# Implementation in Mathematics Education: Fidelity to Implementation Design and Fidelity to Innovation Theory

BOYLAN, Mark <http://orcid.org/0000-0002-8581-1886>

BRILL

IMPLEMENTATION AND REPLICATION STUDIES IN MATHEMATICS EDUCATION 5 (2025) 1–29

IRME

brill.com/irme

# Implementation in Mathematics Education: Fidelity to Implementation Design and Fidelity to Innovation Theory

*Mark Boylan* | ORCID: 0000-0002-8581-1886
Sheffield Institute of Education, Sheffield Hallam University, UK
*m.s.boylan@shu.ac.uk*

Received 21 February 2025 | Accepted 1 April 2025 |
Published online 14 May 2025

## Abstract

The concept of fidelity is important to the implementation of mathematics education innovations. However, viewing fidelity as only adherence to protocols and planned activities may mean that the importance of underpinning causal mechanisms or innovation theory are not attended to fully. To expand and clarify the meaning of fidelity, I consider how the constructs of fidelity and theory are used in implementation methodology, including in mathematics education implementation studies. Alongside consideration of fidelity to implementation design as adherence to planned activity, I propose that fidelity to innovation theory is also important. The construct of fidelity to innovation theory supports assessing whether adaptations in implementation are acceptable or positive. I illustrate how these complementary constructs are applicable depending on the innovation theory, the implementation path characteristics, the fidelity focus, the generative instance, and the actor and activity.

The impact sheet to this article is available online at 10.6084/m9.figshare.28722908.

## Keywords

adaptation – fidelity – implementation – innovation – programme theory

## 1 Introduction

The concept of fidelity is central to the study and evaluation of implementation in education generally and important in mathematics education. The importance of considering fidelity in implementation and replication studies was previously highlighted in an editorial of this journal that identified diverse meanings in mathematics education (Jankvist et al., 2022). Here, I am concerned with the meaning of fidelity in relation to the implementation of innovations in education, innovations being 'programs, interventions, technologies, processes, approaches, methods, strategies, or policies that involve a change (e.g., in behaviour or practice) for the individuals (end users) enacting them' (Century & Cassata, 2016, p. 170).

My central argument in this paper is the need to consider fidelity to innovation theory alongside fidelity to implementation design. Common to multiple definitions of fidelity (O'Donnell, 2008) are related ideas of implementation as proposed, planned, or according to an intended model (Century & Cassata, 2016). Here, I refer to this form of fidelity as *fidelity to implementation design*, which is the most common meaning used in the study of implementation. This meaning of fidelity refers to the implementation of the visible and tangible activities and materials that are planned for, the duration of activity as intended and in the way intended.

Attention to fidelity to implementation design supports identifying essential features of innovations. This is important if an innovation is replicated or scaled (Aguilar et al., 2023), because identification of these features allows developers to provide organisational support structures to ensure these features are implemented. This supports implementation planning.

Additionally, when replicating and scaling, assessing fidelity to implementation design means that if there are challenges to implementation in a new context, then the appropriateness, or not, of the innovation for the context can be understood. Where the programme is not implemented as planned, attending to fidelity brings a focus on key moderators and contextual factors that influence implementation. This is because there is likely to be variation in fidelity depending on these moderators and factors. If implementation is successful but the intended outcomes identified in previous implementation do not occur, the innovation may not be appropriate to the new context.

In the study of implementation, there are different views of the value of fidelity, with two positions that can be broadly summarised as pro-fidelity and pro-adaptation perspectives (Century & Cassata, 2016). The pro-fidelity perspective emphasises the importance of implementing effective practices

without deviations. In this view, users should be encouraged and supported to implement with fidelity and adaptations are viewed as a failure of fidelity.

People who support a pro-adaptation view argue that high fidelity, when understood as adherence to the implementation plan, may be hard to achieve in practice because innovation users will respond to context. Further, those implementing innovations can add effective strategies or improve the innovation. Such adaptations can mean innovations can be more widely applied beyond any one specific context, and this can support sustainability (Century & Cassata, 2016). Rogers (2008) suggests there is a risk of overly focusing on activity to meet targets detailed in the theory of change model rather than the goals of the intervention. For mathematics education researchers concerned with implementation, adaptation may be desirable, acceptable, or inevitable in some innovations, without the innovation departing from the core difference with previous or usual practice.

For new, research-informed innovations, adaptation in use aligns with a different model of the relationship between researchers, designers, and practitioners, in which there is collaboration and co-design. The model positions practitioners as agents, not end users. Underlying the pro-adaptation perspective is an emphasis on the underpinning causal assumptions for a programme or its component parts, often summarised as the programme theory (Munter et al., 2016; Rogers, 2008). The term 'programme theory' may suggest a bounded or highly defined programme. However, my concern here is with the implementation of innovations more broadly (following Century and Cassata's 2016 definition), so I use the term innovation to avoid a confusion that 'programme theory' only refers to manualised programmes. By manualised programmes, I mean programmes in which there are detailed protocols, guides or manuals to follow in implementation. Thus, fidelity to the casual relationships of an innovation or a component of an innovation is fidelity to innovation theory.

Turning again to mathematics education, two tendencies can be observed. The first is that some research on innovations in mathematics education pays insufficient attention to fidelity or to a broader understanding of fidelity beyond adherence to the innovation protocol and implementation plan. For instance, in a systematic review of early mathematics innovations that focused on interventions for elementary students, Bos et al. (2023) identified that reporting of aspects of fidelity often prioritised adherence over other aspects. Thus, collectively, mathematics education could more fully attend to fidelity.

However, alongside this, those who embrace a pro-adaptation perspective go further than accounts of fidelity as adherence to a plan and consider the interplay between mathematical content, teacher activity, the learner, and the

learning environment. Adaptation is particularly important in pedagogical approaches that stress the importance of both teacher and learner agency in the learning process. Beyond individual classrooms, adaptive practice is important to the implementation of mathematical education programmes. This view of fidelity embraces more flexible and fluid aspects of implementation.

However, across these contributions, different reasons for and ways of expanding an understanding of fidelity are found. My aim here is to draw these contributions together, and in so doing, I argue that across them is a shared concern for the importance of articulating and attending to innovation theory in implementation, although the meaning of theory varies. Thus, I build on these contributions towards an expanded understanding of fidelity to justify complementing fidelity *to implementation design* with a complementary construct of *fidelity to innovation theory*. Distinguishing between the forms of fidelity supports the view that fidelity and adaptation are not opposite (O'Donnell, 2008).

The paper now proceeds as follows. In the next section, I describe the approach to reviewing literature and so the basis for the claims made in the paper. I then consider reasons why fidelity is important to study for various actors. I briefly review different definitions of fidelity found in implementation research and methodological literature. Positing fidelity to innovation theory as a concept raises the question as to what sort of theory there might be fidelity to. Addressing this question, I discuss the role of theory in implementation research both outside and inside mathematics education. These theoretical framings inform a hybrid review of a selection of mathematics education research literature. This then supports a fuller articulation and comparison of fidelity to implementation design and fidelity to innovation theory, followed by a discussion of how the model of fidelity is applicable in practice.

