Technology, Knowledge and Learning
https://doi.org/10.1007/s10758-021-09532-3

ORIGINAL RESEARCH

A Framework for Classifying Replication Studies in Educational Technologies Research

Rhonda Christensen[1] · **Charles B. Hodges**[2] · **J. Michael Spector**[1]

Accepted: 20 May 2021

Abstract
Replicating research studies is considered one way of establishing validity and confidence of findings in a field of study. In this paper, we introduce a replication framework for classifying studies conducted in the area of educational technology as a possible guide to conducting and reporting replication studies in the field. The paper includes the benefits as well as challenges of replicating research, and proposes a categorical continuum that might be used to determine the strength of the replication of a study. Examples of replication studies and how they fit the framework are included. Implications for using this framework for conducting studies that are worthy of replication in the field of educational technology are addressed in this paper.

Keywords Education research · Educational technology · Instructional technology · Replication study · Research · Research terminology

1 Introduction

While replication may be seen as the heart of any science, it is associated with the assumption that scientific findings follow laws of nature (Schmidt, 2009). However, from a theoretical point of view, there are no *laws* of nature in the social sciences. Rather, that are patterns of behavior and likely outcomes of interventions. Nevertheless, replication studies have recently been the focus of several scholarly discussions. The Open Science Collaboration (2015) publication reported that experimental findings in 100 psychology studies failed to be duplicated when the experiments were replicated; this has led to scholars in many education disciplines commenting on the notion of *replication studies.* For example,

✉ Rhonda Christensen
Rhonda.christensen@unt.edu

Charles B. Hodges
chodges@georgiasouthern.edu

J. Michael Spector
Mike.spector@unt.edu

[1] University of North Texas, 3940 Elm Street #G-150, Denton, TX 76207, USA

[2] College of Education, Georgia Southern University, Campus, Box 8131, Statesboro, GA 30458, USA

the topic was addressed in engineering education (Benson & Borrego, 2015), music education (Morrison, 2016), and educational technology (Hodges, 2015; Spector et al., 2015). If research results cannot be reproduced, the interpretation and application of research may result in an wrongheaded confidence in their impact, resulting in confusion or inappropriate decisions in learning design. In education, reliance on uncertain research results can be confusing for decision makers, administrators, and practitioners in school environments (Matthews et al., 2018). For a discipline to be considered scientific, it needs to establish stable findings and then build upon those findings. Replication studies are aimed primarily at establishing confidence in findings, or identifying nuances, limitations or problems with prior research.

Currently, few replication studies occur in the field of education (Makel & Plucker, 2014; Plucker & Makel, 2021) and specifically in educational technology where it has been noted that there is a "singular lack of replication studies" (Dron, 2021, p. 1). Thus, there is a need for the replication of research studies in education. Specifically, in the context of educational technologies, why should replication studies be conducted? There is often a faddish approach and overt advocacy for specific technology solutions that tend to detract from the science of learning and instruction. Many meta-analyses related to educational technology studies have been published to allow cross-study comparisons to lead to stronger conclusions and recommendations for improving related studies (e.g. Garzon & Acevedo, 2019; Li et al., 2019). However, the individual studies used in meta-analyses do not typically have commonalities such as shared theoretical frameworks, research designs, implementations, instrumentations, participants, settings, analysis methods or reporting of findings and do not include the same type of comparison that is possible with a replication study (van Alten et al., 2019). In addition to what can be learned from meta-analyses, replication studies can add credibility to educational technology research in a different way with the use of standard validated instruments, strong links to prior research, and relevant theoretical grounding. The point of these types of focused studies is to contribute what the professional and scholarly communities know about using technology to improve learning and instruction. Knowing that a technology worked in one situation, but did not work in another is a contribution; knowing what works or does not work with different groups of learners also is a contribution. Without replication studies that include the sharing of instruments and protocols, there is the risk of having only a series of one-off studies that provide little insight or understanding that can be used systematically to improve learning and instruction at scale. This paper proposes to offer a finer level of detail for labeling replication studies that will allow researchers to classify their work in a more descriptive way and provide a path for researchers to consider as they plan replication studies.

2 Background

Makel and Plucker (2014) defined replication as "the purposeful repetition of previous results to corroborate or disconfirm the previous results" (p. 305). A review of the literature over several years revealed that replication studies in the field of psychology occurred at a rate of 1.07% (Makel et al., 2012). The study included a review of 100 psychology journals with the highest 5-year impact factors. A further study, focused on an analysis of the top 100 education journals ranked by 5-year impact factor, found that only 0.13% of the articles were replications (Makel & Plucker, 2014).

In a large endeavor to replicate research findings in the field of psychology, the project titled "Reproducibility Project: Psychology", suggested that only 30 of the 100 studies reviewed could be reproduced. The researchers in the project created seven categories for how closely the findings reproduced the original study. These seven categories ranged from slightly similar to virtually identical. Of the 39 that might have the ability to be replicated, only 24 were considered "moderately similar" to those of the original experiment. The replication teams followed a highly structured protocol in which criteria included methodology, statistical tests, and sample sizes (Open Science Collaboration, 2012).

Other researchers in the field of psychology have collaborated on the replication of 100 experimental and correlational studies from three psychology journals (Open Science Collaboration 2015). The purpose was to recreate the original research findings to establish reproducibility of the findings with new data. However as a whole, their replications produced weaker results than were reported by the original authors. Researchers attempting to conduct replication studies have noted there is no single indicator for the evaluation of replication success (Schmidt, 2009; Open Science Collaboration 2015).

