.2%	24	72.7%	1	3.0%	33	100.0%
Same journal, Different author	2	8.3%	8	33.3%	13	54.2%	1	4.2%	24	100.0%
Different journal, Same author	6	4.7%	26	20.2%	97	75.2%	0	0.0%	129	100.0%
Different journal and author	28	18.5%	39	25.8%	80	53.0%	4	2.6%	151	100.0%
Total	41	9.3%	95	21.5%	299	67.6%	7	1.6%	442	100.0%

and with sufficiently similar methods to make a direct comparison of the results. With regard to the findings, we judge the replication to be a "success": Like Makel and Plucker, we find very low rates of replication overall (at approximately 0.20% on average across the 2011–2020 period). We also found a small increasing trend, from 0.13% to 0.27% during the decade. It could however be argued that our result is substantially *higher* than that of the original study, with an average rate of 0.20% compared to 0.13% in the original study; it could also be argued to be substantially *lower* because, during the overlapping period in the early 2010s, we estimate a rate of 0.13% to 0.14% while Makel and Plucker estimate a rate in excess of 0.20%. In our view, however, both results have the same substantive significance and interpretation: that rates of replication in education research are very low, albeit slowly increasing. We do not venture a figure for what the ideal rate of replication should be. In our view, all major results (i.e., those that are foundational to subsequent research efforts, or have significant policy or practical implications) would have been replicated several times, and ideally many times, in proportion to the significance of the result. The arithmetic of this places the rate of replication as being in an order of magnitude higher than the rate of seminal studies.

There were other similarities with the original study: The estimated success rate was similar (67.6% vs. 67.4%), and – if we recombine our "approximate" and "conceptual" replication categories – we are in agreement with Makel and Plucker (2014) that approximately two thirds of replications are "conceptual/approximate" (as opposed to "direct"), that is, they vary one or more substantial elements of the original study. Moreover, we find the same relationship between authorship and success, with low rates of replication failure for within-publication replications (2.9%), for replications by the same author in the same journal (in a separate publication; 6.1%), and for the same author in a different journal (4.7%), as compared to replications by different authors in different journals, where 18.5% did not agree with the original result. These results were also consonant with other similar reviews of the replication literature. Other similar reviews find that rates of replication study are very low and the rate of direct replication is even lower.

Table 4. Replication frequency and proportion by journal (2011–2020) (>1 replication and >0.6% rate).

Source Title	Total publications	Replications	Replication Rate
Journal of Educational Psychology	794	27	3.40%
Journal of Behavioral Education	208	6	2.88%
Exceptional Children	308	8	2.60%
Language Learning	480	12	2.50%
Journal of Early Intervention	204	5	2.45%
Learning and Instruction	635	14	2.20%
Instructional Science	409	9	2.20%
Journal of Research on Educational Effectiveness	289	6	2.08%
Language Teaching	310	6	1.94%
Teaching of Psychology	586	11	1.88%
Journal of Special Education Technology	108	2	1.85%
Learning Disabilities Research & Practice	217	4	1.84%
Behavioral disorders	224	4	1.79%
Topics in Early Childhood Special Education	234	4	1.71%
Journal of Learning Disabilities	488	8	1.64%
Metacognition and Learning	187	3	1.60%
Journal of Positive Behavior Interventions	253	4	1.58%
Focus on Autism and Other Developmental Disabilities	255	4	1.57%
Education and Treatment of Children	275	4	1.45%
Reading and Writing	895	13	1.45%
Educational Psychology Review	428	6	1.40%
Journal of Mathematics Teacher Education	147	2	1.36%
International Review of Applied Linguistics in Language Teaching	153	2	1.31%
Contemporary Education Psychology	536	7	1.31%
Journal of Special Education	233	3	1.29%
School Psychology Review	321	4	1.25%
Learning and Individual Differences	1,420	17	1.20%
Journal of Computer Assisted Learning	601	7	1.16%
Education and Training in Autism and Developmental Disabilities	373	4	1.07%
Elementary School Journal	280	3	1.07%
Journal for Research in Mathematics Education	286	3	1.05%
School Psychology Quarterly	302	3	0.99%
School Effectiveness and School Improvement	303	3	0.99%
Sociology of Education	202	2	0.99%
Journal of Research in Reading	330	3	0.91%
Journal of the Learning Sciences	222	2	0.90%
Zeitschrift für Pädagogische Psychologie	222	2	0.90%
School Mental Health	335	3	0.90%
AERA Open	340	3	0.88%
Perspectives on Medical Education	234	2	0.85%
Discourse Processes	374	3	0.80%
Language Teaching Research	517	4	0.77%
Psychology in the Schools	912	7	0.77%
Journal of Experimental Education	394	3	0.76%
Educational Researcher	531	4	0.75%
Educational Administration Quarterly	268	2	0.75%
Educational Evaluation and Policy Analysis	286	2	0.70%
Journal of Developmental and Physical Disabilities	572	4	0.70%
Reading & Writing Quarterly	288	2	0.69%
Educational Technology Research and Development	734	5	0.68%
Journal of School Psychology	448	3	0.67%
Mind, Brain, and Education	304	2	0.66%
Journal of Educational Measurement	310	2	0.65%
Social Psychology of Education	465	3	0.65%

