Working Across Contexts: Scaling Up or Replicating With Variations?

Author(s): Jinfa Cai, Anne Morris, Charles Hohensee, Stephen Hwang, Victoria Robison, Michelle Cirillo, Steven L. Kramer and James Hiebert

Source: *Journal for Research in Mathematics Education* , Vol. 51, No. 3 (May 2020), pp. 258-267

Published by: National Council of Teachers of Mathematics

Stable URL: https://www.jstor.org/stable/10.5951/jresemtheduc-2020-0007

Journal for Research in Mathematics Education
2020, Vol. 51, No. 3, 258–267

Editorial

Working Across Contexts: Scaling Up or Replicating With Variations?

Jinfa Cai, Anne Morris, Charles Hohensee, Stephen Hwang, Victoria Robison, Michelle Cirillo, Steven L. Kramer, and James Hiebert
University of Delaware

Our March editorial (Cai et al., 2020a) focused on the problem of sharing professional knowledge so that teachers can build on what others have learned. Sharing knowledge from person to person and generation to generation is one way that educational improvements can build over time. However, educators face a complementary problem: Instructional activities that work well in one context do not necessarily work as well in other contexts. Educators cannot assume that an instructional innovation will yield the same outcomes when it is moved across contexts. Thus, even if our earlier problem of sharing professional knowledge were solved, that knowledge would not necessarily apply in the same way in the context to which it has been shared. The consequence is that innovative instructional programs and practices that could maximize learning opportunities for students too often remain local.

This problem has frequently been characterized as a scale-up problem—that is, a problem of taking a locally effective innovation and implementing it broadly so that it is effective in many schools, districts, states, or even throughout a nation (Cobb & Smith, 2008; Maass et al., 2019a). A number of researchers have explored scaling up to understand the process and document the effects of scale-up efforts (e.g., see the recent special issue in *Educational Studies in Mathematics* edited by Maass et al., 2019b). When educators transport innovations that have worked in one setting to other settings, one of the most persistent findings from the impressive body of work of scaling up innovations in mathematics education is that "the contextual factors are crucial" (Maass et al., 2019a, p. 315). We build on this finding because we wrestle with some of the challenges that researchers have frequently reported during their attempts to address the scale-up problem for educational improvement. In our view, the field must continue to grapple with the problem of implementing effective innovations across multiple contexts to enhance future progress in the teaching and learning of mathematics.

Reconsidering the Notion of Scaling Up

In this editorial, we consider this problem of effective innovations too often remaining local–that is, the problem of implementing effective innovations across

This work was supported in part by the National Science Foundation under Grant No. 1941494. Any opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the National Science Foundation.

multiple contexts–as the third of five overarching problems. From a traditional scale-up perspective, this problem can be interpreted as "extending the reach of an exemplary intervention to produce similarly positive effects in different settings and to help a greater number of students" (McDonald et al., 2006, p. 16). From this perspective, an intervention shown to be exemplary in one setting is implemented in a large number of new settings with the aim of confirming its effectiveness. In the spirit of our previous editorials, we will take a different perspective. Rather than improving conventional approaches to solving this overarching problem, we will redefine the problem and offer an alternative approach. Our research questions surrounding the scale-up problem will encourage investigating the potential of this alternative. As in previous editorials, our aim is to stimulate discussion in the field and encourage new ways of thinking about this longstanding problem.

Alternative Conceptions of Scaling Up

We are not the first educators to reconsider the notion of scaling up. For example, Coburn (2003) described the traditional increase-the-numbers approach as conceiving of scaling up in two ways: (a) replicating the reform in more schools and (b) adapting the reform to fit the needs of local schools. Coburn proposed an alternative conceptualization of scale that widens the focus beyond just increasing the numbers of students, teachers, and schools involved. Coburn's alternative includes additional dimensions of scale such as sustainability, the spread of beliefs and norms, and the locus of ownership of reforms. We, too, propose widening the field's perspective on scaling up, but we diverge from Coburn by proposing a different way of thinking about the problem as our starting point.