Plucker and Makel (2021) recently commented "explicit replications are rare in many fields, including education and psychology" (p. 90). They go on to advocate for replication studies as a way to develop confidence in findings and the credibility of fields of study. The authors posit many benefits of replication including providing a more solid foundation for continued research, efficiency for researchers themselves, and increasing opportuties for external funding specifically targeting replication studies.

By contrast, fields such as chemistry and medicine commonly replicate studies in a quest to create credibility in the findings. Specifically, the International Committee for Medical Journal Editors requires that researchers must register their study of a clinical trial prior to the enrollment of patients if the researchers plan to submit their studies to the journals represented by the organization. The purpose is to have findings reported even if they turn out to reflect unfavorably on a sponsored project (DeAngelis et al. 2004). Lynch et al., (2015) noted that "replications are a valuable contribution to science and that there is no personal affront when another scholar reports an 'unsuccessful' replication of one's earlier findings" (p. 340).

The 7th edition of the American Psychological Association Publication (APA) Manual (2020) specifically identifies "replication articles" (p. 6) as a type of research article. Terminology for types of replication are recommended, but are qualified with a statement that "other labels for or descriptions of replication" (p. 7) may be used. APA recommends authors "provide comparisons between the original study and the replication being reported so readers can evaluate the degree to which there may be factors present that would contribute to any differences between the findings of the original study and the findings of the replication being reported" (p. 92). The present article goes beyond the recommendations in the APA manual, listing terminology and details specific to the field of educational technology.

The purpose of this paper is to establish a need for replication studies to be conducted in the field of educational technology with the knowledge that some types of replication are not possible due to multiple and difficult to control variables in authentic educational settings. This paper presents a framework that aims to distinguish three types of possible replication-type research environments that range from very stringent, containing all the elements of the initial study, to more feasible studies that contain some of the most crucial elements necessary for repeating a research design and study. This paper also proposes a vocabulary that can be used to describe and delineate various/between replication studies. The benefits as well as challenges of replicating research will be addressed, followed by a

proposed categorical continuum that might be used to determine the strength of the replication of a study.

3 The Use of Replication Studies

3.1 Benefits of Replication

With regard to new and emerging technologies, it is imperative to have a series of studies, including replication studies, so as to minimize the hype curve problem and build steadily toward deep understanding (Sherer & Teo, 2019). Replication in educational research is useful to examine validity issues, methodology and improve generalizability (Morrison et al., 2010). Multiple iterations of the study of an intervention in different contexts can lead to a more reliable summary of results. For example, Bond (2020) found that with regard to flipped learning, a "majority of research has been undertaken in North American and Asian high schools, heavily focused on student perceptions of flipped learning and achievement within STEM subjects" (p. 1). Bond's observation about a lack of diversity in the learning context is an example of how replication is needed to better understand where the flipped learning design does and does not work well.

Replications also test the boundaries of an intervention (Matthews et al., 2018). According to Mackey and Gass (2005), "replication is a central part of any field of enquiry" (p.21). Some researchers have gone as far to say that it is irresponsible to neglect replication research (Smith, 1970). A recent study focused on special education teachers who are tasked with understanding and implementing evidence-based practices indicating an important need for replication studies to provide more confidence in implementation of an intervention (Matthews et al., 2018). Replication studies are important for many reasons including:

- Creating confidence in research findings (generalizability, decision-making, etc.)
- Incremental refinement to overcome limitations
- Building cumulative knowledge by connecting existing and new knowledge

In a listing of core values intended to guide scientific research, Casti (1989) included the requirement that scientific research must include the ability to be replicated by other researchers.

Empirical studies, especially in the area of educational technology, become dated such that the findings may need to be replicated by updating the type of technology studied or the measurement instruments used for the study or even the context in which the study is to be conducted. For example, a survey instrument for a study ten years ago may have included "learning with computers" or "how many hours do you spend on a computer?" Most students today use their smartphones or tablets for the same purpose that computers were used ten years ago indicating that survey items might need to be updated to reflect these types of changes.

Replication, as noted by Plucker and Makel (2021), can strengthen the theory on which scholarship is grounded and allow for better understandings for theory building moving forward. Hew et al. (2019) observed that theory advancement in educational technology is lacking. More replication, specifically in educational technology, could aid in much needed

theory building and expand the generalizability of findings. However, there are several factors that make replication studies a challenge for researchers.

3.2 Challenges of Replication

Multiple factors may contribute to a lack of replication studies in the field of education. Some researchers may prefer not to engage in work that they regard as less innovative and ground-breaking in preference to doing something new and different. According to Barshay (2014), Steve Graham, professor of education, and an editor of multiple educational journals, reported he receives such a large volume of articles that he has the ability to be selective. He admits his bias to publish studies that are new and have significant, new impact rather than replicating other research work (Barshay, 2014). Hannafin et al., (2005) noted that "instructional technology researchers have attempted to conduct controlled, scientifically valid studies in natural settings, but 'real' settings are often so chaotic, unpredictable and idiosyncratic that traditional experimental and quasi-experimental approaches are rendered impractical" (p. 9). Some difficulties in replicating studies in educational settings are related to the opportunity to actually recreate the same environment as the original study because of a multitude of variables, including teacher quality, classroom climate, type and specific uses of technology, dosage of intervention, different resources, revised learning activities, sample demographics, etc., that can differ extensively.