## 2 Approach

This paper is primarily a theoretical paper to develop the construct of fidelity to innovation theory as a complement to fidelity to implementation design. Three interrelated activities supported this development. Firstly, I examined the concept of fidelity in implementation research, specifically focussing on education. The outcomes of this part of the review are organised by a consideration of the reasons for studying fidelity, the variation in the importance of fidelity for different users and a summary of fidelity models. Secondly, I examined the conceptual construction of innovation theory both within and beyond the realm of mathematics education. Given the large bodies of

relevant literature, the selection of material was shaped and bounded by the frameworks and models of fidelity and theory that were considered or used in the corpus of literature identified within the third activity.

The third activity was to undertake a hybrid literature review (Turnbull et al., 2023). A hybrid literature review is a narrative review with systematic elements. I had previously posited the distinction between fidelity to innovation theory and fidelity to implementation design to consolidate and concretise previous insights about implementation, particularly in mathematics education (Boylan, 2025). I aimed to identify related concepts to the initial construction of fidelity to innovation theory in mathematics education implementation research studies. The purpose of the review was to identify the different ways constructs relevant to fidelity to innovation theory appear in mathematics education implementation research. The review strategy did not intend to address many other important questions related to fidelity in mathematics education implementation, nor was it appropriate for doing so.

I used the search terms 'implementation' AND 'fidelity' AND 'mathematics education'. The review was not a fully systematic approach given the limited number of search terms used and the use of only two databases — Scopus and Google Scholar — as well as because I drew on additional sources already known to me or derived from citations in the texts identified in the searches.

The Scopus search identified 15 texts, all of which were retrieved. The nature of the Google Scholar search algorithm meant 178,000 hits were identified. However, the Google Scholar algorithm means that those likely to be more relevant are returned earlier in the list of hits. Those near the top of the list would be placed there because, for example, search terms were in the title or used frequently in the text. Therefore, I retrieved and read texts in sequence and included them in an initial list if they met the following criteria:

- mathematics education focused
- implementation as a central theme (either of a programme, innovation, curriculum or textbook)
- a clear conceptual definition of fidelity explicitly stated, or that could be inferred from a sustained discussion of the meaning of fidelity beyond the specific programme, or from a description of the instruments, measures or methods to assess fidelity.

For the Google Scholar search, the likelihood that the text met these criteria decreased further down the returned list of hits. In the first 10 items, nine appeared relevant; in the 41st to 50th items, only two were included, and these two search results were positioned as 41 and 42 in the list. Following a full, more in-depth text reading, only one of these two was included in the final corpus for review. This approach to filtering identified a further 27 sources. After

resolving duplicates between Scopus hits and Google Scholar hits, I identified and retrieved 35 texts. Additionally, I retrieved and added 5 texts to the corpus by following up citations in the retrieved papers or texts already known.

In total, 40 texts were retrieved and read fully, and a second filtering process took place using the same criteria as initially used to filter texts from Google searches with the additional criteria:

– that a meaning of fidelity was explicit or implicit in the paper
– and the meaning of fidelity was different or broader than a focus on innovation and implementation structure and process or consideration of adherence and dosage (Dane & Schneider et al., 1998; Mowbray et al., 2003; O'Donnell, 2008).

There were some boundary cases where discussion of quality appeared to point to both adherence to design and consideration of constructs similar or related to innovation theory. These boundary cases were included. In total, 15 texts were identified. Each paper was read, and fidelity and related constructs used in the paper were identified and classified. In the introduction to this section, I described the activities as interrelated. I simultaneously considered fidelity and innovation theory as constructs in implementation research and interpreted them in the context of mathematics education implementation research. Towards the end of the review process, additional searches were undertaken to identify if aspects of fidelity to innovation theory and the term itself had been used in educational research beyond mathematics education.

## 3 Fidelity and Implementation

In this section, I expand the consideration of fidelity to implementation design in implementation research. Three things are looked at: models of fidelity and their parts; the significance of fidelity in implementation and research on implementation; and how the priorities of different actors in innovations change over time.

### 3.1 *Reasons for Studying Fidelity for Different Actors*

A core reason for studying fidelity is that it allows us to determine whether outcomes are due to the innovation by distinguishing the extent of implementation from the effects of the innovation (Century & Cassata, 2016; Dulak & DuPre, 2008; Lendrum & Humphrey, 2012). It can support identifying and defining the effective components and the degree or strength of implementation needed for efficacy (e.g., Lendrum & Humphrey, 2012).

The significance of examining fidelity varies depending on the stage of innovation. Some benefits focus on the initial stages of innovation, where the concern may be more with the 'what, why, and how' of implementation early in development (Crawford et al., 2019; Hulleman & Cordray, 2009). Later, the focus may be on how much (Crawford et al. 2019; Woodbridge et al., 2014). This focus may include identifying core components or the determinants of the innovation (Ahl et al., 2022; Chen, 2012). Focusing on fidelity when an innovation is being taken from its early stages of development to use in more common situations could help establish why it works differently in different situations (Hulleman & Cordray, 2009). However, this requires a reference point of the assessment of fidelity in an initial or earlier implementation (Ahl et al., 2022).

The rationale and motivation to formulate fidelity criteria, assess fidelity, and study reasons for variation in fidelity varies for different actors in innovation and programme implementation. Indeed, this may extend to different meanings and understandings of fidelity and its components depending on actors' roles. For example, such meanings may differ between practitioners and evaluators (Hansen, 2014; Jankvist et al., 2022). Table 1 presents reasons cited in discussion of fidelity in implementation methodological literature as to why considering fidelity may be important. It classifies these reasons by the concerns of four types of actors: innovation designers and developers, implementers and delivery teams, researchers and evaluators, and funders and sponsors. In practice, depending on the stage of an innovation's evolution, individuals, groups, or organisations may have more than one role concurrently. The relationships between defining or studying fidelity and other innovation, design, implementation, and evaluation activities are not linear. So, for example, defining fidelity may support the articulation of innovation theory and so precedes it, but it may be that innovation theory needs articulating to define fidelity.

### 3.2 *Fidelity Models*

A common meaning of fidelity is the extent to how far an intervention was conducted according to the original protocol or how the program developers had intended it to be carried out (e.g., Dusenbury et al., 2003; Proctor et al., 2011). The concept of intention is central to the concept of fidelity and frequently appears in definitions (e.g., Carroll et al., 2007; Century & Cassata, 2016).

Researchers often measure fidelity informed by models and frameworks that are based on conceptual theories and reviews of how researchers and others use and apply fidelity (see, O'Donnell, 2008; Mowbray et al., 2003; Dane & Schneider, 1998). These frameworks are concerned with the common meaning of fidelity: the implementation of the designed programme or innovation.