In our view, the problem of scaling up an educational innovation can be better conceived as a problem of learning how to improve the original innovation by studying its implementation in multiple contexts rather than a problem of documenting that the original innovation is effective for larger numbers of students. In other words, we propose that the underlying problem of scaling up is not merely a lack of skill at adapting an innovation so that it can be implemented with high enough fidelity to achieve significant mean score differences between large numbers of experimental and control participants; rather, we propose that the underlying problem is a lack of understanding about how an innovation interacts with local conditions. We reconceive the goal of scaling up to be learning from replications with variation. From this perspective, when an educational innovation is implemented in a new context, the inevitable variation is not treated as a lack of fidelity or as a result of interfering factors that need to be controlled but as a fortuitous source of learning that can help researchers and practitioners better understand how the innovation works and what conditions affect the outcomes in different contexts (Morris, 2012). The goal shifts from planting the innovation in as many sites as possible to gathering new information about how the innovation interacts with different contexts so as to continually improve the innovation.

The approach we take here can be viewed as a version of Marton and Pang's (2006) variation theory of learning. Marton and Pang hypothesized that students learn mathematics deeply by engaging in opportunities to compare what they learn in one context with other contexts that vary in specific ways. These repeated comparisons between contexts help students develop fine discriminations about which features are and are not essential. Applied to researchers studying the effects of an innovation, our view is that researchers can learn more and more about an innovation by repeatedly comparing it in contexts that vary. In Marton and Pang's theory, the variation supporting students' mathematics learning is carefully planned. In our adaptation of Marton and Pang's theory for learning about educational innovations, the variation that drives the learning arises from the natural variation among classroom and school contexts. More specifically, researchers studying an innovation across contexts seek to learn how characteristics that vary in the innovation and context contribute to the learning opportunities students experience. The natural variation provides researchers with opportunities to form and test hypotheses about how characteristics or combinations of characteristics make a difference for the learning opportunities in individual contexts, across a group of contexts, or even across most (perhaps all) contexts.

Does our reconceptualization of scaling up mean that we have entirely abandoned the goal of spreading the innovation across an increasing number of sites? No. We retain the goal of scaling up. However, we see the extended reach of an innovation as a by-product of testing the innovation in more and more sites. Although we foreground the learning opportunities that the sites provide researchers, we believe a continuously improved innovation will provide improved learning opportunities for students in all the sites that have participated. Moreover, and this is a key point, we hypothesize that this process keeps the focus on improving the quality of the innovation, even for sites where the innovation has already been tested. Improvements benefit all sites, even those that participated early in the process, because all sites always have access to continually improving versions of the innovation. Implicit in our description of this process is that a final version of the innovation, a version that could be packaged and distributed widely, is never achieved. Every implementation provides the potential for additional learning and further improvement of the innovation.

Why Alternative Conceptions of Scaling Up Are Warranted

Before proposing research questions that are prompted by our alternative view of scaling up, we review two challenges associated with the traditional interpretation of scaling up, which provide the warrants for our alternative conception. These challenges are found both in the actual scaling process and in studying this process. A first challenge arises from concerns about the fidelity of implementation. Most traditional approaches require the educational innovation to be well defined and fully designed for implementation. To implement the innovation in multiple sites, researchers have usually worked out the bugs in the original site to prepare the innovation for large-scale assessment. Confirming the effectiveness of the innovation requires it to be implemented in every site as it was developed. Naturally, this prompts concerns about the fidelity of the implementations. If the fidelity is low, the large-scale assessment is of little value. But what counts as fidelity is not easy to define (Century et al., 2010). Because the process of scaling up and the assessment of its success require that the original version of the innovation be implemented with fidelity, concerns about defining and assessing fidelity with sufficient clarity and accuracy can complicate efforts to establish claims about the innovation's effectiveness.

A second challenge for researchers pertains to data collection and analysis. When an innovation is implemented across multiple sites, the effects are typically compared with nonintervention sites by comparing mean scores on an outcome measure to determine if the innovation scaled successfully. Given the variation within and across sites, mean scores will rarely capture the actual effects of the innovation. The innovation might work well at some sites and work poorly at others, whereas the mean score shows that it had no effect. In addition, mean scores do not help researchers use the variation that exists within the sample to learn what might have accounted for the uneven effects. In the conventional approach, the need to average the data across multiple sites often trumps the need to understand why the results from one site differ significantly from those at another site. In our alternative approach, the analysis begins by examining the variation.