As with any type of research method, there are issues related to the use of replication studies. Trying to replicate a study that has flaws in the original study or a lack of reporting guidelines in the publication in which details of the intervention, participants, conditions, etc. are unclear is problematic (López et al., 2015; Twining et al., 2017). The level of rigor of a study to be replicated is an important consideration. Some considerations include baseline equivalence between the treatment and comparison groups, including the assumptions for specific statistical methods, quality of assessment measures, inclusion of effect sizes, adequate sample sizes, and inclusion of statistical power analyses. A systematic review of the experimental designs in mobile learning research revealed a low or medium level of rigor in a majority (72%) of the 342 studies published in refereed journals from 2006 to 2016 (Sung et al., 2019). Some of the recommendations made in the systematic review can apply to most empirical studies to allow for more convincing, replicable, and transferable findings (Sung et al., 2019).

Even with high quality rigorous studies, it is also often difficult to replicate studies that have variation across sites (Berliner, 2002), participants, sample size and/or implementation. Even details such as the time of year the intervention was completed, clearly defined terms, and school context are variables that are not often included in the publication of the studies (Bondie et al., 2019).

In the field of educational technology, there is an overvaluation of novelty studies (dissertation studies, etc.) that lead researchers away from spending their time and resources on replication studies. Typically there is not funding to replicate published studies, but only for new research that is breaking ground with new findings regarding evolving technologies and technologies perceived as innovative.

Credibility may be in question, depending upon who conducts the replication study. While the researchers who conduct the original study have a better understanding of the various issues related to the study, outside peer reviewers may question their motives. Some researchers may not find the same results as their initial study, thereby choosing not to publish the findings. Having extensive details of the original study published in an

article is critical for independent researchers wishing to replicate a study. Unfortunately, the details necessary to conduct a high quality replication often are not included in the original publication (Bondie et al., 2019). This is another area in which journal policies could support this type of research—a set of requirements for inclusion of specific details in original research publications.

Even with the numerous issues related to replication studies, these types of studies are important for the many policy decisions in education that otherwise might be based on a single study. The educational technology field could benefit from guidelines for conducting replication studies in order to promote better educational research. Such recommendations would help to refine the recommendations of others (e.g. Lopez et al., 2015; Twining et al., 2017) who have made suggestions for improving the reporting of research methodologies in educational technology. The challenges discussed in this section underscore the need for a consensus on the criteria that should be used to determine whether a replication study has occurred.

4 Moving Toward Replication Studies in Educational Technologies

Makel and Plucker (2014), in a significant contribution, urge educational researchers to pay more attention to the need for replication studies in the field of education. In response to this *call to action*, journal editors in the field of educational technology began a conversation at the National Technology Leadership Summit in 2014 to address this issue. Following the initial meeting, editorials calling for replication articles were published (Spector et al., 2015; Hodges, 2015). A panel was held at the Society for Information Technology in Teacher Education (SITE) in which educational researchers, editors and practitioners convened to discuss the issue and extend the conversation (Christensen et al., 2015). Following the panel discussion, a working group was formed at the National Technology Leadership Summit 2015 in Washington D.C. in which the discussion was focused on the need for replication studies in educational technology. Two days were spent discussing the types of replication studies and possible topics that need to be addressed in the field to give confidence that the research being published can provide ways to impact policy and practice.

In a positive step toward replication studies, the Institute of Education Sciences (IES) and the National Science Foundation (NSF) jointly issued common guidelines for educational research and development that included the need to increase visibility and value of replication studies among education research stakeholders (NSF and IES 2018). In fact, IES has a specific grant application category for proposing replication studies (Chhin et al., 2018), which is an important for prior neglect of funding for replication studies.

Since the inception of this paper, there have been other examples of journals beginning to value the need to include replication studies as an important part of research. Recently, the journal *Educational Research and Evaluation,* invited a call for replication studies in education for a special edition of the journal (See & Perry, 2021).

4.1 Establishing a Shared Vocabulary for Replication Studies in Education

A second purpose of the current paper is to establish a vocabulary that education researchers can use in their papers. Schmidt (2009) examined the literature regarding replication in scientific investigations and introduced the terminology *direct replication* and *conceptual replication* to describe replication studies (p. 91). By Schmidt's definition a *direct*

replication is a "repetition of an experimental procedure" (p. 91) and a *conceptual replication* is a "repetition of a test of a hypothesis or a result of earlier research work with different methods" (p. 91). These two categories are certainly helpful in labeling or classifying research studies, but they represent only two points on a spectrum of possibilities for how research studies may evolve.

Frequently, education research is conducted in authentic settings that require researchers to work within constraints that are beyond their ability to control. Hence, the dichotomous categorization offered by Schmidt may not provide the level of detail needed to enable education researchers to identify, or describe replication studies in a meaningful way. A finer level of detail for describing replication studies will allow researchers to label their work in a more descriptive manner. It also will provide guidance for researchers as they plan studies, helping them understand if they are replicating an existing study, or whether they are simply conducting research related to prior studies.

