Professional development as a means for implementing mathematics education innovations: results from a systematic review

Linda Marie Ahl, Mario Sánchez Aguilar, Uffe Thomas Jankvist, Morten Misfeldt and Johan Prytz

To get an overview of the characteristics of the studies in mathematics education research that explicitly state that they deal with implementation, we have conducted a systematic review. In this paper, we report on a subset of the identified studies from the review, dealing with large-scale professional development for teachers. For the subset of the 11 identified papers, we ask the question: What designs are used to support teachers to adopt new ideas in their practice and what dimensions of scaling are considered in the studies? To articulate design and dimensions of scaling, we draw on theoretical constructs from both mathematics education research and more general implementation research. Results indicate that the choice of facilitating strategy impacts the dimensions of scaling considered in the implementation.

> Implementation research, by our working definition, is the systematic inquiry of innovations enacted in controlled settings or in ordinary practice, the factors that influence innovation enactment, and relationships between innovations, influential factors, and outcomes.
>
> (Century & Cassata, p. 181)

That implementation research (IR) in mathematics education research (MER) has gained momentum during the past few years is beyond any doubt. Since 2017, a thematic working group (TWG 23) at CERME has been dedicated to the topic. In 2021, a new journal – *Implementation and Replication Studies in Mathematics Education* (IRME) – was launched by the well-established Dutch publishing house, Brill. In addition, in 2021 a special issue of ZDM was dedicated to the topic of implementation research in mathematics education (Koichu et al., 2021). Ongoing discussions in relation to IR in MER concern,

Linda Marie Ahl, Uppsala University
Mario Sánchez Aguilar, National Polytechnic Institute of Mexico
Uffe Thomas Jankvist, Aarhus University
Morten Misfeldt,University of Copenhagen
Johan Prytz, Uppsala University

for example, the use of theoretical constructs from outside the field of MER (e.g. health science, economics, etc.) versus those available inside of MER; what we should take *implementability* to mean in relation to IR in MER; to what extent IR should mainly address large scale studies; etc. (Jankvist et al., 2021). Yet, it seems to us that to engage in these discussions on an enlightened basis, a natural starting point is to get an overview of both the numerosity and type of studies in the MER literature which specifically addresses ”implementation”. We have taken on this task by conducting a systematic literature review of the field of implementation research on educational reforms in mathematics education.

In all educational reforms, teachers are viewed as agents of change and thus are expected to play a key role in changing schools and classrooms (Prawat, 1992). In this paper, we focus on a smaller subset of papers from the systematic review, reporting on professional development (PD) programs for teachers.[1] While almost all papers in our review address some kind of PD for teachers, our sample consist of papers that foreground the PD and also discusses scaling of the PD. Papers touching on PD that foreground a curriculum reform or new curriculum materials are not included in the sample. Furthermore, papers that do not report on any approach to scaling are also not included in the sample.

We are interested in what kind of different designs for facilitating change in teacher practice that are used, and what kinds of scaling with the aim to create lasting change that are planned for, in the implementation projects reported on in our sample. The question under investigation still needs definitions of some notions before it can be formulated coherently. Therefore, the specific research question will be presented after the section ”IR theoretical constructs applied”.

Review methodology

Conducting a systematic review on implementation research involved a few delicate considerations on our behalf, not least since a large portion of the research studies in MER may be considered studies addressing implementation. We settled on two inclusion criteria to avoid a too large number of papers. Firstly, we limited the review to include papers that clearly stated to be dealing with some kind of implementation. Secondly, we limited the review to only consider studies published in the top twenty quality-ranked MER journals following the recent journal categorization by Williams and Leatham (2017).

We conducted the literature searches in ERIC (EBSCO) searching for manuscripts with implement* in the title and/or abstract, journal by journal of our top 20 samples (Williams & Leatham, 2017). The advantage of doing the entire search in one database is that it is easy to collect the results in one folder. To ensure that no article had been overlooked, we repeated the search implement* in the title and/or abstract on each journal’s website. We found 1,093 peer-reviewed articles fitting the search criteria. We used the software *Covidence* to manage our literature review.

