On the implicitness of the 'theory of change' in implementations in mathematics education research

Ola Helenius,[1] Uffe Thomas Jankvist,[2] Linda Marie Ahl,[3] & Mario Sánchez Aguilar[4]

[1]University of Gothenburg, Sweden; ola.helenius@ncm.gu.se

[2]Danish School of Education, Aarhus University, Denmark;

[3]Department of Education, Uppsala University, Sweden;

[4]Instituto Politécnico Nacional, CICATA Legaria, Mexico;

Articles in mathematics education that report on the implementation of innovations do not always report models for how the innovation program is supposed to attain its effects. We analyze an article with no such explicit theory of change with the aim of providing a background case for discussing the role of theories of change in implementation research. Our method consists of finding hypothetical determinants, which are the retrospectively deduced and explicated mechanisms for how the program reaches its effects. For the problem-solving promoting professional development program we analyze, we find a set of hypothetical determinants that are nested in a complex way. The use of predefined problems for both the lesson model and the professional development model and the collegial development of guiding questions for teachers to use with the predefined problems seems to form the core of a hypothetical theory of change for the program.

Keywords: Implementation research, theory of change, professional development.

Introduction

Within general implementation research (IR), it is often stressed that IR studies should make the 'Theory of Change' (ToC) of the innovation explicit and that this should play an active part in the final evaluation of a given implementation (Chen, 2012; Weiss, 1995). In IR within mathematics education, however, it has sometimes been found that the ToC is not made explicit—and furthermore that there is not always alignment between the ToC that underpinned the design of the innovation and that which was applied in the final evaluation, as exemplified by Jankvist et al., (2021) when examining the large-scale Swedish professional development project, the Boost of Mathematics (Helenius, 2021). Our observation of this 'implicitness' of a ToC has been further strengthened during an ongoing literature review on IR in mathematics education research concerned with studies that self-identify as dealing with implementations by using that term in the title or abstract. Preliminary results of this review have been presented elsewhere (Ahl, Aguilar et al., 2022a, 2022b, 2023). In our work of summarizing the whole review (Ahl, Helenius et al., 2023), we find that in order to identify a ToC, we often have to 'work backward', meaning that we have to analyze the evaluations with respect to finding the underlying objects of concern, i.e., the implicit ToC. This led us to question: *what potential consequences it might have had for the evaluation that a ToC was not explicated in the first place?* In particular, successful studies, or at least studies reported by the authors as being successful, were often much less explicit in identifying factors of influence, making it harder for us to produce hypotheses on what could be central elements in the ToC of the implemented innovation reported in the study.

To contribute to the discussion on what it entails to construct a ToC, in this paper, we analyze what we find to be a critical case (Flyvbjerg, 2006) with the purpose of reconstructing a possible ToC. We see this case as critical since the selected article reports a successful implementation and does not explicitly discuss a ToC. The chosen article for analysis is both well-written and presents an implementation case that is complex enough to make it possible to illustrate phenomena that are represented in a substantial selection of articles presenting implementations in mathematics education. The case concerns the reporting of a professional development program for promoting problem-solving teaching that turned out to have positive effects on students' problem-solving skills. Before we get to the case, we first provide a bit of background on the notion of theory in mathematics education, ToC in general, and theory-driven evaluations.

Briefly on the notion of theory

The literature, including the mathematics education literature, is rich in definitions and descriptions of the notion of 'theory'. For example, Radford (2008) considers a theory a triplet (P, M, Q), where P is a set of principles, M is a methodology, and Q is a set of paradigmatic research questions. Also, in the context of mathematics education research, Niss and Jankvist (2022) suggest the following "*minimalist* definition":

> A *theory* is a theory *of something*, i.e., it deals with certain sorts of *objects* and *phenomena* and includes terms for these. Its purpose is to produce *corroborated claims* about these objects and phenomena, typically in response to *questions* posed about them. These claims are generated by some *means*, on some *grounds*, involving some fundamental *methodology/ies*. (p. 17)

From a more general perspective, they do, however, state that in writings on the notion of 'theory', there are mainly three meanings that are prevalent, whether explicit or implicit:

- A theory can be a hypothesis that cannot be, or has not yet been, substantiated;
- A theory can be a collection of beliefs, rules or principles that are meant to guide action or behaviour;
- A theory can be a more or less connected edifice of claims intended to explain or predict phenomena occurring within some domain. (p. 17)

As can be seen, the first of these is in line with Mason and Waywood (1996), who describe a theory as a "hypothesis, or possibility such as a concept that is not yet verified but that if true would explain certain facts or phenomena" (p. 1055).

