BRILL

IMPLEMENTATION AND REPLICATION STUDIES IN MATHEMATICS EDUCATION 5 (2025) 137–150

IRME

brill.com/irme

Editorial

"Solid Findings" in Mathematics Education: an Inspiration for Replication

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

Mario Sánchez Aguilar | ORCID: 0000-0002-1391-9388
CICATA Legaria, Instituto Politécnico Nacional, Mexico City, Mexico
mosanchez@ipn.mx

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

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

Published online 5 November 2025

1 Introduction

Some fifteen years ago now, the Education Committee of the European Mathematical Society (Education Committee of the EMS) launched its series on "solid findings" in mathematics education (Education Committee of the EMS, 2011a). In our second editorial of the journal *Implementation and*

Published with license by Koninklijke Brill BV | DOI:10.1163/26670127-00502001
© UFFE THOMAS JANKVIST ET AL., 2025 | ISSN: 2667-0135 (print) 2667-0127 (online)

Replication Studies in Mathematics Education (IRME) (Jankvist et al., 2021), we mentioned this series as a potential inspirational source for replication studies. In this editorial, we revisit that idea, but describe in more detail the initiative of the "solid findings" series and its publications and make more specific suggestions for potential replication studies based on the initiative.

Let us first recapitulate our arguments in the previous editorial, so as to update readers who are not already familiar with this. So, the idea of the solid findings initiative was to highlight research results that are both rigorous and broadly useful. Following Schoenfeld's (2007) framework, the Education Committee of EMS argued that a solid finding should be judged on trustworthiness, importance, and generalizability. Here, trustworthiness and generalizability are especially tied to the role of replication, of course. Since classrooms and learners differ, replication in education is often less about repeating identical experiments and more about testing whether findings hold true across varied contexts. A solid finding is therefore something that is robust, confirmed in multiple studies, and applicable beyond the original setting (Education Committee of EMS, 2011a).

As we shall dig deeper into below, examples of solid findings in mathematics education include observations about students' proof schemes, conditions that support teacher and student learning, as well as theoretical constructs like the didactical contract that travel across contexts. Importantly, most solid findings to date have not come from explicit replication studies but rather from repeated confirmation across different research projects. One of our main theses in the previous editorial was that deliberate, intentional replication studies — both close and conceptual — could strengthen the field by clarifying the scope, limitations, and robustness of existing findings and by helping to identify new candidates for solid findings (Jankvist et al., 2021). This would contribute to accelerating progress in building a more reliable and cumulative knowledge base in mathematics education.

2 The "Solid Findings" Initiative

The Education Committee of EMS defined "solid findings" in mathematics education as findings that:

- result from trustworthy, disciplined inquiry, thus being sound and convincing in shedding light on the question(s) they set out to answer.
- are generally recognised as important contributions that have significantly influenced and/or may significantly influence the research field.

– can be applied to circumstances and/or domains beyond those involved in this particular research.
– can be summarised in a brief and comprehensible way to an interested but critical audience of non-specialists (especially mathematicians and mathematics teachers). (Education Committee of EMS, 2011a, p. 46)

As mentioned above, the first three criteria are directly inspired by Schoenfeld (2007). The fourth, however, is to do with the intended broad audience of the publications on “solid findings”.

The publications were aimed not only at mathematicians, i.e., readers of the *Newsletter of the European Mathematical Society*, but also at researchers from related fields such as psychology, sociology, and anthropology who engaged with mathematics or mathematical thinking. Other important audiences included teacher educators, curriculum designers, policy makers, and test developers, as well as non-specialists interested in understanding the nature and relevance of mathematics education research. Hence, in addition to Schoenfeld’s three criteria of trustworthiness, importance, and generalizability, the series adopted a fourth criterion tailored to this audience. Findings needed to (a) attract the attention of mathematicians and teachers by demonstrating practical usefulness for teaching, and (b) be presented in a concise and accessible way, without technical language that might alienate those outside the field.