Each paper was screened by two reviewers. The screening was made in two steps. First, we screened the title and abstract. In cases where we were hesitant, e.g. because the abstract did not provide sufficient information, we chose to forward the paper to full-text screening. A total of 138 papers were forwarded to full-text screening. In the full-text screening, papers were included if they were in line with Century and Cassata's (2016) definition of IR:

> [...] the systematic inquiry of innovations enacted in controlled settings or in ordinary practice, the factors that influence innovation enactment, and relationships between innovations, influential factors, and outcomes.
>
> (Century & Cassata, p. 181)

As evident from this quote, another central term in IR is that of *innovation*. Innovation refers to the practical implementation of ideas resulting from research that involve a change (e.g. in behavior or practice) for the individuals enacting them (Century & Cassata, 2016).

Of the 138 papers, 95 remained after the full-text screening. To obtain smaller and more manageable units, these were categorized as: Instructional sequences on mathematical concepts, and/or competencies, (19); Curriculum materials (21); Professional development (PD) projects (25); and Curriculum reform (30). There are no clear cuts between the categories. An instructional sequence may stem from a new curriculum material that is implemented through a PD project due to curriculum reform. The category foregrounded in each of the 95 papers was decisive for how the categorization was done.

The data extraction from the papers included general information on the author(s), title, purpose statement(s), country where the study was conducted, research question(s), methods, target group, and results. The specific information about the implementations included what kind of innovation from mathematics education the study concerned, specific or general goals of short term or long term, phase of the implementation studied, stakeholders responsible for the implementation, and identified factors of influence for the outcomes of the implementation.

From the subset of the 25 papers concerning professional development programs, we report on the 10 empirical papers from the subset, discussing design for scaling. We also included 1 theoretical paper that, in line with our focus, discusses support for teachers' long-term use of research-based instruction in large-scale projects (Cobb & Jackson, 2015). In the next section, we will elaborate on design for implementing instructional change through PD programs and define long-term goals as well as four dimensions of scaling.

IR theoretical constructs applied

While the form and content of PD programs come in different shapes, the unifying goal is to implement innovations that increase students' knowledge

through improved teaching practice. We use Kennedy's (2016) classification of PD designs, *prescription, strategies, insight,* and *body of knowledge*, to categorize the form of PD programs in our review. *Prescription* refers to the implementation of scripted instructions for teachers to follow, for changing practice. *Strategies* refer to "toolboxes" from which teachers can choose different strategies to address specific goals for practice, e.g. a problem-solving strategy. *Insight* rests on the idea that increased knowledge exemplified by teaching practices can give teachers tools to change practice. *Body of knowledge* carries the idea that if teachers gain more knowledge about mathematics, didactics, and pedagogy (e.g. MKT, PCK) they will be able to plan and implement better teaching. Body of knowledge often consist of regular university courses or lectures. The different designs for helping teachers enact new ideas within their own ongoing systems of practice set different requirements for how innovations may survive over time. Regardless of the choice of facilitating design for creating change in teaching practice, it is necessary to plan for how the innovation shall survive over time, if one has long-term goals for the implementation.

Long-term goals refer to innovations that intend to change the nature of mathematics teaching practice in a sustained way. For example, as a result of alarms from international tests, politicians may plan for increasing the mathematics teachers' general content knowledge at scale and/or state-wide curriculum reforms to be implemented with long-term goals. On the other hand, small-scale and pilot studies with the goal to try out innovations over a given time span provide valuable insights for future action, implementation studies with long-term goals may need to consider dimensions of scaling (Coburn, 2003).