As described in the previous paragraphs, Schmidt (2009) suggested two fundamental types of replication: *direct replication* and *conceptual replication*. However, Lykken (1968) distinguished between *literal*, *operational* and *constructive* replication and subsequently Darley (2000) expanded the classifications of Lykken and created a typology for replication studies in which he determined four types: *strict*, *partial*, *operational* and *conceptual*. He defines *strict* replication as a study that adheres to most aspects of the first study while a *partial* replication introduces change into one aspect of the original study. He further describes an *operational* replication as less of a replication and more of a testing for validity. His fourth type, *conceptual* replication, is the least stringent and includes seeking to confirm findings from the original study but with different populations, methodologies and/or analyses (Morrison et al., 2010). While helpful in describing different types of replication studies, it is difficult to ascertain which components can vary and by how much to determine the appropriate category. For example, in an educational context, direct replications, as defined as an exact replication, are almost impossible to conduct in education. In addition, conceptual replications have been considered by some to be too flexible to be allow for any comparative conclusions to be made (Plucker & Makel, 2021).

The analysis of Chhin et al. (2018) of grants funded by the Institute of Education Sciences (IES) found that, although many were intended to be replication studies with the goals of determining efficacy and effectiveness, it was still unclear how to determine the number of parameters that could change in the conceptual replication without being considered a new evaluation (Chhin et al., 2018). However, even in their study of replication research grants, they did not use a specific criterion for how many dimensions could vary and still be called a conceptual replication (Chhin et al., 2018).

The use of the proposed vocabulary in the following section will aid researchers as they study existing, published research. If researchers describe their work in these terms and include some explanatory text, it will help others who review their work by showing a clear link to relevant preceding literature. For example, Hodges and Cowan (2012) published a research article in the field of learning technologies that included the phrase: "The present study is a replication of a study by Sheridan and Kelly (2010). Sheridan and Kelly's study focused almost exclusively on graduate students as participants, whereas this study focuses on the perceptions of undergraduate preservice teachers" (p. 143). The overt use of the replication vocabulary in the body of the article, the abstract, or in the list of keywords will help researchers develop a full picture of how a particular study is situated in the body of literature. Makel and Plucker (2014), for example, identified studies to include in their research by searching databases for the text "replicat*" (p. 307). Perhaps significantly more replication studies have occurred, but the lack of specifying the terms related to replication

did not enable the authors to identification of the studies with a keyword search. Since replication studies borrow completely or heavily from the design of former studies, this clear identification provides an acknowledgment of the work done by the researchers whose work is being replicated. Providing more guidelines for how much of the original study must be included to be considered a replication study is also warranted. In the following sections, a suggested categorical continnum for replication studies is described.

5 Providing a Categorical Continuum for Replication Studies in Educational Technology

There is not a consensus on how stringent a replication study must be, and may vary by research field and the type of decisions to be made based on the outcomes. However, creating agreed upon levels in a useful format may help guide researchers in realizing whether or not they are able to replicate research in a strict sense or in a more practical sense. This paper presents three categories of replication studies defined by the extent to which they include six components (forming a continuum) determined to be important for research studies. These types include *reproduction*, *reiteration* and *refocus* and are defined in a general way and more clearly shown in terms of categories of research in Table 1. The first category, *reproduction*, is an almost exact replication and rarely occurs in educational technology research due to many uncontrollable variables. The highest level of reproduction on this continuum (highest on all elements) is close to Darley's (2000) *strict* replication category.

The next category, *reiteration*, is less stringent than *reproduction*. It is also less strict than Darley's (2000) *partial* type because his type only includes a change in one aspect of the original study. This category of replication study should have the same research question and hypothesis but less stringent on reproduction of analysis methods and implementation procedures.

The third category, *refocus*, is the least stringent of the three categories and is defined as having the same general goals as the original research but with new areas of emphasis.

Table 1 Categorical continuum for three types of replication studies

Research design components	Refocus		Reiteration		Reproduction	
	1	2	3	4	5	6
Theoretical/conceptual framework	O	O	O	O	O	O
Research question/hypothesis	O	O	O	O	O	O
Research design	O	O	O	O	O	O
Implementation procedures (intensity, frequency, duration, treatment materials, etc.)	O	O	O	O	O	O
Instrumentation	O	O	O	O	O	O
Participants (profile of participants)	O	O	O	O	O	O
Setting (informal, rural, education level)	O	O	O	O	O	O
Analysis methods	O	O	O	O	O	O
Reporting of results (display of findings)	O	O	O	O	O	O

This category most closely resembles Darley's (2000) *operational* or *conceptual* types. The type of replication study most common in educational technology replication research tends to be *reiteration* (Adams et al., 1992; Oh et al., 2013) including some level of reiteration of the conceptual framework, research plan and design, instrumentation, etc.

Table 1 includes a framework of components often included in research studies. Each of the components is numbered from one (1) to six (6) with two categories in each of the replication types to form a range or continuum. For example, *reproduction* (an almost exact replication) includes the highest level of replication of prior research. While complete duplication is never possible (Schmidt, 2009), each of the components must be closely reproduced in the replication study.

There are two levels in each of the categories to allow for more variance in each component of the categories. The closer to 6 in the table, the more likely the type of research is deemed to be considered in the *reproduction* replication category. For example, reproduction is the highest level of replication study in this categorization and each of the criteria must be 5 or 6 to be considered a *reproduction* type of replication study. While the most stringent reproduction research would have each category rated as a 6, that type of replication is rare in educational settings, yet is necessary to be included to contrast with the other types of replication studies. However, this table is constructed as a continuum so that even if there are 5 s and 6 s, it can still be considered to be a reproduction type of replication. The descriptors are from 1 (may have similar components in the new study) to 6 (an exact replication of a previous study). *Refocus* is the least stringent and does not require a strict following of each of the components as the original study. To be considered a *refocus* replication, some of the components might fall into other categories but the majority would be rated as 1 s or 2 s on the continuum.