TABLE 1 The reasons for defining and studying fidelity for different actors

| | |
|---|---|
| Innovation designers and developers | Supports planning by identifying core activities and the resources needed to enable those activities (O'Donnell, 2008).<br>Supports articulation of the innovation theory, why and how an innovation works and clarifies the evidence base for the innovation (Durlak & DuPre, 2008; Greenberg et al., 2005; Lendrum & Humphrey, 2012; O'Donnell, 2008).<br>Identifies potential revisions for early-stage innovations (O'Donnell, 2008).<br>Helps to design effective monitoring systems and processes for (further) dissemination (O'Donnell, 2008; Lendrum & Humphrey, 2012).<br>Supports replication and scaling (Mowbray et al., 2003). |
| Implementers and delivery teams | Understanding the innovation rationale, what it is intended to achieve and how it works (O'Donnell, 2008).<br>For programmes, breaks delivery processes into steps, maps important links between programme components and helps maintain programme effectiveness by ensuring components are implemented (O'Donnell, 2008).<br>Guides appropriate and acceptable adaptation, revisions, and deletions (Century & Cassata, 2016; O'Donnell, 2008).<br>Identifies potential influences on programme delivery and allows mitigation of barriers to successful delivery, including through support systems (Century & Cassata, 2016; Durlak & DuPre, 2008). |
| Researchers and evaluators | Helps establish internal validity between the innovation and its outcomes (Carroll et al., 2007; Century & Cassata, 2016; Dane & Schneider, 1998; Dusenbury et al., 2003; Lendrum & Humphrey, 2012; Mowbray et al., 2003; O'Donnell, 2008).<br>Identifying core components and relative strength, the strength needed for efficacy and effects of differential implementation and threshold effects (Carroll et al., 2007; Century & Cassata, 2008; Dane & Schneider, 1998; Durlak & DuPre, 2008; Dusenbury et al., 2003; Hulleman & Cordray, 2009; Mowbray et al., 2003; O'Donnell, 2008).<br>Identifying appropriate measures for impact evaluation and statistical analysis that accounts for moderators (Mowbray et al., 2003) and so identifying critical components (Mowbray et al., 2003, O'Donnell, 2008). |

TABLE 1 The reasons for defining and studying fidelity for different actors (*cont.*)

| | |
|---|---|
| | Identifying key dimensions for implementation and process evaluation, such as fidelity and adherence criteria and differences, including across sites in multisite implementation (Mowbray et al., 2003).<br>Supports assessment of differentiation of the innovation from usual practice, including in control conditions (Century & Cassata, 2016; Hulleman & Cordray, 2009; Mowbray et al., 2003) and from similar innovations and implementation (Carroll et al., 2008).<br>Supports identification of contextual influences, moderators, and mediators and causal mechanisms (Century & Cassata, 2016; Dusenbury et al., 2003; Hulleman & Cordray, 2009; Lendrum & Humphrey, 2012).<br>Support assessments of the feasibility and implementability of an innovation (Century & Cassata, 2016; Dusenbury et al., 2003; Greenberg et al., 2005; O'Donnell, 2008).<br>Supports validity in meta-analysis (Carroll et al., 2007). |
| Funders and sponsors | Helps identify performance indicators to support accountability and quality assurance (Mowbray et al., 2003).<br>Allows comparisons across programs and innovations, as well as the relative extent of available evidence (Mowbray et al., 2003).<br>Supports decisions for initial funding, improvement, scaling, or discontinuation (Century & Cassata, 2016; O'Donnell, 2008). |

Education researchers have taken up and adapted these concepts, developed in health and related fields.

A long-standing distinction is between fidelity to structure and fidelity to process, with structure focused on what happens and process on the way in which it happens (Mowbray et al., 2003). A five-component model of fidelity (Dane & Schneider, 1998) comprises:

– adherence of delivery to the prescribed programme
– exposure of recipients to the programme activities
– quality of delivery
– responsiveness of participants
– differentiation of participant experience by planned activities only

Some aspects of implementation, such as how participants respond to the innovation or aspects of it, may have, it is argued, both structural and processual

aspects (O'Donnell, 2008; Mowbray et al., 2003). O'Donnell (2008) relates the five components to structure and process by suggesting that adherence and exposure are structural features, and quality and programme differentiation from previous or usual practice are processual aspects. For O'Donnell, quality refers to the way the programme is delivered. Participant responsiveness combines elements of both. For example, the use of materials relates to adherence, whereas the attitude to using the materials would be processual. Other models may have different or additional components — for example, programme complexity and delivery support as moderators of fidelity (Carroll et al., 2007) and whether the implemented programme differed across study groups (Mowbray et al., 2003). To address these issues, developers and evaluators may develop programme specific fidelity criteria (Mowbray, 2003; Dulak & DuPre, 2008).

## 4 Programme Theory: Outside and Inside Mathematics Education Research

As previously stated, this paper argues for considering fidelity to innovation theory alongside fidelity to implementation design. I now turn to innovation theory and discuss it in relation to evaluation and implementation methodology, and then to the types of theory and mechanisms found in mathematics education programmes and innovation.

In implementation science, programme theory refers to a conceptual framework that explains how and why an intervention or programme is expected to work. Many different terms are used (Rogers, 2007). However, broadly, it outlines the underlying mechanisms, processes, and assumptions that link programme activities to intended outcomes. 'Theory' is used in multiple ways with nuanced meanings, including when considering programme implementation (Helenius et al., 2023). It is often used as part of a compound phrase, 'Theory of Change', but this term too has different interpretations, often being used to mean the visual representation of a programme or innovation's causal structure that links inputs and activities to outcomes (Chen, 2012; Rogers, 2008) and often in the form of linear logic models (Boylan, 2023). Other terms are used instead of or alongside 'theory' to refer to the same or similar phenomena, such as determinants (Chen, 2012; Helenius et al., 2023), although this term is also used with reference to contextual determinants that are necessary for programme implementation (McLoughlin et al., 2021).

Chen (2012) distinguishes between 'the change model' and determinants — the causal processes — and 'the action model': 'An action model is a systematic plan for arranging staff, resources, settings, and support organizations

to reach a target group and deliver intervention services' (p. 18). Weiss (1997) makes a similar distinction between programme theory and implementation theory (Coldwell & Maxwell, 2018); using this distinction, programme theory is the theoretical basis for an innovation and implementation theory is how it is implemented (Weiss, 1997). The programme theory and the implementation theory together represent the full theory of change. Coldwell and Maxwell (2018) note that 'theory' in programme theory and implementation theory have different meanings and so they suggest that 'programme theory' and 'implementation path' is a clearer choice of terms. Implementation design echoes this meaning of implementation path.

Others use 'mechanism' to identify causal relationships (Lewis et al., 2018), often drawing on realist social theory (Coldwell & Maxwell, 2018). This view is rooted in a realist evaluation tradition that places mechanisms as central to programme theory, with one definition being the mechanisms that mediate between the delivery (and receipt) of the program and the emergence of the outcomes of interest (Weiss, 1998). 'Mechanism' is used to refer to causal relationships, although the specific definition and emphasis may depend on paradigms and other theoretical commitments. A generic definition of causal mechanism is the explanation/s for why programme activities lead to sought-for outcomes (Coldwell & Maxwell, 2018). Conversely, statisticians may more commonly consider 'assumptions' in inferential models (Colnet et al., 2024) as a means to identify causal relationships. The meaning of assumption in theory-based evaluation (Chen, 2012) is similar but distinct. The different terms represent important, if subtle, differences that are rooted in paradigmatic perspectives. However, regardless of the language used, these different terms draw attention to how designed programme processes and structure may cause intended outcomes.

Turning to mathematics education implementation studies specifically, it is important to consider implementation in connection with the specifics of the discipline and the mathematics education ecology (Koichu et al., 2021), and this extends to the implicit or explicit causal mechanisms — the innovation theory. However, often the intended causal mechanisms are not explicated (Munter et al., 2016; Helenius et al., 2023).

