Chapter 1

Introduction

**Arnoud Oude Groote Beverborg, Tobias Feldhoff, Katharina Maag Merki,
and Falk Radisch**

Schools are continuously confronted with various forms of change, including changes in students' demographics, large-scale educational reforms, and accountability policies aimed at improving the quality of education. On the part of the schools, this requires sustained adaptation to, and co-development with, such changes to maintain or improve educational quality. As schools are multilevel, complex, and dynamic organizations, many conditions, factors, actors, and practices, as well as the (loosely coupled) interplay between them, can be involved therein (e.g. professional learning communities, accountability systems, leadership, instruction, stakeholders, etc.). School improvement can thus be understood through theories that are based on knowledge of systematic mechanisms that lead to effective schooling in combination with knowledge of context and path dependencies in individual school improvement journeys. Moreover, because theory-building, measuring, and analysing co-develop, fully understanding the school improvement process requires basic knowledge of the latest methodological and analytical developments and corresponding conceptualizations, as well as a continuous discourse on the link between theory and methodology. The complexity places high demands on the designs and methodologies from those who are tasked with empirically assessing and fostering improvements (e.g. educational researchers, quality care departments, and educational inspectorates).

A. Oude Groote Beverborg (✉)
Radboud University Nijmegen, Nijmegen, The Netherlands
e-mail: a.oudegrootebeverborg@fm.ru.nl

T. Feldhoff
Johannes Gutenberg University, Mainz, Germany

K. Maag Merki
University of Zurich, Zurich, Switzerland

F. Radisch
University of Rostock, Rostock, Germany

Traditionally, school improvement processes have been assessed with case studies. Case studies have the benefit that they only have to handle complexity within one case at a time. Complexity can then be assessed in a situated, flexible, and relatively easy way. Findings from case studies can also readily inform practice in those schools the studies were conducted in. However, case studies typically describe one specific example and do not test the mechanisms of the process, and therefore their findings cannot be generalized. As generalizability is highly valued, demands for designs and methodologies that can yield generalizable findings have been increasing within the fields of school improvement and accountability research. In contrast to case studies, quantitative studies are typically geared towards testing mechanisms and generalization. As such, quantitative studies are increasingly being conducted. Nevertheless, measurement and analysis of all aspects involved in improvement processes within and over schools and over time would be unfeasible in terms of the amount of measurement measures, the magnitude of the sample size, and the burden on the part of the participants. Thus, by assessing school improvement processes quantitatively, some complexity is necessarily lost, and therefore the findings of quantitative studies are also restricted.

Concurrent with the development towards a broader range of designs, the knowledge base has also expanded, and more sophisticated questions concerning the mechanisms of school improvement are being asked. This differentiation has led to a need for a discourse on how which available designs and methodologies can be aligned with which research questions that are asked in school improvement and accountability research. In our point of view the potential of combining the depth of case studies with the breadth of quantitative measurements and analyses in mixed-methods designs seems very promising; equally promising seems the adaptation of methodologies from related disciplines (e.g. sociology, psychology). Furthermore, application of sophisticated methodologies and designs that are sensitive to differences between contexts and change over time are needed to adequately address school improvement as a situated process.

With the book, we seek to host discussion of challenges in school improvement research and of methodologies that have the potential to foster school improvement research. Consequently, the focus of the book lies on innovative methodologies. As theory and methodology have a reciprocal relationship, innovative conceptualizations of school improvement that can foster innovative school improvement research will also be part of the book. The methodological and conceptual developments are presented as specific research examples on different areas of school improvement. In this way, the ideas, the chances, and the challenges can be understood in the context of the whole of each study, which, we think, will make it easier to apply these innovations and to avoid their pitfalls.

1.1 Overview of the Chapters

The chapters in this book give examples of the use of Measurement Invariance (in Structural Equation Models) to assess contextual differences (Chaps. 4 and 5), the Group Actor Partnership Interdependence Model and Social Network Analysis to

assess group composition effects (Chaps. 6 and 7, respectively), Rhetorical Analysis to assess persuasion (Chap. 8), logs as a measurement instrument that is sensitive to differences between contexts and change over time (Chaps. 9, 10, 11 and 12), Mixed Methods to show how different measurements and analyses can complement each other (Chap. 10), and Categorical Recurrence Quantification Analysis of the analysis of temporal (rather than spatial or causal) structures (Chap. 11). These innovative methodologies are applied to assess the following themes: complexity (Chaps. 2 and 7), context (Chaps. 3, 4, 5 and 6), leadership (Chaps. 7, 8 and 9), and learning and learning communities (Chaps. 4 and 10, 11 and 12).

In Chap. 2, Feldhoff and Radisch present a conceptualization of complexity in school improvement research. This conceptualization aims to foster understanding and identification of strengths, and possible weaknesses, of methodologies and designs. The conceptualization applies to both existing methodologies and designs as well as developments therein, such as those described in the studies in this book. More specifically, the chapter can be used by those who are tasked with empirically assessing and fostering improvements (e.g. educational researchers, departments of education, and educational inspectorates) to chart the demands and challenges that come with certain methodologies and designs, and to consider the focus and omissions of certain methodologies and designs when trying to answer research questions pertaining to specific aspects of the complexity of school improvement. This chapter is used in the last chapter to order the discussion of the other chapters.