Defining replication

In our literature review, we identified a workable set of definitions that has been established in the literature, but also outlined inconsistency in their use and application, with both substantive and methodological studies identifying challenges in both locating and reliably categorising replications. In our Methods section, we discussed issues surrounding the lack of precision of specification about what (in terms of studies or results) is being replicated, and lack of transparency around pre-planning of comparisons linking results back to the literature (i.e., post hoc determination of study implications). As well as being characteristics commonly advocated for qualifying research as "scientific", these issues made categorisation of papers as replications challenging in this study.

In our Previous reviews in this area section, we identified profound issues concerning the reliability of locating and categorising replication studies. One particular problem, which we consider further here, is the question of whether a paper that does not self-identify as a replication can be considered to be one. Framing this question more broadly, one might wonder whether replication rates are considerably higher than reported here, but did not use the term "replicat*". Strictly, the literature we have mapped represents a group of studies aligning themselves with a scientific discourse and paradigm geared towards the cumulation of a particular type of evidence. We doubt however, that our results have missed a substantial body of de facto replications; and we hold that greater clarity, transparency, pre-planning, and pre-registration, around which studies or results that an original study builds on, is needed, whether or not the language around replication is used. On the other hand, we might argue that most studies within a topic area perform a replication function to a degree: Results that are profoundly mistaken are likely to be identified through research into adjacent topics even in the absence of deliberate replication efforts.

Replications may also be present in the form of interventions that "scale up" intervention studies, by rolling out a programme across a larger or more diverse population. For example, funders such as the IES, in the US, and the Education Endowment Foundation (EEF) in England, can fund projects at multiple stages (e.g., Pilot Study, Efficacy Trial, Effectiveness Trial, Scale-Up; EEF, 2021). Interventions at later stages are delivered to more schools, across more regions, in more ecologically valid ways (i.e., in realistic classroom and implementation conditions). Presumably, the intervention programmes retain their original methods and theory throughout, and thus trials at each stage could potentially be described as "approximate" or "conceptual" replications of each other. However, such research does not typically self-describe as "replication" and is not always published in academic journals, so is likely to have been missed in "replicat*" searches. Linking this back to the short discussion of the ideal rate

of replication above, factors influencing the value and optimal rate of replication include the significance of the original result, including the consequences of error, and the extent to which alternative corroborating evidence from related research might perform the same function with the robustness and precision deemed necessary.