Learning from the variation that occurs as an innovation is implemented depends on accumulating the data in ways that allow researchers to keep the connections between the data and local sites while still detecting trends across sites. What researchers can learn from these data depends on knowing both how the data from each site are connected with the ways in which the innovation was implemented at that site and how the outcomes at each site vary for classrooms, teachers, or students with a variety of characteristics. Although our field has developed impressive statistical techniques for detecting mean differences and controlling for interfering variation, we have not yet developed robust mechanisms for accumulating data on variations as innovations are implemented at multiple sites and over multiple time periods.

We believe these challenges provide a partial explanation for the difficulty that researchers have had in understanding the mechanisms that support and undermine attempts to scale up promising educational innovations. These challenges might also explain the difficulty that educators have had with successfully scaling up innovations that looked so effective in the original site. We argue that it might be worth considering an alternative view of scaling up and asking an alternative set of research questions.

Before previewing these questions, we want to clarify what we mean by *innovation*. Often, the literature on scaling up innovations fails to fully define an innovation, which is another potential challenge for implementing and studying innovations. The innovations in which we are interested are those that aim to enrich the mathematics learning opportunities for all students. These include changes to the curriculum, instructional activities, pedagogical approaches, or other classroom practices that directly affect the nature of students' learning opportunities. The emphasis on learning opportunities aligns with our perspective in the two previous editorials in which we described the first two of five overarching problems for mathematics education researchers.

We believe that two sets of research questions could guide the field as it strives to address the overarching problem discussed in this editorial—that innovative solutions too often remain local. The first set of questions focuses on the process of testing an innovation across multiple sites while continuously revising the innovation so that it is increasingly effective for a wider range of students. To coin a term, these questions focus on *replimentation*, by which we mean conceiving of the fundamental research activity as creating learning opportunities for researchers to better understand and improve an innovation as it repeatedly interacts with different local contexts. In other words, replimentation means that researchers implement innovations in new contexts to scale up, but they also treat new implementations as replications to learn more about and improve innovations. The primary question highlighted by the concept of replimentation is: How can researchers maximize their own learning opportunities when testing an innovation across local sites? Other ways of asking the same basic question include these: What research approaches are most sensitive to the identification of relevant contextual factors and to the study of their influences on an implemented innovation? What research approaches can best take advantage of variation across sites as the innovation is replicated over multiple trials? What kinds of research designs best reveal the association between contextual factors and the nature of learning opportunities created by innovation? Which analytical techniques are best suited to detect patterns in the effects of an innovation when the effects vary across contexts?

Our second set of questions focuses on how innovations actually spread across contexts to impact larger numbers of teachers and students. What are the forces (e.g., incentives for participants, school-level leadership, and outside pressures) that encourage innovations to be implemented in multiple contexts? How can what is learned about the interaction between an innovation and a local context be stored and accumulated to steadily improve and spread innovations over time? Which analytical techniques are best suited to yield information from the growing accumulation of results that researchers can use to improve the innovation over time and for particular profiles of sites? In general, how can testing and revising innovations in individual contexts yield cumulative improvements to the innovation? The next two sections describe the contexts in which these two sets of questions could be productively addressed.

Replimentation: An Alternative Approach to Understanding Implementation Variations and Innovation Improvements

To understand the context in which we believe the first set of research questions could be productively addressed, we need to describe the ways in which the work of researchers would change if they used replimentation as an approach to studying innovations at scale. Rather than conducting a standard sequence of increasingly scaled studies, researchers would anticipate an ongoing series of conceptual replications. Researchers working within this replimentation process would expect that implementing an innovation in different contexts would yield different outcomes because of different conditions that affect the implementation. Because the contextual conditions shape or even determine the effect of innovations, researchers would focus their attention on understanding the ways in which these conditions affect outcomes. The research agenda would consist of identifying the conditions in the new context that differ from the original one, making predictions about how the conditions will affect the innovation, and testing these predictions.

The process of replimentation, as we have defined it, reconceives a distinction that is an unavoidable aspect of scaling up. Specifically, researchers must deal, in some way, with the question of what counts as the core of an innovation and what features can be changed to fit the local context. This distinction is inherent in the persistent issue of fidelity. Is the form of an innovation being implemented at a particular site a version of the innovation, or is it so different that it no longer counts as an example? Earlier, we noted that fidelity is extremely difficult to define and assess clearly when studying the scale up of an innovation. It seems to require specifying the core of an innovation that defines its essence and a set of features that could be altered to implement the innovation more faithfully. In the replimentation process that we envision, we reconceptualize the core of an innovation plus its implementation features as inextricably linked. In the following paragraphs, we propose a way to study an implementation without separating these two fundamental aspects of an innovation.

