Chapter 2

Why Must Everything Be So Complicated? Demands and Challenges on Methods for Analyzing School Improvement Processes

Tobias Feldhoff and Falk Radisch

2.1 Introduction

In the recent years, awareness has risen by an increasing number of researchers, that we need studies that appropriately model the complexity of school improvement, if we want to reach a better understanding of the relation of different aspects of a school improvement capacity and their effects on teaching and student outcomes, (Feldhoff, Radisch, & Klieme, 2014; Hallinger & Heck, 2011; Sammons, Davis, Day, & Gu, 2014). The complexity of school improvement is determined by many factors (Feldhoff, Radisch, & Bischof, 2016). For example, it can be understood in terms of diverse direct and indirect factors being effective at different levels (e.g., the system, school, classroom, student level), the extent of their reciprocal interdependencies (Fullan, 1985; Hopkins, Ainscow, & West, 1994) and at least the different and widely unknown time periods as well as the various paths school improvement is following in different schools over time to become effective. As a social process, school improvement is also characterized by a lack of standardization and determination (ibid., Weick, 1976). For many aspects that are relevant to school improvement theories, we have only insufficient empirical evidence, especially considering the longitudinal perspective that improvement is going on over time. Valid results depend on plausible theoretical explanations as well as on adequate methodological implementations. Furthermore, many studies could be found to reach contradictory results (e.g. for leadership, see Hallinger & Heck, 1996). In our view, this can at least in part be attributed to the inappropriate consideration of the complexity of school improvement.

T. Feldhoff (✉)

Johannes Gutenberg University, Mainz, Germany

e-mail: feldhoff@uni-mainz.de

F. Radisch

University of Rostock, Rostock, Germany

© The Author(s) 2021

A. Oude Groote Beverborg et al. (eds.), *Concept and Design Developments in School Improvement Research*, Accountability and Educational Improvement, https://doi.org/10.1007/978-3-030-69345-9_2

So far, respective quantitative studies that consider that complexity appropriately have hardly been realized because of the high efforts of current methods and costs involved (Feldhoff et al., 2016). Current elaborate methods, like level-shape, latent difference score (LDS) or multilevel growth models (MGM) (Ferrer & McArdle, 2010; Gottfried, Marcoulides, Gottfried, Oliver, & Guerin, 2007; McArdle, 2009; McArdle & Hamagami, 2001; Raykov & Marcoulides, 2006; Snijders & Bosker, 2003) place high demands on study designs, like large numbers of cases at school-, class- and student-level in combination with more than three well-defined and reasoned measurement points. Not only pragmatic research reasons (benefit-cost-relation, a limit of resources, access to the field) conflict with this challenge. Often, also the field of research cannot fulfil all requirements (for example regarding the needed samples sizes on all levels or the required quantity and intensity of measurement points to observe processes in detail). It is obvious to look for new innovative methods that adequately describe the complexity of school improvement, which at the same time present fewer challenges in the design of the studies. Regarding quantitative research, in the past particularly methods from educational effectiveness research were borrowed. Through this, the complexity of school improvement processes and the resulting demands were not sufficiently taken into account and reflected. Therefore, we need an own methodological and methodical analysis. It is not about inventing new methods but about systematically finding methods in other fields that can adequately handle specific aspects of the overall complexity of school improvement, and that can be combined with other methods that highlight different aspects and, in the end, be able to answer the research questions appropriately. To conduct a meaningful search for new innovative methods, it is first essential to describe the complexity of school improvement and its challenges in detail. This more methodological topic will be discussed in this paper. For that, we present a further development of our framework of the complexity of school improvement (Feldhoff et al., 2016). It helps us to define and to systemize the different aspects of complexity. Based on the framework, research approaches and methods can be systematically evaluated concerning their strong and weak points for specific problems in school improvement. Furthermore, it offers the possibility to search specifically for new approaches and methods as well as to consider even more intensively the combination of different methods regarding their contribution to capturing the complexity of school improvement.

The framework is based upon a systematic long-term review of the school improvement research and various theoretical models that describe the nature of school improvement (see also Fig. 2.1). For that, it might be not settled. As a framework, it shows a wide openness for extending and more differentiating work in the future.

Following this, we will try to draft questions that contribute to classification and critical reflection of new innovative methods, which shall be presented in that book.

Fig. 2.1 Framework of Complexity

2.2 Theoretical Framework of the Complexity of School Improvement Processes

School improvement targets the school as a whole. As an organizational process, school improvement is aimed at influencing the collective school capacity to change (including change for improvement relevant processes, like cooperation, processes, etc.), the skills of its members, and the students’ learning conditions and outcomes (Hopkins, 1996; Maag Merki, 2008; Mulford & Silins, 2009; Murphy, 2013; van Velzen et al., 1985). In order to achieve sustainable school improvement, school practitioners engage in a complex process comprising diverse strategies implemented at the district, school, team and classroom level (Hallinger & Heck, 2011; Mulford & Silins, 2009; Murphy, 2013). School improvement research is interested in both, which processes are involved in which way and what their effects are.