Replication and the fundamentals of education science

As we discussed in the opening section, the practice of, and discourse around, replication study connects with wider debates about prevailing practice and best practice in scientific research. We close this paper with a short discussion of fundamental ideas within this discussion and how they link to replication. There are several principles at play within common accounts of "science". First is the assumption that knowledge is or can be cumulative. Through adherence to the scientific method, research should build a cumulative knowledge base that will hold true to a significant degree over time and context. Second is the principle of scepticism, and the need to put the knowledge base to an empirical test (i.e., subject it to potential falsification). The UK's Royal Society's motto *Nullius in verba* (take nobody's word for it) invokes the spirit of seeking verification rather than accepting results on authority, plausibility, or acceptability. Combining these two points, the scientific paradigm maintains that the scientific method can distinguish truth from falsehood, and that the knowledge it produces was both established through scientific tests and – in principle *and* in practice – can be replicated through the same means. Framed in this way, replication is placed at the heart of science, and, as long as we maintain that a single result is insufficient to establish a claim as knowledge, both a way of creating a knowledge base and verifying it. Moreover, as touched on earlier, and discussed at greater length in Morrison (in press), if conceptual replication is a tool for teasing out the "who", "what", "why", "how", and "in what circumstances" something holds (i.e., generalising a practice or knowledge claim), conceptual replication positions replication as a way of extending as well as verifying knowledge.

In our view, at the heart of the matter is this balance between verification and extension of knowledge. If we lack confidence in a result or it rests on too few studies, it ought to be replicated; if verification is unsuccessful, the replication study frequently surfaces variables of interest that might lead the way to refining the theory (Kim, 2019); if verification is successful, then it is through conceptual replication and systematic, critical variation that knowledge can be extended *while retaining coherence with the original result*. Studies that prematurely seek novelty and originality render much of the educational research knowledge base disparate and fail to identify (sometimes long-standing) erroneous results and misconceptions in our current understanding. To the extent to which this is a valid way of framing the problem and the role of replication

in addressing it, a rate of replication in the vicinity of 1 in 500 studies is concerning indeed.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Thomas Perry is an assistant professor in the Department of Education Studies at University of Warwick. His work is focused on the use of research and evidence to improve education. He has expertise in research synthesis and review; quantitative methods and secondary data analysis; and evaluation, enquiry, and improvement.

Rebecca Morris is an assistant professor in the Department of Education Studies at University of Warwick. Her research focuses on issues related to social justice and equity in education, teaching, and learning in schools and higher education, and the teacher workforce.

Rosanna Lea is a postdoctoral researcher in the Department of Education at the University of Oxford. Her research interests include the psychology of education, the relationship between emotion (in particular, stress) and cognition, and the application of cognitive science to teaching and learning.

ORCID

Thomas Perry https://orcid.org/0000-0002-6124-467X
Rebecca Morris https://orcid.org/0000-0002-1699-4172
Rosanna Lea https://orcid.org/0000-0002-2309-0948

References

Aguilar, M. S. (2020). Replication studies in mathematics education: What kind of questions would be productive to explore? *International Journal of Science and Mathematics Education*, *18*(S1), 37–50. https://doi.org/10.1007/s10763-020-10069-7

Altmejd, A., Dreber, A., Forsell, E., Huber, J., Imai, T., Johannesson, M., Kirchler, M., Nave, G., & Camerer, C. (2019). Predicting the replicability of social science lab experiments. *PLoS One*, *14*(12), Article e0225826. https://doi.org/10.1371/journal.pone.0225826

Amemiya, J., & Wang, M.-T. (2018). African American adolescents' gender and perceived school climate moderate how academic coping relates to achievement. *Journal of School Psychology*, *69*, 127–142. https://doi.org/10.1016/j.jsp.2018.05.001

Banerjee, R., Movahedazarhouligh, S., Millen, K., & Luckner, J. L. (2018). Building evidence in early childhood special education: A systematic review of replication intervention studies. *Topics in Early Childhood Special Education*, *37*(4), 246–256. https://doi.org/10.1177/0271121417748197

Benson, L., & Borrego, M. (2015). The role of replication in engineering education research. *Journal of Engineering Education*, *104*(4), 388–392. https://doi.org/10.1002/jee.20082

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., … Wu, H. (2018). Evaluating the replicability

of social science experiments in *Nature* and *Science* between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Chhin, C. S., Taylor, K. A., & Wei, W. 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*, *47*(9), 594–605. https://doi.org/10.3102/0013189×18788047

Collins, T. A., Cook, C. R., Dart, E. H., Socie, D. G., Renshaw, T. L., & Long, A. C. (2016). Improving classroom engagement among high school students with disruptive behavior: Evaluation of the class pass intervention. *Psychology in the Schools*, *53*(2), 204–219. https://doi.org/10.1002/pits.21893

