Teaching and Teacher Education 93 (2020) 103083

Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate

Paving the way for sustainable educational change: Reconceptualizing what it means to make educational changes that last

Mireille D. Hubers, Dr.

University of Twente, PO Box 217, 7500 AE Enschede, the Netherlands

HIGHLIGHTS

- This position paper provides an alternative theoretical framework of sustainable second-order educational change.
- Four characteristics are used to define sustainable second-order educational change.
- The essence of this definition lies not just in the individual characteristics, but also in what they represent as a whole.
- The underlying core premise is that the development of the whole system and its underlying components should be considered.
- Initial methodological considerations are provided to support future research.

ARTICLE INFO

Article history:
Received 23 May 2018
Received in revised form
30 March 2020
Accepted 1 April 2020
Available online 23 April 2020

Keywords:
Educational change
Second-order change
Sustainability
Professional development
School improvement
Organizational routines
Individual learning
Organizational learning

ABSTRACT

This position paper provides an alternative theoretical framework of sustainable second-order educational change. Here, sustainability refers to: 1) substantial changes made that affect the core of educators' everyday practice; 2) a longitudinal process that begins when educators contemplate making changes and ends when satisfactory achievement on the other characteristics is reached and overt learning efforts are stopped; 3) a process of individual and organizational learning as well as changes in behaviors; resulting in 4) significant positive effects on student outcomes. Thus, this definition focuses on educators and their schools, not on professional development programs. Methodological considerations are provided.

Change is ubiquitous in education. Such changes can be classified as first-order (e.g., a teacher who wants to increase his use of collaborative activities in his curriculum), second-order (e.g., data-based decision making practices are introduced to a school organization), and third-order change (e.g., the US Common Core; based on e.g., Lewis & Sahay, 2017; Tsoukas & Papoulias, 2005). The present paper focuses on second-order change, in which the school organization attempts to make a transformational change. The exact nature of such change efforts will be described in the theoretical framework, but examples include substantial changes that can occur within the boundaries of the school organization such as data-based decision making and research-informed practice (e.g., Brown, Schildkamp, & Hubers, 2017; Datnow & Hubbard, 2015); implementing inquiry-based learning (e.g., National Research Council, 2011); and the implementation of new technology in schools (e.g., Frank, Zhao, Penuel, Ellefson, & Porter, 2011). Teachers' professional development is often seen as a key component in implementing such changes and further improving the quality of education (Sleegers & Leithwood, 2010; Van Driel, Meirink, van Veen, & Zwart, 2012).

However, professional development programs vary extensively in their short-term effectiveness. Research has indicated that teachers make superficial changes in their teaching behavior or revert to their 'old' ways entirely after funding and support are withdrawn (e.g., Coburn, 2004; Hubers, Schildkamp, Poortman, & Pieters, 2017; McLaughlin & Mitra, 2001; Spillane, 2000). Even when teachers try to continue innovating their teaching behavior,

E-mail address: m.d.hubers@utwente.nl.

https://doi.org/10.1016/j.tate.2020.103083
0742-051X/

there is a heightened risk for misuse and 'lethal mutations,' thereby endangering the quality of the education they provide (Hubers et al., 2017; McLaughlin & Mitra, 2001). Moreover, previous research has documented the challenges faced by schools, including changes in their priorities, educational initiatives that compete for attention and resources, and shifting conditions in schools and districts, which all make it difficult to make lasting changes (Coburn, 2004; Cuban, 1993; Datnow, Hubbard, & Mehan, 2002; Fullan, 2016; Tyack & Cuban, 1995).

Consequently, it is not surprising that in the long run, most educational changes are not sustained (Baglibel, Samancioglu, & Crow, 2018; Breault, 2013; Fullan, 2016; Hubers et al., 2017; Wolthuis, Van Veen, De Vries, & Hubers, 2020). This lack of sustainability is problematic because second-order change presumes that it is the school's own ambition to change and resources such as time and money were invested in the change process. Therefore, sustainable change is deemed to be one of the biggest challenges in education (Fullan, 2016; Hargreaves, 2002; Harris & Jones, 2010; Hubers et al., 2017; Van Veen, Zwart, Meirink, & Verloop, 2010).

To face the sustainability challenge, there is an increasing demand for more knowledge about how and why changes were (not) sustained over time (Anderson, 2010; Coburn, Russell, Kaufman, & Stein, 2012; Sleegers & Leithwood, 2010). Such knowledge is essential in order to come to a deeper understanding of change processes and the dynamic between teachers' professional development and the resulting on-the-ground responses and actions (Coburn & Turner, 2011; Hubers, 2016). For example, it aids our understanding of what changes teachers and school leaders are willing to make (or not) and why.

Although knowledge about sustainability is crucial, this topic has not yet been addressed systematically. Research on sustainability is both scarce (Cohen & Mehta, 2017) and scattered, in that studies typically do not build on each other's theoretical frameworks and corresponding methodologies. To help the field move forward and to provide an alternative perspective on the sustainability challenge, the present position paper aims to achieve two goals. The first goal is to compare and contrast and elaborate on previous research in order to provide an alternative comprehensive theoretical framework of sustainable educational change. The second goal is to provide initial methodological considerations that correspond with the formulated definition of sustainability.

1. Defining sustainability

1.1. The definition of educational change and its challenging nature

Various classifications exist of the size and scope of change (e.g., Bartunek & Moch, 1987; Lewis & Sahay, 2017; Tsoukas & Papoulias, 2005). The present position paper uses a classification of first-, second-, and third-order change. Here, first-order change refers to small predictable interruptions in everyday practice. This type of change is an extension of the past and can be implemented with educators' existing knowledge and skills (Waters, Marzano, & McNulty, 2003), such as when a teacher wants to increase his use of collaborative assignments. Second-order change refers to large transformational or radical changes that call organizational assumptions into question (Lewis & Sahay, 2017). This is a 'paradigmatic' change that requires new knowledge and skills as well as adaptations in the prevailing values and norms (Tsoukas & Papoulias, 2005; Waters et al., 2003). Finally, third order change "not only involve[s] the transformation of the organization in focus but, through it, impacts on the broader institutional environment of which the organization is a member: the organization changes its institutional environment as it changes itself" (Tsoukas & Papoulias, 2005, p. 81). An example of third-order change is the implementation of the US Common Core. Thus, both second- and third-order change refer to transformational changes, but the former occurs within the boundaries of the organizations whereas the latter occurs (partially) outside the boundaries of the organization (Kuipers et al., 2014; Tsoukas & Papoulias, 2005).

The present position paper focuses on second-order changes in education. Another way to classify this type of change would be planned and episodic (Weick & Quinn, 1999). An example of such second-order change, given that a school is relatively unfamiliar with it and that the change can occur within the boundaries of their organization, is data-based decision making. In order to use data properly, most teachers need to gain data use skills (e.g., Jimerson & Wayman, 2015). Such learning processes can add up to organizational change, for example through changes in school policy, the organization of work, and work practices themselves (Coburn & Turner, 2011).

The difficulties associated with sustaining second-order changes is by no means education specific. For example, it is also evident in health care (e.g., Wiltsey Stirman et al., 2012) and commercial organizations (e.g., Sackmann, Eggenhofer-Rehart, & Friesl, 2009). The explanation as to *why* it is challenging to implement and sustain second-order changes differs per theory. For example, from the perspective of the Behavioral Theory of the Firm (originating from Cyert & March, 1963), the cause is attributed to the decision-making processes, in which elements such as problemistic (myopic) search and unresolved conflicts within the organization play a large role. In contrast, other scholars find this explanation in the attitudes towards organizational change (e.g., Choi, 2011). What makes the matter even more complex is that organizations are multilevel systems that do not change in a linear manner (Kozlowski & Klein, 2000). Thus, inherent to making second-order changes is the fact that not only individuals need to change, but that the entire organizational system needs to change as well. It is this multilevel and dynamic thinking that guided the current framework on sustainable second-order change in education.

1.2. Concepts related to sustainability

Before the framework is presented, it is important to address concepts related to sustainability as one easily gets entangled in the plethora of related definitions, see Table 1. The present paper focuses on the definition of sustainability.

1.3. Towards a definition of sustainable second-order educational change

There are several reasons to provide an alternative theoretical framework about sustainable educational change. The first is that there is conceptual ambiguity about *what* needs to be sustained. Of the studies that have contributed to the sustainability discussion (e.g., through determining the sustainability of a certain professional development program) the majority have lacked an explicit definition of sustainability, thereby causing some conceptual ambiguity. It is therefore challenging to determine exactly what these studies measured or investigated. These studies seemed to follow the general description that 'something' is continued in 'some way' (e.g., Adams, 2007; Drits-Esser, Gess-Newsome, & Stark, 2017; Labone & Long, 2016; Stringfield, Reynolds, & Schaffer, 2008). In general, such studies have focused exclusively on whether or not teachers changed their teaching behavior at post-test in comparison to the pre-test in a way that aligns with the professional development program that was implemented. One is left to infer that in these studies, sustainability seems to refer to replicating the behavior taught to them during these programs. However,

Table 1
Definitions related to Sustainability of Educational Change.

Concept	Definition
Initiation or adoption	Initiation and adoption both refer to the decision to use the change initiative (Fullan, 2016; Rogers, 2003).
Implementation or initial use	Implementation and initial use both refer to educators' first experiences in their attempt to put a change into practice, and this usually refers to a two- or three-year period (Fullan, 2016).
Continuation or institutionalization	An educational change is considered to be continued or institutionalized when it becomes a taken-for-granted feature of everyday practice (Datnow, 2005; Fullan, 2016) and loses its 'special project' status (Berman & McLaughlin, 1974).
Scale	Originally, scale was perceived from a merely quantitative perspective (e.g., Hargreaves & Fink, 2000; Slavin & Madden, 1994; Stringfield, Datnow, Ross, & Snively, 1998). This entailed a focus on increasing the number of teachers, schools, and/or districts involved. Coburn (2003) elaborated on this by arguing that scale is composed of four interrelated dimensions: depth, sustainability, spread, and shift in reform ownership. Depth refers to the nature and quality of the change implementation. Sustainability refers to changes that persist over time. Spread means that the reform is transmitted to greater numbers of classrooms and schools. Finally, ownership refers to the idea that, over time, the educators involved gain authority over the reform (Coburn, 2003).
Sustainability	In the present paper, sustainability of second-order change is defined as: 1) substantial changes made that affect the core of educators' everyday practice; 2) a longitudinal process that begins when educators contemplate making changes and ends when satisfactory achievement on the other characteristics is reached and overt learning efforts are stopped; 3) a process of individual and organizational learning as well as changes in behaviors; resulting in 4) significant positive effects on student outcomes.

questions such as why this reference was made or why the measurement is a valid and reliable indication of this replication are not addressed.

