BRILL

IMPLEMENTATION AND REPLICATION STUDIES IN MATHEMATICS EDUCATION 3 (2023) 1–24

IRME

brill.com/irme

Towards Characterization of Scale and Scaling in Implementation Research within Mathematics Education

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

Linda Marie Ahl | ORCID: 0000-0002-1928-7537
Department of Education, Uppsala University, Uppsala, Sweden
linda.ahl@edu.uu.se

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

Ola Helenius | ORCID: 0000-0002-0609-7064
Department of Pedagogical, Curricular and Professional Studies, University of Gothenburg, Gothenburg, Sweden
ola.helenius@ncm.gu.se

Abstract

Scale and scaling are central concepts in projects aimed at implementing innovations for improving mathematics education. This study aims to provide conceptual clarity on the notions of scale and scaling in mathematics education research. The meanings of these terms are handled tacitly, and there is diversity in how they are used in the literature. To address this, we gathered opinions from experienced researchers in the area of the large-scale implementation of educational innovations, and we also conducted a literature review of articles explicitly addressing aspects of scale and scaling in mathematics education. We analyzed the content of these sources using Coburn's (2003) conceptualization of scale. Our findings provide insight into how experts define and understand the concepts of *small scale*, *large scale*, *at scale* and *scaling*, and how these concepts relate to the four dimensions of scale proposed by Coburn. This

Published with license by Koninklijke Brill NV | DOI:10.1163/26670127-BJA10012
 | ISSN: 2667-0135 (print) 2667-0127 (online)

study contributes to developing a more precise and nuanced understanding of scale and scaling in mathematics education research by offering an operational distinction between small-scale, large-scale, and medium-scale project implementation. This distinction is based on three key factors: the contact factor, the material factor, and the organizational factor.

The impact sheet to this article can be accessed at 10.6084/m9.figshare.22339609.

Keywords

implementation research – small scale – medium scale – large scale – scaling up – at scale

1 Introduction

Although the notions of "scale" and "scaling" are widely used in educational research (e.g., Coburn, 2003; Century & Cassata, 2016), their meanings tend to be handled tacitly, and there is great diversity in how they are used (Morel et al., 2019). The field of mathematics education research is not an exception in this regard, particularly in terms of the implementation of (educational) innovations. With a few exceptions (e.g., Krainer, 2015), notions such as "scaling up," "large scale" and "at scale" are used implicitly, sometimes even interchangeably, thus giving rise to a variety of interpretations and potential misinterpretations of the terms. By "innovation," we refer to Rogers's (2003) definition of "an idea, practice, or object that is perceived as new by an individual or other unit of adoption" (Rogers, 2003, p. 12), and "implementation takes place when an individual puts an innovation into use" (Rogers, 2003, p. 20).

In pursuit of conceptual clarity, we have undertaken a Delphi-inspired study exploring the notions of large-scale and small-scale implementation projects, as well as implementation projects at scale, which implicitly involve aspects of scaling up as well. We refer to these in unison as "scale and scaling." A *Delphi study* is a research method that involves collecting and summarizing opinions from a panel of experts to achieve consensus or identify trends on a particular topic. The Delphi-inspired study is supplemented with a literature review, taking into account the literature specifically addressing aspects of scale and scaling in mathematics education research. The following questions guide the study:

– *How might we think about and understand matters of scale and scaling in relation to innovation implementation in mathematics education research?*

– *What characteristics of scale and scaling might we differentiate in relation to various size implementation projects?*

Although it may be impossible to answer these questions in a definite manner, we use this opportunity to provide an approximation based on the analysis of papers published in specialized journals and the opinions of experienced researchers. To understand the different meanings of scale and scaling at play, we first provide a conceptual framework. As a lens to capture different understandings and descriptions of scale in our data, we use Coburn's (2003) conceptualization of scale, which embraces both quantitative and qualitative aspects. In the section following the presentation of the conceptual framework, the Delphi-inspired method, the literature review, and the data analysis are described. Next, we present our results on the definitions of small scale, large scale, and scaling. We also clarify whether "large scale" and "at scale" are used synonymously in our data. Based on the discussion of our results, we propose a matrix for identifying characteristics of small-scale and large-scale projects, drawing on three factors of implementation projects: the contact factor, the material factor, and the organizational factor. Lastly, the identified characteristics lead us to introduce the notion of medium scale to fill an existing gray area on the border between small-scale and large-scale projects.

2 Conceptual Framework

According to Coburn (2003), the most common take on scale in research deals with scaling up rather than scale as such, and the common understanding of scale in most research reviewed by Coburn involves "solely or predominantly the expansion of numbers of schools reached by a given reform effort" (Coburn, 2003, p. 3). This is also exemplified by Stringfield and Datnow (1998), who described scaling as "the deliberate expansion to many settings of an externally developed school restructuring design that previously has been used successfully in one or a small number of school settings" (p. 271). Of note is that this notion of scaling up posits an initial small-scale implementation. In addition, it deals only with the quantity of participants, specifically of participating schools. Based on a review of publications on scaling as well as examples of reform implementation, Coburn suggested a conceptualization of *scale* that goes beyond merely spreading to a larger number of participants. In addition to the spread dimension, Coburn suggested three further dimensions to consider when discussing scale. The additional dimensions concern the quality of scale rather than just the quantity: *sustainability*, which deals with a temporal aspect; *depth*, which deals with the innovation permeating professional life;

and *shift in reform ownership*, which deals with who is responsible for the continued implementation of the innovation. Below, we explain Coburn's take on the four dimensions.

