

# Micro Policy Making in Schools

*Use of National Test Results in a Norwegian Context*

Ann Elisabeth Gunnulfsen



Thesis submitted for the degree of Philosophiae Doctor (PhD)

Department of Teacher Education and School Research  
Faculty of Educational Sciences

University of Oslo  
January 2018

© Ann Elisabeth Gunnulfson, 2018

*Series of dissertations submitted to the  
Faculty of Educational Sciences, University of Oslo*  
No. 287

ISSN 1501-8962

All rights reserved. No part of this publication may be  
reproduced or transmitted, in any form or by any means, without permission.

Cover: Hanne Baadsgaard Utigard.  
Print production: Reprosentralen, University of Oslo.

# Acknowledgements

At the end of my PhD project, I feel greatly indebted to all the people who have taken an interest in my work. My thanks go to all who have been a part of the process in various ways, and to everyone who encouraged me. In particular, I want to express my gratitude to the following.

Thanks to my main supervisor Jorunn Møller and my co-supervisor Astrid Roe for believing in me, challenging me, and giving me courage. Your support and valuable comments in all phases of the project have been invaluable. I learned so much from your model and from co-authoring the first and third articles of my project. Your doors and minds have always been open to me. You are my stars!

I appreciate my school informants for welcoming me to their meetings and for sharing their practices and experiences. Without them, the MPM project would not have been possible.

Thanks to Judith Warren Little, who warmly welcomed me to the University of California, Berkeley, to Kriz Gutierrez and Bruce Fuller for letting me take part in their research groups, and to the PhD students whom I got to know and learned from. These meetings were valuable aspects of my work. Thanks to Tina Trujillo, who commented on my work for the second article, patiently read all my drafts, and supported me during my year of research abroad. Thanks to P. David and Barbara for offering their friendship and cultural knowledge. You made me feel at home! Special thanks also to my Norwegian writing colleagues and friends at UC Berkeley: Kristin, Heidi, Camilla, Anniken, and Siri. Our shared experiences, reflections, lunches, coffee talks, and fun trips to San Francisco were precious to me. I will never forget our “Californian motto”: lean forward!

I am grateful for the National Graduate School in Educational Research (NATED) for influencing my development as a researcher and for funding six months of my year abroad in California. In particular, I want to thank Jorunn Møller, Kirsten Sivesind, and Gunn Søreide for leading my member period for the final two years in track 3 and for offering opportunities to socialize in the international field of educational research.

Thank you to my fellow PhD candidates at NATED, Marianne, Inger Lise, Maike, Nora, Øyvind, and Sverre. Our collaboration was, in my view, an excellent model for mixing academic and social experiences. In particular, Jeff, I appreciate all previous, present, and future collaboration with you!

I am grateful to the Department of Teacher Education and School Research for being an engaging and supportive workplace in which I am surrounded by dedicated and inspiring colleagues. In particular, I would like to thank all my colleagues in the master's Educational Leadership program for creating a stimulating environment for research, teaching, learning, and social gatherings. Special thanks to Kristin, Ruth, Kirsten, Marit, Guri, Silje, Hedvig, Tor, Jeff, Eli O., Anne Kristin, Eli L., Eyvind, Christian, Dijana, Eivind, and Jorunn.

Thanks to my colleagues in the Curriculum Studies, Leadership, and Educational Governance (CLEG) research group. In addition to the colleagues also present in the master's program, I send my gratitude to Berit Karseth and Kristen Sivesind. Your comments are always constructive and essential.

I want to acknowledge Kristin F.S., Toril, Eli K., Hilde, and Sandra for their administrative and professional support and for keeping track of me and all parts of my work. I also want to give thanks to Kenneth, Bjørn, and Torgeir. I appreciate all your help, technical support, and, most of all, patience.

