Editorial

What to Replicate?

Uffe Thomas Jankvist | ORCID: 0000-0002-4089-0242
Danish School of Education, Aarhus University, Aarhus, Denmark
utj@edu.au.dk

Mario Sánchez Aguilar | ORCID: 0000-0002-1391-9388
CICATA Legaria, National Polytechnic Institute, Mexico City, Mexico
mosanchez@ipn.mx

Morten Misfeldt | ORCID: 0000-0002-6481-4121
Department of Science Education, University of Copenhagen,
Copenhagen, Denmark
misfeldt@ind.ku.dk

Boris Koichu | ORCID: 0000-0001-7298-8175
Department of Science, Weizmann Institute of Science, Rehovot, Israel
boris.koichu@weizmann.ac.il

1 Introduction

In this second editorial of Implementation and Replication Studies in Mathematics Education (IRME), we address the question of what to look for in research results and findings from mathematics education research in terms of sources for replication studies. Surely, a joke in the mathematics education community goes that mathematics education researchers do not want to replicate, they want to get replicated! Probably this is true for many research fields, not only mathematics education. Still, we deem this not to be the major reason for the relatively small number of replication studies in our field (Aguilar, 2020;
Jankvist et al. (2021). Rather we believe this has to do with the research culture. As some of us recently pointed out in the context of studying the use of digital technologies in mathematics education (Jankvist & Misfeldt, 2021), researchers in mathematics education seem more likely to prefer introducing new theoretical constructs for “new problems” rather than looking into the back catalogue of our research field in order to spot already well-developed constructs that may readily provide insights into these new problems (Mason, 2016). This is of course in line with the (mis)interpretation of the old saying “If you want to get ahead, get a theory” (Karmiloff-Smith & Inhelder, 1975) to mean that if you want to get ahead, develop a new theory (and have others replicate it). Still, as we know from Schoenfeld (2014), “If you really want to get ahead, get a bunch of theories ..., and data to test them” (p. 7). This quote is at the heart of replication in mathematics education, we find. Namely, to gather (different types of) data to “test” the theoretical constructs that we apply and adhere to in our research. Still, this does not answer the questions of what to replicate and why.

Now, another reason for not resorting to constructs from the back catalogue of our research field may of course be that new coming researchers are not necessarily well familiar with the back catalogue. Surely, we have some collections of “classical papers” (e.g. Bishop, 2010), but maybe some “pre-digestion” of our fifty years of research results is needed in order to direct and guide (young) mathematics education researchers in a pursuit of research findings to potentially replicate. We find that the listing by the Editorial Committee of the European Mathematical Society of so-called “solid findings” (in mathematics education research) to some extent illustrates this.

2 Solid Findings

In 2011, the Education Committee of the European Mathematical Society launched a collection of articles called “solid findings in mathematics education”. These articles aimed at disseminating, among mathematicians and other scholars interested in research in mathematics education, brief syntheses of research on different topics of importance. The first article in the collection (Education Committee of the European Mathematics Society, 2011a) discusses what is meant by a “solid finding” in mathematics education. To determine what a solid finding is, Schoenfeld’s (2007) criteria of quality of research in mathematics education was adopted. According to Schoenfeld (2007), the contributions of a research study may be characterized and judged along three important dimensions: trustworthiness, importance, and generalizability.
intertwined dimensions of trustworthiness and generalizability are instrumental in the characterization of a solid finding – but they are also of special relevance for the discussion of the potential roles of replication studies in mathematics education and of the kind of replication studies that may be developed.