6 Applying the Framework

A framework is only as good as the ability to use it for the purpose for which it was intended. Based on a list of "edtech research journals you should read" (Lynch, 2018, p. 1), the authors of this paper searched the journal databases using the terms "replication" or "replicate" within the title or body of the articles. The search yielded very few articles that were actual studies. Two of the journals included editorials calling for replication studies (Hodges, 2015; Spector et al., 2015). Other articles, upon reading the abstracts, only replicated some types of methods or statistical analyses but not enough of the criteria to be considered a replication study. Most of the studies found in the simple search were replications of their own work, which is admirable in that most of the variables could be controlled. However, the field needs more replication studies of other researchers' work to establish stronger credibility in the eyes of other researchers as well as practitioners and policy makers in the field (Schmidt, 2009). In addition to the editorials found using the selected search terms, seven research articles were deemed to be replication studies. Two of the articles were published in *Computers & Education* (2009 and 2013), one in the *Journal of Digital Learning in Teacher Education* (2012), one in the *Journal of Research on Technology in Education* (2014), two in *ETR&D* (2017 and 2019) and one was in the *British Journal of Educational Technology* (1993). Two of the articles were selected for examples using the proposed categorical continuum found in Table 1.

The first example (Knezek and Christensen, 2020) was selected because it is more recent and was written in response to the call for replication articles (Spector et al., 2015) in that

journal. This example is a large study that focused on science and technology curriculum and was conducted over a five year period on different groups of students each year, but with the same teachers. The schools were located in multiple states across the U.S., but the teachers were trained together each summer at teacher institutes before implementing the curriculum with their students. As a part of the project, students took home handheld meters that allowed them to monitor the amount of standby power being used when an appliance was plugged in but not being used. The students measured multiple appliances that may or may not use standby power in their homes, collected the data, entered it into a spreadsheet and compared it to other student data. In addition, students participated in an online simulation of the measures in order to successfully unplug "vampires" (standby power) in the virtual game environment, mirroring what they had just completed hands-on in their homes.

In this example study, findings from data gathered in 2009–2011 indicated middle school students gained ($p < 0.05$) in content knowledge and became more positive in their dispositions toward STEM (science, technology, engineering, and mathematics) as a result of participating in hands-on, technology-enhanced, science inquiry activities (Knezek et al., 2013). Findings from data gathered in 2013–2014 replicated the original findings from 2009 to 2011 (Knezek and Christensen, 2020). Specifically, matched pre-post data from 2013 to 2014 confirmed large gains in knowledge of environmental science and vampire power (also known as standby power) ($p < 0.0001$, effect size = 0.86). Aggregate dispositions toward science, mathematics, engineering and technology became more positive for treatment versus comparison group students ($p = 0.022$) (Christensen et al., 2014).

The above example can be considered a replication study falling into the *reiteration* category based on the criteria presented above. As shown in Table 2, many of the criteria are in the *reproduction* range as well as the *reiteration* range. However, the study would be considered *reiteration* in this three category approach due to some criteria being lower than *reproduction* and falling into the column for *reiteration*. In the example study, the classroom teachers remained the same in both studies but the students were different, although in the same target grade levels. There were multiple schools in the project, so implementation procedures were not always duplicated due to differences

Table 2 Example of a reiteration replication study at multiple locations

Research design components	Refocus		Reiteration		Reproduction	
	1	2	3	4	5	6
Theoretical/conceptual framework	O	O	O	O	O	●
Research question/hypothesis	O	O	O	O	●	O
Research design	O	O	O	O	●	O
Implementation procedures (intensity, frequency, duration, treatment materials, etc.)	O	O	O	●	O	O
Instrumentation	O	O	O	O	O	●
Participants (profile of participants)	O	O	O	●	O	O
Setting (informal, rural, education level)	O	O	●	O	O	O
Analysis methods	O	O	O	O	●	O
Reporting of results (display of findings)	O	O	O	●	O	O

in how much time could be spent on the unit, how many students were in each classroom and the time of year the unit was taught. However, the theoretical framework and the well-validated and reliable instrumentation remained the same for both studies. The theoretical framework remained the same for all five years but some additional research questions with additional analysis methods were added based on formative feedback.

A second example includes a replication of an evaluation study focused on e-learning (Grubisic et al., 2009). The researchers were intentional about creating an internal replication of their study due to their belief that effectiveness studies should be replicated "in order to verify the original results and to indicate evaluated advantages and disadvantages" (p. 591). The controlled experiment investigated the effectiveness of an intelligent authoring system. Undergraduate students, enrolled in a course called "Introduction to Computer Science", were the participants in the study that occurred over two school years (once for the initial experiment and the subsequent school year for the replication of the initial experiment). For both the initial and replication studies, the researchers, theoretical framework, research hypotheses, research design, implementation, instrumentation, participants, setting, analysis methods and reporting of findings were almost exact replication. The only two that were considered to be slightly less than exact were the implementation procedures in which the treatment was the same but the instructors conducted both experiments in succession so there may be some slight differences due to the experiences of the initial implementation. In addition, the students were different and there were different percentages of students completing the entire experiment in the initial versus the replication implementations. However, the researchers were meticulous in controlling for these types of differences and used statistical methods to equalize the groups. This study can be considered a *reproduction* type of replication study as defined in this paper as each of the criteria fits into the *reproduction* band with a 5 or 6 in each area. Table 3 displays the criteria applied to the continuum.