We use the metaphor of a cluster category to represent the variety of forms the same innovation could take while it is implemented across a range of contexts. Rather than trying to specify a set of necessary or sufficient characteristics that would definitively determine what counts as an implementation of a given innovation, we propose that researchers study implementations of an innovation as members of a cluster category—a kind of category whose members bear a *family resemblance* (Wittgenstein, 2009). Members of a family share some characteristics but do not need to satisfy a strict, classical definition to be considered members of the category (Gaut, 2000; Lakoff, 1987). As an example, Czocher and Weber (2020) proposed that a cluster category could be used to characterize the notoriously slippery notion of proof in mathematics education. A cluster category is determined by a set of characteristics, each of which can contribute toward a potential member of the cluster category belonging to the category but none of which are considered necessary or sufficient for membership. The more characteristics a potential member possesses, the more likely it is to be considered part of the category, but different members may possess different subsets of characteristics. Czocher and Weber describe how treating proof as a cluster category could both align with the intuitive idea of what proof should be and provide theoretical machinery helpful for analyzing borderline or peripheral cases that do not include all the desired characteristics.

A similar cluster category approach to studying the replimentation of educational innovations would engage researchers in systematically considering implementations of innovations that share a family resemblance. Rather than attempting to establish a set of features that determine whether a specific implementation counts as an example of the innovation, researchers could posit a set of features or characteristics that contribute to membership in the cluster. These characteristics would be features of implemented forms of an innovation that are considered significant for the creation of the desired learning opportunities. It would no longer be necessary to distinguish the core of an innovation from implementation features and, therefore, no longer necessary to worry about fidelity in the traditional sense. Every member of the innovation's cluster category would count as an example, although researchers might consider some more representative of the innovation than others.

In our view, the advantage of treating the characteristics of implemented innovations in this way is that it shifts researchers' focus from the difficult task of determining which variations are acceptable and which violate fidelity to identifying patterns connecting characteristics of implementations to the kinds of learning opportunities created. Researchers could systematically analyze implementations that differ on the cluster category characteristics and study how those differences affect learning opportunities for students. As researchers look from classroom to classroom, they could begin detecting patterns in the characteristics that were more commonly present when rich learning opportunities were created. Researchers' search for connections between implementation characteristics and desired learning opportunities could uncover characteristics that were not originally anticipated. Over time and multiple trials, we expect that the data would reveal characteristics that appeared to be essential for an innovation to provide the desired learning opportunities. We also expect that different sets of characteristics would match with contexts with distinctly different profiles. For example, one set of characteristics might be outlined for students who have had many experiences working in groups, and a somewhat different set of characteristics would be outlined for students with fewer experiences working in groups.

Realization: Spreading Educational Innovations and Identifying Successful Spreading

The second set of research questions that we proposed earlier addresses the spread of an innovation to more students in more contexts. In our January editorial (Cai et al., 2020b), we identified the first of our set of overarching problems for mathematics education: defining and measuring learning opportunities so as to maximize them for every student. In this editorial, we have built on that idea by suggesting that understanding the ways in which an innovation interacts with local contexts will enable scaling up the innovation and spreading ambitious learning opportunities to more students in more contexts. The current inequitable distribution of ambitious learning opportunities serves to highlight the importance of successfully addressing questions about the spread of innovations.

One of the research questions in the second set of questions proposed earlier calls attention to an issue that we believe has been undervalued in much of the previous work on scaling up: How can what is learned about the interaction between an innovation and a local context be stored and accumulated to steadily improve and spread innovations over time? In practical terms, how can researchers keep track of what they learn over multiple trials and store the accumulating information in a form that is easily accessible and immediately available when needed to inform revisions to the innovation? Usual forms, such as researchers' logs, spreadsheets of data, and internal reports, are inconvenient and unwieldy at best. The information is not linked to relevant aspects of the innovation, which can leave researchers without the data that would be most useful for tweaking the innovation. In addition, the usual forms make it difficult for researchers to share with each other what they are learning. This problem is similar to the one we addressed in our March editorial (Cai et al., 2020a) with respect to teachers. We suggested that the field's problem of professional amnesia was the result, in part, of the lack of mechanisms for storing, accumulating, and sharing what is learned by individual teachers. Here, we apply the same reasoning to what is learned by individual researchers.