In relation to trustworthiness, Schoenfeld (2007) refers to replicability as one of the aspects of research results' trustworthiness:

Every person is different; every classroom is different. How can one possibly speak of replication in education? The idea seems strange. One might replicate experiments of some types – but how many educational researchers do experimental work? And, if every classroom is different, what does it mean to replicate someone else's research study? One way to think about the issue of replicability is to think about generality. (p. 87)

The above quotation points to a relationship between replication and generalization of research findings, and this relationship is reflected in the basic characteristics of a solid finding. As indicated by the Education Committee of the European Mathematics Society (2011a), a solid finding “can be applied to circumstances and/or domains beyond those involved in this particular research” (p. 46). Moreover, “the term ‘solid’ also includes an aspect of ‘robustness’ in the sense that the finding should be repeatedly observed or confirmed in many studies reporting the same or similar results leading to the same (general) conclusions.” (p. 48). Thus, one of the potential roles of replication studies in mathematics education is to assess the scope and limitations of the generality of research findings. Empirical phenomena can be tested for their generalizability by means of replication studies. In turn, theoretical constructs can be used to describe and explain these phenomena. Thus, a “replication” of a theoretical construct would be the test of its robustness and usability across contexts. An example of this kind of replication study can be found in the work of Melhuish et al. (2021), who address the robustness of theoretical relationships underlying instruments for measuring mathematics teachers’ specialized knowledge. All this implies the development of conceptual replications with variations in the contexts and conditions of the original studies (Aguilar, 2020), which could inform the field about the circumstances and the domains in which a research finding holds true or not. This can be instrumental in the identification of potential solid findings in mathematics education.

Two aspects of the solid findings are important to highlight here. First, solid findings can be different in nature. An overview of the different findings identified as “solid” provides evidence of this. For instance, a solid finding can refer
to the observation that many students possess an empirical proof scheme rather than a deductive one (Education Committee of the European Mathematics Society, 2011b). A solid finding may refer to the necessary conditions for mathematical learning or development to take place, such as content, community and context identified as decisive for the development of mathematics teaching and teachers’ own learning (Education Committee of the European Mathematics Society, 2012a). Or it could be related to a theoretical construct that holds true outside the context, where it was originally developed, such as the notion of didactical contract (Education Committee of the European Mathematics Society, 2012b). This suggests that there is a variety of research findings, the scope and limitations of which could be explored through replication studies: e.g., learning obstacles; practices; conditions for learning; theoretical constructs, etc. Second, solid findings are “solid” because they to some extent already have been replicated. However, these findings have not been replicated through studies explicitly and intentionally designed for replication purposes. Rather the solid findings are phenomena and constructs whose robustness has been evidenced over the years through different studies carried out under various conditions and in different contexts.

We hypothesize that explicit and intentional replication studies may be an important means to support and accelerate the identification of new solid findings in mathematics education research. Of course, a pending task then becomes to identify potential candidates, i.e. findings that are not yet replicated and not proven to be “solid”. This leads us to the question of what to look for in potential findings to replicate.

3 What to Look For in Potential Findings to Replicate

The solid findings that have been identified thus far hint on the nature of the research findings that could be subjects for replication studies. Still, there are different ideas and arguments about which research findings should be replicated, and which replication studies should be published. For instance, Melhuish and Thanheiser (2018), as part of their argument for the importance of the research studies that they replicate, refer to the outlets where such research studies were published: “In all cases, the original studies were published in high-quality journals” (p. 107). Schoenfeld (2018) refers to the importance of building up a body of programmatic work, where research that has been carried out previously is reexamined to test its scope and context specificity. In turn, Aguilar (2020) argues that the discussion about the criteria for identifying and selecting the original studies worthy of replicating will be
enriched, if we consider studies that are socially relevant or have the potential to inform the design of public policies.

Star (2018) has proposed three criteria that journals (editors, reviewers) can use for evaluating the merits of a replication study. According to him, an outstanding replication study article (p. 99):

1. Makes a convincing case that the topic of study in the replication is of great importance to the field;
2. Makes a convincing case that the field will learn something significant from the replication that is not already known, and;
3. Convincingly shows that there is reason to believe that the results of the original study may be flawed.