In addition, there is not yet a conceptual consensus when comparing the explicit definitions of sustainability with each other; see Table 2 for examples. Most of these studies also refer to the replication of behavior, but some added the required changes in the school system (e.g., the change becomes an institutionalized feature of the organization; Datnow, 2005). In contrast, some definitions pose that an exact replication of behavior is not required, and that educators can or even should make local adaptations. When which definition might be of use has not yet been explained, and conceptual clarity seems required.

The second reason to provide an alternative theoretical framework is that there are challenges associated with the fidelity and local adaptations paradigms underlying the aforementioned studies. One of these challenges is that through the use of these paradigms one does not gain knowledge about *what* works for *whom* in *what context* and *why*. These challenges are addressed in-depth in the section 'Implementation fidelity versus local adaptation, or neither?'

A third reason to provide an alternative theoretical framework is that authors who study sustainability hardly ever provide a clear-cut definition of the nature of (educational) change to which their work applies (e.g., in terms of first-, second-, or third-order; or planned versus emergent). This is problematic, because the nature of the change that is being studied is likely to show its own distinct rhythm (e.g., Weick & Quinn, 1999) and pose specific requirements on its environment, for example in terms of leadership (Van der Voet, Groeneveld, & Kuipers, 2014). Thus, not taking the nature of change into account will result in an incomplete —or maybe even distorted— understanding of the change process.

In order to move forward, conceptual clarity of sustainability is required. This paper aims to contribute to such conceptual clarity through the development of an alternative theoretical framework. To do so, the definitions presented in Table 2 were compared and

Table 2
Definitions of Sustainability in Previous Research in Education.

Authors	Definition
Anderson (2010, pp. 76–77)	"The basic idea is that some innovations lead to enduring changes in the way educators go about doing their work; that is, they become routine features of ongoing practice.…The key point is the idea that some efforts to change do result in what systems and complexity theorists refer to as a state change for the people and organizations involved."
Coburn et al. (2012, p. 140; p. 165)	"The degree to which teachers use reform-related practices in high-quality ways after support for these practices has dissipated". *Moreover,* "sustaining new instructional approaches is not simply about continuing to do the same thing. It requires that teachers and others make continual adjustments to new conditions and needs at the same time that they maintain the underlying pedagogical approach."
Datnow (2005, p. 123)	"Whether the reform lasts over time and becomes an institutionalized feature of a school."
Detrich, Keyworth, and States (2010, pp. 6–7)	"Sustainability refers to, at a minimum, 'maintaining over time'. There are three different but related types of sustainability: student outcomes, interventions, and systems." *Sustainability of student outcomes refers to* "the maintenance of student gains after the formal intervention has been discontinued." *Sustainability of the intervention refers to situations in which* "an intervention produces positive long-lasting results … so that other students can benefit". *Finally, sustainability of the system refers to the continued* "strength of the systems that support its effective implementation." "In this conceptualization of sustainability, sustainable student outcomes are considered to be a result of how well interventions and systems are implemented and sustained."
Fullan (2005, p. ix)	"The capacity of a system to engage in the complexities of continuous improvement consistent with deep values of human purpose."
Hargreaves (2002, p. 193)	"Sustainability in educational change comprises five key and interrelated characteristics. These are: 1) improvement that sustains learning; not merely change that alters schooling; 2) improvement that endures over time; 3) improvement that can be supported by available or achievable resources; 4) improvement that doesn't impact negatively on the surrounding environment of other schools and systems; 5) improvement that promotes ecological diversity and capacity throughout the educational and community environment."
Hubers et al. (2017a, p. 511)	"Behaviors [from educators] indicate sustainability (or lack of sustainability) when the rigor or frequency of the behavior changes over time. Each behavior should not necessarily be displayed ….However, where more types of behaviors are displayed, it seems more likely that sustainable changes were made." *In this definition, 'the behavior' refers to what is being targeted with a professional development program.*
McLaughlin and Mitra (2001, p. 304)	"Sustaining theory-based change requires more than maintaining the status quo or merely continuing the level of implementation achieved when special project resources and attention end. Sustaining theory-based change means deepening changes in practice and understanding in ways that keep practice vital, responsive to changes in students, subject area content and classroom contexts."

Note. This list is of definitions not exhaustive, but these definitions were chosen to illustrate the variability that currently exists.

contrasted, which guided the subsequent analysis and synthesis of previous research.

1.3.1. Similarities between definitions

When the definitions from Table 2 are compared with each other, they show a few similarities. Two of those similarities will be highlighted here.

1.3.1.1. Substantial changes. The first similarity is that all of these definitions seem to speak of substantial changes being made that affect the core of educators' everyday practice, for example, in terms of the content they provide and/or the pedagogies they use to provide that content. This aligns with the transformational nature of second-order educational change. To further specify what 'core of everyday practice' is affected by such substantial change, Eisner's (1992) five dimensions of schooling might be of use. The first dimension is the intentional, which refers to the educational values of the school. The second dimension is the structural, which refers to how subjects, time and roles are organized. The third is the dimension of the curriculum, which includes the content, the organization of this content and the learning activities. Fourth, the pedagogical dimension refers to how the curriculum is being taught to the students. Finally, the evaluative dimension of schooling refers to how the outcomes of schooling are assessed. Deep, meaningful and substantial results are most likely to be achieved when a change effort affects all of these dimensions. That way, one thinks comprehensively about educational change and integrates the change effort within the ecology of schooling (Eisner, 1992).

Thus, simply refining or replacing existing practices, or adding new practices does not necessarily 'count' as, and will not result in, a substantial change (Anderson, 2010). Examples of changes that do *not* affect the core of educators' everyday practice are increasing/decreasing class size and distributing the educational content differently across periods and days (Elmore, 1996). However, examples of changes that deliberately affect the dimensions of schooling also exist (e.g., Hodgkinson, 2013). Taken together, the first characteristic of the new sustainability framework is: Sustainable second-order change refers to substantial changes that are made by the schools' educators. These changes affect the core of educators' everyday practice.

1.3.1.2. Notion of time. The second similarity across the definitions in Table 1 is their notion of time. All definitions seem to agree that sustainability is about maintaining something over time. This not surprising, as previous work has indicated that it can take several years, some say even five to ten years, before an educational change is fully implemented (Darling-Hammond, 1988; Desimone, 2002; Fullan, 2016).

In several change models and practitioner manuals, sustainability is regarded as one of the last phases of the change process (e.g., Adelman & Taylor, 2003; Fullan, 2016; Hodges & Gill, 2015; Lewin, 1947). For example, Fullan (2016) indicated that sustaining change is preceded by the initiation and implementation stages. Indeed, this final phase in which formal learning ended and/or outside funding and support have been withdrawn is of special interest to those who study sustainability. That is the moment of truth, as it will illustrate the extent to which the school(s) or school district(s) can continue the change process without external support. Conversations with educators from various schools (in relation to Hubers, 2016; Hubers, Endedijk, & Van Veen, 2020) have indicated that in educational practice, sustainability is also often conceived as something that comes at the end of a professional development program, or even after the program has ended. Statements such as: "we still need to determine how we want to continue with these changes" were common, even though these schools were facing the end of their professional development program or support had already been withdrawn.

However, even though the end stage of a change trajectory is indeed of special interest, by itself, it does not provide an accurate demonstration of sustainability. The reason for this is that the extent to which educators are able to make sustainable changes is mainly determined by the progress they have made up until that point. For example, schools that were not able to sustain their data use over time already struggled to form practice and policy early on (Hubers et al., 2017). In addition, previous work has illustrated that the chances of achieving long-term changes increase when educators think early on about what roads they want to take to do so, for example, through developing longer-term models for continuing change activities and ensuring that the required resources will remain available (Dede, Rockman, & Knox, 2007). Taking all of this together, the current framework provides an alternative to previous definitions of sustainability by conceptualizing sustainability not as an end stage, but the actual path of change itself.

One might notice that a clear end point of this path has not yet been defined. This is problematic because second-order educational change is episodic in nature: at one point, the real transformational effort is over, although learning and fine-tuning might still occur after that. Most scholars seem to agree that this end point is reached when the change initiative has become an established (institutionalized) practice (e.g., Datnow, 2005; Fullan, 2016). This means that the change has lost its 'special project' status (Berman & McLaughlin, 1974), and becomes part of 'the way we do things around here.'

It is argued that the 'satisficing principle' can be used to determine such an end point (Foss, Heimeriks, Winter, & Zollo, 2012). This refers to a level of satisfactory achievement after which the organization decides to stop overt learning efforts (Foss et al., 2012). This means that, even though the change has not been implemented perfectly, the involved educators agree that it has been implemented to an acceptable degree in relation to the other characteristics of this definition. After this point, one can argue that all fine-tuning and maintenance of the change is no longer of second-order nature, through which the current framework might no longer apply. Taken together, the second characteristic of the new sustainability framework is: Sustainable second-order change refers to a longitudinal process that starts as early as when educators contemplate whether or not changes need to be made and ends when satisfactory achievement in relation to the other characteristics is reached and overt learning efforts are stopped.

1.3.2. Differences between definitions

The sustainability definitions presented in Table 2 also contradict each other in several aspects, two of which will be highlighted and resolved here.

1.3.2.1. Implementation fidelity versus local adaptation, or neither?. As stated before, making a second-order change in education means that the school transformed its way of working. This implies that the change is not limited to a couple of enthusiastic teachers. Thus, it seems timely to turn our attention towards the educators and their knowledge, skills, attitudes and behavior.

When analyzing the definitions in Table 2, one can distinguish a difference across their focus on either an implementation fidelity perspective or one that allows for local adaptation. This means that scholars seem to agree that professional development programs are required to bring about change, but that they disagree about how the program should be implemented and studied. In contrast, the present paper posits that, although professional development is

required to bring about second-order change, such development is not necessarily limited to formal programs. Moreover, neither the implementation fidelity perspective nor the local adaptation perspective sufficiently address the learning and development that takes place in the context of second-order change. An in-depth exploration of the matter is required to explain the challenges that arise when trying to define sustainable second-order change from either of these perspectives. After that, an alternative perspective is presented that aims to resolve these issues: a focus on change itself.

1.3.2.1.1. Implementation fidelity. Implementation fidelity (also known as treatment integrity) refers to the degree to which a professional development program is carried out as intended (Keller-Margulis, 2012). It is an explicit part of Detrich et al.'s (2010) definition (and implicitly studied in Hubers et al., 2017). Its proponents argue that achieving a high level of implementation fidelity is one of the best ways of replicating the success of change initiatives that were achieved in the original effectiveness studies (Carroll et al., 2007). Thus, it poses that as long as educators follow the 'recipe' provided by the program, they can implement and sustain the changes within their own school. Though this seems to make intuitive sense, when thinking it all the way through, one is left with quite a few challenges.