"Appendix" includes guidance to assist with the application of our proposed categorical continuum framework for types of replication studies.

Table 3 Example of a reproduction study within the same location

Research design components	Refocus		Reiteration		Reproduction	
	1	2	3	4	5	6
Theoretical/conceptual framework	O	O	O	O	O	●
Research question/hypothesis	O	O	O	O	O	●
Research design	O	O	O	O	O	●
Implementation procedures (intensity, frequency, duration, treatment materials, etc.)	O	O	O	O	●	
Instrumentation	O	O	O	O	O	●
Participants (profile of participants)	O	O	O	O	●	O
Setting (informal, rural, education level)	O	O	O	O	O	●
Analysis methods	O	O	O	O	O	●
Reporting of results (display of findings)	O	O	O	O	O	●

7 Implications for Research in Educational Technology

Using a framework to help define and guide researchers in conducting replication research in the field of educational technology may be a good starting point. Perhaps many researchers have already done this but have not overtly stated it or recognized it as replication research. In the ever-changing field of learning technologies, what topics might benefit from replication studies? As early as 2003, Roblyer and Knezek (2003) created a list of topics that might benefit from replication studies. In their published paper they listed the following topics:

- Research to establish a relative advantage
- Research to improve technology implementation methods
- Research to monitor technology's impact on important societal goals
- Studies that monitor and report on current technology uses that help us shape desired directions

Other sources to find topics that could be strengthened with replication studies include analyses of publications in top educational technology journals. For example, one article published in the *Journal of Educational Technology and Society* was focused on the trends in educational technology as determined through the lens of highly cited articles published in the journal for a specified number of years (Kinshuk et al., 2013). Kinshuk, (2013) reported that highly cited articles published between 2007 and 2010 included topics such as mobile and ubiquitous learning, e-learning, dynamic media, technology adoption, blended learning and collaborative learning. A similar analysis undertaken over five decades of research published in the *British Journal of Educational Technology* which focused on trends across the decades (Bond et al., 2019). Notable over time was the change in focus from technology to a focus on the learning processes. The most recent decade (2010–2018) included a focus on learning analytics, online collaboration in higher education, and the development of new tools to support mobile learning.

A search of the most cited journal articles from *Computers & Education* since 2017 also revealed topics that may be of interest to replicate, and included topics such as the impact of flipped learning, problem-based gaming, augmented reality, models of technology integration, and impact of mobile learning including the impact on students dispositions and attitudes as well as learning. The high number of citations that some published studies accumulate may indicate the popularity of their ideas or methodologies, but it is unclear from citation count alone how stable the findings in those studies are across time and context. The studies that are receiving a lot of attention are ripe for replication to ensure that researchers are working on a stable foundation (Bond et al., 2019).

8 Summary and Conclusions

For a discipline to be considered scientific, it needs to establish a level of certainty in findings and then build upon those findings (Schmidt, 2009). Replication studies are aimed primarily in establishing confidence in findings or identifying nuances or even

problems with prior research. Specifically, in the context of educational technology, there may be a different approach based on the rapid pace at which technologies are evolving and the need to introduce them into education.

This paper aimed at creating a way to look at replication studies as a continuum in which there are categories of replication studies defined by the extent to which they fall into the three categories. A continuum may provide a useful format to help guide researchers in realizing whether or not they are able to replicate research in a strict sense or in a more practical sense. The three categories presented in this paper include *reproduction*, *reiteration* and *refocus*. Also, the use of the terminology introduced in this paper will explicitly alert readers of published research to the previous work on which new research was based, as well as what may have been changed in the newer study allowing for a more complete understanding of the evolution of research between studies.

There are several limitations of this proposed framework including the lack of application from a beginning research point of view, but developed with existing studies after they were published. In addition, although two of the authors are journal editors, the way in which the authors selected the ratings may be somewhat subjective and open to different interpretation.

Further research to refine these rubrics is needed and would aid the community of learning technologies researchers. Specifically, the language and the proposed continuum are likely to be refined through adoption and further scrutiny by the educational technology community. Many topics were proposed in this paper that may be worthy of replication studies by researchers in the field of educational technology.

Appendix

The categorical continuum is intended to be use as a guide for either determining what should be included in a replication study or whether a completed study might be considered a replication study of some type. Each of the research design components are listed below with questions that might be asked in determining in which category the study might belong. In each of the replication types, there are two categories that allow a little more "wiggle" room for determining the type of study.

See Appendix (Table 4).