Melhuish and Thanheiser (2018) suggest an alternative version of the third criterion proposed by Star (2018): “An outstanding replication study article convincingly shows that there is a reason to test the generalizability or validity of the original study results” (p. 109). To this end, IRME welcomes those replication studies in mathematics education that are explicit as to what is being replicated, and where the importance of the replication lies.

We now turn to the question of why it is important for the mathematics education community to develop more replication studies? There is of course a variety of arguments about the importance of replication studies in mathematics education. Still, it is possible to broadly classify them according to the purposes of the replication study, relying on the proposal by an anonymous reviewer mentioned in a footnote by Star (2018): “An anonymous reviewer of this commentary identified two distinct purposes for conducting a replication study, which were referred to as ‘understanding’ and ‘truth’” (p. 101).

In relation to understanding, it is argued that replication studies are a means to understand more deeply some of the phenomena and results that have been identified in the field of mathematics education (Aguilar, 2020). In particular, replication studies can help us learn more about the conditions that make a particular research finding true or not, and can be instrumental in investigating the effects of particular treatment under different conditions or sample populations (Cai et al., 2018). Investigating the effects of particular didactic treatments through replication creates a connection with implementation studies in mathematics education and enriches the variety of research products whose scope and limitations could be explored through replication studies:

Replication studies can inform us about how particular treatments, interventions, or didactic designs work in different contexts and with different populations. Also, replication studies can help us to identify the enablers,
obstacles, and general conditions that favor or inhibit the implementation of research products in practice.

AGUILAR, 2020, p. S45–S46

In connection to truth, it is possible to consider all those arguments that refer to replication studies as an empirical source that provides research results with objectivity, certainty, credibility, and generality (e.g., Schoenfeld, 2007).

We now briefly address the question of how to conduct a replication study in mathematics education. Although examples of replication studies are not abundant in our field, there are some studies that can be used as a reference to approximate an attempted answer to this question (e.g., De Bock et al., 2011; Jacobson & Simpson, 2019; Melhuish, 2018; Star et al., 2011; Thanheiser, 2018; Van Der Auwera et al., 2021). It is fair to claim that the replication studies developed in the field of mathematics education tend to be conceptual replications (Schmidt, 2009) that attempt to test a hypothesis or a result of earlier research work with some variations in conditions and methods. Such studies may follow quantitative methods (e.g., Van Der Auwera et al., 2021). However, as we shall see in the next section, there are examples of replication studies with an orientation towards qualitative methods as well.

4 Two Replication Studies in This Issue

The second issue of IRME contains two papers that contribute to illustrating the nature and potential of replication studies in mathematics education.

Let us first consider the replication study conducted by Melhuish et al. (2021). This is a study strongly related to the question “What is the mathematical knowledge that teachers need to provide good quality mathematical instruction?”, which has been of central interest in the area of mathematics teacher education. This area of research has produced theoretical models that characterize this specialized knowledge of teachers (e.g., Ball et al., 2008), but it has also developed instruments to measure this specialized knowledge relating it to instructional practice (e.g., Hill et al., 2004). Two examples of these instruments – which are the focus of Melhuish et al.’s replication study – are the Mathematical Knowledge for Teaching (MKT) measures and the Mathematical Quality of Instruction (MQI) instrument.

Melhuish et al. (2021) point out that “underlying the use of these instruments are some fundamental assumptions about how knowledge, instruction, and student learning relate”. In particular, Hill et al. (2012) assume that there is a relationship between Mathematical Knowledge for Teaching (MKT),
Melhuish and colleagues use replication as a means to explore the extent to which this supposed theoretical relationship is maintained across different settings. For instance, one of the replication hypotheses tested in their study is whether the MKT and MQI are positively correlated with value-added scores (as reflected in state standardized assessments). The researchers find evidence that the acknowledged relationship between MKT, MQI and student assessment scores is not always warranted across different settings. This paper illustrates how replication studies can be used to explore the validity and generality of research instruments when applied in different settings, while at the same time addressing the robustness of the underlying theoretical relationships.