Cook, B. G. (2014). A call for examining replication and bias in special education research. *Remedial and Special Education*, *35*(4), 233–246. https://doi.org/10.1177/0741932514528995

Cook, B. G., Collins, L. W., Cook, S. C., & Cook, L. (2016). A replication by any other name: A systematic review of replicative intervention studies. *Remedial and Special Education*, *37*(4), 223–234. https://doi.org/10.1177/0741932516637198

Cooper, M. M. (2018). The replication crisis and chemistry education research [Editorial]. *Journal of Chemical Education*, *95*(1), 1–2. https://doi.org/10.1021/acs.jchemed.7b00907

Coyne, M. D., Cook, B. G., & Therrien, W. J. (2016). Recommendations for replication research in special education: A framework of systematic, conceptual replications. *Remedial and Special Education*, *37*(4), 244–253. https://doi.org/10.1177/0741932516648463

Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, *66*, 93–99. https://doi.org/10.1016/j.jesp.2015.10.002

Duvendack, M., Palmer-Jones, R., & Reed, W. R. (2017). What is meant by "replication" and why does it encounter resistance in economics? *American Economic Review*, *107*(5), 46–51. https://doi.org/10.1257/aer.p20171031

Education Endowment Foundation. (2021). Pipeline of EEF trials. *Evaluation Information for Grantees*. https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/process-and-people/pipeline-of-eef-trials

Flores, M. M., Hinton, V., & Strozier, S. D. (2014). Teaching subtraction and multiplication with regrouping using the concrete-representational-abstract sequence and strategic instruction model. *Learning Disabilities Research & Practice*, *29*(2), 75–88. https://doi.org/10.1111/ldrp.12032

Froiland, J. M., & Worrell, F. C. (2016). Intrinsic motivation, learning goals, engagement, and achievement in a diverse high school. *Psychology in the Schools*, *53*(3), 321–336. https://doi.org/10.1002/pits.21901

Hedges, L. V. (2018). Challenges in building usable knowledge in education. *Journal of Research on Educational Effectiveness*, *11*(1), 1–21. https://doi.org/10.1080/19345747.2017.1375583

Hoogeveen, S., Sarafoglou, A., & Wagenmakers, E.-J. (2020). Laypeople can predict which social-science studies will be replicated successfully. *Advances in Methods and Practices in Psychological Science*, *3*(3), 267–285. https://doi.org/10.1177/2515245920919667

Ioannidis, J. P. A. (2005a). Contradicted and initially stronger effects in highly cited clinical research. *JAMA*, *294*(2), 218–228. https://doi.org/10.1001/jama.294.2.218

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

Johnson, M. D., Mercado, L., & Acevedo, A. (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing*, *21*(3), 264–282. https://doi.org/10.1016/j.jslw.2012.05.011

Johnson, M. D., & Nicodemus, C. L. (2016). Testing a threshold: An approximate replication of Johnson, Mercado & Acevedo 2012. *Language Teaching*, *49*(2), 251–274. https://doi.org/10.1017/S0261444815000087

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

Key, E. M. (2016). How are we doing? Data access and replication in political science. *PS: Political Science & Politics*, *49*(2), 268–272. https://doi.org/10.1017/S1049096516000184

Kim, J. S. (2019). Making every study count: Learning from replication failure to improve intervention research. *Educational Researcher*, *48*(9), 599–607. https://doi.org/10.3102/0013189×19891428

Lemons, C. J., King, S. A., Davidson, K. A., Berryessa, T. L., Gajjar, S. A., & Sacks, L. H. (2016). An inadvertent concurrent replication: Same roadmap, different journey. *Remedial and Special Education*, *37*(4), 213–222. https://doi.org/10.1177/0741932516631116

Leppink, J. (2017). Revisiting the quantitative–qualitative-mixed methods labels: Research questions, developments, and the need for replication. *Journal of Taibah University Medical Sciences*, *12*(2), 97–101. https://doi.org/10.1016/j.jtumed.2016.11.008

Liew, T. W., & Tan, S.-M. (2016). The effects of positive and negative mood on cognition and motivation in multimedia learning environment. *Journal of Educational Technology & Society*, *19*(2), 104–115. https://www.jstor.org/stable/jeductechsoci.19.2.104