Thank you to my PhD friends and colleagues on the 11<sup>th</sup> floor during construction and hefty core drilling in our building. We got some really good exercise and playful competition in counting stairs! Thanks also to Merete, Eli L., Eivind, Vanja, Silje, Kaja, and Drita in the fellow office on the 4<sup>th</sup> floor for showing interest in my work, good talks, and laughs. Special thanks to Lisbeth M. Brevik for encouraging me to apply for the position and supporting me generously on the way. You are a great friend and an exceptionally skilled colleague!

I appreciate Sirpa Lilja for being my friend and running companion. Now I am ready to speed up with a map and compass in the years to come!

In particular, thanks to my family, which anchors my world. To my "Fantastic Four" children, Karsten, Maren, Helge, and Eirik, you are my greatest treasure and my everything. To my companion, Hans Christian, you are my rock. Without you, I would not have been able to do this. To my two supportive brothers, Ole-Erik and Petter, being your little sister has made me strong and confident. To my father Terje, who taught me the history and importance of fighting for all to have free and equal opportunities for education, you are my role model and my hero!

I dedicate this PhD thesis to my mother Kristi, who did not live to see me finish my project. Your support and encouragement was crucial. I know that you are proud!

# Summary

This PhD project investigates the micro policy making of school principals, senior leaders, and teachers in Norwegian lower secondary schools. The thesis, coined the MPM project (Micro Policy Making in Schools), emphasizes the relatively new national policy demand to make use of national test results in Norwegian schools. The concept of “making use of” national test results is explained as school professionals’ construction of discourses regarding the central government’s demands for national testing policies; how they deal with the intentions of such policies, and how power and talk play into these policies. The MPM project is particularly concerned with policy makers’ increasing transnational emphasis on the relations between school quality and large-scale student assessment and how new modes of accountability influenced by new public management have entered the educational context of Norway, which has long been dominated by social democratic values.

Existing international research on data use in the last two decades has largely relied on conceptual work, document analyses, or interviews. This multiple-case study applied a mixed-methods approach to the empirical data, which include 12 individual interviews with school principals and senior leaders from four schools, on-site observations, and informant-conducted video recordings of meetings of school principals and senior leaders and of teacher teams in two lower secondary schools. In addition, 176 lower secondary teachers were surveyed. The empirical data are related to perspectives on micro policy making, including perspectives on micro policy enactment, actors, talk, and discursive roles as well as on the crafting of policy coherence. All of these perspectives contribute to the field of education policy scholarship. Taken together, they enable discussions about the relationships between school professionals’ micro policy making, how it is related to accountability perspectives, and the intention of national testing policies. The findings of the MPM project are reported in three research articles.

*The first article* investigates the stories of four school principals and eight senior leaders as they respond to questions regarding the school’s and the teachers’ work with national test results. The stories illuminated that national test results served mostly as tools for symbolic responses to questions. The school principals and senior leaders mainly focused on low-performing students when talking about national test results. Internal accountability perspectives were most prominent, and the school principals and senior leaders were not especially concerned about report meetings at the superintendent level. The most significant

finding is principals' and senior leaders' disbelief that information obtained from national test results could enhance individual student learning.

*The second article* elaborates how the school principals, senior leaders, and teachers construct discourses about the central government's demands for national testing policies, how they deal with the intentions of policies, and how power and talk are at play. By using a discourse analytical approach to analyze talk in the meetings, different perspectives on power relations as well as the various discursive roles of the school professionals were identified. The findings show that the school principals functioned mostly as narrators and enthusiasts, senior leaders served as messengers and enforcers, and teachers act as critics and preventers of "overburdening." Targeted collaborative talk about the use of national test results did not occur, and the justifications for the school's practices were supported by practical outcomes.

*The third article* quantitatively examined how 176 Norwegian lower secondary school teachers responded to their work with national test results, how they perceived the school principal's facilitation of this work, and their attitudes towards national testing policies. The results illuminated different practices at the school, teacher, and teacher team levels. The teachers felt that the principal exhibited high confidence regardless of whether they actively safeguarded their work. Their perception of the school principal's "clarity" correlated strongly with the teachers' work with and attitudes towards the results. The article argues that teachers and teacher teams feel that they lack information and understanding about how to use national test results in their work to enhance individual student learning.