2.1 *Depth*

With the depth dimension, Coburn intended to reinforce that for a reform to be at scale, the changes in classroom practice must go beyond surface structures, such as the use of specific materials or a specific teaching procedure. The depth dimension thus focuses on the quality and impact of reforms within individual classrooms or schools. In fact, the depth dimension even considers the impact within individuals, since Coburn here referred to changes in teachers' beliefs concerning, for example, the social interaction norms and pedagogical principles that underlie the enacted curriculum. Coburn further emphasized that taking the depth dimension seriously affects researching scaling efforts, since classroom changes as well as the teachers' underlying beliefs should be assessed.

2.2 *Sustainability*

The dimension of sustainability refers to aspects of reform implementation that act to maintain their impact over time, even in the face of various changes, such as shifts in politics, finances, or organizational structures. Coburn pointed out that neither theoretical discussions of scale nor most empirical reports have focused on follow-up over extended timeframes. This narrow scope of investigation has left many questions unanswered regarding the long-term effectiveness of reforms and the factors that contribute to sustainability. The measurement of sustainability involves assessing the potential for reforms to endure beyond their initial implementation and to continue to create a lasting impact. For sustainability to be achieved, critical conditions must be present within schools and school districts to support and maintain classroom changes over an extended period of time.

2.3 *Spread*

In relation to the dimension of spread, research on scale has traditionally focused on external expansion to a larger number of schools. Coburn suggested broadening the notion of spread to include spread within the system, that is, to also consider the spread of reform or innovation within a classroom, a school, or a district. Coburn noted that among the research she reviewed, a few models could be drawn upon for guidance when conceptualizing such within-system spread, considering that within-system spread might pose new challenges for reform implementers and program designers, as well as for researchers. One

strategy suggested for researching this dimension of spread is to use insights from implementation research, which investigates the enactment of pedagogical principles and shifting norms of classroom interaction, thereby also giving the spread dimension a tighter connection to the depth dimension. By applying these insights to different levels of the system, researchers can develop a more nuanced understanding of how reforms can spread effectively within and across different levels of the education system.

2.4 *Shift in Reform Ownership*

A shift in reform ownership concerns the transfer of control and ownership of reforms from external organizations to teachers, schools, or districts. Coburn argued that such a shift is a critical aspect of the dimension of scale. This ownership dimension concerns whether the reforms are becoming ingrained within the organizational culture and whether they are being led and sustained by those who are directly affected by them. Placing reform ownership at the center of scale raises the priority of directing reform attention and resources toward strategies that enable schools and districts to assume ownership of the reform over time. Implementation program designers (and reform developers) need to identify effective strategies to cultivate the capacities necessary for assuming authority for reform, and consider whether these strategies should vary at different levels of the system and for different kinds of reforms. Additionally, implementation program designers need to lay the groundwork for a shift from external to internal ownership from the early days of engagement with a school or district. The shift in reform ownership has, according to Coburn's literature review, rarely been incorporated into studies of attempts to implement reform initiatives at scale, but existing research exists, some of which include an assessment of:

> (a) the presence of structures and mechanisms for ongoing teacher learning about reform (e.g., professional development, teacher study groups); (b) the presence of established strategies to provide continued funding for reform activities; (c) the degree to which districts have taken responsibility for continued spread of reform; and (d) the use of reform-centered ideas or structures in school or district decision making. (Coburn, 2003, p. 8)

Using this four-dimensional conceptualization of scale as a reference, we explored the interpretations of scale and scaling present in mathematics education research. The following section explains the method used in conducting such a conceptual exploration.

3 Methodology

Data for our investigation on the conceptualizations of scale and scaling were collected from two samples. First, we used a Delphi study-inspired method to collect opinions from experts in the field of large-scale implementation of educational innovations (mainly, but not only, in mathematics education). Second, we complemented the gathered information with a literature review based on two special issues related to the notion of scale. We will first describe the adapted Delphi technique and provide an account of how we carried out the literature review. We then describe how we operationalize Coburn's concepts of scale to analyze the answers from our participating experts and the articles in our literature review.

3.1 *The Delphi Technique and the Procedures Implemented*

The Delphi technique is a method used to gather expert opinions or predictions on a specific issue. Its use as a research technique is not new in mathematics education (e.g., Kallia et al., 2021; Manizade & Mason, 2011). It is a systematic, interactive forecasting process that utilizes a panel of experts to provide input through several rounds of anonymous questionnaires. The purpose of the Delphi technique is to achieve a consensus or to converge toward a common understanding, based on the experts' knowledge and experience (Barrett & Heale, 2020).

The Delphi technique typically involves the following methodological procedures:

1. Identify and select a panel of experts in the relevant field.
2. Send an initial questionnaire to the experts to gather their initial opinions and predictions.
3. Compile and summarize the results from the first round of questionnaires.
4. Send a second questionnaire to the experts, incorporating the results of the first round and asking them to revise their opinions in light of the results.
5. Repeat this process until a consensus is reached or there is no further change in the opinions of the experts.

The current study followed Procedures 1, 2 and 3 of the Delphi technique.

In Procedure 1, based on our experience and knowledge of the educational field and its literature, we identified and selected a group of 14 educational researchers from different regions of the world who had experience in leading or investigating the large-scale implementation of educational innovations or reforms. This set of identified experts was invited to participate in developing the second procedure of the Delphi technique.

In Procedure 2, the 14 identified researchers were invited via email to answer two scale-related questions. All invited researchers agreed to participate in the study. They come from Europe, Latin America, and North America. Some decided to participate anonymously, while others consented to us using their names. We explained to the researchers that the questions posed and their answers were part of the development of a review of implementation projects that included small-scale and large-scale projects (Ahl et al., 2023). However, deciding whether an implementation project is small-scale or large-scale was difficult, and some could fall between the two categories. With this context as a background, the researchers were asked the following two questions:

– How do you define a large-scale project or a project at scale?
– Are "large-scale" and "at scale" synonyms?