Our definition of scaling follows Coburn's (2003) notions of *depth, sustainability, spread,* and *shift in reform ownership. Depth* refers to change in classroom practice that goes beyond a shift in teaching resources and the introduction of specific activities. Coburn argues that scaling in depth includes a shift in teachers' beliefs, norms for communication, and pedagogical practices. *Sustainability* concerns the scaffolding that is left to maintain the vitality of the innovation after the support of the reform leaders is withdrawn from the organization. When Coburn considers *spread*, she, in addition to scaling to other schools and classrooms, also includes spread within the organization. Finally, Coburn adds the dimension of a *shift in reform ownership* to the notion of scale. When reform is launched, the ideas and activities are owned by the creators of the reform. According to Coburn, the authority to scale the implementation needs to shift to the districts, schools, and teachers. Only then can scaling in-depth, sustainability, and spread be maintained.

We are now ready to present our research question. Following Kennedy's (2016) designs for PD and Coburn's (2003) dimensions for scaling, we ask: *What designs are used to support teachers to adopt new ideas in their practice and what dimensions of scaling are considered in the studies?*

Results

In table 1, we summarize the answers to the posed question above for the 10 empirical papers in our sample. Due to space limitations, it is not possible to summarize the research design for each study. Instead we provide information about the target group to give some idea of the context in which the innovation was implemented. We follow up by describing what kind of scaling dimension of PD that is in play in terms of each of the four facilitating designs. We close the result section by reflecting on scale from the view put forward in the theoretical paper in our sample.

Table 1. *Characteristics of PD-programs*

Facilitating Strategy for PD	Characteristics of Scale	Author(s)	Target population(s)	Innovation
Prescription	Depth Spread	(Clements et al., 2011)	Preschool teachers	A research-based curriculum ”the building blocks” carries the prescription of learning trajectories for teaching number sense and geometry.
Prescription	Depth Spread	(Corcoran, 2018)	In-service teachers in K-5	Teaching for conceptual understanding and procedural fluency.
Strategies	Depth Sustainability Spread A shift in reformed ownership	(Clark-Wilson & Hoyles, 2019)	In-service teachers in lower secondary school	Algebraic patterns and expressions, linear functions, and geometric similarity.
Strategies	Depth Spread	(Swan, 2007)	In-service teachers with post 16 students	Reformed teaching
Insight	Depth Spread	(Ferrini-Mundy et al., 2007)	In-service teachers in grades K-8	Mathematical knowledge for teaching (MKT), curricular coherence, and learning trajectories.
Insight	Depth Sustainability Spread	(Higgins & Parsons, 2011)	In-service teachers in primary and middle school	Numeracy
Insight	Depth Spread	(Jankvist & Niss, 2015)	Upper secondary school teachers with a master’s degree	Conceptual knowledge, modeling and, reasoning and proof.
Insight	Depth Sustainability Spread A shift in reformed ownership	(Prediger et al., 2019)	In-service middle school teachers	Basic conceptual understanding, research-based materials, and community-based collaboration.
Body of knowledge	Spread	(Buchholtz & Kaiser, 2013)	Prospective teachers for secondary school	Mathematical content knowledge (MCK) and mathematical pedagogical content knowledge (MPCK).
Body of knowledge	Depth Spread	(Gainsburg, 2013)	Prospective K-12 teachers	Reformed teaching

Scaling in programs using prescription

Scaling in-depth, in terms of fidelity to the program theory, is discussed in both PD programs using prescription as a facilitating design for change in practice (Clements et al., 2011; Corcoran, 2018). Both studies also scale by spread to many schools. The sample from Clements et al. (2011) comes from 10 school districts, 42 schools, and 106 classrooms. The study is a scaling up of the program *Building blocks*. Corcoran's sample includes 11 schools from one school district, but the program, the *ORIGO Stepping stone*, is used by more than 100 school districts with more than 450,000 students. Scaling by sustainability and shift in reform ownership are not discussed in these studies.