Theory of change and theory-driven evaluation

A ToC may be defined as "a collection of theoretical constructs that may be used to guide the design and implementation of a program and to evaluate its success or failure" (Jankvist et al., 2021, p. 1048). These theoretical constructs may originate from both inside and outside of mathematics education research. This is to say that "a ToC may be something developed locally as part of an implementation, for example, one that relies on subject-specific theoretical constructs" (p. 1048). Hence, a ToC may often be a 'local' theory. Furthermore, a ToC may also involve descriptions of a set of assumptions that explain mini-steps leading to long-term goals and connections between

program activities and outcomes occurring at each step (Chen, 2012). Hence, a ToC may often be closer to the definition of 'theory', which considers theory a kind of hypothesis rather than the minimalist definition (cf. section above).

When talking about theory-driven evaluations, the 'theory' in theory-driven may refer to the ToC, but also sometimes to a larger model encompassing the ToC, i.e., an "explicit theory or model of how the program causes the intended or observed outcomes and an evaluation that is at least partly guided by this model" (Rogers et al., 2000, p. 5). Still, according to Jankvist et al. (2021),

> a theory-driven evaluation presupposes an explicit ToC. Theory-driven evaluations have two key components: a conceptual one, that is, that the evaluation should explicate a program theory or model, and an empirical one, that is, that the evaluation should investigate how the programs (or innovations) cause intended or observed changes and outcomes. (p. 1048)

In Chen's (2012) account of theory-driven evaluation, the corresponding concept to a ToC is called a change model. Chen described that "each program must identify a leverage or mechanism upon which it can develop a treatment or intervention to meet a need. That leverage or mechanism is variously called the determinant or the intervening variable" (p. 18). When we analyze our critical case, we will do so by looking for such mechanisms and referring to them as *hypothetical determinants*. The term hypothetical is relevant for two reasons. First, we do not know exactly what the designers of the program under scrutiny consider to be the most important mechanisms of the program. There are several aspects of the program that we do not choose to discuss that might, in fact, be important. Second, before an evaluation of the role of possible determinants of the program has been made, any presentations of important mechanisms will remain hypothetical.

A critical case

To illustrate the difference between reporting some given result of an implemented program and approaching the evaluation and reporting from an implementation perspective, we look at the professional development (PD) program focusing on problem-solving (PS) reported by Saadati and Felmer (2021). The article is published in a ZDM special issue on empirical research on problem-posing and problem-solving. So, while the authors use the term 'implementation' when discussing the PD-program ARPA, the authors do not use concepts or constructs from IR to frame their implementation. Note that this is a well-written article, and our decomposition of it is a result of our intention to illustrate the potential of taking an implementation approach and the efforts involved when trying to reconstruct a ToC. We are not intending to criticize the choices made by the authors when writing the article.

The PD program under scrutiny was designed as a series of nine collegial workshops and seven monthly intermediate problem-solving lessons that teachers would carry out with their classes. From the program attendants, two teachers who taught two grade five classes each were chosen. The teachers implemented the monthly problem-solving lessons in one of their classes, while the other class acted as a control. By administering a problem-solving test in a pre/post fashion and analyzing the variance, the authors showed that the experimental group increased their problem-solving skills more than the control group (statistically significant at the $p<0.05$ level). The authors point out that it is important to analyze the student results in this fashion (and we agree). If improved student results

in the experimental group relative to the control group can be established, then the project can be considered to be successful. But, from an implementation perspective, questions remain: *What is it that was successful and why?*

Since Saadati and Felmer describe the organizational details of the PD project quite well, it is possible for us to extract elements of a ToC from their description. We do that as a series of if-then statements, which are the hypothetical determinants that are supposed to explicate "causal processes [that] are expected to happen to attain program goals" (Chen, 2012, p. 18).