Within our framework the complexity of school improvement as a social process can be described by six characteristics: (a) the longitudinal nature, (b) the indirect nature, (c) the multilevel phenomenon, (d) the reciprocal nature, (e) differential development and nonlinear effects and (f) the variety of meaningful factors (Feldhoff et al., 2016). Explanations of these characteristics are given below:

(a) *The Longitudinal Nature of School Improvement Process*

As Stoll and Fink (1996) pointed out, “Although not all change is improvement, all improvement involves change” (p. 44). Fundamental limitations of the cross-sectional design, therefore, constrain the validity of results when seeking to understand ‘school improvement’ and its related processes. Since school improvement always implies a change in organizational factors (processes and conditions, e.g. behaviours, practices, capacity, attitudes, regulations and outcomes) over time, it is most appropriately studied in a longitudinal perspective.

It is important to distinguish between changes in micro- and macro-processes. The distinction between micro- and macro-processes is the level of abstraction with which researchers conceptualise and measure practices of actors within schools. Micro-processes are the direct interaction between actors and their practices in the daily work. For example, the cooperation activities of four members of a team in one or more consecutive team meetings. Macro-processes can be described as a sum of direct interactions at a higher level of abstraction and, for the most part, over a longer period of time. For example, what content teachers in a team have exchanged in the last 6 months or about the general way of cooperation in a team (e.g. sharing of materials, joint development of teaching concepts, etc.). While changes of micro processes are possible in a relatively short time, changes of macro-processes can often only be detected and measured after a more extended period (see, e.g. Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010; Fullan, 1991; Fullan, Miles, & Taylor, 1980; Smink, 1991; Stoll & Fink, 1996). Stoll and Fink (1996) assume that moderate changes require 3–5 years while more comprehensive changes involve even more extended periods of time (see also Fullan, 1991). The most school improvement studies analyse macro-processes and their effects. But it must also be considered that concrete micro-processes can lead to changes faster due to the dynamical component of interaction and cooperation being more direct and immediate in these processes. Regarding macro-processes, the common way of “aggregation” in micro-processes (usually averaging of quality respectively quantity assessments or their changes) leads to distortions. One phenomenon described adequately in the literature is the one of professional cooperation between teachers. Usually, there are several – parallel – settings of cooperation that can be found in one school. It is highly plausible that already the assessment of the micro-processes in these settings of cooperation turns out to be high-graded different and that this is true in particular for the assessment of changes of micro-processes in these cooperation settings. For example, in individual settings will appear negative changes while in meantime there will be positive changes in others. The usual methods of aggregation to generate characteristics of macro-processes on a higher level are not able to consider these different dynamics – and therefore inevitably lead to distortions.

The rationale for using longitudinal designs in school improvement research is not only grounded in the conceptual argument that change occurs over time (e.g. see Ogawa & Bossert, 1995), but also in the methodological requirements for assigning causal attributions to school policies and practices. Ultimately, school improvement

research is concerned with understanding the nature of relations among different factors that impact on the productive change in desired student outcomes over time (Hallinger & Heck, 2011; Murphy, 2013). The assignment of causal attributions is facilitated by substantial theoretical justification as well as by measurements at different points in time (Finkel, 1995; Hallinger & Heck, 2011; Zimmermann, 1972). “With a longitudinal design, the time ordering of events is often relatively easy to establish, while in cross-sectional designs this is typically impossible” (Gustafsson, 2010, p. 79). Cross-sectional modeling of causal relations might lead to incorrect estimations even if the hypotheses are excellent and reliable. For example, a study investigating the influence of teacher cooperation as macro-processes on student achievement in mathematics demonstrates no effect in cross-sectional analyses, while positive effects emerge in longitudinal modeling (Klieme, 2007).

Recently, Thoonen, Slegers, Oort, and Peetsma (2012) highlighted the lack of longitudinal studies in school improvement research. This lack of longitudinal studies was also observed by Klieme and Steinert (2008) as well as Hallinger and Heck (2011). Feldhoff et al. (2016) have systematically reviewed how common (or rather uncommon) longitudinal studies are in school improvement research. They find only 13 articles that analyzed the relation of school improvement factors and teaching or student outcome longitudinal. Since school improvement research that is based on cross-sectional study designs cannot deliver any reliable information concerning change and its effects on student outcomes, a longitudinal perspective is a central criterion for the power of a study.

Based on the nature of school improvement, the following factors are relevant in longitudinal studies:

Time Points and Period of Development

To investigate a change in school improvement processes and their effects, it is pertinent to consider how often and at which point in time data should be assessed to model the dynamics of the reviewed change appropriately.

The frequency of measurements strongly depends on the different dynamics of change regarding factors. If researchers are interested in the change of micro-processes and their interaction, a higher dynamics is to be expected than those who are interested in changes of macro-processes. A high level of dynamics requires high frequencies (e.g., Reichardt, 2006; Selig et al., 2012). This means that for changes in micro-processes, sometimes daily or weekly measurements with a relatively large number of measurement times (e.g. 10 or 20) are necessary, while for changes of macro-processes, under certain circumstances, significantly less measurement times (e.g. 3–4) suffice, at intervals of several months. Within the limits of research pragmatics, intervals should be accurately determined according to theoretical considerations and previous findings. Furthermore, a critical description and clear justification needs to be given. To identify effects, the period assessed needs to be determined in a way that such effects can be expected from a theoretical point of view (see Stoll & Fink, 1996).