Scaling in programs using strategies

The PD-program *Cornerstone maths* addresses all dimensions of scaling (Clark-Wilson & Hoyles, 2019). Sustainability of the innovation and spread within the school is catered for by a web-based professional development toolkit, to maintain scaling beyond the timeline of the funded project. The design of the toolkit aims to provoke a rethinking of mathematics and challenge existing beliefs. Altogether, the design aims to facilitate a shift in reform ownership. Scaling in depth by challenging teachers' beliefs is also one of the goals with the task-based PD presented by Swan (2007). By pull-out workshops for facilitators from different schools, and a toolbox of tasks to use in practice, the program aims to spread to new schools and within schools.

Scaling in programs using insight

Scaling by depth is a component of all programs in our subset using the facilitating design insight, namely *PROMSE* (Ferrini-Mundy et al., 2007), *Numeracy development project* (Higgins & Parsons, 2011), *Maths counsellor* (Jankvist & Niss, 2015), and *Mastering math* (Prediger et al., 2007). The core of the programs is that the new knowledge should provide insights that give participants a changed, or deepened, view of what mathematics teaching is. All four programs use facilitators for spread within schools. Either the program aims at educating facilitators (Ferrini-Mundy et al., 2007; Jankvist & Niss, 2015) or use external facilitators employed to assist teachers in practice (Higgins & Parsons, 2011; Prediger et al., 2007).

The report from the Mastering maths program has been designed for sustainability and a shift in reform ownership. Design for scaling up with sustainability is reported from the Numeracy development project (Higgins & Parsons, 2011), but not a shift in reform ownership. Sustainability and shift in reform ownership are not discussed in the studies by Ferrini-Mundy et al. (2007) and Jankvist and Niss (2015).

Scaling in programs using a body of knowledge

Innovations implemented in university courses for prospective teachers have the body of knowledge as design for facilitating change in practice, i.e., the studies by Buchholtz and Kaiser (2013) and Gainsburg (2013). Scaling, as the spread of innovations, goes through the graduated prospective teachers into their classrooms. Depth, e.g. changed beliefs on what mathematics teaching should be, is discussed as an important factor for the spread in the study by Gainsburg (2013). Sustainability and shift in reformed ownership are not discussed.

Scaling as a theoretical reflection on experience

In their theoretical paper, based on experiences from different PD programs and existing literature on the subject, Cobb and Jackson (2015) address all dimensions of Coburn's scaling. The authors suggest that the depth of teachers' learning involves:

> (a) clarifying the goals for teachers' learning with respect to the products of the classroom design study, and (b) documenting teachers' current instructional practices and relevant forms of knowledge and conceptions about teaching and learning. (p. 1029)

For the sustainability of innovation, the same group of teachers must get opportunities to continue to collaborate. For spread, an organizational level for extending dissemination designs is important. An extending dissemination design on the organizational level is also necessary for an adequate plan for a shift in ownership.

The authors conclude that PD programs, where teachers leave their classrooms to learn how to change their practice, are not enough for a successful implementation. "We argue that high-quality pull-out professional development is essential but not sufficient, and go on to consider teacher collaboration and one-on-one coaching in the classroom as additional supports." (Cobb & Jackson, 2015, p. 1027)

Discussion

We asked the question: What designs are used to support teachers to adopt new ideas in their practice and what dimensions of scaling are considered in the studies? We found that the facilitating design for supporting teachers to enact new ideas within their practice spread over Kennedy's (2016) four categories. The papers were distributed as follows: prescription (2 papers); strategies (2 papers); insight (4 papers); and, a body of knowledge (2 papers). We can of course only hypothesize based on this small sample, but from the dimension of scaling that was addressed in these papers, scaling seems to align with the choice of design for change in teaching practice.