3 The “Solid Findings” Publications

So, the “solid findings” that were eventually chosen were research results that had converged on clear conclusions, contributed to a broader line of disciplined inquiry, and were convincing in addressing the questions they set out to answer. They were also expected to be generalizable beyond the original studies, relevant to both specialists and non-specialists, and expressible in a brief and comprehensible manner. In total, the Education Committee published ten reports on “solid findings” over a four-year period from 2011 to 2014:

(1) Do Theorems Admit Exceptions? Solid Findings in Mathematics Education on Empirical Proof Schemes. (Education Committee of EMS, 2011b)
(2) It is Necessary that Teachers are Mathematically Proficient, but is it Sufficient? Solid Findings in Mathematics Education on Teacher Knowledge. (Education Committee of EMS, 2012a)

(3) What are the Reciprocal Expectations between Teacher and Students? Solid Findings in Mathematics Education on Didactical Contract. (Education Committee of the EMS, 2012b)
(4) Models and Modelling in Mathematics Education. (Education Committee of the EMS, 2012c)
(5) Solid Findings in Mathematics Education: Living with Beliefs and Orientations –Underestimated, Nevertheless Omnipresent, Factors for Mathematics Teaching and Learning. (Education Committee of the EMS, 2013a)
(6) Sociomathematical Norms: In Search of the Normative Aspects of Mathematical Discussions. (Education Committee of the EMS, 2013b)
(7) Solid Findings on Students' Attitudes to Mathematics. (Education Committee of the EMS, 2013c)
(8) Why is University Mathematics difficult for students? Solid findings about the secondary-tertiary transition. (Education Committee of the EMS, 2013d)
(9) Solid Findings in Mathematics Education: The influence of the use of digital technology on the teaching and learning of mathematics in schools. (Education Committee of the EMS, 2014a).
(10) Solid Findings: Concept images in students' mathematical reasoning. (Education Committee of the EMS, 2014b)

Surely not all of the above "solid findings" are equally suited for replication, and some might be better suited for certain types of replication studies than others. Before we touch on this, let us first recall our discussion on "What to Replicate" in our previous editorial (Jankvist et al., 2021).

4 What to Replicate (Updated Recap)

So, back to where we left off in relation to the discussion of potential replications. Our previous editorial on the matter addressed which research findings might be chosen for replication in mathematics education and how such studies could be evaluated. We observed that different scholars offer different criteria: Melhuish and Thanheiser (2018) emphasize replicating studies originally published in high-quality journals, while Schoenfeld (2018) stresses the value of systematically reexamining prior research to test its scope and context-dependence. Aguilar (2020) highlights the importance of including studies with social relevance or policy implications. Star (2018) proposes three criteria for assessing replication studies: the topic's importance to the field, the likelihood of generating new insights, and the possibility that the original study's results may be flawed. Melhuish and Thanheiser (2018) suggest softening the last point, replacing it with the need to test the generality or validity of earlier results.

Following Star (2018, p. 101, footnote) we noted that the reasons for undertaking replication studies can be grouped into two broad purposes: "understanding" and "truth". Replications serve understanding by providing deeper insight into conditions under which findings hold, such as testing educational treatments or interventions in different contexts and populations. This connects replication to implementation studies, which explore how designs or practices function in varied environments. Replications serve truth by strengthening the objectivity, credibility, and generalizability of research findings, making them more reliable.

At the time of the previous editorial, we noted that examples of replication studies in mathematics education remained relatively few. Existing work (e.g., Giberti & Maffia, 2022; Hernández et al., 2023; Hohensee & Borji, 2024; Melhuish, 2018; Van Der Auwera et al., 2021) shows that these are usually conceptual replications — retesting prior hypotheses with variations in conditions, methods, or populations. While many adopt quantitative approaches, there are also instances of qualitative replication studies, reflecting the diversity of methods suited to exploring robustness and generality in educational research. To a large extent, this characterization of replication studies in mathematics education still holds. However, the five published volumes of IRME have added eleven empirical replication studies (e.g., Cooper et al., 2024; García-Cerdá et al., 2024; Liebendörfer & Wagner, 2024) and one theoretical article (Star, 2021) to the corpus of replication studies, providing a stronger basis for these claims and a clearer sense of the methodological range of replication work in mathematics education.