As part of Procedure 3, the researchers' email responses were concentrated into a shared document for data analysis.

3.2 *Literature Review*

To approximate how the notions of "large scale," "small scale," "at scale" and "scaling" are interpreted in mathematics education research, 22 articles included in two special issues related to implementing and scaling innovations were reviewed. These special issues were published in the journals *Educational Studies in Mathematics* and *ZDM — Mathematics Education*. More precisely, we refer to the special issues on "*Different ways to large-scale implementation of innovative teaching approaches*" (Maass et al., 2019a, 8 articles) and "*Evidence-based CPD: Scaling up sustainable interventions*" (Roesken-Winter et al., 2015a, 14 articles). These issues were selected because they include international studies involving the notion of scale in mathematics education — with particular emphasis on large scale. Furthermore, to the best of our knowledge, these are the most recent special issues on the topic.

In addition, the book *Large-Scale Studies in Mathematics Education* was published in 2015 (Middleton et al., 2015). Here, what is meant by large scale is discussed. However, the focus is on whether the *evaluation* of implementation projects is done by large-scale studies, following Anderson and Postlethwaite's (2007) factors of scale: "(1) Sample size, (2) purpose of the research, (3) generalizability of results, (4) type and complexity of data analysis, and (5) cost" (Middleton et al., 2015, p. 3). Since we are primarily interested in determining whether an implementation project is small or large, Middleton et al.'s (2015) discussion does not help our cause. Hence, we have left it out of this review.

The articles were read and analyzed by the first and second authors. We used the software tool *Covidence* to manage our review. We extracted data on definitions of scale and explanations of what was referred to when talking

about small scale, large scale, scaling, and at scale. We also extracted data on how the scale was described in the studies in relation to Coburn's four dimensions of scale.

3.3 *Data Analysis*

As previously mentioned, the analysis of the empirical data (the experts' responses and the articles from the special issues) was guided by the four dimensions that constitute the conceptualization of the scale proposed by Coburn (2003). First, the indicators for each dimension were defined (see below). Afterwards, the presence of the indicators in the explicit definitions or characterizations of the concepts of "large scale," "small scale," scaling, and at scale stated by the experts or declared in the articles were sought for and located. In the case of the articles in which an explicit definition of these concepts was not found, the indicators embedded in the implicit formulations of these concepts were identified. The indicators used to identify the four dimensions of scale, depth, sustainability, spread, and shift in reform ownership are given in Table 1.

TABLE 1 Indicators for dimensions of scale in the data analysis

Coburn's dimensions	**Indicators**	**Examples**
Depth	– Change in beliefs – Changes in social norms of interaction – Changes in underlying pedagogical principles	"Large scale" means "big" (often costly as well) but something may be big without necessarily permeating a system. (Expert 14) To what extent can resources for mathematics classroom teaching, assessment, and professional development lead to changes in day-to-day classroom teaching? (Maass et al., 2019b, p. 307)
Sustainability	– Utterances of sustainability of innovation over time – Conditions present in context and organization supporting lasting change	The expected outcomes of PD [professional development] programs are not only focused on short-term effects that occur during or at the end of the project, but also on long-term effects that emerge (even after some years) after the project's termination. (Zehetmeier, 2015, p. 118) What comes after the professional development program is over? How can we sustain

TABLE 1 Indicators for dimensions of scale in the data analysis (*cont.*)

Coburn's dimensions	Indicators	Examples
		what was gained and how can the impact be scaled up? (Tirosh et al., 2015, p. 153)
Spread	– Innovation reaching many classrooms, schools or districts	Of course, no one would deny that mathematics education research studies involving thousands of students or teachers are large-scale studies. (Expert 6)
	– Innovations spread within schools	I think that what matters in this respect is the intent to enlarge the size of the project. (Expert 7)
Shift in reform ownership	– Innovation designers are no longer responsible for the sustainability of innovation	Effective scaling requires teachers to obtain ownership of the reform and ultimately transfer the new approaches into an "internal reform" in their own schools. (Potari et al., 2019, p. 419)
	– The presence of structures or mechanisms favoring ongoing learning about the reform or innovation	We examine what are the crucial aspects that enable participants to spread improvements in depth in their own schools and what preconditions allow for a *shift in reform ownership*. (Weißenrieder et al., 2015, p. 28)
	– Internal actors take responsibility for continued support for sustained use of the innovation.	

The absence or presence of these indicators in the empirical data were considered as markers of the dimensions of scale that are considered in the interpretations of small scale, medium scale, large scale, scaling, and at scale in mathematics education research. The results of this analysis are shown in the following section.

4 Results

4.1 *Definitions of Small Scale, Large Scale and Scaling*

The experts' responses regarding the definitions of large scale and small scale can be categorized into two distinct groups. The first group of responses acknowledged that "large scale" and "small scale" are vague and challenging to define precisely. Seven experts cited the difficulties associated with defining these concepts:

> There is no absolute border between small and large-scale (Expert 3).
>
> There is no clear definition (Expert 11).
>
> The difference between small-scale and large-scale projects is quite challenging (Expert 13).

The second group refers to nine experts whose responses contained an explicit characterization of small-scale and large-scale projects in their responses, and which were analyzed to identify indicators of Coburn's (2003) four dimensions of scale. The results showed that all explicit references to these dimensions contained *spread indicators*. For instance:

> Projects that focus on instructional improvement at scale can be quite small scale (i.e., require limited funding and involve just a few people) (Expert 1).
>
> Of course, no one would deny that mathematics education research studies involving thousands of students or teachers are large-scale studies (Expert 6).
>
> We consider our PD project to be a large-scale project, as we offer PD for facilitators who themselves conduct PD for teachers (Expert 11).