If the usual forms for storing data do not work, what might? This will be a critical part of answering the research question that we posed above. We can provide some beginning hypotheses based on what has been proposed for teachers to preserve and share the knowledge they acquire (Cai et al., 2018, 2020a; Gu et al., 2017; Morris & Hiebert, 2011; Stigler & Hiebert, 1999). We envision a tangible instructional artifact that both accumulates the knowledge for improving an innovation and presents the best current version of the innovation. The artifact could take different forms, and it could also include multiple forms. Examples include: videos of classroom instruction, annotated lesson plans, teacher editions of student textbooks, instructional materials, and online resources for teaching. The artifact must have several properties, including being capable of preserving knowledge gained by studying implementations, communicating innovations directly to teachers, affording endless revisability based on new information, and linking revisions with the information that prompted them in a way that is user-friendly. This kind of artifact shares many features with the group-developed knowledge products described by Bereiter (2002).

An immediate consideration when choosing artifacts is their alignment with the unit of innovation (an instructional activity, a lesson, a unit of lessons, etc.). We do not endorse any particular unit. However, there are reasons to consider a daily lesson, or a series of daily lessons, as the unit. A lesson is small enough to analyze in detail but large enough to capture most of the classroom interactions that create learning opportunities for students (Stigler & Hiebert, 1999). It is also a unit that is familiar to teachers. If the daily lesson is the unit used to communicate and implement the innovation, then annotated lesson plans can function well as the tangible, instructional artifact.

That said, it may be necessary to have multiple tangible artifacts to address different needs. Teachers, who must make use of the artifact to directly support implementation of the innovation, might want a relatively brief artifact that outlines the characteristics of the innovation that represents the current "best understanding" of what characteristics matter most. The accumulated data from replicating the innovation in multiple contexts would likely overwhelm teachers. We suspect that one form of artifact will be needed to support the creation of learning opportunities for students and another form to support the creation of learning opportunities for researchers. The forms that these take will likely evolve simultaneously with empirical work that demonstrates the advantages of different forms.

Learning From Replimentation: The Current Support Landscape

Conventional scale-up models typically follow a linear sequence. Such models assume that the first phase is to develop an innovation that works in one (exploratory) setting, followed by implementing the same innovation in multiple settings, perhaps adapting it in some ways on the basis of feedback from trial sites. Because of its linear logic, this process is not designed to use the learning that occurs at later sites to improve the innovation for earlier sites. In fact, some conventional approaches do not build learning from variation into the scale-up process because improving the innovation itself is not viewed as the goal; rather, the goal is improving the implementations at different sites. This linear path is explicated in the sequence of stages for studying educational innovations typically outlined by funding agencies, for example, by the Institute of Education Sciences and the National Science Foundation (2013).

Despite the historical dominance of linear scale-up sequences in funded research, during the past decade, there has been movement toward supporting scale-up research of the type that we advocate in this article, which focuses on learning from variation with the goal of continuous improvement. For example, in 2015, the National Science Foundation DRK-12 program began funding Implementation and Improvement Studies as an alternative to the traditional linear sequence. As described in their most recent solicitation, such studies "capitalize on variation in educational contexts to address the sources of variability in outcomes to understand what works, for whom, and under what conditions" (National Science Foundation, 2017, p. 6). Additionally, large educational research organizations have begun to advocate for research of the type that we have termed replimentation. Examples include the Real-time Scaling Labs supported by the Brookings Institution (Robinson & Curtiss, 2018) and multiple initiatives supported by the Carnegie Foundation for the Advancement of Teaching (see, e.g., https://www.carnegiefoundation.org/our-work/).

Conclusions

In this editorial, we addressed the stubborn problem of scaling up an innovation from one site to multiple sites. We posed this problem as the third of five overarching problems that we believe the field must address to enhance future progress in the teaching and learning of mathematics. Rather than describing potentially better ways to apply conventional approaches to solving this problem, we proposed an alternative perspective. We drew on the idea of replication with variations to redefine the problem and reconceive the work of researchers who study this problem. The research questions that we posed to address this third problem would be most productively addressed within this alternative approach for taking innovations to scale.