Longitudinal Assessment of Variables

Not only the number of measurement points and the time spans between are relevant for longitudinal studies, but also which of the variables are investigated longitudinally. In many cases, studies often focus solely on a longitudinal investigation of the so-called dependent variable in the form of student outcomes – but concerning conceiving school improvement as change, it is also essential to measure the relevant school improvement factors longitudinally. This is especially important when considering the reciprocal nature of school improvement (see 2.2.4).

Measurement Variance and Invariance

It is highly significant to consider measurement invariance in longitudinal studies (Khoo, West, Wu, & Kwok, 2006, p. 312), because if the meaning of a construct changes, it is empirically not possible to elicit whether change of the construct causes an observed change of measurement scores, change of the reality or an interaction of both (see also Brown, 2006).

For that, the prior testing of the quality of the measuring instruments is more critical and more demanding for longitudinal than for cross-sectional studies. For example, it has to cover the same aspects as well, but in addition with a component that is stable over time. For example, a change of construct-comprehension of the test persons (through learning effects, maturing, etc.) has to be taken into account, and the measuring instruments need to be made robust against these changes for using with common methods. Before the first testing, it is essential to consider which aspects the longitudinal studies should evaluate concerning the improvement processes. Especially more complex school improvement studies present challenges because dynamics can arise and processes gain meaning that cannot be foreseen. That particular dynamic component of the complexity of school improvement can explicitly lead to (maybe intended) changing meanings of constructs by the school improvement processes itself. For example, it is plausible that due to school improvement processes single aspects and items acquiring cooperation between colleagues change concerning their value for participants. Regarding an ideal-typical school improvement process, in the beginning cooperation for a teacher means in particular division of work and exchange of materials and in the end of the process these aspects lost their value and those of joined reflection and preparing lessons as well as trust and a common sense increase. With the help of an according orientation and concrete measures this effect can even be a planned aim of school improvement processes but can also (unwantedly) appear as a side effect of intended dynamical micro-processes. Depending on the involvement and personal interpretation of the gathered experiences, different changes and displacements of attribution of value can be found. – At a moment that will mostly hinder a longitudinal measurement by a lack of measurement invariance across the measurement time points, since most of the methods analysing longitudinal data need a specific measurement invariance.

Many longitudinal studies use instruments and measurement models that were developed for cross-sectional studies (for German studies, this is easily viewable in

the national database of the research data centre (FDZ) Bildung, <https://www.fdz-bildung.de/zugang-erhebungsinstrumente>). Their use is mostly not critically questioned or carefully considered in connection with the specific requirements of longitudinal studies. For psychological research, Khoo, West, Wu and Kwok (2006) recommend more attention to the further consideration of measuring instruments and models. This can be simultaneously transfer to the improvement of measuring instruments for school improvement research.

Measurement invariance touches upon another problem of the longitudinal testing of constructs: The sensitivity of the instruments towards changes that should be observed. The widely used four-level or five-level Likert scales are mostly not sensitive enough towards the different theoretical and empirical expectable developments. They were developed to measure the manifestation or structure of a characteristic on a specific point of time – usually aiming to analyse differences and connections of these manifestation. How and in which dynamic a construct changes over time was not considered in creating Likert scales. For example, cooperation between colleagues, intensity of joined norms and values, the willingness of being innovative are all constructs which are developed out of a cross-sectional perspective in school improvement research. It might be more reasonable to operationalize the construct in a way that can depict various aspects through the course of development, by using the help of different items. Looking at these constructs, for example those of cooperation between colleagues (Gräsel et al., 2006; Steinert et al., 2006) you will often find theoretical deliberations of distinguishing between forms of cooperation and the underlying beliefs. Furthermore, evidences for actual frequency and intensity of cooperation remaining behind their significance are being found again and again not only in the German-speaking field. Concerning school improvement, it is highly plausible that exactly aimed measures can lead to not only increasing amount and intensity of cooperation but also changes in beliefs regarding cooperation which then also lead to a different assessment of cooperation and a displacement of significance of single items and the whole construct itself. It is even assumable that this is the only way of sustainably reaching a substantial increase of intensity and amount of cooperation. A quantitative measure of changes with cross-sectional developed instruments and usual methods is demanding to impossible. We either need instruments, that are stabile in other dimensions to be able of displaying the necessary changes comparably – or methods which are able to portray dynamic construct changes.

(b) *Direct and Indirect Nature of School Improvement*

School improvement can be perceived as a complex process in which changes are initiated at the school level to achieve a positive impact on student learning at the end. It is widely recognized that changes at the school level only become evident after individual teachers have re-contextualized, adapted and implemented them in their classrooms (Hall, 2013; Hopkins, Reynolds, & Gray, 1999; O'Day, 2002). Two aspects of the complexity of school improvement can be deduced from this description, i.e., the direct and indirect nature of school improvement on one hand and the multilevel structure on the other (see 2.2.3).