Nevertheless, some researchers cautioned that the issue of size in large-scale projects is relative. For instance, Expert 3 stated, 'Above 100 teachers is not small scale anymore, but what is really large might also depend on the size of the country.'

Only one of the mathematics education researchers provided *indicators of depth and sustainability* in their responses. Expert 14 noted that large-scale projects should "permeate the system," which is interpreted as a reference to a significant change in the system in which an innovation is implemented:

> Large scale means "big" (often costly as well) but something may be big without necessarily permeating a system (Expert 14).

Additionally, Expert 14 mentioned "stability" when referring to large-scale projects, which indicates the potential of implemented projects to remain stable over time, that is, sustainability:

> There are teaching support programs called 'math circles'. There are many such organizations — see https://mathcircles.org/. The number of organizations and their stability indicate that they are large-scale (Expert 14).

Regarding how the concepts of large scale and small scale are addressed in the literature, the research articles in the selected special issues explicitly use the terms "large scale" and "small scale" without explicitly defining them, as illustrated in the following two exemplary quotes.

> The *small-scale* school-based intervention, which resulted from the participation of four teachers in a professional development project (EPMT) in each of the two schools, was scaled up in the respective schools by the teachers. (Kaur, 2015, p. 114, our emphasis)

> This is among the first random assignment *large-scale* studies of teacher learning from several types of professional learning experiences. (Heck et al., 2019, p. 337, our emphasis)

An approximation of an explicit characterization of the notions of large scale and small scale is the idea of the "levels of scaling" of an innovation that is conceptualized in different works. One of those conceptualizations is that of *levels of scaling up*, which refers to three different levels at which an innovation could be implemented: at a micro level with a few teachers, at a meso level with tens of teachers, or at a macro level with hundreds (or even thousands) of teachers (see Krainer, 2015; Maass et al., 2019b). Similarly, Schoenfeld (2015) used *levels of scale* to refer to various organizational levels, from the individual classroom to a national system: classroom level, site level, district level, state, or province level, and national level. Prediger et al. (2019) referred to different *levels of institutional contexts* in which material- and community-based implementation strategies could be applied. These levels are: the level of students, level of classrooms, level of teacher communities in schools, level of schools as institutions, and level of school districts (Prediger et al., 2019, p. 368).

Our analysis of the implicit characterizations of large-scale and small-scale implementations contained in the articles through Coburn's (2003) scale

dimensions revealed that the *spread dimension* is dominant in such characterizations, as they emphasize quantitative elements of the scope of implementation. For instance:

> This extends to around 40,000 teachers spread over the country's 290 municipalities and nearly 6,000 schools. The government has set aside SEK 650,000,000 (EUR 75,000,000) for the 4-year programme. This is one of several large-scale initiatives to improve mathematics teaching in Sweden. (Boesen et al., 2015, p. 129)

Although the spread dimension of scale is a common element in these implicit characterizations, other scale dimensions identified by Coburn (2003) are also present in 13 of the 22 articles in our sample, for example, sustainability (Clark-Wilson et al., 2015; Jackson et al., 2015; Roesken-Winter et al., 2015b; Tirosh et al., 2015; Weißenrieder et al., 2015; Zehetmeier, 2015; Clark-Wilson & Hoyles, 2019; Heck et al., 2019), depth (Clark-Wilson et al., 2015; Kaur, 2015; Roesken-Winter et al., 2015b; Tirosh et al., 2015; Weißenrieder et al., 2015; Clark-Wilson & Hoyles, 2019; Ryve & Hemmi, 2019), and shift in reform ownership (Clark-Wilson et al., 2015; Jackson et al., 2015; Kaur, 2015; Weißenrieder et al., 2015; Krainer et al., 2019).

We also identified conceptualizations of scaling that are not explicitly captured in Coburn's (2003) conceptualization of scale. Potari et al. (2019), and Clark-Wilson et al. (2015), citing Hung et al. (2010), framed scaling to be both the products and processes in the implementation projects: 'The products are the tangible outcomes of the reform (e.g., better learning outcomes), while the processes are the main actions taken for the products to be developed' (Potari et al., 2019, p. 419). Additionally, in the experts' answers, we found ideas of scale and scaling as differences in methodological conditions and tools, and they also highlighted notions such as representativeness and generalization:

> There is no doubt for me that whatever distinctions are made, studies at different scales differentiate in terms of methodological conditions and constraints, which has an impact on their possible outcomes (Expert 6).

> One criterion could be if you can do a quantitative statistical analysis on the group? Versus small scale, if only qualitative analysis makes sense? (Expert 12).

> Large scale is rather vague; I may suggest, is it all the population or large enough so that then a generalization is accepted? Actually, it is the

> question of generalization that is important and must be the criterion (Expert 8).

In summary, we can report that there are no explicitly agreed-upon definitions of small scale, large scale, and scaling. Although it is not particularly difficult to identify that interventions involving only one or a few teachers are small-scale or that nationwide curriculum reforms are large-scale, there is a large gray area. We can identify this gray area in the descriptions of scale by Krainer (2015), Schoenfeld (2015) and Prediger et al. (2019). In all three descriptions of levels, no doubt, the first levels illustrate the small-scale projects, and the highest levels illustrate the large-scale projects. However, what about Krainer's meso level, Schoenfeld's (2015) site and district level, and Prediger's (2019) level of teacher communities in schools and the level of schools as institutions? It is by no means obvious whether they describe small- or large-scale projects, which is why we see a need to introduce another level of scale in implementation research: medium scale.