On the surface, the program follows Desimone's (2009) suggested change model: a PD with the elements of content focus, active learning, coherence, duration, and collective participation that leads to changes in teachers' knowledge, skills, attitudes, and beliefs, which again leads to changes in the instruction, which in turn leads to improved student learning. However, the components of the change model proposed by Desimone are not detailed enough to capture which ideas and teacher behaviors actually lead to change. Saadati and Felmer describe the design of the program in much more detail, and several related and quite intricate determinants can be explicated. First, the program involves seven PS lessons that are supposed to follow a detailed model involving work in randomized groups with problems specific to the program. Also, a core idea is that the teacher should not give hints to struggling students but instead respond to student questions with guiding questions. A hypothetical determinant is, hence, that *if students attend lessons following this model, then students' PS skills will improve* (D1). Since students are evaluated individually but practice problem-solving in groups, a further determinant is that *if students improve their PS skills in groups, then their individual PS skills will improve* (D2). We can also explicate the sub-aspect of the previous hypothetical determinants. For example, *if teachers support students with high-quality guiding questions, then students' PS skills will improve* (D3), or *if students practice problem-solving with well-chosen tasks, then PS skills will improve* (D4). Note that D3 and D4 are (hypothetically) independent of each other and also independent of the PS model as a whole and are thus instead assumed to have independent effects on students' skill development. However, the guiding questions and specific problems are elements of the teaching model, so another hypothetical determinant involving these elements is: *if teachers are prepared with guiding questions for a particular problem, then they will be able to conduct the PS lesson with fidelity* (D5).

We also need to look at the role of the PD sessions. Here, one possible hypothetical determinant is *if teachers participate in PD sessions according to the model, then teachers will be able to carry out the PS lessons with fidelity* (D6). The PD sessions also follow a model, so the role of the expert PD session monitor might be important*: if an expert monitor oversees the PS session, then it is carried out according to the model* (D7). We can also look at more specific elements tying the lesson design and PD design; for example, *if teachers work collegially with a problem, then they can develop good guiding questions* (D8). Moreover, the role of the specifically chosen problems might not only have a role in being good for practicing problem-solving, but these may also play a role in the PD: *if problems are pre-chosen for the PS lessons, then the same problems can form the basis for the PD sessions* (D9). Figure 1 gives an indication of how the hypothetical determinants D1 to D9 are nested in a hypothetical ToC within the general change model, where the PD program supports a particular

lesson model, which in turn leads to the development of students' PS skills and how the different determinants relate to each other.

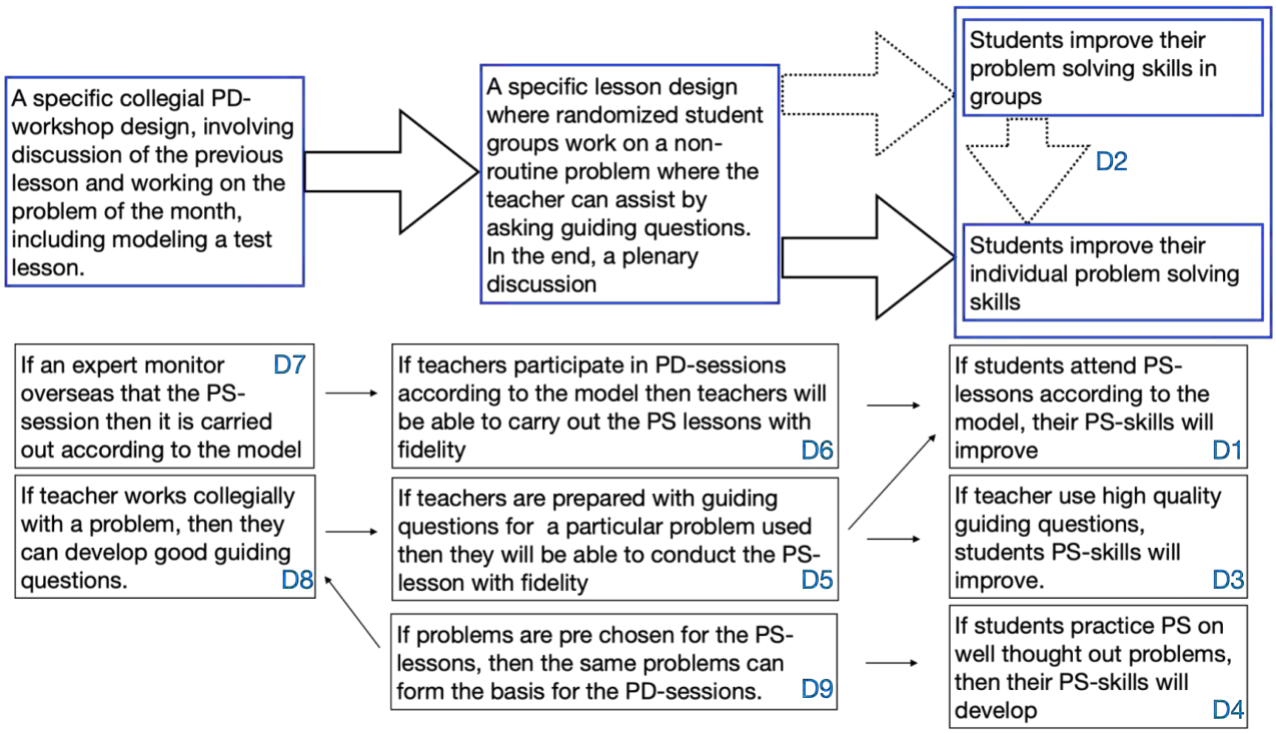

Figure 1: Elements of a hypothetical ToC in relation to the PD program's implementation described by Saadati and Felmer (2021)