Depth as in shift in teachers' beliefs, norms for communication, and pedagogical practices is crucial for studies using prescription as facilitating design. Prescription calls for fidelity to program theory in teachers' implementation. Yet, while teachers who both believe in and know the program theory will be likely to implement the program with fidelity to its core ideas, teachers that disbelieve or lack knowledge will focus on surface manifestations and circumvent the pedagogical ideas (Gregoire, 2003). Conversely, striving to understand the intentions behind a program and implement it with fidelity seem to enhance a shift in teachers' beliefs, norms for communication, and pedagogical practices (Guskey, 1986). It is uncertain what precedes what, but it may be the case that depth can arise both due to the teachers already having beliefs in line with the program ideas, or because the teachers carry out the program according to the prescriptions, i.e., in line with the model by Clarke and Hollingsworth (2002).

PD programs facilitating by applying the designs strategies and insight seem to have a broader approach to the dimensions of scale. All four dimensions of scale are discussed in the PD program Mastering Math (Prediger et al., 2019). In the other studies, facilitating by applying the design strategies or insights, one gets implicit information of considerations of scaling in all four dimensions. It is of course reasonable to assume that the authors have chosen to foreground certain parts of the program, leaving others out.

For PD programs using a body of knowledge as a design for teacher change, the scaling opportunities seem to narrow down. Sustainability and shift of reform ownership are not discussed. Perhaps this is due to to the implementers losing control of future development of the innovation, when a university course is completed.

Results from the students to the participating teachers are not measured and reported in six of the studies. One reason may be that while it is doable to observe a change in teaching practice or teacher knowledge and beliefs, it is more complicated to measure what effect the changed practice has on students. Or it is the focus of another paper. For example, both studies using prescription (Clements et al., 2011; Corcoran, 2018) report on students' outcomes. Also, the studies on PD programs relying on a body of knowledge reports on students (Buchholtz & Kaiser, 2013; Gainsburg, 2013).

The study from Clements et al. (2011) shows positive results on student-level together with acceptable fidelity to the program theory in teachers' implementation. The study from Corcoran (2018), on the other hand, reports weak to nonexisting differences between the experimental groups and the control group. A lack of fidelity to the program theory may, according to the author, explain these results.

Buchholtz and Kaiser (2013) measure teacher students' mathematical content knowledge and mathematical pedagogical content knowledge after taking innovative courses for prospective teachers. Gainsburg (2013) observes whether or not recent graduates implement program-emphasized teaching

practices in their classrooms. Both studies working with the design body of knowledge report on weak results for university courses to equip teacher students for reformed teaching.

There are no clear cuts between the four categorizations used in our review: Instructional sequences on mathematical concepts, and/or competencies, (19); curriculum materials (21); professional development (PD) projects (25); and curriculum reform (30). Since most PD programs include more than one category, we needed to decide on the most prominent one in the paper. The same reasoning can be applied to Kennedy's designs and Coburn's categorizations. Large programs often cover several designs, which is why our categorization is grounded in what is written in the papers. Further information in the PD programs than what is provided in the papers might of course change the picture.

Of the 25 papers categorized as PD programs, only 10 empirical papers and 1 theoretical paper discussed scaling of the innovation. It is reasonable to assume that many of the other PD programs reported on in the literature had dimensions of scale as well, yet without discussing it explicitly. In the next cycle of our review, we shall leave the systematic approach and move on to a heuristic approach. The insights from the systematic review shall guide us to search for theoretical papers together with all the 95 papers from the systematic review. With this fuller picture as a basis, we hope to be able to address the question of how to potentially create an empirically founded theoretical framework for IR in MER concerning large-scale development programs.

Acknowledgments

This paper is part of grant 2020-04090 under the Swedish Research Council.