Munter et al. (2016) offer the beginning of a framework for considering different types of mechanisms relevant to mathematics education programmes and innovations. This framework is relevant to fidelity to innovation theory in mathematics education, as it illustrates the type of causal theories that might be examined when applying the construct. Their framework has three aspects: goals for student learning, the process of student learning, and the

organisation of instruction. The goals of student learning are subdivided using an earlier model (Kilpatrick, Swafford & Findell, 2001):

– conceptual understanding
– procedural fluency
– strategic competence
– adaptive reasoning
– productive disposition

The process of student learning refers to the relationship between the goals and mediating constructs, for example, mathematical authority. Organisation of instruction points to pedagogical mechanisms such as degree of explicitness and the teacher's role. They apply this framework to identify the programme theory for a selection of mathematics curricula and programmes in the USA, so demonstrating its potential. The framework might be further developed by considering and refining it in relation to a wider range of programmes and curricula, particularly those developed in other contexts and with different pedagogical traditions.

Further, the focus on learning mechanisms of the students might be complemented by considering mechanisms related to curriculum resources and materials, professional development and teacher learning mechanisms, and potentially other categories. Lindvall et al. (2022) consider both a theory of instruction and a theory of teacher change. From this perspective, Munter et al.'s (2016) framework provides a theory of instruction, but lacks a corresponding theory of teacher change.

## 5 Mathematics Education: beyond Fidelity as Adherence and Dosage

Against this background, I turn to the examples of mathematics education research texts identified through the hybrid literature review. Across the 15 texts considered, I identified two broad themes, each with subthemes. One of these themes comprised arguments or reasons to extend consideration beyond fidelity to the implementation design. The second broad theme relates to 'what' type of extensions were needed and, specifically, as to what types of theory were relevant. These studies help to address the question as to what types of innovation theory are important in an extended understanding of fidelity.

### 5.1 *Why? The Need for an Extended Understanding of Fidelity*

In this broad theme, there were two related issues. The first is that when applying the five-component model of fidelity (Dane & Schneider, 1998; O'Donnell, 2008), what constitutes implementation quality is somewhat open, and for

Crawford et al. (2019) and McNaught et al. (2010), implementation quality is related to programme theory. Innovation theory was an important reference in the design of instruments. Crawford et al. (2019) studied the implementation of a technology-enhanced fraction intervention for students with learning disabilities. Some of the intervention's components were relatively more prescribed, with content structured into levels. It also included components that provided more room for teacher flexibility, such as having students describe and defend their mathematical thinking. The theory of socially mediated learning informed this element. Gauging instructional quality required consideration of teacher practices in relation to such theory. For Crawford et al. (2019), this was considered through the construct of the 'core intent' in a technology-based innovation.

For McNaught et al. (2010) in a textbook based curriculum innovation 'presentation fidelity' was defined as follows:

> Presentation fidelity indicated the enacted curriculum being consistent with the expectations of the authors as expressed in author interviews and the author philosophy as reflected in the notes and suggestions to teachers in the Teacher's Edition of the textbook. (p. 6)

A second issue in this theme echoes the earlier discussion about the relationship between adaptation and fidelity, that is that adaptation is inevitable in implementation (e.g., Ahl et al., 2022; Hull et al., 2018; Koichu et al., 2022) and often desirable (Jacobs et al., 2017; Koichu et al., 2022; Munter et al., 2014) and that modifications can be useful and support achieving intended innovation outcomes. There are different terms used to describe the value of adaptation and variation. Considering the Maths Recovery intervention, Munter et al. (2014) refer to 'positive infidelity'. They define this as practices that had the potential to impact positively on outcomes for pupils that are explicitly or implicitly prohibited by the program developers' specification but, from the evaluators' perspective, compatible with (and potentially improving) the Maths Recovery programme (p. 107).

Jacobs et al. (2017) identify productive adaptations made by facilitators to meet the local and contextual needs of mathematics teachers. Such productive adaptations, they argue, are possible not only in relatively unprescribed professional development programmes but also in ones in which fidelity also means closely following the professional development materials and facilitation guidance. Following Seago (2007), they argue that professional development facilitators' adaptations may be productive, neutral, or fatal. The later adversely impacts fidelity by undermining the core principles of the

professional development. What is important in assessing the nature of adaptations are the central learning goals. Similarly, Ahl et al. (2022) provide an example of adaptations that may appear to be in tension with the program components but may still be aligned with the innovation's core determinants (Ahl et al., 2022).

These examples also demonstrate how adaptation that aligns with an innovation's core principles or theory can improve and develop it. Koichu et al. (2022) argue that the adaptations of innovation designers (in their case, mathematics education researchers) should also be considered. In their report, both teachers and designers realign the intended with the enacted in the case of their problem-solving innovation to maintain productive struggle as the central principle.

### 5.2 *What? Fidelity beyond Implementation Design*

In addition to considering why an extended understanding of fidelity is useful, the nature of the extension needed is also pointed to in the reviewed literature. There are three aspects of this, each providing a different emphasis on what constitutes innovation theory in relation to fidelity:

- fidelity to the designers or, in the case of curriculum materials, authors' theory of how and why the innovation could lead to desired change (Brown et al., 2009; Heck et al., 2012; Jankvist et al., 2022; McNaught et al., 2010; Superfine et al., 2015)
- fidelity to a body of theory or research (Ahl et al., 2022; Borrego et al., 2013; Clements et al., 2011, 2015; Clements & Samara, 2008; Koichu et al., 2022; Munter et al., 2014)
- fidelity to causal mechanisms intrinsic to the innovation (Ahl et al. 2022; Munter et al., 2016).

Brown et al. (2009) consider fidelity in whole-number lessons as enactments of the United States standards-based curriculum. They distinguish between fidelity to written lessons and fidelity to lesson authors' intended curriculum. Fidelity to written lessons means the extent of alignment between the instructional materials, including guidance. Fidelity to the intended curriculum refers to alignment with the authors' intended opportunities to learn. The assessment of the intended opportunities to learn lacks clarity. McNaught et al. (2010) consider mathematics textbook implementation and posit 'presentation fidelity,' meaning that the enacted curriculum is consistent with the expectations of the authors. Superfine et al. (2015) similarly found that adhering to the written curriculum materials did not ensure that the intended curriculum was implemented. These concerns echo concepts such as the core intention of an innovation (Crawford et al., 2019) and the 'original intent' and the 'original idea'

(Jankvist et al., 2022). A similar concept in curriculum enactment is whether there is fidelity to both the mathematical and pedagogical 'storylines' intended in the curriculum materials (Heck et al., 2012). In summary, in curriculum innovations, fidelity to the author's intention in the curriculum examples is fidelity to the pedagogical and learning mechanisms the author seeks to stimulate.

The second meaning is fidelity to a body of theory and research. This is found in the aforementioned studies by Ahl et al. (2022), Koichu et al. (2022), and Munter et al. (2014), where the innovation is informed by research on mathematics teaching and learning. Clements and colleagues give another example of this when they consider implementation fidelity in an early mathematics innovation (Clements et al., 2011, 2015; Clements & Samara, 2008). They sought to measure fidelity to 'a body of research about the characteristics and teaching strategies of effective teachers of early childhood mathematics' (Clements et al., 2011, p. 137). They developed measuring instruments to assess the extent of such fidelity in different participants' classrooms. Borrego et al. (2013), on the other hand, use a similar method but a different analytical scale. They look at descriptions of engineering course curricula and how well they match up with research-based ways of teaching.