The second replication study is by Jankvist and Niss (2021). They consider the well-known Students-Professors (S/P) problem, originally formulated by Kaput and Clement (1979). Using S for the number of students and P for that of professors, the S/P-problem consists in writing an algebraic equation equivalent to the statement: “At a certain university, there are six times as many students as there are professors” (Kaput and Clement, 1979: p. 208). A well-known error in students’ answers to this question has been named the “reversal error”, i.e. writing 6S=P instead of 6P=S. Jankvist and Niss apply two versions of the S/P-problem; one resembling the original, and one where students are asked to make sense of the expression 6P=S, knowing what P and S stands for. They analyze Danish upper secondary school students’ answers from a mathematics modelling perspective (augmented with the notion of didactical contract), thus referring to the first version of the S/P-problem as one of mathematization and the second version as one of de-mathematization. Relying on a selection of theoretical constructs not typically applied to the S/P-problem, the authors identify so-called major error types for both versions of this: “The error types have been established by combining an a priori analysis of the two versions of the S/P-problem with our classification of errors found a posteriori in students’ actual responses.” The authors provide illustrative qualitative examples of quantitative tendencies observed, and further state that this “mixed approach of drawing on both quantitative and qualitative data enables us to provide further insight into students’ misconceptions and difficulties concerning the problem, findings which have not been reported before in the rich literature on the S/P-problem.”

The replication study by Jankvist and Niss (2021) can be classified as a “conceptual replication” according to Aguilar (2020); one that involves a methodological variation in both conditions and procedures to test scope and generalizability of previous findings related to the S/P-problem. With relation to the distinction between “truth” and “understanding” (Star, 2018), as
mentioned above, the authors argue: “Although we do to some degree confirm the ‘truth’, although not the completeness, of several previous findings related to the S/P-problem, the main aspiration and contribution of the study described in this paper is to further ‘understanding’ of the phenomena and factors involved”.

5 Briefly on the ZDM Issue on Implementation Research

We are glad to announce in this editorial that a special issue (S1) of ZDM – Mathematics Education on implementation and implementability of mathematics education research was published in September 2021 (Koichu et al., 2021). Three of us have had the honor to serve as its guest editors. The S1 was conceived two years ago as an academic precursor of IRME: the main goal of the S1 was to boost the implementation problématiques in the international mathematics education community, whereas the main aspiration of IRME is to serve as a stable platform for furthering the discussion. Twenty-nine scholars from five continents took part in the S1. It consists of sixteen papers, including six papers that can be tagged as theoretical, nine papers that can be tagged as empirical and a single survey paper.

The S1 papers contain diverse suggestions about what “implementation” and “implementation research” are or should be in mathematics education. Some authors situated their work in one of the existing conceptualizations of “implementation” and “implementation research” (e.g., Century & Cassata, 2016; Nilsen, 2015), whereas others challenged some of the basic assumptions on which the existing conceptualizations rely (e.g., Cai & Hwang, 2021). Different projects from Austria, Chile, Germany, Israel, New Zealand, Norway, Spain, Sweden, Taiwan, UK and the US are presented and analyzed in the S1 papers through various implementation-related lenses.

Overall, the S1 provided fruitful grounds for reflections on four themes of importance for further research and practice: (1) objects of implementation; (2) stakeholders in implementation; (3) implementation vs. scaling up; and (4) implementability of mathematics education research.