Depending on the aim, school improvement processes have direct or indirect effects. An example of direct effects is the influence of cooperation on teachers' professionalization. In many respects, school improvement processes involve mediated effects, for instance concerning processes, located in the classroom or even on the team-level that are initiated and/or managed by the school's principal. In school leadership research, Pitner (1988), at an early stage, already stated that the influence of school leadership is indirect and mediated by (1) purposes and goals; (2) structure and social networks; (3) people and (4) organizational culture (Hallinger & Heck, 1998, p. 171). Similar models we can find in school improvement research (Hallinger & Heck, 2011; Slegers et al., 2014). They are based on the assumption that different school improvement factors reciprocally influence each other; some of them directly and others indirectly through different paths (see also: reciprocity). We, moreover, assume that teaching processes are essential mediators of school improvement effects, especially on student outcomes. Ever since school leadership actions have consistently been modeled as mediated effects in school leadership research, a more similar picture of findings has emerged, and a positive impact of school leadership on student outcomes have been found (Hallinger & Heck, 1998; Scheerens, 2012). Also, Hallinger and Heck (1996) and Witziers, Bosker, and Krüger (2003) showed that neglect of mediating factors leads to a lack of validity of the findings, and it remains unclear which effects are being measured. Similar patterns can be expected for the impact of school improvement capacity (see 2.2.6).

(c) School Improvement as a Multilevel Phenomenon

Following Stoll and Fink (1996), we see school improvement as an intentional, planned change process that unfolds at the school level. Its success, however, depends on a change in the actions and attitudes of individual teachers. For example, in the research on professional communities, the actions in teams have a significant impact on those changes (Stoll, Bolam, McMahon, Wallace, & Thomas, 2006). Changes in the actions and attitudes of individual teachers should lead to changes in instruction and the learning conditions of students. These changes should finally have an impact on the students' learning gain. School improvement is a phenomenon that takes place at three or four different known levels within schools (the school level, the team level, the teacher or classroom level, and the student level). It is essential to consider these different levels when investigating school improvement processes (see also Hallinger & Heck, 1998). For school effectiveness research, Scheerens and Bosker (1997, pp. 58) describe various alternative models for cross-level effects, which offer approaches that are also interesting for school improvement research.

Many articles plausibly point out that neither disaggregation at the individual level (that means copying the same number to all members of the aggregate-unit) nor aggregation of information is suitable for taking the hierarchical structure of the data into account appropriately (Heck & Thomas, 2009; Kaplan & Elliott, 1997). School effectiveness research also has widely demonstrated the issues that arise when neglecting single levels. Van den Noortgate, Opdenakker, and Onghena (2005) carried out analyses and simulation studies and concluded that it is essential to not

only take those levels into account where the interesting effects are located. A focus on just those levels might lead to distortions, bearing a negative impact on the validity of the results. Nowadays, multi-level analyses have thus become standard procedure in empirical school (effectiveness) research (Luyten & Sammons, 2010). And it is only a short step postulating that this should become standard in school improvement research too.

Particularly, the combination of micro and macro-processes can only be deduced on methodical ways which adequately display the complex multilevel structure of school (e.g. parallel structures (e.g. classroom vs. team structure), sometimes unclear or instable multilevel structure (e.g. newly initiated or ending team structures every academic year or changings within an academic year), dependent variables on a higher level (e.g. if it is the overall goal to change organisational beliefs), etc.).

(d) *The Reciprocal Nature of School Improvement*

Another aspect, reflecting the complexity of school improvement, evolves from the circumstance that building a school capacity to change and its effects on teaching and student or school improvement outcomes result from reciprocal and interdependent processes. These processes involving different process factors (leadership, professional learning communities, the professionalization of teachers, shared objectives and norms, teaching, student learning) and persons (leadership, teams, teachers, students) (Stoll, 2009). Reciprocity of micro- and macro-processes set differing requirements to the methods (see 2.2.1, longitudinal nature). In micro-processes, there is reciprocity in the way of direct temporal interactions of various persons or factors (within a session/meeting, or days, or weeks). For example, interactions between team members during a meeting enable sense-making and encourage decision-making. In macro-processes, the reciprocity of interactions between various persons or factors is on a more abstract or general level during a longer course of time (maybe several months or years) of improvement processes.

This means, for example, that school leaders not only influence teamwork in professional learning communities over time but also react to changes in teamwork by adapting their leadership actions. Regarding sustainability and the interplay with external reform programs, reciprocity is relevant as a specific form of adaptation to internal and external change. For example, concepts of organizational learning argue that learning is necessary because the continuity and success of organizations depend on their optimal fit to their environment (e.g. March, 1975; Argyris & Schön, 1978). Similar ideas can be found in contingency theory (Mintzberg, 1979) or the context of capacity building for school improvement (Bain, Walker, & Chan, 2011; Stoll, 2009) as well as in school effectiveness research (Creemers & Kyriakides, 2008; Scheerens & Creemers, 1989). School improvement can thus be described as a process of adapting to internal and external conditions (Bain et al., 2011; Stoll, 2009). The success of schools and their improvement capacity is thus a result of this process.

The empirical investigation of reciprocity requires designs that assess all relevant factors of the school improvement process, mediating factors (for example instructional factors) and outcomes (e.g. student outcomes) at several measurement points, in a manner that allows to model effects in more than one direction.