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/10.3102/0013189×14545513

Makel, M. C., Plucker, J. A., Freeman, J., Lombardi, A., Simonsen, B., & Coyne, M. (2016). Replication of special education research: Necessary but far too rare. *Remedial and Special Education*, *37*(4), 205–212. https://doi.org/10.1177/0741932516646083

Makel, M. C., Smith, K. N., McBee, M. T., Peters, S. J., & Miller, E. M. (2019). A path to greater credibility: Large-scale collaborative education research. *AERA Open*, *5*(4). https://doi.org/10.1177/2332858419891963

Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning*, *68*(2), 321–391. https://doi.org/10.1111/lang.12286

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097

Morrison, K. (in press). Conceptual replications, research, and the "what works" agenda in education. *Educational Research and Evaluation* (this issue).

Open Science Collaboration. (2015, August 28). Estimating the reproducibility of psychological science. *Science*, *349*(6251), Article aac4716. https://doi.org/10.1126/science.aac4716

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, *7*(6), 531–536. https://doi.org/10.1177/1745691612463401

Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist*, *56*(2), 90–100. https://doi.org/10.1080/00461520.2021.1895796

Root, J. R., Jimenez, B. A., Saunders, A. F., & Stanger, C. (2020). Replication research to support mathematical learning of students with extensive support needs. *Exceptionality*, *28*(2), 109–120. https://doi.org/10.1080/09362835.2020.1743708

Sakkal, A., & Martin, L. (2019). Learning to rock: The role of prior experience and explicit instruction on learning and transfer in a music videogame. *Computers & Education*, *128*, 389–397. https://doi.org/10.1016/j.compedu.2018.10.007

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90–100. https://doi.org/10.1037/a0015108

Travers, J. C., Cook, B. G., Therrien, W. J., & Coyne, M. D. (2016). Replication research and special education. *Remedial and Special Education*, *37*(4), 195–204. https://doi.org/10.1177/0741932516648462

van Dijk, W., Schatschneider, C., & Hart, S. A. (2021). Open science in education sciences. *Journal of Learn Disabilities*, *54*(2), 139–152. https://doi.org/10.1177/0022219420945267

Appendix 1. Replication frequency and proportion by journal (2011–2020) (>1 replication and <=0.60% rate)

Source Title	Total publications	Replications	Replication Rate
Reading Research Quarterly	336	2	0.60%
Journal of Creative Behavior	345	2	0.58%
Journal of Educational Research	551	3	0.54%
Remedial and Special Education	368	2	0.54%
Journal of Educational Computing Research	559	3	0.54%
Training and Education in Professional Psychology	378	2	0.53%
Child Development	1,744	9	0.52%
Journal of Higher Education	405	2	0.49%
Journal of Counseling Psychology	621	3	0.48%
Research in Higher Education	428	2	0.47%
Economics of Education Review	888	4	0.45%
Early Education and Development	679	3	0.44%
Studies in Educational Evaluation	453	2	0.44%
European Journal of Teacher Education	467	2	0.43%
British Journal of Educational Psychology	476	2	0.42%
Physical Review Physics Education Research	485	2	0.41%
Research in Autism Spectrum Disorders	1,302	5	0.38%
Assessment & Evaluation in Higher Education	878	3	0.34%
TESOL Quarterly	614	2	0.33%
Computers & Education	2,170	7	0.32%
System	1,245	4	0.32%
Journal of Psychoeducational Assessment	710	2	0.28%
Educational Psychology	718	2	0.28%
Health Education Journal	743	2	0.27%
Research in Developmental Disabilities	2,659	7	0.26%
CBE—Life Sciences Education	781	2	0.26%
Early Childhood Research Quarterly	819	2	0.24%
Journal of College Student Development	819	2	0.24%
The Modern Language Journal	880	2	0.23%
Interactive Learning Environments	992	2	0.20%
Higher Education	1,316	2	0.15%
Teaching and Teacher Education	1,618	2	0.12%
Medical Teacher	3,073	2	0.07%

Note: This table is continued from Table 4.