The first challenge is that defining sustainability from the fidelity perspective means that sustainability equals following the recipe (see for example King-Sears, Walker, & Barry, 2018). This is problematic, because it implies that we know exactly why a certain program was effective and that we have so much control over this, we know for sure that the program will work and will be sustained in exactly the same manner in other schools. However, we do not have that knowledge, because this is not how we study effectiveness. Often, effectiveness is determined through administering a pre-test (and possibly one or more measurements in between) and a post-test, which preferably are part of an experimental Randomized Control Trial. This results in information *that* something is effective (on average), but does not result in knowledge about *why* something works, in which instances it works, and what the underlying working mechanism is (Pawson & Tilley, 1997). This working mechanism is always influenced by the local context (Pawson & Tilley, 1997). So, it is not just a matter of determining that something is effective, but *why*, for *whom*, and *under what circumstances* this is the case (Cohen & Mehta, 2017; Pawson & Tilley, 1997). Though some authors claim the fidelity perspective contextualizes the outcomes of effectiveness studies (e.g., Desimone & Hill, 2017) and results in information about how and why a change strategy did (not) work, such insights are still based on pre-test post-test designs through which the aforementioned limitations hold.

Thus, in-depth considerations of *why*, *how*, and *under what circumstances* a change does (not) work, are missing from the fidelity perspective. The influence of such situational and contextual factors is often underappreciated (Harris, Jones, Sharma, & Kannan, 2013; Kennedy, 2010) even though previous research has indicated that factors such as culture, teachers' beliefs, individuals' sensemaking and agency, the extent to which capacity is built, strong school leadership, establishing shared goals, school climate, and available time/money influence the implementation of educational change (Baglibel et al., 2018; Coburn & Talbert, 2006; Dede et al., 2007; Desimone, 2002; Harris & Jones, 2018; Higgins & Spitulnik, 2008; Hubers et al., 2017b; Hubers, Poortman, Schildkamp, & Pieters, 2019; Kurland, Peretz, & Hertz-Lazarowitz, 2010; Lee & Seashore Louis, this issue; Liou, Canrinus, & Daly, this issue; Penuel, Fishman, Cheng, & Sabelli, 2011). Thus, even if previous research indicated that some schools used a professional development program to implement and sustain, for example, data use, this does not mean that other educators can mindlessly engage in the program and expect to see similar sustained changes. Maybe they need to work harder to positively stimulate teachers' beliefs on data use, develop stronger school policy, or need to take special needs from the student population into account. However, such contextual considerations are either not part of studies using the implementation fidelity perspective or based on insights about what, on average, resulted in a significant difference between conditions in the pre- and post-test design.

The second challenge of the fidelity perspective has to do with the extent of allowable deviation. For example, Detrich et al. (2010, p. 7) stated that: "if an intervention or process is not implemented as designed, then something else is being implemented; so, by default; the intervention or process has not been sustained." This seems to signal that even the slightest deviation, whether intentional or not, means that sustainability is out of the question. Though authors are increasingly arguing for taking the *degree* of fidelity into account (e.g., Anderson, 2017; Desimone & Hill, 2017; Hubers et al., 2017; Wiltsey Stirman et al., 2012), this does not seem to make it substantially easier.

One of the difficulties of taking the degree of fidelity into account is that the enacted practice remains to be compared with the 'benchmark practice' as defined by program designers (see, for example, Anderson, 2017). This means that there is still one fixed desirable outcome, which is that the enacted practice is identical to the benchmark practice, regardless of whether or not such an outcome is within reasonable reach of the participating educators. A second reason is that fidelity scores might be more difficult to interpret than often assumed. For example, 21st century education is about addressing skills such as critical thinking, non-routine problem solving, teamwork and information fluency (Ananiadou & Claro, 2009). Say a professional development program for teachers addresses all of those dimensions effectively. When helping the schools to determine the sustainability of their changes, you find out that at one school, teachers only focus on non-routine problem solving, and do so frequently and well. In contrast, at one of the other schools, you determine that the teachers focus on all the 21st century skills in an adequate manner, yet not as often as you would like to see. Which school displays sustained change in line with the implementation fidelity perspective, the one that addresses one element very well, or the one that addresses the complete set of elements, but only at an adequate level? Or can neither school show sustainable changes because they did not have a perfect match with the original program? And, even if one finds two schools with exactly the same fidelity score, there is a challenge left. The fidelity perspective then suggests that these two schools should, on average, obtain a similar effect with the change. This might be too simplistic due to the abundance of (contextual) factors that are at play in education that partially influence why a similar change can have different effects across contexts.

To summarize, the fidelity perspective reduces sustainability to whether or not (or to what degree) the recipe of a professional development program was followed. Methodological issues aside, it reduces second-order change to replicating what is learned during the program into one's own educational context. Moreover, it presumes that identical replications have similar value across contexts. Taken together, the fidelity perspective seems too simplistic to fully understand and support the transformational nature of sustainable second-order change.

1.3.2.1.2. Local adaptation. The counterpart of implementation fidelity is allowing for local adaptation. Local adaptation refers to the idea that the professional development program itself is adapted to the local context and/or that educators themselves make adaptations in the materials they received (e.g., Clarke & Dede, 2009; Desimone, 2002; Greenberg, 2004; Squire,

MaKinster, Barnett, Luehmann, & Barab, 2003). One way in which this can be done is through improvement science (see for example Hannan, Russell, Takahashi, & Park, 2015). Thus, from the local adaptations perspective, sustainability is not an exact copy of a professional development program; it is one's version of that program. Here too, sustainability equals following a recipe, only it is one's own version of the recipe.

The risk of defining sustainability from a local adaptations perspective is that one loses sight of what changes are actually implemented. When local adaptations are made, there is a heightened risk of misuse and/or "lethal mutations", the latter referring to "[teachers who] modify practice, or extend it and unintentionally violate rudiments of the reform's theoretical base" (McLaughlin & Mitra, 2001, p. 307). An example of a 'lethal mutation' was made by a teacher who was part of a data use program. The grades his students received were considered to be too high, and he used his data to deliberately include difficult questions in his tests to pull down students' grades (see Hubers et al., 2017). Especially when educators do not yet have all the knowledge required for successful implementation of a new practice (McLaughlin & Mitra, 2001) one should be vigilant for the risk of misuse and/or 'lethal mutations'. Another example is that schools that tried to implement data use practices all deviated from the original professional development program (Hubers et al., 2017). The educators no longer evaluated whether the quality of schooling improved through their use of data, although this had been part of their training. As evaluation is an essential aspect of data use (Bryk, Gomez, Grunow, & LeMahieu, 2015), the way in which data was used at these schools can at best be described as being incomplete (Hubers et al., 2017). Thus, by allowing for local adaptations, the teaching practice can, at some point, lose its coherence and result in inaccurate and uneven implementation (Desimone, 2002).

Besides this risk of "lethal mutations", for researchers, this perspective of local adaptations makes it challenging to determine the effectiveness of the professional development program. After all, when a program is not implemented in exactly the same way in each school, any comparison between these schools and the control group becomes void. Add to this the aforementioned contextual differences between schools, and we are left unsure of who sustained what and why, and what causal inferences can be drawn and why.

Taken together, both the implementation fidelity perspective and the perspective of local adaptations bring serious challenges to the definition of sustainability. Although it is argued that the two perspectives can or even should be united (e.g., Webster-Stratton, Reinke, Herman, & Newcomer, 2011), this would not solve the aforementioned issues. Most importantly, it would not lead to additional insights into the working principle of change: *why* and *how* does something work, for *whom* does it work and *under what circumstances* does it work. A possible solution to this could be to switch our focus altogether. Both the program fidelity and the local adaptation perspective take the professional development program as their focal point. Fullan (2016) termed such perspectives as innovation-focused approaches. But, what about learning that occurs outside the formal programs, or at the level of the school? Our main goal as educational scholars is to help (networks of) schools improve their quality of education. So, let us oppose the innovation-focused approaches and take the school (or school system) as our focal point when defining sustainability. To do so, it is important to look at both learning processes and behavioral change processes.

1.3.2.1.3. ***Neither implementation fidelity nor local adaptation: Learning and changing behavior.*** The actual changes that are being made in educational practice can serve as a vantage point to find an alternative for the implementation fidelity and local adaptation perspective. Over time, educational change can affect one or more of the aforementioned dimensions of schooling: the intentional, structural, curricular, pedagogical and evaluative dimensions (Eisner, 1992). For example, individual teachers can make changes in their everyday practice through developing new lesson content, providing different types of instruction to students and through developing new assessment strategies (e.g., Bean, Dole, Nelson, Belcastro, & Zigmond, 2015; Tam, 2009). Making changes in one's teaching practice is a reflection of the change process, which is why such behavior needs to be included in the definition of sustainability. In addition, such behavior is part of teachers' process of professional development (e.g., Clarke & Hollingsworth, 2002; Guskey, 2002).

The professional development process is also composed of gaining knowledge and changing one's beliefs and attitudes (Clarke & Hollingsworth, 2002; Guskey, 2002). So, an integral part of the change process is not just whether changes in behavior were made at the individual level, but also whether and how teachers gained knowledge and/or changed their beliefs and attitudes (see, for example, Mouza, 2009; Ng & Nicholas, 2013; Whitehead, 2010). Therefore, teachers' professional development is often seen as a key component of change processes (Sleegers & Leithwood, 2010; Van Driel et al., 2012). This is especially the case with second-order change, as it calls organizational assumptions into question and requires new knowledge and skills as well as adaptations in the prevailing values and norms (Lewis & Sahay, 2017; Waters et al., 2003). Although formal programs usually play an important role in teachers' learning processes, it would be an oversight to focus our definition of sustainability (and its accompanying research) exclusively on the learning that is tied to such programs. For example, teachers also learn explicitly and implicitly in and from their everyday practice (Eraut, 2000; Horn & Little, 2010; Little, 2012a).

However, limiting our definition of sustainability to teachers' learning and changing behaviors would still underestimate the context in which this learning and behaving is displayed. Therefore, it is important to also take the organizational level at which learning and behaving take place into account. As Opfer and Pedder (2011, p. 379) stated: "any attempt to understand teachers' professional learning at only a subsystem level [the level of the teacher] must be understood as partial, incomplete and biased." Adding the perspective of organizational learning would be beneficial, as this includes more than the sum of all individual learning (Sleegers & Leithwood, 2010). Though different conceptions of organizational learning exist, most authors agree that it refers on the one hand to the exploration of new knowledge or experiments both within and outside of the organization, and on the other hand to the exploitation (use and refinement) of existing knowledge (Crossan, Lane, & White, 1999; Huber, 1991; March, 1991; Supovitz, 2010). It includes activities such as developing a shared vision, collective planning and decision making, and engaging in dialogues to understand others' frame of reference (Mitchell & Sackney, 1998). Ultimately, organizational learning leads to the embedment of "the gained knowledge into the regular routines and practices of the organization so that the new knowledge is codified and practiced by all" (Supovitz, 2010, p. 721).