(e) Differential Paths of Development and Nonlinear Trajectories

The fact that the development of an improvement capacity can progress in very different ways adds to the complexity of school reform processes (Hopkins et al., 1994; Stoll & Fink, 1996). Because of their different conditions and cultures, schools differ in their initial levels, the strength and in the progress of their development. The strength and progress of the development itself depends also from the initial level (Hallinger & Heck, 2011). In some schools, development is continuous while in other cases an implementation dip is observable (e.g., Bryk et al., 2010; Fullan, 1991). Theoretically, many developmental trajectories are possible across time, many of which are presumably not linear.

Nonlinearity does not only affect the developmental trajectories of schools. It can be assumed that many relationships between school improvement processes among themselves or in relation to teaching processes and outcomes are also not linear (Creemers & Kyriakides, 2008). Often curvilinear relationships can be expected, in which there is a positive relation between two factors up to a certain point. If this point is exceeded, the relation is near zero or zero, or it can become negative. The first case, the relation becomes zero or near zero, can be interpreted as a kind of a saturation effect. For example, theoretically, it can be assumed that the willingness to innovate in a school, at a certain level, has little or no effect on the level of cooperation in the school. An example of a positive relationship that becomes negative at some level is the correlation between the frequency and intensity of feedback and evaluation on the professionalization of teachers. In the case of a successful implementation, it can be assumed that the frequency and intensity of feedback and evaluation will have a positive effect on the professionalization of teachers. If the frequency and intensity exceed a certain level, it can be assumed that the teachers feel controlled, and the effort involved with the feedback exceeds the benefits and thus has a negative effect on their professionalization. Where the level is set which is critical for each individual school and when it is reached, is dependent on the interaction of different factors on the level of micro- and macro-processes (teachers feeling assured, frustration tolerance, type and acceptance of the style of leadership, etc.). With this example, it also gets clear that there does not only exist no “the more the better” in our concept but also the type and grade of an “ideal level” is dependent on the dynamical and reciprocal interaction with other factors in the duration of time and on the context of the considered actors. Currently, our understanding of the nature of many relationships of school improvement processes among themselves, or in relation to teaching and outcomes is very low (Creemers & Kyriakides, 2008).

To map this complexity, methods are required that enable modelling of nonlinear effects as well as individual development. In empirical studies, it is necessary to examine the course of developments and correlations – whether they are linear,

curvilinear or better describable and explainable via sections or threshold phenomena (e.g. by comparing different adaptation functions in regressive evaluation methods, sequential analysis of extensive longitudinal sections a variety of measurements, etc.). Particularly valuable are procedures that justify several alternative models in advance and test them against each other. Such approaches could improve understanding of changes in school improvement research. But however, these methods (e.g. nonlinear regressive models) have never been used in school improvement research nor in school effective research. The same applies to the study of the variability of processes, developments and contexts. Particularly in recent years, for example, with growth curve models and various methods of multi-level longitudinal analysis, numerous new possibilities have been established in order to carry out such investigations. They also open up the possibility of looking at and examining the variability of processes as dependent variables.

The analysis of different development trajectories of schools and how these correlate e.g. with the initial level and the result of the development is obviously highly relevant for educational accountability and the evaluation of reform projects. In many pilot projects or reform programs, too little consideration is given to the different initial levels of schools and their developmental trajectories. This often leads to Matthew effects. However, reforms in their implementation can only take those factors into account, if the appropriate knowledge about them has been generated in advance in corresponding evaluation studies.

(f) *Variety of Meaningful Factors*

Many different persons and processes are involved in changes in school and their effects on student outcomes (e.g., Fullan, 1991; Hopkins et al., 1994; Stoll, 2009 see 2.2.2). The diversity of factors relates to all parts of the process (e.g. improvement capacity, student outcomes, and teaching, contexts). Because this chapter (and this book) deals with school improvement, we want to confine ourselves exemplarily to two central parts. On one hand, we focus on the variety of factors of improvement processes, because we want to show that in this central part of school improvement a reduction of the variety of factors is not easily achieved. On the other hand, we focus on the variety of outcomes/outputs, since we want to contribute to the still emerging discussion about a stronger merging of school improvement research and school effectiveness research.

Variety of Factors of Improvement Capacity

As outlined above, school improvement processes are social processes that cannot be determined in a clear-cut way. School improvement processes are diverse and interdependent, and they might involve many aspects in different ways. It is essential to consider the variety and reciprocity of meaningful factors of a school's improvement capacity (e.g., teacher cooperation, shared meaning and values, leadership, feedback, etc.) while investigating the relation of different school improvement aspects and their outcomes. A neglect of this diversity can lead to a false estimation of the effects. Only by considering all meaningful factors of the

improvement capacity, it will be possible to take into account interactions between the factors as well as shared, interdependent and/or possibly contradictory effects.

By merely looking at individual aspects, researchers might fail to identify effects that only result from interdependence. Another possible consequence might be a mistaken estimation of factors.