In Chap. 3, Reynolds and Neeleman elaborate on the complexity of school improvement by discussing contextual aspects that need to be more extensively considered in research. They argue that there is a gap between research objects from educational effectiveness research on the one hand, and their incorporation into educational practice on the other hand. Central to their explanation of this gap is the neglect to account for the many contextual differences that can exist between and within schools (ranging from school leaders' values to student population characteristics), which resulted from a focus on 'what universally works'. The authors suggest that school improvement (research) would benefit from developments towards more differentiation between contexts.

In Chap. 4, Lomos presents a thorough example of how differences between contexts can be assessed. The study is concerned with differences between countries in how teacher professional community and participative decision-making are correlated. The cross-sectional questionnaire data from more than 35,000 teachers in 22 European countries in this study come from the International Civic and Citizenship Education Study (ICCS) 2009. The originality of the study lies in the assessment of how comparable the constructs are and how this affects the correlations between them. This is done by comparing the correlations between constructs based upon Exploratory Factor Analysis (EFA) with those based upon Multiple-Group Confirmatory Factor Analysis (MGCFA). In comparison to EFA, MGCFA includes the testing of measurement invariance of the latent variables between countries. Measurement invariance is seldom made the subject of discussion, but it is an important prerequisite in group (or time-point) comparisons, as it corrects for bias due to differences in understanding of constructs in different groups (or at different

time-points), and its absence may indicate that constructs have different meanings in different contexts (or that their meaning changes over time). The findings of the study show measurement invariance between all countries and higher correlations when constructs were corrected to have that measurement invariance.

In Chap. 5, Sauerwein and Theis use measurement invariance in the assessment of differences in the effects of disciplinary climate on reading scores between countries. This study is original in two ways. First, the authors show the false conclusions that the absence of measurement invariance may lead to, but second, they also show how measurement invariance, as a result in and of itself, may be explained by another variable that has measurement invariance (here: class size). The cross-sectional data from more than 20,000 students in 4 countries in this study come from the Programme for International Student Assessment (PISA) study 2009. Analysis of Variance (ANOVA) was used to assess the magnitude of the differences between countries in disciplinary climate and Regression Analysis was used to assess the effect of disciplinary climate on reading scores and of class size on disciplinary climate. As in Chap. 4, this was done twice: first without assessment of measurement invariance and then including assessment of measurement invariance. The findings of the study show that some comparisons of the magnitude of the differences in disciplinary climate and effect size between countries were invalid, due to the absence of measurement invariance there. Moreover, the authors assessed how patterns in how class size affected disciplinary climate resembled the patterns of the differences in measurement invariance in disciplinary climate between countries. They found that the effect of class size on disciplinary climate varied in accord with the differences in measurement invariance between countries. This procedure could uncover explanations of why the meaning of constructs differs between contexts (or time-points).

In contrast to the previous two chapters that focussed on between-group comparisons, in Chap. 6, Schudel and Maag Merki focus on within-group composition. They use the concept of diversity and assess the effect of staff members' positions within their teams on job-satisfaction additional to the effects of teacher self-efficacy and collective-self-efficacy. They do so by applying the Group Actor-Partner Interdependence Model (GAPIM) to cross-sectional questionnaire data from more than 1500 teachers in 37 schools. The GAPIM is an extended form of multilevel analysis. Application of the GAPIM is innovative, because it takes differences in team compositions and the position of individuals within a team into consideration, whereas standard multilevel analysis only takes averaged measures over individuals within teams into consideration. This allows more differentiated analysis of multilevel structures in school improvement research. The findings of this study show that the similarity of an individual teacher to the other teachers in the team, as well as the similarity amongst the other teachers themselves in the team, affects individual teachers' job satisfaction, additional to the effects of self and collective-efficacy.

In Chap. 7, Ng approaches within-group composition from another angle. He conceptualizes schools as social systems and argues that the application of Social Network Analysis is beneficial to understand more about the complexity of

educational leadership. In fact, the author shows that complexity methodologies are neither applied in educational leadership studies, nor are they taught in educational leadership courses. As such, the neglect of complexity methodologies, and therefore also the neglect of innovative insights from the complex and dynamic systems perspective, is reproduced by those who are tasked with, and taught, to empirically assess and foster school improvement. Moreover, the author highlights the mismatch between the assumptions that underlie commonly used inferential statistics and the complexity and dynamics of processes in schools (such as the formation of social ties or adaptation), and describes the resulting problems. Consequently, the author argues for the adoption of complexity methodologies (and dynamic systems tools) and gives an example of the application of Social Network Analysis.