4.2 *Are Large Scale and at Scale Synonymous?*

There is no consensus on what the term *at scale* means. The experts interviewed had different ideas about what at scale means, and this is also reflected in our selection of articles. Whereas some experts considered "at scale" a process, others interpreted it as synonymous with a large scale, and some conceived it as a final state of implementation. Experts who interpreted "at scale" as a process associated it with the spread, diffusion, or replication of an implementation, as illustrated in the following quotes:

> In my mind, the notion of "project at scale" is linked to the idea of upscaling (Expert 10).

> The difference is relatively clear. "At scale" is usually used with implementation or replication. That is, to scale up based on a tested intervention. In summary, "at scale" is to scale up a tested intervention through implementation and replication in different contexts or sizes (Expert 13).

> A project at scale focuses squarely on improving instruction and learning on a large scale (Expert 1).

Expert 11 considered at scale to be synonymous with large scale:

> Are large scale and at scale synonyms? Yes, from my German language perspective, I would use both expressions synonymously.

An opposite understanding of Expert 11's was put forward by Expert 14:

> Large scale and at scale are not synonyms. "Large scale" means "big" (often costly as well), but something may be big without necessarily permeating a system. The Kumon teaching system is large-scale in the US, reaching across the US (mostly into areas with a high Japanese-American population). But it's not "at scale" in that there's not a system that it permeates. (Expert 14)

By contrast, Expert 4 associated the notion of "at scale" with a final state of an implementation process:

> There is a difference between a scaling project and a project that is 'at scale'. Saying 'at scale' suggests that the project has accomplished the goal of having fully realized all targeted dimensions of scaling. (Expert 4)

The understanding of Expert 14 is supported by Coburn (2003). In her first use of "at scale," Coburn (2003) discussed the shift in reform ownership, that is, when innovation moves from the hands of implementers to the hands of end users.

> Finally, to be considered "at scale" ownership over the reform must shift so that it is no longer an "external" reform, controlled by a reformer, but rather becomes an "internal" reform with authority for the reform held by districts, schools, and teachers who have the capacity to sustain, spread, and deepen reform principles themselves. (Coburn, 2003, p. 7)

Further, in the following excerpt from one of the articles, there is the perception that "at scale" is connected with end users themselves building up communities for ownership of the innovation.

> Reform that aims at more ambitious goals for students' learning often fails *at scale* because minimal attention is given to the school and broader system contexts in which teachers develop and revise their instructional practices. (Jackson et al., 2015, p. 93, our emphasis)

In summary, we identified a lack of consensus in the understanding and use of "at scale," which has the potential to hinder communication between researchers in the field of research on implementing innovations. The perceptions that at scale and large scale are synonymous, that at scale is an upscaling process,

and that at scale means that the innovation has shifted in reform ownership and the implementation process is complete as far as the constructors of the innovation are concerned co-exist and thus make up a basis for misunderstandings among scholars.

5 Discussion of Results: What's Next?

By examining experts' expressed opinions on issues related to scaling using a Delphi technique complemented with a review of specific papers dealing with scaling in mathematics education research, we have tried to shed light on the question of how we should think about and understand matters of scale and scaling. As it turns out, the most common way to discuss scale is through what Coburn (2003) called spread, which, above all, deals with the number of participants involved. However, there is no clear view of where lines should be drawn between small- and large-scale implementation projects. One way to deal with the question suggested by some experts, and used in some articles is to specify scale in terms of organizational levels, differentiating among, for example, school-level implementations, district-level implementations and national-level implementations (Schoenfeld, 2015; Prediger et al. 2019). Still, even such characterizations are not well defined. It is possible to conceive of projects involving a large number of teachers spread over a nation but where the project is not organized at a national level (Barnett et al. 2022). Conversely, an implementation project might be organized by a national-level organization without the project having a national reach. We agree that it is a good idea to specify which organizational levels are involved when presenting or discussing implementation projects, but we also argue that this in itself does not provide conclusive information about scale and scaling.

A different aspect of scaling mentioned by some experts deals with the *potential* for scaling by stipulating that whether the project involves the education of educators, multipliers (e.g., Roesken-Winter et al. 2015a,b), advisors (e.g., Lindvall et al. 2022), or similar stakeholders, who in turn are instrumental in implementing the project with further users; this would indicate the project has been set up with a potential for scaling. This relates to the question we posed to the experts about the term "at scale." As shown earlier, the answers related to inquiries regarding at scale were also divergent. Nevertheless, one line of reasoning, which we think offers an interesting perspective on "at scale," is the process–product perspective. Some of our experts reserved the term "at scale" for when an implementation has reached some form of maturity so that it can be implemented or replicated as a type of product, in which the

developmental stages in some sense have already been completed, or a system that permeates the implementation of the innovation. Some of our experts also discussed scale in relation to a sort of stability in relation to the efforts that uphold the usage of the innovation, which relates to Coburn's (2003) dimensions of sustainability and shift in ownership.

In general, our analysis of results makes it evident that, although scaling at least shares the common understanding of spread, neither knowledgeable experts nor research article authors tend to offer a clear definition of scale. Yet, in particular, in implementation research, it seems important to have a set of relatively agreed-upon principles to refer to when discussing scale. We will therefore turn the discussion toward some suggestions for such principles for small- and large-scale implementation projects and then turn our attention toward what might characterize medium-scale projects.

5.1 *Characteristics of Small-Scale and Large-Scale Implementation Projects*

In the area of medical diagnosis, when some particular condition does not allow a clear distinction, a solution can instead be to present a set of criteria and state that if m of these n conditions are present, then the patient is said to have the condition in question. We suggest a similar approach for dealing with scale and scaling. In the remaining part of this paper, we will refrain completely from using quantitative information regarding the number of users as a criterion. Instead, we will attempt to shed light on what it might entail for an implementation *project* to be of a *small-scale* or *large-scale* type by establishing aspects of an implementation project that are particular for small scale and aspects that are prerequisites for potential large scale in terms of spread.