5 On the Potential Replicability of the "Solid Findings"

Several of the "solid findings" in mathematics education have reached this status not because they were established through explicit replication studies, but rather because independent research conducted in different places and contexts has converged toward similar results. Yet, these converging studies were not declared as replications. This raises a question: *What would be gained by conducting and explicitly labeling studies as replication studies?* Part of the answer lies in the visibility and legitimacy that the act of naming provides. By identifying them as replications, the field could actively promote this type of research, clarify how such studies can be designed and carried out in practice, and offer concrete examples that illustrate diverse forms of replication. Moreover, replication studies would allow us to return to established constructs and findings that risk being overlooked in the constant pursuit of novelty within mathematics education research. After considering these general points, it becomes possible to illustrate this argument with three specific solid findings, showing what could be learned from replicating them and how such replications might be envisioned.

A first example can be drawn from the solid findings on models and modelling (Education Committee of the EMS, 2012c), which stresses that strong knowledge of pure mathematics, while necessary, is far from being sufficient for successful engagement with modelling tasks. Research has documented that students who are proficient in algebra, calculus, or geometry often fail to apply this knowledge when confronted with real-world situations that require assumptions, simplifications, data collection, or the evaluation of extra-mathematical constraints (Ikeda & Stephens, 1998; Kaiser & Maass, 2007). In other words, there is no automatic transfer from mathematical competence to modelling competence. This gap reflects a reluctance — or at times an inability — among learners, and even some teachers, to leave the “safe quarters” of mathematics and engage with the uncertain, messy, and context-dependent decisions that modelling requires. Replication studies could therefore play an important role in clarifying how robust this finding is across different cultural and curricular contexts, and in examining whether recent shifts in education, such as project-based learning, digital simulations, or data-rich environments, alter the boundary between mathematical and extra-mathematical reasoning. By systematically replicating prior studies, the field could better understand not only the persistent obstacles students face when engaging in modelling, but also the pedagogical designs that most effectively help them bridge the gap between possessing mathematical knowledge and using it meaningfully in authentic situations.

A second case involves the well-known finding that teachers’ mathematical proficiency, although necessary, is not sufficient for effective teaching (Education Committee of the EMS, 2012a). Research has already shown that pedagogical content knowledge, collaboration among teachers, and institutional conditions all play essential roles in shaping student learning. The so-called CK–PCK–context triad captures the interplay of these dimensions, yet most of the evidence has been generated in countries with relatively stable professional development infrastructures, such as China, Germany, or Japan (Fernandez & Yoshida, 2004; Huang & Bao, 2006; Krauss et al., 2008).

Replication studies would be particularly valuable in regions where teachers face very different realities, such as high staff turnover, limited access to continuing education, or classrooms marked by strong inequalities in student preparation. For example, one could investigate whether professional learning communities or lesson study approaches produce the same benefits in systems where teachers are under pressure from strict evaluation and monitoring mechanisms or resource scarcity. Replications could also address how institutional policies, including recruitment, initial integration of new teachers, and evaluation frameworks, shape the balance between teachers' mathematical knowledge and their capacity to transform it into meaningful instruction. In this way, systematic replication would not only test the universality of the CK–PCK–context triad, but also refine our understanding of which levers of teacher knowledge and practice are most decisive in varied educational landscapes, offering much-needed guidance for both teacher education and policy design.

A third promising case is the didactical contract, a construct that has proven powerful in explaining the tacit expectations negotiated between teachers and students (Education Committee of the EMS, 2012b). These implicit rules — such as students expecting every problem posed by the teacher to be solvable with the most recent method introduced — have been extensively documented in European research traditions, particularly within the French didactics school (Brousseau et al., 2020). Yet, the empirical basis of the construct remains strongly tied to specific cultural settings and classroom norms, raising the question of how far its explanatory power extends. Replicating studies on the didactical contract in technologically mediated environments, such as online, hybrid, or AI-assisted classrooms, could provide fresh insights into how new modes of communication reshape expectations and ruptures. Likewise, examining the construct in cultural contexts with distinct norms of authority — for example, in classrooms where student initiative is highly valued, or conversely where teacher authority is rarely questioned — would illuminate whether ruptures of the contract continue to serve as productive opportunities for learning. Replication could also focus on how the didactical contract interacts with curricular reforms or assessment regimes that implicitly redefine what students expect from mathematical tasks. However, replication should not be confused with merely applying the same theoretical lens in a new study: two investigations that both draw on the notion of didactical contract are not automatically in a source–replication relationship. Rather, replication would involve taking specific empirical findings and testing whether these phenomena recur under different cultural or institutional conditions. Such work would expand our understanding of the conditions under which the construct remains useful and whether it requires theoretical adaptation to contemporary forms of teaching and learning. In this sense, replicating the didactical contract across diverse contexts would not only test its robustness but also contribute to refining a concept that continues to shape international research on mathematics classrooms.