In Chap. 8, Lowenhaupt assesses educational leadership by focusing on the use of language to implement reforms in schools. Applying Rhetorical Analysis (a special case of Discourse Analysis) to data from 14 observations from one case, she undertakes an in-depth investigation of the language of leadership in the implementation of reform. She gives examples of how a school leader's talk could connect more to different audiences' rational, ethical, or affective sides to be more persuasive. The chapter's linguistic turn uncovers aspects of the complexity of school improvement that require more investigation. Moreover, the chapter addresses the importance of sensitivity to one's audience and attuned use of language to foster school improvement.

In Chap. 9, Spillane and Zuberi present yet another methodological innovation to assess educational leadership with: logs. Logs are measurement instruments that can tap into practitioners' activities in a context (and time-point) sensitive manner and can thus be used to understand more about the systematics of (the evolution of) situated micro-processes, such as in this case daily instructional and distributed leadership activities. The specific aim of the chapter is the validation of the Leadership Daily Practice (LDP) log that the authors developed. The LDP log was administered to 34, formal and informal, school leaders for 2 consecutive weeks, in which they were asked to fill in a log-entry every hour. In addition, more than 20 of the participants were observed and interviewed twice. The qualitative data from these three sources were coded and compared. Results from Interrater Reliability Analysis and Frequency Analyses (that were supported by descriptions of exemplary occurrences) suggest that the LDP log validly captures school leaders' daily activities, but also that an extension of the measurement period to encompass an entire school year would be crucial to capture time-point specific variation.

In Chap. 10, Vanblaere and Devos present the use of logs to gain an in-depth understanding of collaboration in teachers' Professional Learning Communities (PLC). Using an explanatory sequential mixed methods design, the authors first administered questionnaires to measure collective responsibility, deprivatized practice, and reflective dialogue and applied Hierarchical Cluster Analysis to the cross-sectional quantitative data from more than 700 teachers in 48 schools to determine the developmental stages of the teachers' PLCs. Based upon the results thereof, 2 low PLC and 2 high PLC cases were selected. Then, logs were administered to the 29 teachers within these cases at four time-points with even intervals over the course

of 1 year. The resulting qualitative data were coded to reflect the type, content, stakeholders, and duration of collaboration. Then, the codes were used in Within and Cross-Case Analyses to assess how the communities of teachers differed in how their learning progressed over time. This study's procedure is a rare example of how the breadth of quantitative research and the depth of qualitative research can thoroughly complement each other to give rich answers to research questions. The findings show that learning outcomes are more diverse in PLCs with higher developmental stages.

In Chap. 11, Oude Groote Beverborg, Wijnants, Slegers, and Feldhoff, use logs to explore routines in teachers' daily reflective learning. This required a conceptualization of reflection as a situated and dynamic process. Moreover, the authors argue that logs do not only function as measurement instruments but also as interventions on reflective processes, and as such might be applied to organize reflective learning in the workplace. A daily and a monthly reflection log were administered to 17 teachers for 5 consecutive months. The monthly log was designed to make new insights explicit, and based on the response rates thereof, an overall insight intensity measure was calculated. This measure was used to assess to whom reflection through logs fitted better and to whom logs fitted worse. The daily log was designed to make encountered environmental information explicit, and the response rates thereof generated dense time-series, which were used in Recurrence Quantification Analysis (RQA). RQA is an analysis techniques with which patterns in temporal variability of dynamic systems can be assessed, such as in this case the stability of the intervals with which each teacher makes information explicit. The innovation of the analysis lies in that it captures how processes of individuals unfold over time and how that may differ between individuals. The findings indicated that reflection through logs fitted about half of the participants, and also that only some participants seemed to benefit from a determined routine in daily reflection.

In Chap. 12, Maag Merki, Grob, Rechsteiner, Wullschleger, Schori, and Rickenbacher applied logs to assess teachers' regulation activities in school improvement processes. First, they developed a theoretical framework based on theories of organizational learning, learning communities, and self-regulated learning. To understand the workings of daily regulation activities, the focus was on how these activities differ between teachers' roles and schools, how they relate to daily perceptions of their benefits and daily satisfaction, and how these relations differ between schools. Second, data about teachers' performance-related, day-to-day activities were gathered using logs as time sampling instruments, a research method that has so far been rarely implemented in school improvement research. The logs were administered 3 times for 7 consecutive days with a 7-day pause between those measurements to 81 teachers. The data were analyzed with Chi-square Tests and Pearson Correlations, as well as with Binary Logistic, Linear, and Random Slope Multilevel Analysis. This study provides a thorough example of how conceptual development, the adoption of a novel measurement instrument, and the application of existing, but elaborate, analyses can be made to interconnect. The results revealed that differences in engagement in regulation activities related to teachers' specific roles, that perceived benefits of regulation activities differed a little between schools, and that those perceived benefits and perceived satisfaction were related.

In Chap. 13, Chaps. 3 through 12 will be discussed in the light of the conceptualization of complexity as presented in Chap. 2. We hope that this book is contributing to the (much) needed specific methodological discourse within school improvement research. We also hope that it will help those who are tasked with empirically assessing and fostering improvements in designing and conducting useful, complex studies on school improvement and accountability.

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.