Not only can learning be defined at the level of the school, behavioral changes can be defined at the level of the school as well. For this, the concept of organizational routines will be of use. Organizational routines refer to 'repetitive, recognizable patterns of interdependent actions, involving multiple actors' (Feldman & Pentland, 2003, p. 96), and much of the work in schools takes place in and through such routines (Nelson & Winter, 1982). Roughly speaking, they refer to policy and practice respectively. The ostensive aspect represents standard operating procedures, taken-

for-granted norms and one's subjective understanding of the routine, whereas the performative aspect represents the everyday practice of specific actions, taken by specific people, at specific times (Feldman & Pentland, 2003). Practically, examples of such organizational behavior include developing and communicating a clear vision, developing standard operating procedures, and holding teachers accountable for their own practice (Sherer & Spillane, 2011).

When looking at learning and change it also makes sense from a practical perspective to include both the individual level and the organizational level. First of all, the ultimate goal of educational change, albeit indirectly, is to improve students' achievement (Stoll, Bolam, McMahon, Wallace, & Thomas, 2006), general well-being, and/or future (educational) opportunities. This is a goal at the level of the school, and therefore, it can be assumed that schools are likely to want to continue the positive changes made by individual teachers (or, at the very least, not cancel them out). Second, it is likely that schools want to change the quality of the education they provide in a way that is not dependent on individual teachers, but lasts even when certain key staff members leave the school (Sherer & Spillane, 2011). Third, previous research has extensively showed how teachers' practice and the changes they are (not) willing to make within that practice are influenced by organizational factors, including: receiving support, guidance and feedback from colleagues; a supportive school climate; and a learning culture (Thurlings, Evers, & Vermeulen, 2015). So, although teachers can perform their everyday tasks in relative isolation from their colleagues, they remain part of the organizational system and will be influenced by what happens in this system.

Taken together, using a perspective of learning and changing behavior helps us appreciate the vastness of sustainable change: it refers to both the individuals as well as the organization as a whole who learn and change their way of doing things. It is through such a multilevel process perspective that we can begin to have confidence that we understand the mechanisms underlying the implemented changes, and that these changes were intentional, effortful and enduring. Thus, perspectives on both individual and organizational learning and change need to be part of the definition. This results in the third characteristic of the new framework: Sustainable second-order change refers to a process of individual and organizational learning as well as of changing behaviors at both the individual and organizational levels.

1.3.2.2. Achieving intended outcomes. The second difference across the definitions in Table 2 concerns whether or not a certain quality standard needs to be achieved. For example, Fullan (2005) and Coburn et al. (2012) indicated that sustainability refers to high quality use of new strategies. Hargreaves (2002) specified this further by indicating that these improvements should not have a negative impact on the surrounding environment. In contrast, the other definitions do not specify the quality of the change as an explicit requirement for sustainable change.

In the present paper, it is argued that at least some level of good quality is required for sustainability because the goal of a change strategy is never to just have that change strategy in place (e.g., Morrissey, 2000). This is especially the case with second-order change, which centers around the idea that a transformational effort is required to achieve a certain goal. Therefore, in the end, students must profit from it, albeit indirectly (Stoll et al., 2006). Moreover, whether or not educators perceive the benefits of their hard work in the achievements of their students is a known factor that drives and provides input to the change process (Guskey, 2002; Spillane, Reiser, & Reimer, 2002).

However, it should be noted that outcomes such as increased student well-being or equality would be beneficial (and desirable) as well, and change should never be sustained at the cost of such core values. The same holds for teachers: it seems highly undesirable if core values such as their well-being are substantially diminished through the implementation of the change, regardless of the amount of learning, change and positive outcomes that can be detected. For additional insights into teachers' well-being during educational change processes, see, for example Hargreaves, Shirley, & Wanee, (this issue).

The criterion of achieving improved student outcomes is quite strict, yet also extremely important. After all, when an educational change does not bring about the desired positive outcome, why would all parties involved continue to invest time, effort and resources in it? Therefore, the final characteristic of the new sustainability framework is: Sustainable second-order change refers to a change process that results in improved student outcomes.

1.4. A comprehensive definition of sustainable second-order change

Taken together, sustainable second-order educational change within a school refers to (see also Fig. 1):

1. substantial changes that are made by the schools' educators. These changes affect the core of their everyday practice.
2. a longitudinal process that starts as early as when the schools' educators contemplate whether or not changes need to be made. It ends when satisfactory achievement on the other three characteristics is reached and overt learning efforts are stopped.

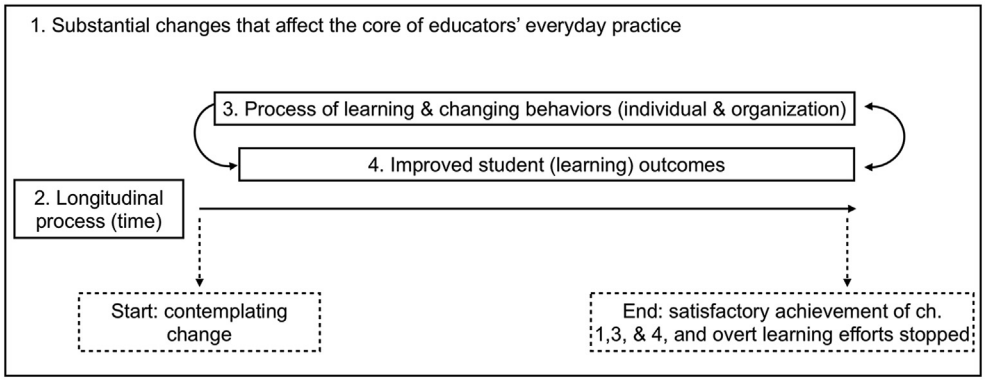

Fig. 1. The Framework of Sustainable Second-Order Educational Change: Four Characteristics and their Underlying Relationships.

3. a process of individual learning (professional development) and organizational learning as well as of changing behaviors at both the individual and organizational levels.
4. a change process that results in improved student outcomes.

It is important to note that the third and fourth characteristics can mutually influence each other. Although educational change starts with learning and making actual changes in behaviors, when this process results in improved student outcomes, these outcomes can give a new impetus to the change process. See also Fig. 1 for an illustration of the relations among the four characteristics of sustainable second-order educational change.

Before addressing the initial methodological considerations that accompany the definition of sustained second-order change, a reflection on the whole definition seems appropriate. Importantly, the essence of this definition lies not just in the individual characteristics, but also in what they represent as a whole. This definition poses that sustainable second-order change reflects the change process itself: it represents the changes that are made in the core of educators' everyday practice by the individual educators and the school (system) as a whole, students' changing (learning) outcomes and how these outcomes feed back into the change process. It poses that sustainability of second-order change cannot be understood as a narrow phenomenon, for example, by focusing only on whether the specific recipe of a certain professional development program was followed to the letter, as so much more is happening in terms of learning and changing in the broader context of the school (system). Neither can sustainability be understood by focusing on the moment in time when the change becomes part of the organizational routines, as this development in itself as well as the accompanying learning and development processes reflect the sustainability process. Sustainability is not an after-thought after some types of changes already occurred, it is the build-up and continuation of the change process. Thus, the core premise underlying the new definition presented here is that our understanding of sustainability remains incomplete and inaccurate when the overall development of the whole system and its underlying components are not taken into account. This dynamic and multilevel perspective on change provides a new starting point to face the sustainability challenge.

2. Initial methodological considerations

Since the definition of sustainable second-order change has been addressed, it is important to discuss the implications for our research methodology. However, defining a single best methodology is not yet feasible. The first reason for this is that research on sustainability is scarce (Cohen & Mehta, 2017). Sustainability is often not explicitly addressed in studies, even though there is an increasing demand for more knowledge about it (Van Driel et al., 2012). Second, as the conceptual definitions of sustainability have differed (see also Table 2), the corresponding methodologies have differed as well. Finally, the new framework provided here requires a comprehensive whole systems approach to study sustainability which has not yet been used. For example, studies focused exclusively on the way in which a specific professional development program resulted in sustainable outcomes, on either the individual level or organizational level in change processes but hardly ever on both (Song & Chermack, 2008), and a process view is particularly underdeveloped (Crossan & Apaydin, 2010).

However, it is of crucial importance that we start working towards commonly accepted research methodologies, as this will help us to build on each other's work and establish a rich understanding of the sustainability challenge. Therefore, the aim of this section is to provide initial methodological considerations which researchers should reflect upon when studying sustainable second-order educational change.

Based on Fig. 1, the following section is divided into four parts: arguing about substantial change, studying change over time, studying the process of learning and changing behavior and studying student outcomes. In addition to discussing the individual characteristics, the relationship between the process of learning and changing behavior and student outcomes is discussed as well. As an exhaustive discussion of each of the four characteristics would require their own paper(s), the aim of the following section is to bring some initial reflections to the table that can provide some useful considerations.

2.1. Arguing about substantial change

As explained in the previous section on defining sustainability, the first characteristic of sustained second-order change refers to the idea that change affects the core of educators' everyday practice. Consequently, it seems important that researchers explain why and how the change affects educational practice, for example, in terms of the intentional, structural, curricular, pedagogical and evaluative dimensions of schooling (Eisner, 1992; see Newmann, Smith, Allensworth, & Bryk, 2001, for an example of a study that determines whether change affects the dimensions of schooling). This argumentation could indicate that the intended change is not transformational in nature, for example, because teachers can build on their existing knowledge and the school already has some policies in place. In that case, researchers should ask themselves whether the framework presented here is suitable. For example, a first-order change perspective might work better. The same holds for changes that are heavily influenced by nation-wide or district wide actions and policies. In that case, a third-order change perspective might work better.

One way to argue the second-order nature of the educational change is to use the theoretical underpinnings for a certain educational change and argue how this affects one or more of the dimensions of schooling. Additionally, data about educators' perception of the pervasiveness of the change can be collected through surveys and/or questionnaires. However, Fullan (1993) indicated that teachers often have false clarity about change strategies, which might make their judgement about the pervasiveness of change, at least in the early stages, less reliable.