A fourth example can be drawn from the solid finding on empirical proof schemes, which highlights the persistent tendency of students to validate universal mathematical statements through examples rather than through deductive arguments. This phenomenon shows how students' reasoning is often influenced by everyday ways of thinking. A well-known illustration is the "Student–Professor problem" (S/P-problem), which previously has been researched extensively (for a review of previous studies, see Jankvist & Niss, 2021): when asked to algebraically express the statement "there are six times as many students as professors" using S and P as the number of students and professors, the vast majority of students write $6S = P$ rather than the mathematically correct $6P = S$ (Rosnick & Clement, 1980). In everyday terms, their expression makes sense — it represents that six students correspond to one professor — but mathematically it inverts the intended multiplicative relationship. The role of language makes this phenomenon even more intriguing. In Danish, for instance, the phrase "six times more students than professors" is not equivalent to "six times as many," but is often interpreted additively, as "six more students in addition to the number of professors," corresponding to a total of "seven times as many students as professors." Here, the everyday use of the word "more" points to an additive relationship, which further complicates students' attempts to mathematize verbal statements as illustrated by Jankvist and Niss (2021), who further observe:

> This problem might be different in different languages, though, (for Spanish see also González-Calero et al., 2015) which would warrant a separate study across countries and languages. It is interesting to note that Lopez-Real (1995) by changing the core idiom in the original problem statement to "6 times P is the same as S", a verbalized equation supposed to carry the very same meaning, helped reduce the number of reversal errors considerably. (p. 219)

This observation underscores that difficulties with the S/P-problem cannot be reduced to syntactic translation or conceptual misunderstanding alone, but involve deeper interactions between the translation of verbal descriptions of quantitative situations into simple equations, cognitive framing, and linguistic conventions. Replication studies across languages and cultural settings could help to further clarify the extent to which the reversal error is universal or language-specific, and to what degree instructional strategies that address it in one linguistic context can be transferred to another. Such studies would expand our understanding of how language mediates mathematical translation, and in doing so, would shed light on the broader question of how students learn to mathematize relationships expressed in natural language.

These examples illustrate that replication is not a mere exercise in repeating what has already been done, but a means of expanding the scope, clarifying the boundaries, and strengthening the relevance of solid findings in mathematics education. Declaring studies as replications and designing them deliberately would thus contribute to building a more cumulative knowledge base, ensuring that enduring insights remain central to our collective agenda even as the field continues to seek new directions.

6 In This Issue

The first paper in this issue "Exploring the enactment of reform mathematics textbooks in Mexican public schools" by Cristian Nava-Guzmán, María del Socorro García-González, and Mario Sánchez Aguilar, explores how mathematics teachers in Mexican public schools are engaging with the reform-oriented textbooks introduced under the New Mexican School (NEM) reform. Drawing on implementation research and the notion of "spheres of influence" (Century & Cassata, 2016), the authors analyze responses from 57 teachers across diverse contexts to examine patterns of use, adaptation, and resistance. Their findings reveal both the widespread adoption of the textbooks and the heterogeneity of their enactment: teachers translated and rewrote tasks into Indigenous languages, supplemented activities with locally designed materials, and adjusted the pacing to meet students' needs. At the same time, they faced significant obstacles, including language mismatches, resource scarcity, and large class sizes. Teachers also voiced concrete proposals for improvement, such as clearer sequencing, a broader repertoire of practice tasks, and editions tailored for Indigenous education. By situating these patterns within the framework of spheres of influence, the study not only provides a situated diagnosis of reform enactment in Mexican classrooms but also highlights the broader policy and editorial steps needed to strengthen implementation in multilingual and low-resource settings.