Table 4 Guidelines for using the categorical continuum framework for types of replication studies

	Refocus	Reiteration	Reproduction
Theoretical/conceptual framework			
Is the topic the same or based on similar concepts?	√		
Is there overlapping theories and/or conceptual frameworks for the study?		√	
Are the theories the same but perhaps additional ones?			√
Research question/hypothesis			
Are the research questions different but similar topic?	√		
Do some of the research questions overlap?		√	
Are the research questions the same?			√
Research design			
Is the design different (for example qualitative versus quantitative?)	√		
Is the research design the same type?		√	
Does the research design match the original design as closely as possible			√
Implementation procedures (intensity, frequency, duration, treatment materials, etc.)			
Do the implementation procedures differ in intensity, frequency, duration and materials used?	√		
Are there similarities in the intensity, frequency, duration and materials used in the study?		√	
Are all implementation procedures followed as closely as possible?			√
Instrumentation			
Are the instruments different from the other study?	√		
Are the instruments using similar concepts but not the same?		√	
Are the instruments the same or almost the same?			√
Participants (profile of participants)			
Is the study on a different population?	√		
Is the study using participants similar to the original study?		√	
Are the participants the same in age, SES backgrounds, ethnicity, etc.?			√
Setting (informal, rural, education level)			
Is the setting in a different type of educational environment?	√		
Is the setting similar to the original study?		√	

Table 4 (continued)

	Refocus	Reiteration	Reproduction
Do each of the setting parameters match the original study?			√
Analyses methods			
Are different types of analyses techniques used for the findings	√		
Are most of the methods from the original study followed?		√	
Are all methods of analyses included in the new study?			√
Reporting of results (display of findings)			
Are results reported in a different format from original study	√		
Are most of the results displayed in a similar way that allows comparison to original study?		√	
Are all results from the original study included in the replicated study?			√

Acknowledgements This paper began as a thread of conversation at the 2015 National Technology Leadership Summit organized by the National Technology Leadership Coalition and hosted by the American Association of Colleges for Teacher Education. The authors wish to thank the working group attendees who helped form this discussion: Lynn Bell, Robert Branch, Arlene Borthwick, Denise Crawford, Gerald Knezek, Meghan Manfra, Natalie Milman, David Rutledge, and Melanie Shoffner.

Author Contributions (optional: please review the submission guidelines from the journal whether statements are mandatory).

Funding None.

Availability of Data and Material There are no data associated with this paper as the examples are from previously published papers which can be accessed for a fee or through subscription. The authors declare no conflict of interest. No ethics review was required to undertake this paper. No human subjects were included.

Declarations

Conflict of interest The manuscript is not under consideration by other journals and the research meets ethical and legal guidelines.

References

Adams, D. A., Nelson, R. R., & Todd, P. A. (1992). Perceived usefulness, ease of use, and usage of information technology: A replication. *Management Information Systems (MIS) Quarterly, 16*(2), 227–247.

APA. (2020). *Publication manual of the American Psychological Association* (7th ed.). Washington, D.C.: American Psychologial Association.

Barshay, J. (2014). Education researchers don't check for errors – dearth of replication studies. *The Hechinger Report.* Available: http://educationbythenumbers.org/content/education-researchers-dont-check-errors-dearth-replication-studies_1762/

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

Berliner, D. C. (2002). Educational research: The hardest science of all. *Educational Researcher, 31*(8), 18–20.

Bond, M. (2020). Facilitating student engagement through the flipped learning approach in K-12: A systematic review. *Computers & Education*. https://doi.org/10.1016/j.compedu.2020.103819

Bond, M., Zawacki-Richter, O., & Nichols, M. (2019). Revisiting five decades of educational technology research: A content and authorship analysis of the British journal of educational technology. *British Journal of Educationa Technology, 50*(1), 12–63.

Bondie, R. S., Dahnke, C., & Zusho, A. (2019). How does changing "one-size-fits-all" to differentiated instruction affect teaching? *Review of Research in Education, 43*, 336–362.

Casti, J. L. (1989). *Paradigms lost: Images of man in the mirror of science*. William Morrow.

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 educational sciences. *Educational Researcher, 47*(9), 594–605.

Christensen, R., Knezek, G., & Tyler-Wood, T. (2014). Student perceptions of science, technology, engineering and mathematics (STEM) content and careers. *Computers in Human Behavior*, *34*, 173–186. https://doi.org/10.1016/j.chb.2014.01.046.

Christensen, R., Spector, J. M., Thompson, A., Schmidt-Crawford, D., Bull, G. & Knezek,G. (2015). Innovation versus replication in research findings: Has the novelty of newresearch findings worn off? In D. Slykhuis & G. Marks (Eds.), *Proceedings of Society forInformation Technology & Teacher Education International Conference 2015* (pp. 1126–1129). Chesapeake, VA: Association for the Advancement of Computing in Education(AACE).

Darley, W. K. (2000). Status of replication studies in marketing: A validation and extension. *Marketing Management Journal, 10*(2), 121–132.

De Angelis, C., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine, C., Marusic, A., Overbeke, A. J. P. M., Schroeder, T. V., Sox, H. C., & VanDer Weyden, M. B. (2004). Clinical trial