It should be noted that there are many more possible determinants to explicate for the PD program, but our intention here is not to explicate all possible determinants, only a few that are enough to make our point. We also want to note that we believe some of the determinants above really are vital as to why ARPA worked, while others are examples that we do not think matter much but which cannot be ruled out either because no evaluation of them is reported in this analyzed paper.

Discussion

In this paper, we analyzed elements of an implicit ToC, in terms of hypothetical determinants, in a critical case (Flyvbjerg, 2006) reporting the implementation and student-level results of a PD program focusing on problem-solving. This exercise aims to illustrate the difference between reporting PD results with a focus on the problem-solving component and reporting on the connections between the implementation determinants in relation to the program's impact, i.e., having implementation research in focus.

The case we analyzed is interesting since PD programs promoting problem-solving are relatively common in mathematics education research. The program reported by Saadati and Felmer (2021), ARPA, was reported as successful. But, to answer our research question, from an implementation research perspective, we need to know more about the program determinants to be able to adapt the innovations to other contexts. Unless internal mechanisms in a program are explicated and preferably evaluated, it is difficult to identify why the program works. The more one knows about the

determinants and, more generally, the theory of change, the easier it will be to plan for replication. Transparency with determinants is also important for comparisons with other problem-solving programs, showing more mixed results. This is the general idea behind the theory-driven evaluation movement (Chen, 2012). The article by Saadati and Felmer (2021) represents a good example of a relatively typical article within the problem-solving subfield of mathematics education research. It is not written as an article in implementation research. The work we presented in this paper can be seen as an illustration of what researching the ARPA project from an implementation perspective could entail.

Of all the particularities of ARPA, we identified and focused on a set of hypothetical determinants, of which two are of particular interest: the role of the guiding questions and the role of the specific problems. In our view, it is likely that these hypothetical determinants have a core role in the functioning of the program. First, it is well-known from mathematics education research that a challenge when implementing problem-solving is that teachers often feel tempted to give too many hints or even give away the solution, in particular when students struggle (Heibert & Grouws, 2007). Since it is often argued that it is the struggle that induces the learning effect in problem-solving-based teaching, extensive support would ruin the learning effect (Heibert & Grouws, 2007). Unless one has great experience with a problem, providing students with the right guiding questions without giving away the solution is hard. In ARPA, teachers prepare productive guiding questions in collegial PD sessions. This brings us to the second aspect: having the problems for the problem-solving lessons fixed beforehand. It is this design choice that allows the teachers to prepare guiding questions. The role of the guiding questions and the role of the chosen problems are thus nested together and also nested with both the PD-session design and the lesson design in ARPA. This nesting is, in our view, probably an important aspect of the ARPA design.

Can ARPA be considered an implementation of research results from mathematics education? Yes, because the role of a productive struggle for learning, as well as the tendency of teachers to see such struggle as unproductive, together with a preference to instead explain solutions or give hints that lead to the avoidance of struggle, are well-documented empirical findings (e.g., Hiebert et al., 1996, Heibert & Grouws, 2007). Therefore, we can say that the theory of change for ARPA is, in fact, an explanatory type of theory that connects the important role of struggle with teachers' tendency to avoid it and suggests that well-prepared guiding questions are a way of providing both teachers and students with more acceptance for the struggle phase. If we posit the role of teachers being prepared with guiding questions as an enabling factor for teachers being likely to conduct problem-solving lessons where students are kept in enough productive struggle, then the set of determinants we have focused on can be seen as part of a more conclusive theory of change. The interdependent roles of the pre-specified problems and associated guiding questions in both the PD sessions and PS lessons then fit as a concrete realization of such a theory of change.

It should be noted again that our analysis of ARPA is theoretical and based only on the reporting of the program by Saadati and Felmer (2021). Because the article we analyzed did not report any evaluation other than on student problem-solving skill development, we cannot know if the determinants we explicated are, in fact, the important ones. We also, on purpose, chose to discuss only a few of the possible determinates of the program and, for example, did not say anything about

the role of group work in the PS lessons or the role of teachers spending time in the PD- sessions to work on and discuss the solution to the problems. In replications of ARPA, it would be worthwhile to include measures of the determinants we have highlighted, as well as other possible determinants.