2.2. Studying change over time

The second characteristic of sustainable second-order change is that it refers to a longitudinal process that starts as early as when educators contemplate whether or not changes need to be made and ends when satisfactory achievement on the other three characteristics is achieved and overt learning efforts are stopped. Though understanding how change unfolds over time is one of the key purposes of sustainability research, such a process view is particularly underdeveloped in previous research (Crossan & Apaydin, 2010). Therefore, it is essential to consider the use of a longitudinal research design, preferably with multiple points of data collection, as this gives the most information about the change process. Moreover, as sustainability is conceptualized in this paper as a process that starts as early as when educators contemplate whether or not changes need to be made, it would be beneficial when researchers are involved at the early stages of the change process, and are already trying to systematically collect as much data as possible about the change process and the student outcomes from that moment on. Moreover, it would be beneficial when they are involved at least until the overt learning efforts are stopped. Examples that signal that such efforts are stopped include

that teachers no longer participate in the change-related professional development program and that the 'change committee' no longer exists. This roughly means being involved for as long as the change effort still holds a 'special project' status (Berman & McLaughlin, 1974). During this time, data could be collected about the learning processes and behavioral change processes that take place, which reflects the third characteristic of sustainable second-order change.

2.3. Studying learning processes and behavioral change processes

2.3.1. The learning processes of teachers

In studying educational change, it is important to pay greater attention to the teachers' actual learning process, because, as argued in the beginning of this paper, teachers' professional development is often seen as a key component in order to implement changes in education (Sleegers & Leithwood, 2010; Van Driel et al., 2012). Insight into their learning process is scarce and has often relied on self-reports of positive impacts (e.g., De Vries & Pieters, 2007; McFadyen & Cannella, 2004; Van den Bossche, Geijselaers, Segers, Woltjer, & Kirschner, 2011; Vescio, Ross, & Adams, 2008). Moreover, studies have focused on specific activities, processes, or programs in isolation from the complex teaching and learning environments in which teachers live (Opfer & Pedder, 2011). As a result, teachers' learning processes remain relatively unknown: a "black box." Because of that, studies of how teachers acquire knowledge or what they actually do within a professional development context is underdeveloped (Gerzon, 2015; Little, 2012b).

Therefore, one of the most important knowledge gaps related to sustainability is insight into both the individual and collective learning processes that are required to successfully implement change. To do so, analyzing micro-processes within specific cases is likely to be helpful (see, for example, Hubers, Poortman, Schildkamp, Pieters, & Handelzalts, 2016). Such analyses might not only focus on what is learned during a specific professional development program, as educators can also learn explicitly and implicitly at the workplace or during other programs (Eraut, 2000; Horn & Little, 2010; Little, 2012a). This might be one of the most important recommendations, as researchers usually determine whether their professional development program and the behaviors associated with this program were replicated in educational practice (see the 'defining' section above).

In addition to gaining insight into the learning processes, social network analyses (see, for example, Shirrell & Spillane, this issue) and instruments such as the Structured Learning Report (Endedijk, Brekelmans, Sleegers, & Vermunt, 2016), which measures teachers' regulation of learning in multiple contexts for various learning experiences, can be used. Again, this is not an exhaustive list of ways to measure the learning process, but a list of initial reflections.

2.3.2. The learning processes of the school organization

Besides paying attention to individual teachers' actual learning processes, it is important to study organizational learning processes, which covers more than the sum of all individual learning processes (Sleegers & Leithwood, 2010). Here too, studying micro-processes is likely to be helpful. This could include detailed analyses of, for example, meetings during which teachers and school leaders collectively develop a shared vision of the change, make decisions and concrete plans, and engage in dialogues to understand each other's frame of reference (Mitchell & Sackney, 1998). Importantly, such analyses should, when and where possible, cover the dynamics and balance between on the one hand, educators' exploration of new knowledge and the experiments they conduct both within and outside of the organization, and on the other hand, educators' exploitation (use and refinement) of existing knowledge (Crossan et al., 1999; Huber, 1991; March, 1991; Supovitz, 2010).

Though various theories on teachers' individual learning processes and organizational learning processes exist, limited attention has been paid to determining how such individual and organizational learning are related to each other (Song & Chermack, 2008). In general, organizational learning is conceptualized as including more than the sum of all individual learning (Sleegers & Leithwood, 2010). However, what exactly the difference is between such organizational and individual processes and how they influence each other remains unknown, though theories exist (Kozlowski & Klein, 2000; Wee & Taylor, 2018). Finding empirical evidence for these theories is an important endeavor for future research.

2.3.3. The behavioral change processes of teachers

In studying the sustainability of a second-order educational change, it is also important to determine whether or not teachers are actually changing their everyday practice. Again, the premise of second-order change is a transformation in the organization, and one needs to ensure that this transformation in fact took place within individual teachers. However, the measures used to determine the development of teachers' practice are often too general and are not aligned with the learning activities teachers used to support their implementation of the change (Hattie, 2009). Moreover, many studies have relied on teachers' self-reported effects of participating in professional development programs (Van Driel et al., 2012). Both of these approaches do not yet provide us with a full understanding of how teachers' practice changed over time.

Therefore, the appropriate measures should be selected with special care, and more objective measures should be used to study the way in which teachers have changed their behavior. For example, Brekelmans, Mainhard, den Brok, and Wubbels (2011) found that students can provide reliable information about their teachers' classroom behavior. Another way to document the way in which teachers changed their behavior could be to use classroom observations (Panayiotou et al., 2014). Since the process of change is of main interest, the behavioral measures should be used repeatedly in order to understand how teachers' practice unfolds over time.

2.3.4. The behavioral change processes of the school organization

In addition to studying how individual teachers did or did not change their behavior, it is important to determine how changes were made on the level of the school. As explained in the section about the third characteristic of sustainability, the concept of organizational routines will be of use here.

Studying the development of organizational routines results in a nuanced image of how practices change or persist over time. No one decision, formal structure or person is responsible for this; they all mutually influence each other (Spillane, 2012). Typically, the development of organizational routines is determined with a case study design (Foss et al., 2012). Preferably, observation techniques are used within such a design, as they help to understand how practice unfolds over time (Spillane, 2012). Other possible information sources include interviews, policy documents, and questionnaires (see, for example, Hubers et al., 2017).

2.4. Studying the student outcomes

The fourth characteristic of sustainable second-order change is that the change process results in improved student outcomes. In the past, measures determining the effectiveness of educational change often reflected changes in the teachers (e.g., whether they showed significant learning and/or changes in their behavior),

whereas measures of student outcomes were rarely included (Hattie, 2009; Van Driel et al., 2012). Fortunately, studies that include changes in teachers as well as in student outcomes are becoming more prevalent (e.g., Basma & Savage, 2018; Desimone, 2009; Slavin, 2008), though the goal of these studies is often to evaluate one or more programs. Thus, whereas such studies are essential in determining the effect of specific program(s), they do not aim to help us understand the change process within an entire school, during which a specific program is one of the many efforts that takes place. Moreover, pre-test post-test designs with randomized control trials are used, and fidelity and/or local adaptations perspective are often taken as a vantage point. Thus, similar to the argumentation in the beginning of this paper, this results in knowledge that, on average, a change was (not) successful, but specific knowledge of the working mechanism (e.g., how and why a certain change resulted in certain outcomes) is missing.

In order to switch our perspective to the school and student outcomes in relation to the change process within that school, it is critical to learn about the specific happenings of an educational change and how it is expected to bring about the required changes in students (Weiss, 1998). As explained in the first section of this article, an educational change could affect, among other things, the curricular, pedagogical and evaluative dimensions of schooling (Eisner, 1992). Therefore, scholars should make specific assumptions about how educators use resources and activities and the subsequent responses from the students to which this should lead (Weiss, 1998). Thinking about this chain of events is helpful to determine *what* needs to be measured, but of course one also needs to determine *how* it will be measured.

According to Fenwick (2001), no single measure is a valid indicator of student outcomes, and multiple methods are always preferred. Moreover, rich assessments of student learning are required, whereas often, student achievement tests serve as the primary outcome (Whitcomb, Borko, & Liston, 2009). Rich assessments not only include standardized achievement data, but also could include, for example, student products (e.g., portfolios, designs, and presentations), observations of students performing a specific task, classroom observations, student reports about their own learning experiences and/or students' self-assessment (Fenwick, 2001). Moreover, Scheerens (2015) indicates that approaches such as analyzing communication and interaction patterns, for example through sociometric methods or simulation studies based on dynamic models, could also guide one's study.

2.5. Studying the relationship between the change process and student outcomes

It was already mentioned that the third and fourth characteristics, learning/changing behaviors and student outcomes respectively, can mutually influence each other, see also Fig. 1. Therefore, the section below provides some initial reflections that address the relationship between those two characteristics.

2.5.1. Causality

First of all, it is important to determine that the positive effects on student outcomes are in fact caused by the process of educational change. However, this is not an easy task (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). One important step in linking the change process with student outcomes is to consider the context in which such changes are made. For example, was there a change in school leadership or policy that might (partly) explain the positive effects that were found? Moreover, forming a chain of evidence in which several critical links are argued for (Cochran-Smith, 2005) is likely to be helpful. These links include demonstrations of how engaging in (learning) activities or professional development programs resulted in educators' learning processes, how these learning processes resulted in changes in their everyday practice, how this provided input to and was influenced by the overall school development and how the changes in educators' everyday practice as well as the schools' changes in policy and practice resulted in changes in student outcomes. To build such a convincing chain, we need, among other things, precise measures of teacher knowledge, reliable and valid observation protocols, and rich assessments of student learning (Whitcomb et al., 2009). A logic model could be used to define what needs to be measured in the chain of evidence (Cooksy, Gill, & Kelly, 2001).

2.5.2. Understanding the why

As argued at the beginning of this paper, providing hard evidence *that* change is effective is not enough. It is also of critical importance to understand *why* it is (or is not) effective and *under what conditions* this is the case. There are several methodological approaches to cover such a perspective, including 'Theories of Change' and 'Realistic Evaluation' (Pawson & Tilley, 1997; see Blamey & Mackenzie, 2007, for a comparison between the two approaches). The core idea of both approaches is that context is part of the evaluation process and can give crucial insights into why successful changes were (or were not) made. Paying such detailed attention to situational factors is important, yet often overlooked (Kennedy, 2010; see the implementation fidelity section in the beginning of this article for examples of such situational factors). Moreover, an understanding of the context is helpful for generalizing to the settings and organizations for which the educational change is likely to work. Examples of studies that include such a framework are Secker, Bowers, Webb, and Llanes (2005) and Ehren, Eddy-Spicer, Bangpan, and Reid (2017). This initial consideration is listed together with all aforementioned considerations in Table 3.