In the next editorial, we will address the fourth of five overarching problems that we believe the field must address to make future progress in the teaching and learning of mathematics: Despite increasing pressure to make data-informed instructional decisions, teachers lack timely, useful student data. Building on our conjecture that the daily classroom lesson might be a useful unit for tackling the first three problems discussed in our editorials for 2020, we consider the kinds of data that teachers and researchers might collect to impact mathematics teaching and learning. Questions that guide our discussion in the next editorial include the following: What kinds of data do teachers most need to make more informed instructional decisions? How can researchers gather and process these data to make them actually usable by teachers?

References

Bereiter, C. (2002). *Education and mind in the knowledge age*. Erlbaum.

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., Cirillo, M., Kramer, S. L., Hiebert J., & Bakker, A. (2020a). Addressing the problem of always starting over: Identifying, valuing, and sharing professional knowledge for teaching. *Journal for Research in Mathematics Education*, *51*(2), 130–139.

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., Cirillo, M., Kramer, S. L., Hiebert J., & Bakker, A. (2020b). Maximizing the quality of learning opportunities for every student. *Journal for Research in Mathematics Education*, *51*(1), 12–25. https://doi.org/10.5951/jresematheduc.2019.0005

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., & Hiebert, J. (2018). Building and structuring knowledge that could actually improve instructional practice. *Journal for Research in Mathematics Education*, *49*(3), 238–246. https://doi.org/10.5951/jresematheduc.49.3.0238

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, *31*(2), 199–218. https://doi.org/10.1177/1098214010366173

Cobb, P., & Smith, T. (2008). The challenge of scale: Designing schools and districts as learning organizations for instructional improvement in mathematics. In K. Krainer & T. Wood (Eds.), *International handbook of mathematics teacher education: Vol. 3. Participants in mathematics teacher education: Individuals, teams, communities and networks* (pp. 231–254). Sense.

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

Czocher, J. A., & Weber, K. (2020). Proof as a cluster category. *Journal for Research in Mathematics Education*, *51*(1), 50–74. https://doi.org/10.5951/jresematheduc.2019.0007

Gaut, B. (2000). “Art” as a cluster concept. In N. Carroll (Ed.), *Theories of art today* (pp. 25–44). University of Wisconsin Press.

Gu, F., Huang, R., & Gu, L. (2017). Theory and development of teaching through variation in mathematics in China. In R. Huang & Y. Li (Eds.), *Teaching and learning mathematics through variation: Confucian heritage meets Western theories* (pp. 13–41). Sense. https://doi.org/10.1007/978-94-6300-782-5_2

Institute of Education Sciences & National Science Foundation. (2013). *Common guidelines for education research and development*.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* University of Chicago Press.

Maass, K., Cobb, P., Krainer, K., & Potari, D. (2019a). 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

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

Marton, F., & Pang, M. F. (2006). On some necessary conditions of learning. *Journal of the Learning Sciences*, *15*(2), 193–220.

McDonald, S.-K., Keesler, V. A., Kauffman, N. J., & Schneider, B. (2006). Scaling-up exemplary interventions. *Educational Researcher*, *35*(3), 15–24.https://doi.org/10.3102/0013189x035003015

Morris, A. K. (2012). Using “lack of fidelity” to improve teaching. *Mathematics Teacher Educator*, *1*(1), 71–101. https://doi.org/10.5951/mathteaceduc.1.1.0071

Morris, A. K., & Hiebert, J. (2011). Creating shared instructional products: An alternative approach to improving teaching. *Educational Researcher*, *40*(1), 5–14. https://doi.org/10.3102/0013189X10393501

National Science Foundation. (2017). *Discovery research preK-12 (DRK-12) program solicitation (NSF 17-584)*. Retrieved from https://www.nsf.gov/pubs/2017/nsf17584/nsf17584.pdf.

Robinson, J. P., & Curtiss, M. (2018). *Millions learning Real-time Scaling Labs: Designing an adaptive learning process to support large-scale change in education*. Center for Universal Education at Brookings.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world’s teachers for improving education in the classroom*. Free Press.

Wittgenstein, L. (2009). *Philosophical investigations* (4th ed.). Wiley-Blackwell.

doi:10.5951/jresemtheduc-2020-0007