To begin with, let us think simplistically of the implemented innovation as a product: a product made by producers (typically researchers or designers) and applied by users (typically teachers). "Product" should here be interpreted in a wide sense and could be any educational resource, such as a set of instructional strategies or material resources. Note that this use of the term product is different from the product as an outcome, a meaning used by Hung et al. (2010) and Potari et al. (2019). We are not saying that an implementation project can be a product in the sense of a package ready to use, but we are saying that the producer–product–user metaphor is still meaningful for our discussion because of how the scale of a project can both open potentials and set limits to the producer–user relationship. The producer–user relationship can, on the one hand, be thought of as the relationship between the proponent or designer of innovation and the adapters, "who adapt the resource while taking agency

over it" (Koichu et al., 2022, p. 79). The usage of the product is then a "secondary production hidden in the process of its utilization" (de Certeau, 1984, p. xiii) by means of "manipulation by users who are not [the products] makers" (de Certeau, 1984, p. xiii). On the other hand, the user and the producer can be co-producers of the product. The producer–user relationship hence concerns the responsibility and authority of adaptation. When a product is in development, it is important for producers and users to have a functioning feedback loop. Information from users on how the product works as well as how users adapt the product should be taken into account by the producers and used to refine the product. Users and producers are co-responsible for the adaptation process and hence need a well-developed contact system between users and producers. In this sense, the users are co-producers.

A second reason for the need for direct contact between the user and producer may be that the instructions needed to understand the product are not yet fully realized in some material form. By material form, we mean some set of manifest resources, such as text or video. At this stage, the developer may not even have enough understanding of how to produce such instructions, and so again will need to have direct contact with the user to explain the product. Lastly, in the developmental stage, the producer may want to have good control and involvement in the organizational structure itself to be able to have good contact with users. We associate high contact between developers and participants, a material support that is not yet fully developed, and a strong involvement by the developer in the project organization with small scale. Now, we turn to a discussion of large scale.

Since we limit ourselves to innovations that build on mathematics education research, we assume that mathematics education researchers are among the producers. In the context of implementation, and particularly when discussing scaling, we can assume that the producer at some point wants the product to impact many learners. However, being heavily involved in a particular implementation effort is not the main job of a researcher, and it is also impossible to be a central person in implementation projects running at many different sites at once. Therefore, at some point, the material support needs to be so developed that the users can be introduced to the innovation without the direct involvement of the producer. Similarly, unless implementation efforts can be run by organizations, where the producer is no longer in the central position, it will be impossible for the product to spread and in the end influence many learners.

This leads us to three factors that we think are worth considering when distinguishing between small-scale and large-scale projects. The first factor deals

TABLE 2 The contact, material, and organizational factors according to small or large scale projects

	Contact factor	Material factor	Organizational factor
Small-scale implementation projects	Participants in the project have direct contact with the innovation designers.	The material support structures for participating in the project are not yet so developed that it is possible to start working on the project without direct contact with the innovation developers.	The organizational support structure is set up in a way that involves the developers in a central position
Large-scale implementation projects	Most participants in the project have no or at most sporadic direct contact with the innovation designers.	The material support structures for participating in the project is developed to a degree where it is possible to start working on the project without direct contact with the innovation developers.	An organizational support structure can be set up by organizations that do not involve the developers in central positions.

with developer–user contact, the second factor deals with the character of the material support, and the third factor deals with the organization. Table 2 gives the three factors for both small-scale and large-scale implementation projects.

Obviously, implementation projects can exist that are small without all three small-scale factors in Table 2 being fulfilled, or large projects can exist without all three large-scale characterizations holding true. We still claim that the contact factor, material factor, and organizational factor are relevant to consider when thinking about the scale of implementation projects. Note again that we are discussing project size, not the size of the associated implementation studies. Even a large-scale implementation project can be meaningfully studied with small-scale studies.

5.2 *Introducing the Notion of Medium-Scale Implementation Projects*

A consequence of the choice to view implementation projects through the three factors above is that the responsibility for adaptations to the product shifts to the user in large-scale projects. Thus, it becomes harder for developers to control adaptations. Therefore, we also suggest that it is meaningful to characterize *medium-scale studies*. However, there can be several different ways of thinking about medium-scale studies. We provide three examples. First, let us consider a small-scale study with a selection of teachers in a school, municipality, or district, fulfilling the contact, material, and organizational small-scale

criteria from Table 2, but where the school, municipality, or district decides to use the participating teachers as resources for engaging, educating, and advising new participating teachers. This is an extension of the type of implementation effort described by Coburn (2003) with her scale dimension of *spread within*. The developer may keep in contact with the previous teachers but not engage much with the new teacher. In such a situation, the material support might have developed but could also be supported with locally produced instructional materials. When small-scale projects transition to medium-scale projects in this sense, they also include the scale dimension of a shift in ownership (Coburn, 2003).

A second way to conceive of medium-scale implementation is when the developer takes on the responsibility of acting as or educating multipliers. Assuming there is a developed organizational structure and a relatively developed material structure, new organizations may start up new implementation projects and use a developer, or one or several persons educated by the developer, as advisors and experts to support the participating teachers. This type of *deliberate multiplier* setup can be quite close to a small-scale implementation, not least in situations in which developers simply begin a replication of a previous project at a new site. Alternatively, it can be close to a large-scale project if the developers have a thought-out multiplier structure and a developed accompanying material structure (e.g., Prediger et al., 2019).