3. Discussion

There is an increasing demand for more knowledge about sustainable educational change (Anderson, 2010; Coburn et al., 2012; Sleegers & Leithwood, 2010). Currently, studies typically do not build on each other's theoretical frameworks and corresponding methodologies. The present study aimed to help the field move forward through providing an alternative framework, specifically in relation to second-order educational change. Whereas first-order change refers to an extension of the past and can be implemented with educators' existing knowledge and skills (Waters et al., 2003), and third-order change refers to a transformational effort that occurs (partially) outside the boundaries of the organization (Tsoukas & Papoulias, 2005), second-order change refers to a transformational change within the boundaries of the organization that calls organizational assumptions into question, requires new knowledge, as well as adaptations in the prevailing values and norms (Lewis & Sahay, 2017; Tsoukas & Papoulias, 2005; Waters et al., 2003). It is important to take this second-order change context into account, because ones' definition of sustainability might differ depending on whether first-, second- or third-order changes are being implemented.

To help the field move forward and to provide an alternative perspective on the sustainability challenge, the present position paper aimed to achieve two goals. The first goal was to compare and contrast and elaborate on previous research in order to provide an alternative comprehensive theoretical framework of sustainable second-order educational change. The second goal was to provide initial methodological considerations that correspond to the new definition of sustainability. The conclusions from both sections are summarized below, see also Table 3.

Table 3
Characteristics of sustainable second-order educational change and the recommended initial methodological considerations.

Characteristic	Recommended initial methodological considerations
1. Substantial changes that are made that affect the core of educators' everyday practice.	- Determine why and how the educational change affects the core of educators' everyday practice (e.g., through using its theoretical underpinnings, collecting data about educators' perception).
2. A longitudinal process that starts as early as when educators contemplate whether or not changes need to be made. It ends when satisfactory achievement on the other three characteristics is reached and overt learning efforts are stopped.	- Gain a process view of how change develops over time. If possible, collect data at multiple time points and try to be involved from the beginning to the end of the change process.
3. A process of individual learning (professional development) and organizational learning as well as of changing behaviors at both the individual and organizational levels.	- Gain insights into learning processes at the individual and organizational level (e.g., through micro-process case studies or social network analyses). If possible, consider how these learning processes relate to each other. - Determine what changes were made at the individual and organizational level (e.g., through students' perceptions, observations, and studying the development of organizational routines).
4. A change process that results in improved student outcomes.	- Determine *what* needs to be measured by determining how educators' engagement in the change influenced students outcomes. - Determine *how* student outcomes should be measured. Using multiple measurement methods is preferred. - Determine how characteristic 3 and 4 can mutually influence each other through determining the causality within the change process (e.g., creating a causal chain of evidence that illustrates that the positive effects on student outcomes are caused by the educational change). - Determine what steps can be taken to understand *why* the change occurred in a certain way and *under what conditions* this is the case.

3.1. Defining sustainability

The first characteristic of sustainable second-order change is that it refers to substantial changes that are made that affect the core of educators' everyday practice. Moreover, sustainability refers to a change process that starts as early as when educators are contemplating a certain change, and ends when satisfactory achievement on all characteristics is reached and overt learning efforts are stopped. After that, all fine-tuning of the change might no longer be of second-order nature.

During the change process, both the individual educators and the organization as a whole learn about the change and change their behavior. This change process should result in a significant positive effect on student outcomes. The definition of sustainability as posed in this paper does not refer to whether or not a professional development program resulted in sustained outcomes. Instead, it refers to whether or not the educators at a certain school and the school organization as a whole changed over time during which professional development programs can be of added value. Thus, the core premise underlying the definition presented here is that the overall development of the whole system and its underlying components need to be studied from a dynamic and multilevel perspective in order to understand sustained second-order change.

3.2. Methodological considerations

When studying sustainability, all of its characteristics should be addressed. Thus, a research methodology should aim to include an argument for the substantial nature of the change (e.g., using Eisner's dimensions of schooling), ways to study the change process in terms of learning and behavior from beginning to end, ways to study student outcomes, and ways to study the relationship between the change process and the change outcomes. Concrete suggestions were provided for each of these. Moreover, the relationship between the change process and the change outcomes needs to be studied, which requires a convincing chain of evidence which demonstrates that the process caused the outcomes. In addition to making such a claim for causality, scholars should gain understanding of why and under what conditions this chain of events took place.

3.3. Directions for future research

Beyond the methodological considerations offered above, it would be beneficial to conduct a systematic review analysis of previous research addressing the sustainability challenge. Such an analysis would result in in-depth insights into the specific challenges of sustaining certain types of educational change (e.g., first-, second- and third-order changes) and ways of overcoming these challenges. Moreover, such an analysis could help scholars identify which research paradigm and corresponding methodology match the specific questions at hand.

Second, research (or at least the majority of research that is being published) often focuses on instances in which educational change was (at least partially) successful. However, understanding intended changes that were unsuccessful is just as important (Hattie, 2009; Scheerens, 2015). It is impossible to come to a true understanding of successful educational change if we do not understand why change efforts are sometimes not successful. Thus, it is critical to publish studies of both success and failure.

Third, it is important to join forces with a wider community of scholars and build on each other's findings. The reason for this is that the challenge of how to sustain changes within an organization is not limited to education, but also evident in, for example, health care (e.g., Wiltsey Stirman et al., 2012), and commercial organizations (e.g., Sackmann et al., 2009). Though the contexts of these professions differ, the sustainability challenge itself and the mechanisms underlying it are likely to show overlap. Using each other's lessons learned and research methodologies can inspire our work and thinking.

3.4. A final note

Though this paper defined sustainability as an effective change, this is not meant as license to mindlessly implement such a change. Sometimes, the costs of change are just not worth the benefits (Hattie, 2009). For example, when teachers implement a new practice that leads to tremendous positive outcomes for all students, but teachers are under too much pressure and their work

satisfaction decreases significantly, one needs to be mindful of the next step. Continuing on in the same way is highly undesirable for the teachers. In that case, one needs to determine what elements of the new practice can be used to retain as much of the positive outcomes as possible while reducing teachers' burden significantly. Thus, scholars should always consider whether the opportunities for learning that follow from the change are educationally desirable (Biesta, 2007).

Taking this all together, there is much more for us to learn about sustaining changes in education. To do so, the present paper provided a theoretical framework of sustainable second-order educational change, and provided related methodological considerations. Both the framework and the considerations can serve as initial building blocks to pave the way for sustainable educational change.

CRediT authorship contribution statement

Mireille D. Hubers: Conceptualization, Investigation, Writing - original draft, Writing - review & editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tate.2020.103083.

References

Adams, D. (2007). Implementing and sustaining educational reforms: The case for little "r". *Prospects, 37*, 385–401. https://doi.org/10.1007/s11125-008-9042-5.

Adelman, H. S., & Taylor, L. (2003). On sustainability of project innovations as systemic change. *Journal of Educational and Psychological Consultation, 14*(1), 1–25. https://doi.org/10.1207/S1532768XJEPC1401_01.

Ananiadou, K., & Claro, M. (2009). In *21st century skills and competences for new millennium learners in OECD countries (OECD Education Working Papers, No. 41)*. Paris: OECD Publishing.

Anderson, S. E. (2010). Moving change: Evolutionary perspectives on educational change. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins (Eds.), *Second international handbook of educational change* (pp. 65–84). Dordrecht: Springer.

Anderson, E. R. (2017). Accommodating change: Relating fidelity of implementation to program fit in educational reforms. *American Educational Research Journal, 54*(6), 1288–1315. https://doi.org/10.3102/0002831217718164.

Baglibel, M., Samancioglu, M., & Crow, G. M. (2018). Factors affecting the sustainability of educational changes: A mixed method research. *Cogent Education, 5*(1), 1–14. https://doi.org/10.1080/2331186X.2018.1502395.

Bartunek, J. M., & Moch, M. (1987). First order, second order, and third order change and organization development interventions: A cognitive approach. *The Journal of Applied Behavioral Science, 23*, 483–500. https://doi.org/10.1177/002188638702300404.

Basma, B., & Savage, R. (2018). Teacher professional development and student literacy growth: A systematic review and meta-analysis. *Educational Psychology Review, 30*, 457–481. https://doi.org/10.1007/s10648-017-9416-4.

Bean, R. M., Dole, J. A., Nelson, K. L., Belcastro, E. G., & Zigmond, N. (2015). The sustainability of a national reading reform initiative in two states. *Reading & Writing Quarterly, 31*(1), 30–55. https://doi.org/10.1080/10573569.2013.857947.

Berman, P., & McLaughlin, M. W. (1974). *Federal programs supporting educational change: A model of educational change* (Vol. I). Santa Monica, CA: Rand.

Biesta, G. (2007). Why "what works" won't work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory, 57*(1), 1–22. https://doi.org/10.1111/j.1741-5446.2006.00241.x.

Blamey, A., & Mackenzie, M. (2007). Theories of change and realistic evaluation: Peas in a pod or apples and oranges? *Evaluation, 13*(4), 439–455. https://doi.org/10.1177/1356389007082129.

Breault, D. A. (2013). The challenges of scaling-up and sustaining professional development school partnerships. *Teaching and Teacher Education, 36*, 92–100. https://doi.org/10.1016/j.tate.2013.07.007.

Brekelmans, M., Mainhard, M. T., den Brok, P. J., & Wubbels, T.h. (2011). Teacher control and affiliation: Do students and teachers agree? *Journal of Classroom Interaction, 46*, 17–26.

Brown, C., Schildkamp, K., & Hubers, M. D. (2017). Combining the best of two worlds: A conceptual proposal for evidence-informed school improvement. *Educational Research, 59*(2), 154–172. https://doi.org/10.1080/00131881.2017.1304327.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*, 1–9. https://doi.org/10.1186/1748-5908-2-40.

Choi, M. (2011). Employees' attitudes toward organizational change: A literature review. *Human Resource Management, 50*(4), 479–500. https://doi.org/10.1002/hrm.20434.

Clarke, J., & Dede, C. (2009). Design for scalability: A case study of the river city curriculum. *Journal of Science Education and Technology, 18*(4), 353–365. https://doi.org/10.1007/s10956-009-9156-4.

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

Coburn, C. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher, 32*(6), 3–12.

Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education, 77*(3), 211–244. https://doi.org/10.1177/003804070407700302.

Coburn, C. E., Russell, J. L., Kaufman, J. H., & Stein, M. K. (2012). Supporting sustainability: Teachers' advice networks and ambitious instructional reform. *American Journal of Education, 119*(1), 137–182. https://doi.org/10.1086/667699.

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education, 112*, 469–495. https://doi.org/10.1086/505056.

Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives, 9*(4), 173–206. https://doi.org/10.1080/15366367.2011.626729.

Cochran-Smith, M. (2005). Studying teacher education: What we know and need to know. *Journal of Teacher Education, 56*(4), 301–306. https://doi.org/10.1177/0022487105280116.

Cohen, D. K., & Mehta, J. D. (2017). Why reform sometimes succeeds: Understanding the conditions that produce reforms that last. *American Educational Research Journal, 54*(4), 1–47. https://doi.org/10.3102/0002831217700078.

Cooksy, L. J., Gill, P., & Kelly, P. A. (2001). The program logic model as an integrative framework for a multimethod evaluation. *Evaluation and Program Planning, 24*(2), 119–128. https://doi.org/10.1016/S0149-7189(01)00003-9.

Crossan, M. M., & Apaydin, M. (2010). A multi-dimensional framework of organizational innovation: A systematic review of the literature. *Journal of Management Studies, 47*(6), 1154–1191. https://doi.org/10.1111/j.1467-6486.2009.00880.x.

Crossan, M. M., Lane, H. W., & White, R. E. (1999). An organizational learning framework: From intuition to institution. *Academy of Management Review, 24*(3), 522–537. https://doi.org/10.5465/AMR.1999.2202135.

Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms*. New York, NY: Teachers College Press.

Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. Malden, MA: Blackwell.

Darling-Hammond, L. (1988). Policy and professionalism. In A. Lieberman (Ed.), *Building a professional culture in schools* (pp. 57–77). New York, NY: Teachers College Press.

Datnow, A. (2005). The sustainability of comprehensive school reform models in changing district and state contexts. *Educational Administration Quarterly, 41*(1), 121–153. https://doi.org/10.1177/0013161X04269578.

Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record, 117*(4), 1–26.

Datnow, A., Hubbard, L., & Mehan, H. (2002). *Extending educational reform: From one school to many*. London: Routledge/Falmer.

De Vries, B., & Pieters, J. M. (2007). The knowledge-creating school proudly presents: A school community sharing its story. In C. Kimble, P. Hildreth, & I. Bourdon (Eds.), *Communities of practice: Creating learning environments for educators* (pp. 355–370). Charlotte, NC: Information Age Publishing.

Dede, C., Rockman, S., & Knox, A. (2007). Lessons learned from studying how innovations can achieve scale. *Threshold, 5*(1), 4–10.

Desimone, L. M. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research, 72*(3), 433–479. https://doi.org/10.3102/00346543072003433.

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.

Desimone, L. M., & Hill, K. L. (2017). Inside the black box: Examining mediators and moderators of a middle school science intervention. *Educational Evaluation and Policy Analysis, 39*(3). https://doi.org/10.3102/0162373717697842, 551-536.

Detrich, R., Keyworth, R., & States, J. (2010). Treatment integrity: A fundamental unit of sustainable educational programs. *Journal of Evidence-Based Practices for Schools, 11*(1), 4–29.

Drits-Esser, D., Gess-Newsome, J., & Stark, L. A. (2017). Examining the sustainability of teacher learning following a year-long science professional development programme for inservice primary school teachers. *Professional Development in Education, 43*(3), 375–396. https://doi.org/10.1080/19415257.2016.1179664.

Ehren, M. C. M., Eddy-Spicer, D., Bangpan, M., & Reid, A. (2017). School inspections in low- and middle-income countries: Explaining impact and mechanisms of impact. *Compare: A Journal of Comparative and International Education, 47*(4), 468–482. https://doi.org/10.1080/03057925.2016.1239188.

Eisner, E. W. (1992). Educational reform and the ecology of schooling. *Teachers College Record, 93*(4), 610–626.