The second paper is by Julia Tsygan — "Learning by Implementing: What Small-scale Implementation Studies Reveal About Instructional Innovations" — and examines what can be learned from small-scale, teacher-led implementations of research-based innovations in the teaching and assessment of mathematical proof. The study reports on two cases where secondary school teachers implemented innovations originally designed for tertiary education: a teaching sequence aimed at shifting students from empirical to deductive proof schemes, and an assessment framework for evaluating proof comprehension. Using a retrospective multiple case study approach, the author analyzes how these teacher-led implementations revealed not only contextual factors influencing success but also opportunities for refinement of the innovations themselves. While both innovations were grounded in research, their effectiveness and practicality varied in secondary school settings, highlighting issues of implementability and clarity of published guidance. The findings underscore the value of implementation-integral research (Cai & Hwang, 2021), where the interplay between teaching and theory development allows classroom practice to inform the refinement of innovations. Small-scale implementations, the paper argues, can generate rich insights by exposing challenges, adaptations, and teachers' perspectives. They not only test the feasibility of innovations in context but also inform theory-building, providing a bidirectional account of how research and practice shape one another.

The third paper is by Linda Marie Ahl, Ola Helenius, Maria Kirstine Østergaard, Uffe Thomas Jankvist, and Johan Prytz — "Beyond Beliefs: Using Scheme Theory to Understand Implementation in Mathematics Education" — and addresses the role of teachers' beliefs in implementation research in mathematics education. While such beliefs are often cited as decisive for the success or failure of innovations, the construct is frequently defined too broadly to offer concrete guidance for implementation design. The authors argue that although research on beliefs has provided valuable insights, it does not fully explain why teachers may support reforms in principle yet still struggle to enact them in practice. To address this limitation, the paper proposes the use of scheme theory (Vergnaud, 1998, 2009), which conceptualizes teachers' practices as organized structures of action consisting of goals, routines, operational invariants, and inferences. Reframing implementation outcomes in terms of schemes allows for the identification of more specific and actionable factors, such as clashes between established routines and new instructional demands, difficulties with content sequencing, or tensions with external testing regimes. By re-analyzing earlier studies, the authors demonstrate how scheme theory provides a more operationalizable perspective than beliefs alone, highlighting influences that extend beyond individual cognition to organizational and contextual domains. They conclude that incorporating scheme theory, alongside other fine-grained perspectives, can yield actionable insights that strengthen the design, support, and sustainability of implementation projects in mathematics education.

Finally, the fourth paper is a replication study by Eirini Geraniou and Nicola Bretscher called "Investigating Teachers' Beliefs on Teaching Mathematics with Technology: A Replication Study in England". This paper is focused on pre-service and early-career mathematics teachers' beliefs about teaching with technology in England. Designed as a close replication of Thurm and Barzel's (2022) study with in-service teachers in Germany, the paper examines three dimensions of teacher beliefs: pedagogical beliefs about technology, self-efficacy, and epistemological orientations, alongside teachers' reported modes of technology use. Using the same survey instrument and analytical procedures, the authors compare the English data with the German findings to test their validity and generalisability in a contrasting educational context. The study confirms that beliefs about multiple representations are central to technology use in both contexts, while self-efficacy beliefs show strong but context-dependent variation, and epistemological beliefs play only a peripheral role. By situating these results within the distinct institutional and curricular conditions of England, the paper highlights the contextual factors that shape technology integration and provides valuable insights for teacher education and professional development aimed at supporting effective and sustainable use of digital tools in mathematics classrooms.

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*(1, Suppl.), S37–S50. https://doi.org/10.1007/s10763-020-10069-7.

Brousseau, G., Sarrazy, B., & Novotná, J. (2020). Didactic contract in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (2nd ed., pp. 197–202). Springer. https://doi.org/10.1007/978-3-030-15789-0_46.

Cai, J., & Hwang, S. (2021). What does it mean to make implementation integral to research? ZDM–Mathematics Education, 53(5), 1149–1162. https://doi.org/10.1007/s11858-021-01301-x.

Century, J., & Cassata, A. (2016). Implementation research: Finding common ground on what, how, why, where, and who. *Review of Research in Education*, *40*(1), 169–215. https://doi.org/10.3102/0091732X16665332.

Cooper, J., Hershkowitz, R., & Koichu, B. (2024). Teachers' pedagogical sensemaking of mathematics education research through replication: A way of increasing implementability of research findings. *Implementation and Replication Studies in Mathematics Education*, *4*(2), 243–281. https://doi.org/10.1163/26670127-bja10024.