The third meaning is fidelity to causal agents. Munter et al. (2016) refer to 'program theory' as active ingredients and mechanisms. Ahl et al. (2022) refer to 'programme determinants. The scale of the focus for their analysis is different. Munter et al. reviewed evaluations of programs that met the USA What Works Clearing House criteria, including a control group. In contrast, Ahl et al. (2022) are concerned with comparing case studies of implementation by two teachers. In both cases, the researchers argue that attention to fidelity can support understanding the underlying processes that may lead to change in practice and outcomes. Of these three meanings of theory, these two examples are the most aligned with the meaning of program theory as used in implementation research beyond education.

## 6 Fidelity Design and Fidelity to Innovation Theory

The different ways that the word 'fidelity' is used in research on mathematics education provide insights into the differences between fidelity to implementation design and fidelity to innovation theory. Fidelity to implementation design is the extent to which programme structure and activities are implemented as planned. It is concerned with operational and procedural aspects of the programme or intervention. The focus is how far delivery corresponds with the intended design. Fidelity to innovation theory, on the other hand, looks at

how closely the program follows the causal logic and the ways that inputs are linked to outputs and outputs are linked to outcomes. These principles and mechanisms explain how and why outcomes happen. Attending to fidelity to innovation theory ensures that core theoretical components and causal pathways are implemented and that the enacted programme is consistent with the innovation theory. The construct of adherence found in fidelity to implementation design is also central to fidelity to innovation theory, but rather than adherence to the planned activities, fidelity to innovation theory focuses on adherence to the mechanisms and causal processes that underpinned the planned design. Table 2 contrasts the two forms of fidelity with the constructs of adherence, exposure, and quality, drawing on Dane and Schneider's (1998) model.

For any given programme or component of a programme, three possibilities are possible in terms of the importance of the two different forms of fidelity:
– fidelity to implementation design
– fidelity to innovation theory
– fidelity to implementation design and fidelity to innovation theory.

TABLE 2 Fidelity to implementation design and fidelity to innovation theory compared

| **Aspect of fidelity** | **Fidelity to implementation design** | **Fidelity to innovation theory** |
|---|---|---|
| Focus | Structure, materials, protocols, planned design and processes | Underpinning theory, principles and mechanisms |
| Purpose | Ensures consistency with the planned design and the intended implementation path | Assures alignment with the innovation theory |
| Adherence | Implementation as planned with prescribed protocols, and activities | Implementation is congruent with the programme mechanisms and assumptions |
| Exposure | The 'dosage' — e.g., number of sessions, duration and materials used | Sufficient engagement with the causal mechanisms |
| Quality | How and how well the programme or components are implemented | How and how well implementation enacts the mechanisms |

## 7 Fidelity to Implementation Design and Fidelity to Innovation Theory in Programme Design, Implementation and Research

In this section, I consider how the distinction between fidelity to implementation design and fidelity to innovation theory applies to program design, implementation, and research. This provides an initial framework of dimensions for designers, implementers, and evaluators to consider when applying the design-theory distinction in practice. The discussion is organised into the following categories: the innovation's theory and content, the characteristics of the implementation path, the fidelity focus, the generative instance, and the activity and actor. Each category interrelates with the others. I illustrate these applications with reference to literature previously discussed.

### 7.1 *The Innovation's Theory and Content*

The causal mechanisms embedded in innovations imply different degrees of prescription. Relatively prescribed programmes explicitly or implicitly assume that a high level of adherence is needed for the causal mechanisms to operate (Munter et al., 2014). In these cases, fidelity to innovation theory may be redundant as a concept because innovation theory is embedded in the process and structure of a program. Examples of this are direct instruction programmes in which adherence to carefully designed scripts is how the innovation theory can be enacted. Theory is embedded in the programme structure and processes. In these cases, fidelity to implementation design may effectively be fidelity to innovation theory. This contrasts with less prescriptive programmes in which there is a greater expectation of professional experimentation (Clarke & Hollingsworth, 2002) or adaptation.

Studies in the above review that argue for the importance of productive adaptations were concerned with innovations of this second type. Jacobs et al. (2017) were specifically concerned with supporting professional development facilitators to adapt professional development materials productively and so emphasise their agency. Similarly, the Maths Recovery programme requires teachers to adapt to the needs of the individual learners they work with, informed by relevant learning theories (Munter et al., 2014). There are aspects of mathematics education curricula, pedagogy, and professional development where flexibility, agency, and an adaptive orientation are essential to intended practice. Both Koichu et al.'s (2022) problem-solving pedagogy and Hull et al.'s (2018) introduction of inquiry-informed practices exemplify this, and, in both cases, fidelity to innovation theory would potentially be a useful construct to support evaluation.

### 7.2 *The Characteristics of the Implementation Paths*

Fidelity to implementation theory is likely to be more relevant to complex rather than simple linear implementation paths. A simple design may have a single linear implementation pathway. This lends itself to a logic model representation of the Theory of Change. In logic model representations, such simple designs might be shown graphically as having a professional development or training component, followed by a path modelling implementation by teachers or other practitioners. However, such path models posit that different parts of the innovation form a linear sequence from inputs to outcomes (Boylan et al., 2018). In contrast, complicated designs have multiple components and multiple simultaneous causal strands, and complex designs may have feedback loops, tipping points, and emergent outcomes (Rogers, 2008).

Elsewhere, I have examined a complicated innovation in England — the MX program — that has four related sets of program components (Boylan, 2025). Some components are well developed with embedded programme theory, so fidelity to implementation design is important in implementation and evaluation. Others are being developed during implementation, so theory fidelity is more important because the process of implementation influences this development. There is not a fixed, linear implementation path that can be adhered to.

The national scale innovation of inquiry-informed pedagogy in Belize (Hull et al., 2015) was a complex programme, involving four macro components: inquiry-based instructional practices, math manipulatives, teacher professional development through courses and training in inquiry pedagogy, and ongoing mentoring for teachers in combination with a lesson study. In this innovation, there are interconnections between these different factors, as well as relatively high levels of variation in implementation in different national districts. In complex innovations, an important part of the innovation theory may be the mechanism of interaction and interconnection between different components. Ahl et al.'s (2022) consideration of determinants across a number of different components of an innovation show how attention to fidelity to innovation theory can support examining how different innovation components may interact — providing feedback loops.

### 7.3 *The Fidelity Focus: the Whole Innovation or Components*

The concept of fidelity focus follows on from understanding innovations as having multiple components and that they can be viewed at different levels of granularity. 'Fidelity focus' points to what component or aspect of a programme is of interest at a particular time. From a systems perspective, it may be relevant to consider different system levels (Munter et al., 2014). Alternatively,

different units of the innovation may be the focus (Cai et al., 2020). Units may be a module of a professional development component, a set of curriculum materials or one specific resource, a sequence of lessons, or a single lesson or a teaching interaction within a lesson.

Identifying forms of fidelity that are most relevant depending on the focus can support Theory of Change development and evaluation activity. This is especially true when determining if and how to measure fidelity to the focus. Where fidelity to implementation design is relevant, it is important to consider appropriate fidelity measures (O'Donnell, 2008). Where fidelity to innovation theory is indicated as being important, fidelity measures may not be appropriate, or they will need to be developed in relation to the programme theory rather than adherence. Clements and colleagues (Clements et al., 2011, 2015; Clements & Samara, 2008) provide examples of this in relation to developing structured classroom observation protocols to assess fidelity of implementation (to innovation theory and design). Rather, a better focus in Theory of Change development and evaluation activity in such cases is examining the underpinning programme theory relevant to the component under consideration. The concepts of mathematical and pedagogical 'storylines' in curriculum enactment are examples where the fidelity focus is on a longer arc of implementation than a single use of resources or materials (Heck et al., 2012).