Elmore, R. F. (1996). Getting to scale with good educational practice. *Harvard*

Educational Review, 66(1), 1–26. https://doi.org/10.17763/haer.66.1.g73266758j348t33.
Endedijk, M. D., Brekelmans, M., Sleegers, P., & Vermunt, J. D. (2016). Measuring students' self-regulated learning in professional education: Bridging the gap between event and aptitude measurements. *Quality and Quantity, 50*, 2141–2164. https://doi.org/10.1007/s11135-015-0255-4.
Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology, 70*(1), 113–136. https://doi.org/10.1348/000709900158001.
Feldman, M. S., & Pentland, G. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly, 48*, 94–118.
Fenwick, T. J. (2001). Using student outcomes to evaluate teaching: A cautious exploration. *New Directions for Teaching and Learning, 88*, 63–74. https://doi.org/10.1002/tl.38.
Foss, N. J., Heimeriks, K. H., Winter, S. G., & Zollo, M. (2012). A Hegelian dialogue on the micro-foundations of organizational routines and capabilities. *European Management Review, 9*, 173–197. https://doi.org/10.1111/emre.12001.
Frank, K. A., Zhao, Y., Penuel, W. R., Ellefson, N., & Porter, S. (2011). Focus, fiddle, and friends: Experiences that transform knowledge for the implementation of innovations. *Sociology of Education, 84*(2), 137–156. https://doi.org/10.1177/0038040711401812.
Fullan, M. (1993). *Change forces: Probing the depths of educational reform*. London: Falmer Press.
Fullan, M. (2005). *Leadership and sustainability. System thinkers in action*. Thousand Oaks, CA: Corwin Press.
Fullan, M. (2016). *The new meaning of educational change*. New York, NY: Teachers College Press.
Gerzon, N. (2015). Structuring professional learning to develop a culture of data use: Aligning knowledge from the field and research findings. *Teachers College Record, 117*(4), 1–28.
Greenberg, M. T. (2004). Current and future challenges in school-based prevention: The researcher perspective. *Prevention Science, 5*(1), 5–13. https://doi.org/10.1023/B:PREV.0000013976.84939.55.
Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice, 8*(3/4), 381–391. https://doi.org/10.1080/1354060021000005 12.
Hannan, M., Russell, J. L., Takahashi, S., & Park, S. (2015). Using improvement science to better support beginning teachers: The case of the building a teaching effectiveness network. *Journal of Teacher Education, 66*(5), 494–508. https://doi.org/10.1177/0022487115602126.
Hargreaves, A. (2002). Sustainability of educational change: The role of social geographies. *Journal of Educational Change, 3*, 189–214. https://doi.org/10.1023/A:1021218 71.
Hargreaves, A., & Fink, D. (2000). The three dimensions of reform. *Educational Leadership, 57*(7), 30–34.
Hargreaves, A., Shirley, D., & Wanee, S. (2020). The sustainability and unsustainability of teachers' and leaders' well-being (this issue). *92*, 102987. https://doi.org/10.1016/j.tate.2019.102987.
Harris, A., & Jones, M. (2010). Professional learning communities and system improvement. *Improving Schools, 13*(2), 172–181.
Harris, A., & Jones, M. (2018). Why context matters: A comparative perspective on education reform and policy implementation. *Educational Research for Policy and Practice, 17*(3), 195–207. https://doi.org/10.1007/s10671-018-9231-9.
Harris, A., Jones, M., Sharma, S., & Kannan, S. (2013). Leading educational transformation in Asia: Sustaining the knowledge society. *Asia Pacific Journal of Education, 33*(2), 212–221. https://doi.org/10.1080/02188791.2013.782802.
Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.
Higgins, T. E., & Spitulnik, M. W. (2008). Supporting teachers' use of technology in science instruction through professional development: A literature review. *Journal of Science Education and Technology, 17*, 511–521. https://doi.org/10.1007/s10956-008-9118-2.
Hodges, J., & Gill, R. (2015). *Sustaining change in organizations*. London, UK: Sage publications.
Hodgkinson, T. (2013). The design of a secondary charter school. In D. J. Flinders, P. B. Uhrmacher, & C. M. Moroye (Eds.), *Curriculum and teaching dialogue* (Vol. *15*, pp. 63–82). Charlotte, North Carolina: Information Age Publishing, 1 and 2.
Horn, I. S., & Little, J. W. (2010). Attending to problems of practice: Routines and resources for professional learning in teachers' workplace interactions. *American Educational Research Journal, 47*(1), 181–217. https://doi.org/10.3102/0002831209345158.
Huber, G. P. (1991). Organizational learning: Papers in honor of (and by) James G. March. *Organization Science, 2*(1), 88–115.
Hubers, M. D. (2016). *Capacity building by data team members to sustain schools' data use (Doctoral dissertation)*. The Netherlands: Gildeprint: Enschede.
Hubers, M. D., Endedijk, M. D., & Van Veen, K. (2020). *Effective characteristics of professional development programs for science and technology education. Professional Development in Education*. https://doi.org/10.1080/19415257.2020.1752289.
Hubers, M. D., Moolenaar, N. M., Schildkamp, K., Daly, A. J., Handelzalts, A., & Pieters, J. M. (2017b). Share and succeed: The development of knowledge sharing and brokerage in data teams' network structures. *Research Papers in Education*, 1–23. https://doi.org/10.1080/02671522.2017.1286682.
Hubers, M. D., Poortman, C. L., Schildkamp, K., & Pieters, J. M. (2019). Spreading the word: Boundary crossers building collective capacity for data use. *Teachers College Record, 121*(1), 1–45.
Hubers, M. D., Poortman, C. L., Schildkamp, K., Pieters, J. M., & Handelzalts, A. (2016). Opening the black box: Knowledge creation in data teams. *Journal of Professional Capital and Community, 1*(1), 41–68. https://doi.org/10.1108/JPCC-07-2015-0003.
Hubers, M. D., Schildkamp, K., Poortman, C. L., & Pieters, J. M. (2017). The quest for sustained data use: Developing organizational routines. *Teaching and Teacher Education, 67*, 509–521. https://doi.org/10.1016/j.tate.2017.07.007.
Jimerson, J. B., & Wayman, J. C. (2015). Professional learning for using data: Examining teacher needs and supports. *Teachers College Record, 117*(4), 1–36.
Keller-Margulis, M. (2012). Fidelity of implementation framework: A critical need for response to intervention models. *Psychology in the Schools, 49*(4), 342–352. https://doi.org/10.1002/pits.21602.
Kennedy, M. M. (2010). Attribution error and the quest for teacher quality. *Educational Researcher, 39*(8), 591–598. https://doi.org/10.3102/0013189X10390804.
King-Sears, M. E., Walker, J. D., & Barry, C. (2018). Measuring teachers' intervention fidelity. *Intervention in School and Clinic, 54*(2), 89–96. doi: 10.1 177.177/1053451218765229.
Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein, & S. W. J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). San Francisco: Jossey-Bass.
Kuipers, B. S., Higgs, M., Kickert, W., Tummers, L., Grandia, J., & Van der Voet, J. (2014). The management of change in public organizations: A literature review. *Public Administration, 92*(1), 1–20. https://doi.org/10.1111/padm.12040.
Kurland, H., Peretz, H., & Hertz-Lazarowitz, R. (2010). Leadership style and organizational learning: The mediate effect of school vision. *Journal of Educational Administration, 48*(1), 7–30. https://doi.org/10.1108/09578231011015395.
Labone, E., & Long, J. (2016). Features of effective professional learning: A case study of the implementation of a system-based professional learning model. *Professional Development in Education, 42*(1), 54–77. https://doi.org/10.1080/19415257.2014.948689.
Lee, M., & Seashore Louis, K. (2020). Mapping a strong school culture and linking it to sustainable school improvement (this issue). *81*, 84–96. https://doi.org/10.1016/j.tate.2019.02.001.
Lewin, K. (1947). Frontiers in group dynamics. *Human Relations, 1*, 5–41.
Lewis, L., & Sahay, S. (2017). Change, organizational. In C. R. Scott, & L. Lewis (Eds.), *The international encyclopedia of organizational communication* (pp. 221–243). Wiley Blackwell.
Liou, Y., Canrinus, E., & Daly, A. J. (2020). Activating the implementers: The role of organizational expectations, teacher beliefs, and motivation in bringing about reform (this issue). *79*, 60–72. https://doi.org/10.1016/j.tate.2018.12.004.
Little, J. W. (2012a). Professional community and professional development in the learning- centered school. In M. Kooy, & K. Van Veen (Eds.), *Teacher learning that matters: International perspectives* (pp. 22–43). London: Routledge.
Little, J. W. (2012b). Understanding data use practice among teachers: The contribution of microprocess studies. *American Journal of Education, 118*(2), 143–166. https://doi.org/10.1086/663271.
March, J. G. (1991). Exploration and exploitation in organization learning. *Organization Science, 2*, 71–87.
McFadyen, M. A., & Cannella, A. A. (2004). Social capital and knowledge creation: Diminishing returns of the number and strength of exchange relationships. *Academy of Management Journal, 47*(5), 735–746.
McLaughlin, M. W., & Mitra, D. (2001). Theory-based change and change-based theory: Going deeper, going broader. *Journal of Educational Change, 2*, 301–323. https://doi.org/10.1023/A:1014616908334.
Mitchell, C., & Sackney, L. (1998). Learning about organizational learning. In K. Leithwood, & K. S. Louis (Eds.), *Organizational learning in schools* (pp. 177–199). Lisse, the Netherlands: Swets & Zeitlinger.
Morrissey, M. S. (2000). *Professional learning communities: An ongoing exploration*. Austin, TX: Southwest Educational Development Laboratory.
Mouza, C. (2009). Does research-based professional development make a difference? A longitudinal investigation of teacher learning in technology integration. *Teachers College Record, 111*(5), 1195–1241.
National Research Council. (2011). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.
Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.
Newmann, F. M., Smith, B. A., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis, 23*(4), 297–321. https://doi.org/10.3102/01623737023004297.
Ng, W., & Nicholas, H. (2013). A framework for sustainable mobile learning in schools. *British Journal of Educational Technology, 44*(5), 695–715. https://doi.org/10.1111/j.1467-8535.2012.01359.x.
Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research, 81*(3), 376–407. https://doi.org/10.3102/0034654311413609.
Panayiotou, A., Kyriakides, L., Creemers, B. P. M., McMahon, L., Vanlaar, G., Pfeifer, M., et al. (2014). Teacher behavior and student outcomes: Results of a European study. *Educational Assessment, Evaluation and Accountability, 26*(1), 73–93. https://doi.org/10.1007/s11092-013-9182-x.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
Penuel, W. R., Fishman, B. J., Cheng, B. H., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher, 40*(7), 331–337. https://doi.org/10.3102/0013189X11421826.
Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York, NY: Free Press.
Sackmann, S. A., Eggenhofer-Rehart, P. M., & Friesl, M. (2009). Sustainable change. Long-term efforts toward developing a learning organization. *The Journal of Applied Behavioral Science, 45*(4), 521–549. https://doi.org/10.1177/0021886309346001.
Scheerens, J. (2015). Theories on educational effectiveness and ineffectiveness. *School Effectiveness and School Improvement, 26*(1), 10–31. https://doi.org/10.1080/09243453.2013.858754.
Secker, J., Bowers, H., Webb, D., & Llanes, M. (2005). Theories of change: What works in improving health in mid-life? *Health Education Research, 20*(4), 392–401. https://doi.org/10.1093/her/cyg133.
Sherer, J. Z., & Spillane, J. P. (2011). Constancy and change in work practice in schools: The role of organizational routines. *Teachers College Record, 113*(3), 611–657.
Shirrell, M., & Spillane, J. P. (2020). Opening the door to sustainable improvement: Educational infrastructure, physical infrastructure, and work-related social interactions in schools (this issue). *88*, 102846. https://doi.org/10.1016/j.tate.2019.05.012.
Slavin, R. E. (2008). Perspectives on evidence-based research in education—what works? Issues in synthesizing educational program evaluations. *Educational Researcher, 37*(1), 5–14. https://doi.org/10.3102/0013189X08314117.
Slavin, R. E., & Madden, N. A. (1994). Whenever and wherever we choose: The replication of Success for All. *Phi Delta Kappan, 75*, 639–647.
Sleegers, P., & Leithwood, K. (2010). School development for teacher learning and change. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (Vol. 7, pp. 557–562). Oxford: Elsevier.
Song, J. H., & Chermack, T. J. (2008). A theoretical approach to the organizational knowledge formation process: Integrating the concepts of individual learning and learning organization culture. *Human Resource Development Review, 7*(4), 424–442. https://doi.org/10.1177/1534484308324983.
Spillane, J. P. (2000). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction, 18*(2), 141–179. https://doi.org/10.1207/S1532690XCI1802_01.
Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *Teachers College Record, 118*(2), 113–141.
Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research, 72*, 387–431.
Squire, K. D., MaKinster, J. G., Barnett, M., Luehmann, A. L., & Barab, S. (2003). Designed curriculum and local culture: Acknowledging the primacy of classroom culture. *Science Education, 87*(4), 468–489. https://doi.org/10.1002/sce.10084.
Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change, 7*(4), 221–258.
Stringfield, S., Datnow, A., Ross, A., & Snively, F. (1998). Scaling up school restructuring in multicultural, multilingual contexts: Early observations from Sunland County. *Education and Urban Society, 30*(3), 326–357.
Stringfield, S., Reynolds, D., & Schaffer, E. C. (2008). Improving secondary students' academic achievement through a focus on reform reliability: 4- and 9-year findings from the high reliability schools project. *School Effectiveness and School Improvement, 19*(4), 409–428. https://doi.org/10.1080/09243450802535190.
Supovitz, J. (2010). Knowledge-based organizational learning for instructional improvement. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins (Eds.), *Second international handbook of educational change* (pp. 707–723). Dordrecht: Springer.
Tam, F. W. (2009). Sufficient conditions for sustainable instructional changes in the classroom: The case of Hong Kong. *Journal of Educational Change, 10*, 315–336. https://doi.org/10.1007/s10833-008-9091-9.
Thurlings, M., Evers, A. T., & Vermeulen, M. (2015). Towards a model of explaining teachers' innovative behavior: A literature review. *Review of Educational Research, 85*(3), 430–471. https://doi.org/10.3102/0034654314557949.
Tsoukas, H., & Papoulias, D. B. (2005). Managing third-order change: The case of the public power corporation in Greece. *Long Range Planning, 38*, 79–95. https://doi.org/10.1016/j.lrp.2004.11.015.
Tyack, D., & Cuban, L. (1995). *Tinkering toward Utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.
Van Der Voet, J., Groeneveld, S., & Kuipers, B. S. (2014). Talking the talk or walking the walk? The leadership of planned and emergent change in a public organization. *Journal of Change Management, 14*(2), 171–191. https://doi.org/10.1080/14697017.2013.805160.
Van Driel, J. H., Meirink, J. A., van Veen, K., & Zwart, R. C. (2012). Current trends and missing links in studies on teacher professional development in science education: A review of design features and quality of research. *Studies in Science Education, 48*, 129–160. https://doi.org/10.1080/03057267.2012.738020.
Van Veen, K., Zwart, R., Meirink, J., & Verloop, N. (2010). *Professionele ontwikkeling van leraren. Een reviewstudie naar effectieve kenmerken van professionaliseringsinterventies van leraren* [Teachers' professional development. In *A review of effective characteristics of teachers' professional development interventions*. Leiden: ICLON/Expertisecentrum Leren van Docenten.
Van den Bossche, P., Gijselaers, W., Segers, M., Woltjer, G., & Kirschner, P. (2011). Team learning: Building shared mental models. *Instructional Science, 39*(3), 283–301.
Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education, 24*(1), 80–91.
Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. Retrieved from https://files.eric.ed.gov/fulltext/ED481972.pdf.
Webster-Stratton, C., Reinke, W. M., Herman, K. C., & Newcomer, L. L. (2011). The Incredible Years Teacher Classroom Management Training: The methods and principles that support fidelity of training delivery. *School Psychology Review, 40*(4), 509–529.
Wee, E. X. M., & Taylor, S. M. (2018). Attention to change: A multilevel theory on the process of emergent continuous organizational change. *Journal of Applied Psychology, 103*(1), 1–13. https://doi.org/10.1037/apl0000261.
Weick, K. E., & Quinn, R. E. (1999). Organizational change and development. *Annual Review of Psychology, 50*, 361–386. https://doi.org/10.1146/annurev.psych.50.1.361.
Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies*. Upper Saddle River, NJ: Prentice Hall.
Whitcomb, J., Borko, H., & Liston, D. (2009). Growing talent: Promising professional development models and practices. *Journal of Teacher Education, 60*(3), 207–212. https://doi.org/10.1177/0022487109337280.
Whitehead, D. (2010). The year after: Sustaining the effects of literacy professional development in New Zealand secondary schools. *Language and Education, 24*(2), 133–149. https://doi.org/10.1080/09500780903350273.
Wiltsey Stirman, S., Kimberly, J., Cook, N., Calloway, A., Castro, F., & Charns, M. (2012). The sustainability of new programs and innovations: A review of the empirical literature and recommendations for future research. *Implementation Science, 7*, 1–19. https://doi.org/10.1186/1748-5908-7-17.
Wolthuis, F., Van Veen, K., De Vries, S., & Hubers, M. D. (2020). Between lethal and local adaptation: Lesson study as an organizational routine. *International Journal of Educational Research, 100*. https://doi.org/10.1016/j.ijer.2020.101534.
Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement (Issues & Answers Report, REL 2007–No. 033)*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.