Variety of Outcomes

Given the functions, schools hold for society and the individual, a range of school-related outputs and outcomes can be deduced. The effectiveness of school improvement has been left unattended for a long time. Different authors and sources claim that school effectiveness research and school improvement research should cross-fertilize (Creemers & Kyriakides, 2008). One of the central demands is to make school improvement research more effective in a way that includes all societal and individual spheres of action. Under such a broad perspective that is necessarily connected with school improvement research, it is clear, that focusing on student-related outcomes (what itself means more than cognitive outcomes) is only exemplary (Feldhoff, Bischof, Emmerich & Radisch, 2015; Reezigt & Creemers, 2005). Scheerens and Bosker (1997) distinguish short-term outputs and long-term outcomes (pp. 4). Short-term outputs comprise cognitive as well as motivational-affective, metacognitive and behavioural criteria (Seidel, 2008). The diversity of short-term outputs suggests that the different aspects of the capacity are correlated in different ways to individual output criteria via different paths. Findings on the relation of capacity to one output cannot automatically be transferred to other output aspects or outcomes. If we wish to understand school improvement, we need to consider different aspects of school output in our studies. Seidel (2008) has demonstrated that school effectiveness research at the school level is almost exclusively limited to cognitive subject-related learning outcomes (see also Reynolds, Sammons, De Fraine, Townsend, & Van Damme, 2011). Seidel indicates that the call for consideration of multi-criterial outcomes in school effectiveness research has hardly been addressed (see p. 359). In this regard, so far little if anything is known about the situation in school improvement research.

2.3 Conclusion and Outlook

The framework systematically shows the complexity of school improvement processes in its six characteristics and which methodological aspects need to be considered when developing a research design and choosing methods. Like we drafted in the introduction it is for example due to limited resources and limited access to schools not always possible to consider all aspects similarly. Nevertheless, it is important to reflect and reason: Which aspects can not or only limited be considered, what effects emerge on knowledge acquisition and the results out of this non-consideration or limited consideration of aspects and why a limited or non-consideration is despite limits in terms of knowledge acquisition still reasonable.

In this sense, unreflect or inadequate simplification and thus inappropriate modelling might lead to empirical results and theories that do not face reality or that are leading to contradictory findings. In sum it will lead to a stagnation in the further development of theoretical models. A reliable and further development would require the recognition and the exclusion of inappropriate consideration of the complexity as a cause of contradictory findings. Our methods and designs influence our perspectives as they are the tools by which we generate knowledge, which in turn is the basis for constructing, testing and enhancing our theoretical models (Feldhoff et al., 2016).

Therefore, it is time to search for new methods that make it possible to consider the aspects of complexity, and that has not been made use of in the research of school improvement so far. Many quantitative and qualitative methods have emerged over the last decades within various disciplines of social sciences that need to be reflected for their usefulness and practicality for the school improvement research. To ease the systematic search for adequate and useful methods, we formulated questions based on the theoretical framework, that helps to review the methods' usefulness overall critically and for every single aspect of the complexity. They can also be used as guiding questions for the following chapters.

2.3.1 Guiding Questions

Longitudinal

- Can the method handle longitudinal data?
- Is the method more suitable for shorter or longer intervals (periods between measurement points)?
- How many measurement points are needed and how many are possible to handle?
- Is it affordable to have similar measurement points (comparable periods between the single measurement points and same measurement points for all the individuals and schools)?
- Is the method able to measure all variables of interest in a longitudinal way?
- Is the method able to differentiate the reasons for (in-)variant measurement scores over time, or does the method handle the possible reasons for the (in-) variation of measurements in any other useful way?

Indirect Nature of School Improvement

- Is the method able to evaluate different ways of modeling indirect paths/effects (e.g., mediation, moderation in one or more steps)?
- Is the method able to measure different paths (direct and indirect) between different variables at the same time?

Multilevel

- Is the method able to handle all the different needed levels of school improvement?
- Is the method able to consider effects at levels that are not of interest?

- Is the method able to consider multilevel effects from a lower level of the hierarchy to a higher level?
- Is the method able to handle more complex structures of the sample (e.g., single or maybe multiple-cross and/or multi-group-classified data)?

Reciprocity

- Is the method able to model reciprocal or other non-single-directed paths?
- Is the method able to model circular paths with unclear differentiating between dependent and independent variables?
- Is the method able to analyze effects on both side of a reciprocal relation at the same time?

Differential Paths and Nonlinear Effects

- Is the method able to handle different paths over time and units?
- What kind of effects can the method handle at the same time (linear, non-linear, different positioning over the time points at different units)?

Variety of Factors

- Is the method able to handle a variety of independent factors with different meanings on different levels?
- Is the method able to handle different dependent factors and does not only focus on cognitive or on another measurable factor at the student level?

In addition to these questions on the individual aspects of complexity, it is also essential to consider to what extent the methods are also suitable for capturing several aspects. Alternatively, with which other methods the method can be combined to take different aspects into account.

Overall Questions

Strengths, Weaknesses, and Innovative Potential

- In which aspects of the complexity of school improvement are the strengths and weaknesses of the method (general and comparable to established methods)?
- Does the method offer the potential to map one or more aspects of the complexity in a way that was previously impossible with any of the “established” methods?
- Is the method more suitable for generating or further developing theories, or rather for checking existing ones?
- What demands does the method make on the theories?
- With which other methods can the respective method be well combined?