### 7.4 *The Generative Instance*

A further consideration is the generative instance of an innovation. Every innovation has its own context, history, and potential or intended future instance. Some may be replications of earlier innovations (Cai et al., 2020), or if not a replication, may be informed by other, similar innovations or draw on some component previously implemented. The term generative instance aims to embrace the stage or life cycle of an innovation and its relationship to other instances or related instances of other innovations.

In approaches to innovation that favour randomised controlled trial evaluations, a common path is from initial piloting to an efficacy trial, then an effectiveness trial, and beyond that to scaling (Lendrum & Humphrey, 2012). In an efficacy trial, the study of implementation, including studying fidelity, enables testing the innovation theory to inform programme modifications and improvements (Greenberg et al., 2005). In an effectiveness trial, conditions will be less controlled, and studying fidelity supports understanding contextual influences. When scaling an innovation, the issues of adaptation and variation become more important, and the focus may shift to the relationship between fidelity and sustainability (Lendrum & Humphrey, 2012). Similar concerns are found where other design, development, and evaluation approaches

are used in scaling and replicating innovations (Augilar et al., 2023; Cai et al., 2020; Jankvist et al., 2022).

When designing and piloting innovations in the early stages of innovation development, fidelity to innovation theory may be particularly important. Examples are the early application of research findings. In such cases, adaptation will be common and often desirable to develop an effective innovation, as materials, systems, and processes will be tested in practice and refined. Here, articulating innovation theory and fidelity criteria for the enactment of the theory will be important. As innovations develop, fidelity to implementation design is likely to become more important as innovation protocols are developed and the innovation is manualised. The innovation structure and process may become codified, and the author's intention (in a curriculum scheme or textbook) or the designer's, developer's, or researcher's goal will be explicit and embedded. Fidelity to innovation theory may once again be an important consideration when innovations are implemented at scale, as variation and adaptation increase due to a more distant relationship between the program designers and implementers, and greater contextual variation.

Jankvist et al. (2022) note that there is a lack of agreement about what counts as a replication study. The construct of fidelity to innovation theory may help to identify whether an intended replication mobilises the same mechanisms as a previous study and so may support the classification of studies as replications or not.

### 7.5 *The Actor and Activity*

In discussing the reasons for studying fidelity, I organised these reasons in relation to different types of actors and, by implication, types of activity (in brackets). The actors were innovation designers and developers (design and development), implementers and delivery teams (implementation and delivery), researchers and evaluators (research and evaluation), and funders and sponsors (resourcing and monitoring). These categories are not mutually exclusive, and roles may be combined for some innovations or at different stages of development. Indeed, in research-based innovations in the earliest stages of development, a research team may effectively undertake all these types of activities.

Fidelity to implementation design and fidelity to innovation theory are complementary, and the extent to which one should be stressed will vary in relation to the role and purpose of actors. It is likely that designers, developers, researchers, and evaluators will need to attend to fidelity to innovation theory alongside fidelity to implementation design. An important interest for them is developing the innovation theory of change. Conversely, implementers and

delivery teams and funders and sponsors may give greater attention to fidelity to implementation design. However, such distinctions represent tendencies, and the previous four categories — particularly the content of the program and innovation and the characteristics of the implementation path — may be more important. The examples in the review of mathematics education literature are limited as they reflect the concerns of researchers and evaluators rather than other actors. However, there are indications within these studies of how research methods shape fidelity concerns and how fidelity to implementation design and fidelity to innovation theory might be operationalised. In quantitative evaluation, measures of fidelity support establishing whether higher degrees of fidelity are associated with higher levels of intended impact and so support establishing the casual theory of the innovation (see, for example, Clements et al., 2011, 2015; Clements & Samara, 2008, Hull et al., 2015; Munter et al., 2014).

## 8 Conclusion

I have argued that two forms of fidelity should be considered in implementation studies in mathematics education: fidelity to implementation design and fidelity to innovation theory. This builds on a distinction in implementation research between a focus on the designed intervention and the underpinning intentions (e.g., Proctor et al., 2011). Fidelity to implementation design corresponds to common meanings of implementation fidelity. In mathematics education implementation research, the word 'theory' can mean different things. In some cases, it refers to the designers' intentions — the assumptions that connect the innovation activity to the desired outcomes. At other times, it refers to larger bodies of mathematics education or other theory, and in some cases, it refers to clearly stated factors that cause outcomes. The expanded model of fidelity could potentially be used for implementation research in other subject areas in education or beyond education. Similar arguments informed by social practice theory have been made in science education (Buxton et al., 2015). In implementation in clinical psychology, certain types of treatment protocols may require flexibility within fidelity (Kendall & Frank, 2017). The concept of fidelity to innovation theory might guide appropriate flexibility.

There is a risk that attempting to achieve greater conceptual clarity about the meaning of fidelity in implementation research could create confusion in other ways. In an earlier formulation of the distinction between two forms of fidelity (Boylan, 2025), I referred simply to the shorter form of the terms, which are fidelity to design and fidelity to theory. However, because one way

of innovating is through design research, the short form, if used alone, may create some confusion, particularly given that the notion of fidelity is not one used much in design research. The longer form — 'fidelity to implementation design' — may help to draw attention to the implementation of a designed programme or intervention rather than design as a research activity.

Similarly, using the word 'theory' alone is potentially ambiguous, as it could refer to more universal social theory as much as the local theory of causality in an innovation. Therefore, the term 'fidelity to innovation theory' rather than 'fidelity to theory' may avoid a potential confusion. To add to this, the idea of program integrity (Meland & Brion-Meisels, 2023) has been put forward (Tsygan, 2025), which fits in well with the idea of fidelity to innovation theory. However, Dane and Schneider's (1998) framework for program integrity is a key source of the fivefold model of fidelity, and they use integrity and fidelity as synonyms and use both to mean adherence to the implementation design.

In general, whether to consider fidelity to implementation design or fidelity to innovation theory or both should be informed by considering the purpose of the actors (designers, implementers, researchers, evaluators, and sponsors), the nature of the programme theory, the implementation paths, the generative instances, and the focus of interest. Fidelity to innovation theory is particularly important for programmes in early stages of development where essential components are not yet identified. It is also important for programmes in which adaptation, agency, and flexibility of mathematics teachers are central to programme aims and important to intended outcomes.

A counterargument to adopting this broader fidelity framework is that a proliferation of concepts may not be helpful. Such a view might point to a lack of attention to all aspects of design fidelity beyond adherence and exposure in mathematics education implementation literature. An alternative to identifying fidelity to innovation theory as a distinct construct might be to consider it an aspect of the quality dimension (Dane & Schneider, 1998; O'Donnell, 2008). However, this does not account for cases where maintaining integrity to the programme theory may mean not adhering strictly to the innovation protocol — cases where productive adaptation is important. Fidelity to innovation theory is arguably more neutral than considering adherence to design as a mark of quality, as it does not assume that the underlying causal mechanisms are necessarily more effective than business as usual. It is also applicable across different stages or life cycles of an innovation, a particular implementation, or a generative instance. Promoting fidelity to innovation theory mitigates the risk of activity focused on meeting targets detailed in the model rather than the actual goals of the intervention (Rogers, 2008).