The review of the past literature and of the S1 papers led us to make a few observations about the state of the art. Five of them are presented here as an appetizer for further reading. First, research concerned with implementation of mathematics education innovations is highly diverse methodologically speaking. It embraces studies that can be characterized as design research, intervention studies, controlled experiments, teaching experiments, and more. It is actually not a particular methodology, but a scholar’s decision to study enactment of a resource/innovation/approach, and not only in the context...
in which it was originally developed, but also in new contexts that make the study fit a paradigm of implementation-related research. Second, the authors frequently self-identify their studies as “research accompanying an implementation project”, but rarely can we find “implementation research” as an explicitly stated type of research. This is in contrast to other fields of research (cf., Century & Cassata, 2016; Eccles & Mittman, 2006). Third, studies on long-term consequences of implementation of innovations – that is, what happens when the innovation proposers step out and the innovation remains essentially in the hands of its end-users – are extremely rare. Fourth, types of implementation-related research vary depending on the objects of implementation. Roughly speaking, two objects of implementation are usually considered: material-centered objects (e.g., a collection of published instructional materials) and/or interaction-centered objects (e.g., a particular PD model or strategy). It looks as if studies on implementation of material-centered objects more readily discuss implementation outcomes, whereas studies focusing on interaction-centered objects are more deliberate about processes of implementation. In addition, it seems that material objects of implementation favor the “improvement of practice at scale” problématique, while interactional objects of implementation favor “bridging research and practice” agendas. Fifth, the SI papers essentially differ in relation to the scopes of projects and research accompanying them. It looks as though a distinction between implementation-as-dissemination and implementation-across-contexts can be substantiated and further explored in the future research.

6 Two Implementation Studies in This Issue

The first implementation study in this issue of IRME is by Helenius (2021) and concerns the relationships between policy driven implementation and global and institutional stakeholders that surround such a change. The object of Helenius’ paper is the evolution of design recommendations for mathematics teacher professional development (PD) in the context of the Swedish “Boost for Mathematics” (BfM) project. Helenius focuses on an extended set of stakeholders, including teachers, researchers and policymakers. The paper shows how the various stakeholders influenced these recommendations over time. Furthermore, Helenius traces the reasons for these changes. He does so by developing an approach for understanding stakeholder influence on BfM by using Chevallard’s Anthropological Theory of the Didactic (ATD).

The other implementation study in this issue of IRME, by Misfeldt et al. (2021), also focuses on stakeholders. Yet, where Helenius is concerned with global and institutional stakeholders, Misfeldt et al. (2021) present a local
case focusing on how (individual) teachers, supervisors, consultants and school management relate internally. The case is that of the municipality of Copenhagen and their implementation of the approach “Action Learning”. The emphasis is on the local supervisors that are running the program, and describes how they are met with conflicting expectations from managers and colleagues. In the paper, the authors show how the supervisors are expected to both facilitate processes from an equal position, act as the more knowledgeable peer, and at the same time be part of the line management. The paper is both an exploration of the implementation of an action learning program as well as a case study of an organizational implementation problem; namely the attendance to systemic and organizational interplays between various participants in implementation processes, and that these should always be an attention point. Methods consisting of mapping out the relations between the supervisors and the other participants offer one way to approach this problem.

As mentioned, both of the implementation papers concern stakeholders. Yet, they do so at rather different levels, although the questions of scaling intersect. By looking at the way these two papers discuss implementation, we may obtain an idea of how a global institutional stakeholder analysis (as done by Helenius) displays important aspects of why and how innovations are implemented, while smaller and more qualitative investigations of specific cases (as done by Misfeldt et al.) reveal other insights about implementation processes. In relation to the five foci developed in the ZDM SI described above, it seems that Helenius’ study can be seen as one of the – thus far – rare examples of implementation research, whereas the paper by Misfeldt et al. (2021) is closer to what is described as research in relation to implementation projects. Still, both papers contain methodological innovations – which was also the case for several of the ZDM SI papers. The authors of the two papers have found it necessary to develop or adapt approaches in order to focus on the implementation aspect of their object. The institutional focus in Helenius’ paper led him to take Chevallard’s ATD as a point of departure, whereas the interpersonal focus in the work by Misfeldt et al. (2021) made relational mappings the primary departure point.

References

What to Replicate?