registration: A statement from the international committee of medical journal editors. *New England Journal of Medicine, 251*(12), 1250–1251.
Dron, J. (2021). Educational technology: What it is and how it works. *AI & Society*. https://doi.org/10.1007/s00146-021-01195-z
Garzon, J., & Acevedo, J. (2019). Meta-analysis of the impact of augmented reality on students' learning gains. *Educational Research Review, 28*(1), 244–260. https://doi.org/10.1016/j.edurev.2019.04.001
Grubisic, A., Stankov, S., Rusic, M., & Zitko, B. (2009). Controlled experiment replication in evaluation of e-learning system's educational influence. *Computers & Education, 53*, 591–602.
Hannafin, M., Orrill, C., Kim, H., & Kim, M. (2005). Educational technology research in postsecondary settings: Promise, problems, and prospects. *Journal of Computing in Higher Education, 16*(2), 3–22. https://doi.org/10.1007/BF02961472
Hodges, C. B. (2015). Replication studies in educational technology. *TechTrends*, *59*(4), 3–4. https://doi.org/10.1007/s11528-015-0862-2.
Hodges, C. B., & Cowan, S. F. (2012). Preservice teachers' views of instructor presence in online courses. *Journal of Digital Learning in Teacher Education*, *28*(4), 139–145. https://doi.org/10.1080/21532974.2012.10784694.
Hew, K. F., Lan, M., Tang, Y., Jia, C., & Lo, C. K. (2019). Where is the "theory" within the field of educational technology research? *British Journal of Educational Technology, 50*(3), 956–971. https://doi.org/10.1111/bjet.12770
Kinshuk Huang, H.-W., Sampson, D., & Chen, N.-S. (2013). Trends in educational technology through the lens of the highly cited articles published in the journal of educational technology and society. *Journal of Educational Technology and Society, 16*(2), 3–20.
Knezek, G., & Christensen, R. (2020). Project-based learning for middle school students monitoring standby power: replication of impact on stem knowledge and dispositions. *Educational Technology Research & Development (ETR&D)*, *68*(1), 137–162. https://doi.org/10.1007/s11423-019-09674-3.
Knezek, G., Christensen, R., Tyler-Wood, T., & Periathiruvadi, S. (2013). Impact of environmental power monitoring activities on middle school student perceptions of STEM. *Science Education International*, *24*(1), 98–123.
Li, J., Antonenko, P. D., & Wang, J. (2019). Trends and issues in multimedia learning research in 1996–2016: A bibliometric analysis. *Educational Research Review, 28*(1), 100282. https://doi.org/10.1016/j.edurev.2019.100282
López, X., Valenzuela, J., Nussbaum, M., & Tsai, C.-C. (2015). Some recommendations for the reporting of quantitative studies [Editorial]. *Computers & Education, 91*, 106–110. https://doi.org/10.1016/j.compedu.2015.09.010
Lynch, M. (2018). *Ten edtech research journals that you should read*. https://www.thetechedvocate.org/10-edtech-research-journals-that-you-should-read/
Lynch, J. G., Bradlow, E. T., Huber, J. C., & Lehmann, D. R. (2015). Reflections on the replication corner: In praise of conceptual replications. *International Journal of Research in Marketing, 32*(4), 333–342. https://doi.org/10.1016/j.ijresmar.2015.09.006
Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70*, 151–159.
Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Erlbaum.
Makel, M C., & Plucker, J.A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Research, 20*(10), 1–13. Retrieved from http://edr.sagepub.com/content/early/2014/07/23/0013189X14545513
Makel, M. C., Plucker, J., & Hegarty, C. B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science, 7*(6), 537–542. https://doi.org/10.1177/1745691612460688
Matthews, H. M., Hirsch, S. E., & Therrien, W. J. (2018). Becoming critical consumers of research: Understanding replication. *Intervention in School and Clinic, 53*(5), 267–275.
Morrison, S. J. (2016). Forum. *Journal of Research in Music Education, 63*(4), 395–396. https://doi.org/10.1177/0022429415625485
Morrison, R., Matuszek, T., & Self, D. (2010). Preparing a replication or update study in the business disciplines. *European Journal of Scientific Research, 47*(2), 278–287.
National Science Foundation and Institite of Education Sciences. (2018). *Common guidelines on replication & reproducibility in education research: A supplement to the commone guidelines for education research and development*. National Science Foundation and the Institute of Education Sciences, U.S: Department of Education, Washington DC.

Oh, Y. J., Jia, Y., Lorentson, M., & LaBanca, F. (2013). Development of the educational and career interest scale in science, technology, and mathematics for high school students. *Journal of Science Educational and Technology, 22*, 780–790.

Open Science Collaboration. (2012). Perspect. *Psychological Science, 7*, 657–660.

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

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

Roblyer, M. D., & Knezek, G. A. (2003). New Millennium Research for educational technology: A call for a national research agenda. *Journal of Research on Technology in Education, 36*(1), 60–71.

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.

See, B.H. & Perry, T. (2021). A call for replication studies in education special issue of educational research and evaluation. *Educational Research and Evaluation*. https://think.taylorandfrancis.com/special_issues/replication-studies/

Sherer, R., & Teo, T. (2019). Unpacking teachers' intentions to integrate technology: A meta-analysis. *Educational Research Review, 28*(1), 90–109.

Sheridan, K., & Kelly, M. A. (2010). The indicators of instructor presence that are important to students in online courses. *MERLOT Journal of Online Learning and Teaching, 6*(4), 767–779.

Smith, N. C., Jr. (1970). Replication studies: A neglected aspect of psychological research. *American Psychologist, 25*, 970–975.

Spector, J. M., Johnson, T. E., & Young, P. A. (2015). An editorial on replication studies and scaling up efforts. *Education Technology Research & Development*, *63*(1), 1–4.

Sung, Y.-T., Lee, H. Y., Yang, J. M., & Chang, K.-E. (2019). The quality of experimental designs in mobile learning research: A systemic review and self-improvement tool. *Educational Research Review, 28*, 100279. https://doi.org/10.1016/j.edurev.2019.05.001

Twining, P., Heller, R. S., Nussbaum, M., & Tsai, C. C. (2017). Some guidance on conducting and reporting qualitative studies. *Computers & Education, 106*, A1–A9. https://doi.org/10.1016/j.compedu.2016.12.002

van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review, 28*(1), 1–18. https://doi.org/10.1016/j.edurev.2019.05.003