Requirements/Cost-Benefit-Ratio

- Which requirements (e.g., numbers of cases at school-, class- and student-level, amount and rigorous timing of measurement points, data collection) put the methods to the design of a study?
- Are the requirements realistic to implement such a design (e.g., concerning finding, enough schools/realizing data collection, get funding)?
- What is the cost-benefit-ratio compared to established methods?

References

- Argyris, C., & Schön, D. (1978). *Organizational learning – A theory of action perspective*. Reading, MA: Addison-Wesley.
- Bain, A., Walker, A., & Chan, A. (2011). Self-organisation and capacity building: Sustaining the change. *Journal of Educational Administration*, 49(6), 701–719.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research: Methodology in the social sciences*. New York, NY: Guilford Press.
- Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.
- Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London, UK/New York, NY: Routledge.
- Feldhoff, T., Bischof, L. M., Emmerich, M., & Radisch, F. (2015). Was nicht passt, wird passend gemacht! Zur Verbindung von Schuleffektivität und Schulentwicklung. In H. J. Abs, T. Brüsemeister, M. Schemmann, & J. Wissinger (Eds.), *Governance im Bildungswesen – Analysen zur Mehrebenenperspektive, Steuerung und Koordination* (pp. 65–87). Wiesbaden, Germany: Springer.
- Feldhoff, T., Radisch, F., & Bischof, L. M. (2016). Designs and methods in school improvement research. A systematic review. *Journal of Educational Administration*, 2(54), 209–240.
- Feldhoff, T., Radisch, F., & Klieme, E. (2014). Methods in longitudinal school improvement research: State of the art. *Journal of Educational Administration*, 52(5), 565–736.
- Ferrer, E., & McArdle, J. J. (2010). Longitudinal Modeling of developmental changes in psychological research. *Current Directions in Psychological Science*, 19(3), 149–154.
- Finkel, S. E. (1995). *Causal analysis with panel data*. Thousand Oaks, CA: Sage.
- Fullan, M. G. (1985). Change processes and strategies at the local level. *The Elementary School Journal*, 85(3), 391–421.
- Fullan, M. G. (1991). *The new meaning of educational change*. New York, NY: Teachers College Press.
- Fullan, M. G., Miles, M. B., & Taylor, G. (1980). Organization development in schools: The state of the art. *Review of Educational Research*, 50(1), 121–183.
- Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., Oliver, P. H., & Guerin, D. W. (2007). Multivariate latent change modeling of developmental decline in academic intrinsic math. *International Journal of Behavioral Development*, 31(4), 317–327.
- Gräsel, C., Fussangel, K., & Parchmann, I. (2006). Lerngemeinschaften in der Lehrerfortbildung. *Zeitschrift für Erziehungswissenschaft*, 9(4), 545–561.
- Gustafsson, J. E. (2010). Longitudinal designs. In B. P. M. Creemers, L. Kyriakides, & P. Sammons (Eds.), *Methodological advances in educational effectiveness research* (pp. 77–101). Abingdon, UK/New York, NY: Routledge.
- Hall, G. E. (2013). Evaluating change processes: Assessing extent of implementation (constructs, methods and implications). *Journal of Educational Administration*, 51(3), 264–289.
- Hallinger, P., & Heck, R. H. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980–1995. *Educational Administration Quarterly*, 32(1), 5–44.
- Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement*, 9(2), 157–191.
- Hallinger, P., & Heck, R. H. (2011). Conceptual and methodological issues in studying school leadership effects as a reciprocal process. *School Effectiveness and School Improvement*, 22(2), 149–173.
- Heck, R. H., & Thomas, S. L. (2009). *An introduction to multilevel modeling techniques (Quantitative methodology series)* (2nd ed.). New York: Routledge.
- Hopkins, D. (1996). Towards a theory for school improvement. In J. Gray, D. Reynolds, C. T. Fitz-Gibbon, & D. Jesson (Eds.), *Merging traditions. The future of research on school effectiveness and school improvement* (pp. 30–50). London, UK: Cassell.