The distinction between the two forms of fidelity helps to explain why, for some mathematics education practices, fidelity and adaptation are a false dichotomy. This is because fidelity to the underlying principles of the programme or activity may require adaptation of materials or practice. From a pro-adaptation perspective, considering fidelity to innovation theory as a separate construct is helpful in addressing what adaptations are acceptable or desirable for each specific innovation or component. Acceptable adaptations support the enactment of the programme theory, and desirable adaptations are those that help to amplify the programme theory mechanisms. The distinction between fidelity to innovation theory and fidelity to implementation design offers an alternative perspective to the view that less structured innovations would benefit from a pro-adaptation perspective and highly structured and prescribed innovations from a pro-fidelity perspective (Ahl et al., 2022). Considering complementary forms of fidelity suggests that fidelity should be a concern no matter the degree of prescription or structure, but that the form of fidelity may vary.

Designing programmes to promote fidelity or evaluating programmes to assess fidelity is challenging, and the distinction between fidelity to implementation design and fidelity to innovation theory does not erase those challenges. Nevertheless, it is potentially a useful distinction that helps to draw attention to the relationship between planned programme activity, implementation, and enactment and the embedded programme theory. Importantly, it can help to address a long-standing issue of designers and evaluators not articulating programme theory and the causal mechanisms that are intended to lead to desired outcomes (Weiss, 1997; Rogers, 2007; Coldwell & Maxwell, 2018; Munter et al., 2016). However, in mathematics education research, a general weakness is that innovation or programmes theory of change are not always fully described.

In conclusion, fidelity to implementation design and fidelity to innovation theory are two critical but distinct aspects of implementing interventions in education and other fields. Fidelity to implementation design ensures that an intervention is delivered as planned, adhering to the specific protocols, activities, and structures outlined in its design. This consistency is essential for replicability, accurate evaluation, and ensuring that the intervention's operational components are executed effectively. On the other hand, fidelity to innovation theory focuses on maintaining alignment with the underlying theoretical principles and mechanisms that explain how and why the intervention is expected to work. It ensures that the core theoretical components driving the intervention's effectiveness are preserved, even as practical adaptations may be made to fit local contexts.

While fidelity to implementation design emphasises the 'what' and 'how' of implementation, fidelity to innovation theory addresses 'why' and ensures that the intervention remains grounded in its conceptual foundation. Both are essential for achieving desired outcomes: fidelity to implementation design ensures consistency and reliability, while fidelity to innovation theory ensures that the intervention's logic and mechanisms are implemented as intended. Research on mathematics education implementation pays little explicit attention to reporting and considering fidelity. Further, a challenge for putting fidelity to innovation theory to work as a construct is that programme theory in mathematics innovations is often not well described (Munter et al., 2016), and frameworks for identifying types of programme theory in mathematics education are lacking (Helenius et al., 2023).

Given these issues, strong claims would be unwarranted about the relative importance of the two forms of fidelity. However, it could be argued that finding a balance between these two types of fidelity is key to the successful implementation of many innovations, since it lets changes be made without affecting the integrity or effectiveness of the intervention. Together, the two dimensions of fidelity provide a comprehensive framework for implementing interventions that are both theoretically sound and practically feasible, ultimately enhancing their impact and scalability.

## Acknowledgements

An earlier version of this paper was presented at CERME 14, and I am grateful to the members of TWG 23 for their insights and reflections on the model of two forms of fidelity, and to the journal reviewers.

## References

Aguilar, M. S., Ahl, L. M., Jankvist, U. T., & Helenius, O. (2023). Towards characterization of scale and scaling in implementation research within mathematics education. *Implementation and Replication Studies in Mathematics Education*, *3*(1), 99–122. https://doi.org/10.1163/26670127-bja10012.

Ahl, L. M., Helenius, O., & Koljonen, T. (2022). Gauging fidelity to an implemented teaching model through the lens of the documentational approach to didactics. *Implementation and Replication Studies in Mathematics Education*, *2*(1), 45–75. https://doi.org/10.1163/26670127-bja10003.

Borrego, M., Cutler, S., Prince, M., Henderson, C., & Froyd, J. E. (2013). Fidelity of implementation of research-based instructional strategies (RBIS) in engineering science courses. *Journal of Engineering Education*, *102*(3), 394–425. https://doi.org/10.1002/jee.20020.

Bos, S. E., Powell, S. R., Maddox, S. A., & Doabler, C. T. (2023). A synthesis of the conceptualization and measurement of implementation fidelity in mathematics intervention research. *Journal of Learning Disabilities*, *56*(2), 95–115. https://doi.org/10.1177/00222194211065498.

Boylan, M. (2023). Researching implementation in mathematics education innovations: a practice-oriented theory of change model. 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*) (pp. 4284–4291). Alfréd Rényi Institute of Mathematics; ERME.

Boylan, M. (2025). The implementation of mathematics education programmes: fidelity to design and fidelity to theory. Paper presented at the Fourteenth Congress of the European Society for Research in Mathematics Education. Bolsano, Italy.

Boylan, M., Coldwell, M., Maxwell, B., & Jordan, J. (2018). Rethinking models of professional learning as tools: a conceptual analysis to inform research and practice. *Professional Development in Education*, *44*(1), 120–139. https://doi.org/10.1080/19415257.2017.1306789.

Brown, S. A., Pitvorec, K., Ditto, C., & Kelso, C. R. (2009). Reconceiving Fidelity of Implementation: An Investigation of Elementary whole number lessons. *Journal for Research in Mathematics Education*, *40*(4), 363–395. https://doi.org/10.5951/jresematheduc.40.4.0363.

Buxton, C. A., Allexsaht-Snider, M., Kayumova, S., Aghasaleh, R., Choi, Y. J., & Cohen, A. (2015). Teacher agency and professional learning: Rethinking fidelity of implementation as multiplicities of enactment. *Journal of Research in Science Teaching*, *52*(4), 489–502. https://doi.org/10.1002/tea.21223.

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., Cirillo, M., Kramer, S., & Hiebert, J. (2020). Working across contexts: Scaling up or replicating with variations? *Journal for Research in Mathematics Education*, 51(3), 258–267. https://doi.org/10.5951/jresemtheduc-2020-0007.

Carroll, C., Patterson, M., Wood, S. *et al.* (2007). A conceptual framework for implementation fidelity. *Implementation Science 2*(40). https://doi.org/10.1186/1748-5908-2-40.

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.

Chen, H. (2012). Theory-driven evaluation: Conceptual framework, application and advancement. In Strobl, R., Lobermeier, O., & Heitmeyer, W. (Eds.), *Evaluation von*

*Programmen und Projekten für eine demokratische Kultur* (pp. 17–40). Springer vs. https://doi.org/10.1007/978-3-531-19009-9_2.

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*(8), 947–967. https://doi.org/10.1016/S0742-051X(02)00053-7.

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, *45*(2), 443–494. https://doi.org/10.3102/0002831207312908.

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

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2015). Sustainability of a scale-up intervention in early mathematics: A longitudinal evaluation of implementation fidelity. *Early Education and Development*, *26*(3), 427–449. https://doi.org/10.1080/10409289.2015.968242.