References

Buchholtz, N. & Kaiser, G. (2013). Improving mathematics teacher education in Germany: empirical results from a longitudinal evaluation of innovative programs. *International Journal of Science and Mathematics Education*, 11 (4), 949–977. doi: 10.1007/s10763-013-9427-7

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. doi: 10.3102/0091732X16665332

Clarke, D. & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, 18 (8), 947–967. doi: 10.1016/s0742-051x(02)00053-7

Clark-Wilson, A. & Hoyles, C. (2019). A research-informed web-based professional development toolkit to support technology-enhanced mathematics teaching at scale. *Educational Studies in Mathematics*, 102 (3), 343–359. doi: 10.1007/s10649-018-9836-1

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A. & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: a large-scale cluster randomized trial. *Journal for Research in Mathematics Education*, 42 (2), 127–166. doi: 10.5951/jresematheduc.42.2.0127

Cobb, P. & Jackson, K. (2015). Supporting teachers' use of research-based instructional sequences. *ZDM*, 47 (6), 1027–1038. doi: 10.1007/s11858-015-0692-5

Coburn, C. E. (2003). Rethinking scale: moving beyond numbers to deep and lasting change. *Educational Researcher*, 32 (6), 3–12. doi: 10.3102/0013189X032006003

Corcoran, R. P. (2018). Preparing teachers to raise students' mathematics learning. *International Journal of Science and Mathematics Education*, 16 (6), 1169–1185. doi: 10.1007/s10763-017-9819-1

Ferrini-Mundy, J., Burrill, G. & Schmidt, W. H. (2007). Building teacher capacity for implementing curricular coherence: mathematics teacher professional development tasks. *Journal of Mathematics Teacher Education*, 10 (4-6), 311–324. doi: 10.1007/s10857-007-9053-9

Gainsburg, J. (2013). Learning to model in engineering. *Mathematical Thinking and Learning*, 15 (4), 259–290. doi: 10.1080/10986065.2013.830947

Guskey, T. R. (1986). Staff development and the process of teacher change. *Educational Researcher*, 15 (5), 5–12. doi: 10.3102/0013189x015005005

Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational Psychology Review*, 15 (2), 147–179. doi: 10.1023/a:1023477131081

Higgins, J. & Parsons, R. (2011). Improving outcomes in mathematics in New Zealand: a dynamic approach to the policy process. *International Journal of Science and Mathematics Education*, 9 (2), 503–522. doi: 10.1007/s10763-011-9275-2

Jankvist, U. T., Aguilar, M. S., Misfeldt, M. & Koichu, B. (2021). Launching Implementation and replication studies in mathematics education (IRME). *Implementation and Replication Studies in Mathematics Education*, 1 (1), 1–19. doi: 10.1163/26670127-01010001

Jankvist, U. T. & Niss, M. (2015). A framework for designing a research-based "maths counsellor" teacher programme. *Educational Studies in Mathematics*, 90 (3), 259–284. doi: 10.1007/s10649-015-9629-8

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, 86 (4), 945–980. doi: 10.3102/0034654315626800

Koichu, B., Aguilar, M. S. & Misfeldt, M. (2021). Implementation-related research in mathematics education: the search for identity. *ZDM*, 53 (5), 975–989. doi: 10.1007/s11858-021-01302-w

Prawat, R. S. (1992). Teachers' beliefs about teaching and learning: a constructivist perspective. *American Journal of Education*, 100 (3), 354–395. doi: 10.1086/444021

Prediger, S., Fischer, C., Selter, C. & Schöber, C. (2019). Combining material- and community-based implementation strategies for scaling up: the case of supporting low-achieving middle school students. *Educational Studies in Mathematics*, 102 (3), 361–378. doi: 10.1007/s10649-018-9835-2

Swan, M. (2007). The impact of task-based professional development on teachers' practices and beliefs: a design research study. *Journal of Mathematics Teacher Education*, 10 (4-6), 217–237. doi: 10.1007/s10857-007-9038-8

Williams, S. R. & Leatham, K. R. (2017). Journal quality in mathematics education. *Journal for Research in Mathematics Education*, 48 (4), 369–396. doi: 10.5951/jresematheduc.48.4.0369

Note

1 We are aware that there is a plethora of papers discussing PD-programs. However, since it is "implementation" that is in focus here we do not aim to review all papers on PD-programs from our 20-journal sample.