Education Committee of the European Mathematical Society. (2011a). "Solid findings" in mathematics education. *Newsletter of the European Mathematical Society*, *81*, 46–48.

Education Committee of the European Mathematical Society. (2011b). Do theorems admit exceptions? Solid findings in mathematics education on empirical proof schemes. *Newsletter of the European Mathematical Society*, *82*, 50–53.

Education Committee of the European Mathematical Society. (2012a). It is necessary that teachers are mathematically proficient, but is it sufficient? Solid findings in mathematics education on teacher knowledge. *Newsletter of the European Mathematical Society*, *83*, 46–50.

Education Committee of the European Mathematical Society. (2012b). What are the reciprocal expectations between teacher and students? Solid findings in mathematics education on didactical contract. *Newsletter of the European Mathematical Society*, *84*, 53–55.

Education Committee of the European Mathematical Society. (2012c). Models and modelling in mathematics education. *Newsletter of the European Mathematical Society*, *86*, 49–52.

Education Committee of the European Mathematical Society. (2013a). Solid findings in mathematics education: living with beliefs and orientations — Underestimated, nevertheless omnipresent, factors for mathematics teaching and learning. *Newsletter of the European Mathematical Society*, *87*, 42–44.

Education Committee of the European Mathematical Society. (2013b). Sociomathematical norms: In search of the normative aspects of mathematical discussions. *Newsletter of the European Mathematical Society*, *88*, 59–61.

Education Committee of the European Mathematical Society. (2013c). Solid findings on students' attitudes to mathematics. *Newsletter of the European Mathematical Society*, *89*, 51–53.

Education Committee of the European Mathematical Society. (2013d). Why is university mathematics difficult for students? Solid findings about the secondary-tertiary transition. *Newsletter of the European Mathematical Society*, *90*, 46–48.

Education Committee of the European Mathematical Society. (2014a). Solid findings in mathematics education: The influence of the use of digital technology on the teaching and learning of mathematics in schools. *Newsletter of the European Mathematical Society*, *91*, 51–53.

Education Committee of the European Mathematical Society. (2014b). Solid findings: Concept images in students' mathematical reasoning. *Newsletter of the European Mathematical Society*, *93*, 50–52.

Fernandez, C., & Yoshida, M. (2004). *Lesson study: A Japanese approach to improving mathematics learning and teaching*. Lawrence Erlbaum.

García-Cerdá, C., Segura, C., & Ferrando, I. (2024). Influence of the type of mathematical problems on students' and pre-service teachers' interest and performance. A replication and elaboration study. *Implementation and Replication Studies in Mathematics Education*, *4*(1), 125–159. https://doi.org/10.1163/26670127-bja10017.

Giberti, C., & Maffia, A. (2022). Primitive model of partitive division: A replication of the Fischbein and colleagues’ study. *Implementation and Replication Studies in Mathematics Education, 2*(2), 149–173. https://doi.org/10.1163/26670127-bja10007.

González-Calero, J. A., Arnau, D. & Laserna-Belenguer, B. (2015). Influence of additive and multiplicative structure and direction of comparison on the reversal error. *Educational Studies in Mathematics, 89*(1), 133–147. https://doi.org/10.1007/s10649-015-9596-0.

Hernández, J. H., Xolocotzin, U., & Aguilar, M. S. (2023). The natural number bias hampers secondary school students in different ways: A conceptual replication of Van Hoof et al. (2013). *Implementation and Replication Studies in Mathematics Education, 3*(2), 275–303. https://doi.org/10.1163/26670127-bja10016.

Hohensee, C., & Borji, V. (2024). Preparing elementary pre-service teachers to teach early algebra. A conceptual replication study. *The Journal of Mathematical Behavior, 75*, Article 101174. https://doi.org/10.1016/j.jmathb.2024.101174.

Huang, R., & Bao, J. (2006). Towards a model for teacher professional development in China: Introducing Keli. *Journal of Mathematics Teacher Education, 9*(3), 279–298. https://doi.org/10.1007/s10857-006-9002-z.