Previously, we discussed how the responsibility for adaptations could make to the product shift toward the user in large-scale projects, and the two types of medium-scale projects described above also include various forms of such shifts. Obviously, there can be times when the developer wants to keep adaptations small, that is, assume a *pro-fidelity* stance (Century & Cassata, 2016), for example, when conducting effect studies or assessing how a relatively well-developed educational innovation works when implemented with other users compared to those involved in developing and testing the initial product. We see this as a possible third scenario of medium-scale studies. Here, the material structures should preferably be so developed that it is maximally clear what the innovation is and how to implement it in, for example, classrooms. The developers might still want to have a relatively high degree of contact with the users, for example, to enforce and perhaps assess fidelity. In such cases, the developers will probably also want to be relatively central players in the organizational structure.

In Table 3 we summarize these three examples of what can reasonably be called medium-scale implementation projects.

We do not suggest that the three examples in Table 3 are the only cases that could deserve the name medium scale; rather, they are examples of what medium-scale implementation projects might look like. The fact that there

TABLE 3 The contact, material and organizational factors for three types of medium-scale implementation projects

	Contact factor	Material factor	Organizational factor
Medium-scale implementation projects, "spread within"	Participants in the project have different amounts of direct contact with the innovation designers.	The material support structures for participating in the project might be complemented with locally produced instructions.	The organizational support structure includes experienced implementation teachers as advisors.
Medium-scale implementation projects, "deliberate multiplier"	Participants in the project have direct contact with the innovation designers or with other experts assigned by the developers.	The material support structures for participating in the project need to be accompanied by expert developer support when new teachers start to work on the project.	An organizational support structure can be set up by organizations that do not involve the developers in central positions but use developers or persons assigned by the developers as experts.
Medium-scale implementation projects, "pro-fidelity"	Participants in the project have relatively high degrees of direct contact with the innovation designers, mainly for monitoring purposes.	The material support structures for participating in the project are developed to a degree where it is possible to start working on the project without direct contact with the innovation developers.	The organizational support structure is set up in a way that involves the developers in a position where fidelity to the implementation ideas can be overseen and enforced.

are many different ways of "filling up" the space between small-scale and large-scale studies may, in fact, be an explanation for why it is hard for experts to pinpoint the meaning of scale and scaling, as our results revealed.

6 Conclusions

Neither among experts, when asked directly, nor in research articles did we find agreed-upon definitions of scale and scaling — not in terms of what should characterize small scale versus large scale, or what "at scale" would mean in relation to large scale. Among the four scale dimensions suggested by

Coburn (2003), spread was by far the factor mentioned the most by experts, whereas sustainability, depth, and shift in reform ownership were only sporadically apparent. Despite the spread dimension being the most predominant, it still appears challenging to agree upon a way to use spread to characterize what is particular to small-scale versus large-scale implementation projects. As an alternative, we instead suggest focusing on *a set of characteristics* of a given implementation project to draw out aspects of scale and scaling. More precisely, we have suggested a characterization that focuses on three different factors: *the contact factor*, which deals with the level of interaction between users and developers; *the material factor*, which deals with the degree to which instructional materials are useable without associated explanations by the developer; and *the organizational factor*, which deals with the position of the developer in implementation project organizations. Drawing on Coburn (2003), we have exemplified these three factors in terms of small-scale and large-scale implementation projects, also suggesting an intermediate range of medium-scale implementation projects.

Acknowledgement

This paper is part of grant 2022–04811 under the Swedish Research Council.

References

Ahl, L. M., Helenius, O., Aguilar, M. S., Jankvist, U. T., Misfeldt, M., & Prytz, J. (2023). *A theoretical framework for analyzing factors of influence for the success or failure of implementation projects* [Manuscript in preparation]. Uppsala University, University of Gothenburg, Instituto Politécnico Nacional, Aarhus University and University of Copenhagen.

Anderson, L. W., & Postlethwaite, T. N. (2007). *Program evaluation: Large-scale and small-scale studies*. International Academy of Education; International Institute for Education Planning. https://unesdoc.unesco.org/ark:/48223/pf0000181752.

Barnett, J. H., Clark, K. M., & Can, C. (2022). Transforming mathematics instruction via primary historical sources: A study of influential factors on implementation of a curricular innovation at the tertiary level. *Implementation and Replication Studies in Mathematics Education*, *2*(2), 208–240. https://doi.org/10.1163/26670127-bja10006.

Barrett, D., & Heale, R. (2020). What are Delphi studies? *Evidence-Based Nursing*, *23*(3), 68–69. https://doi.org/10.1136/ebnurs-2020-103303.

Boesen, J., Helenius, O., & Johansson, B. (2015). National-scale professional development in Sweden: Theory, policy, practice. *ZDM — Mathematics Education*, *47*(1), 129–141. https://doi.org/10.1007/s11858-014-0653-4.

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.

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. https://doi.org/10.1007/s10649-018-9836-1.

Clark-Wilson, A., Hoyles, C., Noss, R., Vahey, P., & Roschelle, J. (2015). Scaling a technology-based innovation: Windows on the evolution of mathematics teachers' practices. *ZDM — Mathematics Education*, *47*(1), 79–192. https://doi.org/10.1007/s11858-014-0635-6.

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

de Certeau, M. (1984). *The practice of everyday life*. University of California Press.

Heck, D. J., Plumley, C. L., Stylianou, D. A., Smith, A. A., & Moffett, G. (2019). Scaling up innovative learning in mathematics: Exploring the effect of different professional development approaches on teacher knowledge, beliefs, and instructional practice. *Educational Studies in Mathematics*, *102*(3), 319–342. https://doi.org/10.1007/s10649-019-09895-6.