Acknowledgments

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

References

Ahl, L. M., Aguilar, M. S., Jankvist, U. T., Misfeldt, M., & Prytz, J. (2022a). Implementation research on instructional sequences focusing on mathematical concepts and competencies: Results from a review. In J. Hodgen, E. Geraniou, G. Bolondi, & F. Ferretti (Eds.), *Proceedings of the Twelfth Congress of the European Society for Research in Mathematics Education (CERME12)* (pp. 4087–4094). Free University of Bozen-Bolzano and ERME.

Ahl, L. M., Aguilar, M. S., Jankvist, U. T., Misfeldt, M., & Prytz, J. (2022b). Professional development as a means for implementing mathematics education innovations: Results from a systematic review. In L. Mattsson, J. Häggström, M. Carlsen, C. Kilhamn, H. Palmér, M. Perez, & K. Pettersson (Eds.), *The relation between mathematics education research and teachers' professional development. Proceedings of MADIF 13.* (pp. 57–66). SMDF.

Ahl, L. M., Aguilar, M. S., Jankvist, U. T., Misfeldt, M., & Prytz, J. (2023). *Factors of influence for implementing curriculum reform: An analysis based on a systematic review* In P. Drijvers, C. Csapodi, H. Palmér, K. Gosztonyi, & E. Kónya (Eds.), *Proceedings of the Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13)*. Eötvös Loránd University and ERME

Ahl, L M., Helenius, O., Aguilar, M. S., Jankvist, U. T., Misfeldt, M., & Prytz, J. (2023). Implementation research in mathematics education: A systematic mapping review. *Implementation and Replication Studies in Mathematics Education*, *3*(2), 161–225. https://doi.org/10.1163/26670127-bja10015

Chen, H. T. (2012). Theory-driven evaluation: Conceptual framework, application and advancement. In R. Strobl, O. Lobermeier, & W. Heitmeyer (Eds.), *Evaluation von Programmen und Projekten für eine demokratische Kultur* (pp. 17–40). Springer. https://doi.org/ghtd6

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, *12*(2), 219–245. https://doi.org/10.1177/1077800405284363

Helenius, O. (2021). A stakeholder analysis of the development of a large scale professional development program. *Implementation and Replication Studies in Mathematics Education*, *1*(2), 1–30. https://doi.org/10.1163/26670127-01010009

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., Olivier, A., & Wearne, D. (1996). Problem Solving as a Basis for Reform in Curriculum and Instruction: The Case of Mathematics. *Educational Researcher, 25*(4), 12–21. https://doi.org/10.2307/1176776

Hiebert, J., & Grouws, D. A. (2007). The effect of classroom mathematics teaching on students'learning. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Information Age Publishing.

Jankvist, U. T., Gregersen, R. M., & Lauridsen, S. D. (2021). Illustrating the need for a 'Theory of Change' in implementation processes. *ZDM—Mathematics Education*, *53*(5), 1047–1057. https://doi.org/10.1007/s11858-021-01238-1

Mason, J., & Waywood, A. (1996). The role of theory in mathematics education research. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 1055–1089). Kluwer.

Niss, M., & Jankvist, U. T. (2022). On the mathematical competencies framework and its potentials for combining and networking with other theories and theoretical constructs. In U. T. Jankvist, & E. Geraniou (Eds.), *Mathematical Competencies in the Digital Era* (pp. 15–38). Springer. https://doi.org/10.1007/978-3-031-10141-0_2

Radford, L. (2008). Connecting theories in mathematics education: challenges and possibilities. *ZDM—Mathematics Education*, *40*(2), 317–327. https://doi.org/10.1007/s11858-008-0090-3

Rogers, P., Petrosino, A., Huebner, T., & Hacsi, T. (2000). Program theory evaluation: Practice, promise, and problems. *New Directions for Evaluation*, *2000*(87), 5–13. https://doi.org/10.1002/ev.1177

Saadati, F., & Felmer, P. (2021). Assessing impact of a teacher professional development program on student problem-solving performance. *ZDM—Mathematics Education*, *53*(4), 799–816. https://doi.org/10.1007/s11858-020-01214-1

Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. P. Connell, A. C. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), *New approaches to evaluating community initiatives: Concepts, methods, and contexts* (Vol. 1, pp. 65–92). Aspen Institute.