Ikeda, T., & Stephens, M. (1998). The influence of problem format on students’ approaches to mathematical modelling. In P. Galbraith, W. Blum, G. Booker, & I. D. Huntley (Eds.), *Mathematical modelling. Teaching and assessment in a technology-rich world* (pp. 222–232). Horwood.

Jankvist, U. T., Aguilar, M. S., Misfeldt, M., & Koichu, B. (2021). What to replicate? [Editorial]. *Implementation and Replication Studies in Mathematics Education, 1*(2), 141–153. https://doi.org/10.1163/26670127-01010015.

Jankvist, U. T., & Niss, M. (2021). The Students-Professors problem — The reversal error and beyond. *Implementation and Replication Studies in Mathematics Education, 1*(2), 190–226. https://doi.org/10.1163/26670127-01010008.

Kaiser, G., & Maass, K. (2007). Modelling in lower secondary mathematics classroom — Problems and opportunities. In W. Blum, P. Galbraith, H.-W. Henn, & M. Niss. (Eds.), *Modelling and applications in mathematics education. The 14th ICMI Study* (pp. 99–108). Springer. https://doi.org/10.1007/978-0-387-29822-1_8.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology, 100*(3), 716–725. https://doi.org/10.1037/0022-0663.100.3.716.

Liebendörfer, M., & Wagner, K. (2024). Do real-world connections make tasks catch secondary school students’ interest? What preservice teachers expect and what students really experience. *Implementation and Replication Studies in Mathematics Education, 4*(2), 282–314. https://doi.org/10.1163/26670127-bja10023.

Lopez-Real, F. (1995). How important is the reversal error in algebra? In B. Atweh & S. Flauvel (Eds.), *Proceedings of the 18th annual Conference of the Mathematics Education Research Group of Australasia* (pp. 390–396). MERGA.

Melhuish, K. (2018). Three conceptual replication studies in group theory. *Journal for Research in Mathematics Education*, *49*(1), 9–38. https://doi.org/10.5951/jresematheduc.49.1.0009.

Melhuish, K., & Thanheiser, E. (2018). Reframing replication studies as studies of generalizability: A response to critiques of the nature and necessity of replication. *Journal for Research in Mathematics Education*, *49*(1), 104–110. https://doi.org/10.5951/jresematheduc.49.1.0104.

Rosnick, P., & Clement, J. (1980). Learning without understanding: The effect of tutoring strategies on algebra misconceptions. *The Journal of Mathematical Behavior*, *3*(1), 3–27.

Schoenfeld, A. H. (2007). Method. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 69–110). Information Age Publishing.

Schoenfeld, A. H. (2018). On replications. *Journal for Research in Mathematics Education*, *49*(1), 91–97. https://doi.org/10.5951/jresematheduc.49.1.0091.

Star, J. R. (2018). When and why replication studies should be published: Guidelines for mathematics education research journals. *Journal for Research in Mathematics Education*, *49*(1), 98–103. https://doi.org/10.5951/jresematheduc.49.1.0098.

Star, J. R. (2021). In pursuit of a replication culture in mathematics education. *Implementation and Replication Studies in Mathematics Education*, *1*(1), 53–76. https://doi.org/10.1163/26670127-01010003.

Stillman, G. (2002). The role of extra-mathematical knowledge in application and modelling activity. Teaching Mathematics 27(2), 18–31.

Thurm, D., & Barzel, B. (2022). Teaching mathematics with technology: a multidimensional analysis of teacher beliefs. *Educational Studies in Mathematics*, *109*(1), 41–63. https://doi.org/10.1007/s10649-021-10072-x.

Van Der Auwera, S., Mathys, L., De Smedt, B., Torbeyns, J. & Verschaffel, L. (2021). Upper elementary school children's adaptive use of subtraction by addition: A choice/no-choice replication study involving two choice conditions. *Implementation and Replication Studies in Mathematics Education*, *1*(1), 111–138. https://doi.org/10.1163/26670127-01010005.

Vergnaud, G. (1998). Towards a cognitive theory of practice. In A. Sierpinska & J. Kilpatrick, *Mathematics education as a research domain: A search for identity* (pp. 227–240). Springer. https://doi.org/10.1007/978-94-011-5190-0_14.

Vergnaud, G. (2009). The Theory of Conceptual Fields. *Human Development*, *52*(2), 83–94. https://doi.org/10.1159/000202727.