- Hopkins, D., Ainscow, M., & West, M. (1994). *School improvement in an era of change*. London, UK: Cassell.
- Hopkins, D., Reynolds, D., & Gray, J. (1999). Moving on and moving up: Confronting the complexities of school improvement in the improving schools project. *Educational Research and Evaluation*, 5(1), 22–40.
- Kaplan, D., & Elliott, P. R. (1997). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling: A Multidisciplinary Journal*, 4(1), 1–24.
- Khoo, S.-T., West, S. G., Wu, W., & Kwok, O.-M. (2006). Longitudinal methods. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement* (pp. 301–317). Washington, DC: American Psychological Association.
- Klieme, E. (2007). *Von der "Output-" zur "Prozessorientierung": Welche Daten brauchen Schulen zur professionellen Entwicklung ?* Kassel: GEPP-Jahrestagung.
- Klieme, E., & Steinert, B. (2008). Schulentwicklung im Längsschnitt. Ein Forschungsprogramm und erste explorative analysen. In M. Prenzel & J. Baumert (Eds.), *Vertiefende Analysen zu PISA 2006* (pp. 221–238). Wiesbaden, Germany: VS Verlag.
- Luyten, H., & Sammons, P. (2010). Multilevel modelling. In B. P. M. Creemers, L. Kyriakides, & P. Sammons (Eds.), *Methodological advances in educational effectiveness research* (pp. 246–276). Abingdon, UK/New York, NY: Routledge.
- Maag Merki, K. (2008). Die Architektur einer Theorie der Schulentwicklung – Voraussetzungen und Strukturen. *Journal für Schulentwicklung*, 12(2), 22–30.
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60, 577–605.
- McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 139–175). Washington, DC: American Psychological Association.
- Mintzberg, H. (1979). *The structuring of organizations: A synthesis of the research*. Englewood Cliffs, NJ: Prentice-Hall.
- Mulford, B., & Silins, H. (2009). Revised models and conceptualization of successful school principalship in Tasmania. In B. Mulford & B. Edmunds (Eds.), *Successful school principalship in Tasmania* (pp. 157–183). Launceston, Australia: Faculty of Education, University of Tasmania.
- Murphy, J. (2013). The architecture of school improvement. *Journal of Educational Administration*, 51(3), 252–263.
- O'Day, J. A. (2002). Complexity, accountability, and school improvement. *Harvard Educational Review*, 72(3), 293–329.
- Ogawa, R. T., & Bossert, S. T. (1995). Leadership as an organizational quality. *Educational Administration Quarterly*, 31(2), 224–243.
- Pitner, N. J. (1988). The study of administrator effects and effectiveness. In N. J. Boyan (Ed.), *Handbook of research on educational* (pp. S. 99–S.122).
- Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Reezigt, G. J., & Creemers, B. P. M. (2005). A comprehensive framework for effective school improvement. *School Effectiveness and School Improvement*, 16(4), 407–424.
- Reichardt, C. S. (2006). The principle of parallelism in the design of studies to estimate treatment effects. *Psychological Methods*, 11(1), 1–18.
- Reynolds, D., Sammons, P., De Fraine, B., Townsend, T., & Van Damme, J. (2011). *Educational effectiveness research. State of the art review*. Paper presented at the annual meeting of the international congress for school effectiveness and improvement, Cyprus, January.
- Sammons, P., Davis, S., Day, C., & Gu, Q. (2014). Using mixed methods to investigate school improvement and the role of leadership. *Journal of Educational Administration*, 52(5), 565–589.
- Scheerens, J. (Ed.). (2012). *School leadership effects revisited. Review and meta-analysis of empirical studies*. Dordrecht, Netherlands: Springer.

- Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford, UK: Pergamon.
- Scheerens, J., & Creemers, B. P. M. (1989). Conceptualizing school effectiveness. *International Journal of Educational Research*, 13(7), 691–706.
- Seidel, T. (2008). Stichwort: Schuleffektivitätskriterien in der internationalen empirischen Forschung. *Zeitschrift für Erziehungswissenschaft*, 11(3), 348–367.
- Selig, J. P., Preacher, K. J., & Little, T. D. (2012). Modeling time-dependent association in longitudinal data: A lag as moderator approach. *Multivariate Behavioral Research*, 47(5), 697–716. <https://doi.org/10.1080/00273171.2012.715557>
- Slegers, P. J., Thoonen, E., Oort, F. J., & Peetsma, T. D. (2014). Changing classroom practices: The role of school-wide capacity for sustainable improvement. *Journal of Educational Administration*, 52(5), 617–652. <https://doi.org/10.1108/JEA-11-2013-0126>
- Smink, G. (1991). The Cardiff congress, ICSEI 1991. *Network News International*, 1(3), 2–6.
- Snijders, T., & Bosker, R. J. (2003). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Los Angeles, CA: Sage.
- Steinert, B., Klieme, E., Maag Merki, K., Döbrich, P., Halbheer, U., & Kunz, A. (2006). Lehrerkoooperation in der Schule: Konzeption, Erfassung, Ergebnisse. *Zeitschrift für Pädagogik*, 52(2), 185–204.
- Stoll, L. (2009). Capacity building for school improvement or creating capacity for learning? A changing landscape. *Journal of Educational Change*, 10(2–3), 115–127.
- 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.
- Stoll, L., & Fink, D. (1996). *Changing our schools. Linking school effectiveness and school improvement*. Buckingham, UK: Open University Press.
- Thoonen, E. E., Slegers, P. J., Oort, F. J., & Peetsma, T. T. (2012). Building school-wide capacity-for improvement: The role of leadership, school organizational conditions, and teacher factors. *School Effectiveness and School Improvement*, 23(4), 441–460.
- Van den Noortgate, W., Opdenakker, M.-C., & Onghena, P. (2005). The effects of ignoring a level in multilevel analysis. *School Effectiveness and School Improvement*, 16(3), 281–303.
- van Velzen, W. G., Miles, M. B., Ekholm, M., Hameyer, U., & Robin, D. (1985). *Making school improvement work: A conceptual guide to practice*. Leuven and Amersfoort: Acco.
- Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21(1), 1–19.
- Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, 39(3), 398–425.
- Zimmermann, E. (1972). *Das Experiment in den Sozialwissenschaften*. Stuttgart, Germany: Poeschel.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.