Hung, D., Lim, K., & Huang, D. (2010). Extending and scaling technology-based innovations through research. In Organisation for Economic Co-operation and Development (Ed.), *Inspired by technology, driven by pedagogy: A systemic approach to technology-based school innovations* (pp. 87–100). OECD Publishing. https://doi.org/10.1787/9789264094437-en.

Jackson, K., Cobb, P., Wilson, J., Webster, M., Dunlap, C., & Appelgate, M. (2015). Investigating the development of mathematics leaders' capacity to support teachers' learning on a large scale. *ZDM — Mathematics Education*, *47*(1), 93–104. https://doi.org/10.1007/s11858-014-0652-5.

Kallia, M., van Borkulo, S. P., Drijvers, P., Barendsen, E., & Tolboom, J. (2021). Characterising computational thinking in mathematics education: A literature-informed Delphi study. *Research in Mathematics Education*, *23*(2), 159–187. https://doi.org/10.1080/14794802.2020.1852104.

Kaur, B. (2015). What matters? From a small-scale to a school-wide intervention. *ZDM — Mathematics Education*, *47*(1), 105–116. https://doi.org/10.1007/s11858-014-0645-4.

Krainer, K. (2015). Reflections on the increasing relevance of large-scale professional development. *ZDM — Mathematics Education*, *47*(1), 143–151. https://doi.org/10.1007/s11858-015-0674-7.

Krainer, K., Stefan, Z., Hanfstingl, B., Rauch, F., & Tscheinig, T. (2019). Insights into scaling up a nationwide learning and teaching initiative on various levels. *Educational Studies in Mathematics*, *102*(3), 395–415. https://doi.org/10.1007/s10649-018-9826-3.

Koichu, B., Cooper, J., & Widder, M. (2022). Implementation of problem solving in school: From intended to experienced. *Implementation and Replication Studies in Mathematics Education*, *2*(1), 76–106. https://doi.org/10.1163/26670127-bja10004.

Lindvall, J., Helenius, O., Eriksson, K., & Ryve, A. (2022). Impact and design of a national-scale professional development program for mathematics teachers. *Scandinavian Journal of Educational Research*, *66*(5), 744–759. https://doi.org/10.1080/00313831.2021.1910563.

Maass, K., Cobb, P., Krainer, K., & Potari, D. (Eds.). (2019a). Different ways to large-scale implementation of innovative teaching approaches [Special issue]. *Educational Studies in Mathematics*, *102*(3).

Maass, K., Cobb, P., Krainer, K., & Potari, D. (2019b). Different ways to implement innovative teaching approaches at scale. *Educational Studies in Mathematics*, *102*(3), 303–318. https://doi.org/10.1007/s10649-019-09920-8.

Manizade, A. G., & Mason, M. M. (2011). Using Delphi methodology to design assessments of teachers' pedagogical content knowledge. *Educational Studies in Mathematics*, *76*(2), 183–207. https://doi.org/10.1007/s10649-010-9276-z.

Middleton, J. A., Cai, J., & Hwang, S. (Eds.). (2015). *Large-scale studies in mathematics education*. Springer. https://doi.org/10.1007/978-3-319-07716-1.

Morel, R. P., Coburn, C., Catterson, A. K., & Higgs, J. (2019). The multiple meanings of scale: Implications for researchers and practitioners. *Educational Researcher*, *48*(6), 369–377. https://doi.org/10.3102/0013189X19860531.

Potari, D., Psycharis, G., Sakonidis, C., & Zachariades, T. (2019). Collaborative design of a reform-oriented mathematics curriculum: Contradictions and boundaries across teaching, research, and policy. *Educational Studies in Mathematics*, *102*(3), 417–434. https://doi.org/10.1007/s10649-018-9834-3.

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. https://doi.org/10.1007/s10649-018-9835-2.

Roesken-Winter, B., Hoyles, C., & Blömeke, S. (Eds.). (2015a). Evidence-based CPD: Scaling up sustainable interventions [Special issue]. *ZDM — Mathematics Education*, *47*(1).

Roesken-Winter, B., Schüler, S., Stahnke, R., & Blömeke, S. (2015b). Effective CPD on a large-scale: Examining the development of multipliers. *ZDM — Mathematics Education*, *47*(1), 13–25. https://doi.org/10.1007/s11858-014-0644-5.

Rogers, E. M. (2003). Diffusion of innovations (5th edn.). Free Press. (Original work published 1961).

Ryve, A., & Hemmi, K. (2019). Educational policy to improve mathematics instruction at scale: Conceptualizing contextual factors. *Educational Studies in Mathematics*, *102*(3), 379–394. https://doi.org/10.1007/s10649-019-09887-6.

Schoenfeld, A. H. (2015). Thoughts on scale. *ZDM — Mathematics Education*, *47*(1), 161–169. https://doi.org/10.1007/s11858-014-0662-3.

Stringfield, S., & Datnow, A. (1998). Introduction: Scaling up school restructuring designs in urban schools. *Education and Urban Society*, *30*(3), 269–276. https://doi.org/10.1177/0013124598030003001.

Tirosh, D., Tsamir, P., & Levenson, E. (2015). Fundamental issues concerning the sustainment and scaling up of professional development programs. *ZDM — Mathematics Education*, *47*(1), 153–159. https://doi.org/10.1007/s11858-015-0665-8.

Weißenrieder, J., Roesken-Winter, B., Schueler, S., Binner, E., & Blömeke, S. (2015). Scaling CPD through professional learning communities: Development of teachers' self-efficacy in relation to collaboration. *ZDM — Mathematics Education*, *47*(1), 27–38. https://doi.org/10.1007/s11858-015-0673-8.

Zehetmeier, S. (2015). Sustaining and scaling up the impact of professional development programmes. *ZDM — Mathematics Education*, *47*(1), 117–128. https://doi.org/10.1007/s11858-015-0671-x.