Coldwell, M., & Maxwell, B. (2018). Using evidence-informed logic models to bridge methods in educational evaluation. *Review of Education*, *6*(3), 267–300. https://doi.org/10.1002/rev3.3151.

Colnet, B., Mayer, I., Chen, G., Dieng, A., Li, R., Varoquaux, G., Vert J-P, Josse, J. & Yang, S. (2024). Causal inference methods for combining randomized trials and observational studies: A review. *Statistical Science*, *39*(1), 165–191. https://doi.org/10.1214/23-STS889.

Crawford, L., Freeman, B., Huscroft-D'Angelo, J., Quebec Fuentes, S., & Higgins, K. N. (2019). Implementation fidelity and the design of a fractions intervention. *Learning Disability Quarterly*, *42*(4), 217–230. https://doi.org/10.1177/0731948719840774.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45. https://doi.org/10.1016/S0272-7358(97)00043-3.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*, 327–350. https://doi.org/10.1007/s10464-008-9165-0.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health education research*, *18*(2), 237–256. https://doi.org/10.1093/her/18.2.237.

Greenberg, M. T., Domitrovich, C. E., Graczyk, P. A., & Zins, J. E. (2005). *The study of implementation in school-based preventive interventions: theory, practice and research*. Washington, DC: Department of Health and Human Services.

Hansen, W. B. (2014). Measuring fidelity. In. Sloboda, Z. & Petras, H. (Eds.), *Defining prevention science* (pp. 335–359). Springer US.

Heck, D. J., Chval, K., Weiss, I. R, &. Ziebarth, S. W. (2012). Developing measures of fidelity of implementation for mathematics curriculum materials enactment. In Chval, K., Heck, D., Weiss, I. & Ziebarth, S. W. (Eds.), *Approaches to studying the enacted mathematics curriculum* (pp. 67–87). Information Age Publishing.

Helenius, O. Jankvist, U., Ahl L., & Aguilar M. (2023). On the implicitness of the 'theory of change' in implementations in mathematics education research. 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*) (pp. 4308–4315). Alfréd Rényi Institute of Mathematics; ERME.

Hull, D. M., Hinerman, K. M., Ferguson, S. L., Chen, Q., & Näslund-Hadley, E. I. (2018). Teacher-led math inquiry: A cluster randomized trial in Belize. *Educational Evaluation and Policy Analysis*, *40*(3), 336–358. https://doi.org/10.3102/0162373718768430.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, *2*, 88–110. https://doi.org/10.1080/19345740802539325.

Jacobs, J., Seago, N., & Koellner, K. (2017). Preparing facilitators to use and adapt mathematics professional development materials productively. *International Journal of STEM Education,* 4, 1–14. https://doi.org/10.1186/s40594-017-0089-9.

Jankvist, U. T., Aguilar, M. S., Misfeldt, M., & Koichu, B. (2022). How about fidelity?. *Implementation and Replication Studies in Mathematics Education*, 2(2), 131–148. https://doi.org/10.1163/26670127-02012020.

Kendall, P. C., & Frank, H. E. (2018). Implementing evidence-based treatment protocols: Flexibility within fidelity. *Clinical Psychology: Science and Practice*, *25*(4), Article e12271. https://doi.org/10.1111/cpsp.12271.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academies Press.

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

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.

Lendrum, A., & Humphrey, N. (2012). The importance of studying the implementation of interventions in school settings. *Oxford Review of Education*, *38*(5), 635–652. https://doi.org/10.1080/03054985.2012.734800.

Lewis, C. C., Klasnja, P., Powell, B. J., Lyon, A. R., Tuzzio, L., Jones, S., Walsh-Bailey, C. & Weiner, B. (2018). From classification to causality: advancing understanding of

mechanisms of change in implementation science. *Frontiers in Public Health*, *6* Article 136. https://doi.org/10.3389/fpubh.2018.00136.

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.

McLoughlin, G. M., Allen, P., Walsh-Bailey, C., & Brownson, R. C. (2021). A systematic review of school health policy measurement tools: Implementation determinants and outcomes. *Implementation Science Communications*, *2*(1), Article 67. https://doi.org/10.1186/s43058-021-00169-y.

McNaught, M. D., Tarr, J. E., & Sears, R. (2010). *Conceptualizing and Measuring Fidelity of Implementation of Secondary Mathematics Textbooks: Results of a Three-Year Study*. Paper presented at the Annual Meeting of the American Educational Research Association (Denver, CO, Apr 30–May 4, 2010).

Meland, E. A., & Brion-Meisels, G. (2023). Integrity over fidelity: transformational lessons from youth participatory action research to nurture SEL with adolescents. *Frontiers in Psychology*, 14, Article 1059317. https://doi.org/10.3389/fpsyg.2023.1059317.

Mowbray, C., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *The American Journal of Evaluation*, *24*(3), 315–340. https://doi.org/10.1016/S1098-2140(03)00057-2.

Munter, C., Cobb, P., & Shekell, C. (2016). The role of program theory in evaluation research: A consideration of the What Works Clearinghouse standards in the case of mathematics education. *American Journal of Evaluation*, *37*(1), 7–26. https://doi.org/10.1177/1098214015571122.

Munter, C., Wilhelm, A. G., Cobb, P., & Cordray, D. S. (2014). Assessing fidelity of implementation of an unprescribed, diagnostic mathematics intervention. *Journal of Research on Educational Effectiveness*, *7*(1), 83–113. https://doi.org/10.1080/19345747.2013.809177.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33–84. https://doi.org/10.3102/0034654307313793.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., … & Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*, 65–76. https://doi.org/10.1007/s10488-010-0319-7.

Rogers, P. J. (2007). Theory-based evaluation: Reflections ten years on. *New Directions for Evaluation*, 2007(114), 63–81. https://doi.org/10.1002/ev.225.

Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation*, *14*(1), 29–48. https://doi.org/10.1177/1356389007084674.

Seago, N. (2007). Fidelity and adaptation of professional development materials: can they co-exist? *National Council of Supervisors of Mathematics Journal*, *9*(2), 16–25.

Superfine, A. C., Marshall, A. M., & Kelso, C. (2015). Fidelity of implementation: Bringing written curriculum materials into the equation. *The Curriculum Journal*, *26*(1), 164–191. https://doi.org/10.1080/09585176.2014.990910.

Tsygan, J. (2025). *Research-designed, teacher-refined: making sense of teachers' adaptations of an instructional innovation*. Paper presented at the Fourteenth Congress of the European Society for Research in Mathematics Education. Bolsano, Italy.

Turnbull, D., Chugh, R., & Luck, J. (2023). Systematic-narrative hybrid literature review: A strategy for integrating a concise methodology into a manuscript. *Social Sciences & Humanities Open*, *7*(1), 100381. https://doi.org/10.1016/j.ssaho.2022.100381.

Weiss, C. H. (1997) How can theory-based evaluation make greater headway?, *Evaluation Review*, *21*(4), 501–524. https://doi.org/10.1177/0193841X9702100405.

Woodbridge, M. W., Sumi, W. C., Yu, J., Rouspil, K., Javitz, H. S., Seeley, J. R., & Walker, H. M. (2014). Implementation and sustainability of an evidence-based program: Lessons learned from the PRISM applied to First Step to Success. *Journal of Emotional and Behavioral Disorders*, *22*(2), 95–106. https://doi.org/10.1177/1063426613520456.