Empirically, the MPM project contributes to the field by combining data from on-site observation, informant-created video data, interviews, and survey data to document micro policy making, how the intentions of these policies are dealt with, and micro-political "power talk" at schools. Its theoretical contribution is the finding that the principal's discursive role and facilitating practices as a formal leader seem to depend on the principal's relationship with teachers. Moreover, local school context, perspectives on conforming practices, and symbolic responses to national testing policies are important, but great trust in school professionals and internal responsibility for the core purpose of education are essential. In summary, this study contributes new knowledge about micro policy making in schools in a social democratic educational context and how the power at play may take unpredictable and diverse forms both within and between schools.

# Part I: Extended Abstract

<b>1. Introduction.....</b>	<b>1</b>
1.1 Aims and research questions .....	1
1.2 Positioning of the study .....	4
1.3 The Norwegian testing policy context.....	6
1.4 Clarifications and limitations of the empirical case.....	7
1.5 Outline the extended abstract .....	7
<b>2. State of the field.....</b>	<b>9</b>
2.1 Literature search .....	9
2.2 New governing dimensions of education in Norway .....	10
2.2.1 National testing and policy rhetoric .....	11
2.2.2 Tensions around practices related to national tests in Norway .....	11
2.3 Global governing and local pressure .....	13
2.3.1 Large-scale assessment in education.....	14
2.3.2 Large-scale data use for improving schools and student learning.....	15
2.3.3 Performance of accountability .....	17
2.4 Contribution to existing research and practice .....	18
<b>3. Micro policy making in schools.....</b>	<b>19</b>
3.1 Micro policy enactment .....	19
3.2 Micro policy actors.....	21
3.3 Micro policy talk and discursive roles.....	23
3.4 Crafting policy coherence.....	24
3.5 Characterizing micro policy making in schools .....	26
<b>4. Methodology and data .....</b>	<b>28</b>
4.1 Methodology .....	29
4.2 Mixed methods approach.....	30
4.2.1 The multiphase design.....	30
4.3 Participants .....	32
4.4 Data .....	34
4.4.1 Phase 1: Interviews.....	34
4.4.2 Phase 2: Observation .....	35
4.4.3 Phase 3: Teacher survey .....	37
4.5 Analyses.....	37
4.5.1 Analytical concepts .....	38
4.5.2 Analytical process .....	39
4.6 Research credibility.....	42
4.6.1 Reliability .....	42
4.6.2 Validity.....	43
4.6.3 Ethical considerations .....	48
4.7 Short summary.....	49
<b>5. Summary of the articles.....</b>	<b>50</b>
5.1 Article I.....	50
5.2 Article II.....	50
5.3 Article III .....	51
<b>6. Discussion.....</b>	<b>52</b>
6.1 School professionals' engagement with policy initiatives .....	52

6.1.1 Intentions and translations.....	52
6.1.2 Symbolic responses and tensions .....	54
6.2 <i>Micro policy making - constraining and enabling factors</i> .....	55
6.2.1 Values at play .....	56
6.2.2 Power at play .....	57
6.3 <i>Teacher engagement and principal facilitation</i> .....	59
6.3.1 Teacher attitudes and engagement .....	59
6.3.2 The distinct principal.....	60
6.4 <i>Leadership in a social democratic testing context</i> .....	62
6.4.1 Leadership and performativity .....	62
6.4.2 Leadership and the micro politics of leading .....	63
6.5 <i>Short summary</i> .....	64
<b>7. Conclusion</b> .....	<b>66</b>
7.1 <i>Contributions</i> .....	66
7.1.1 Theoretical and conceptual contributions.....	66
7.1.2 Methodological and empirical contributions .....	67
7.2 <i>Implications</i> .....	68
7.2.1 Implications for school professionals.....	69
7.2.2 Implications for policy makers.....	69
7.3 <i>Limitations and need for further research</i> .....	70
<b>References</b> .....	<b>72</b>
<b>Appendices</b> .....	<b>82</b>
<i>Appendix 1: Literature review</i> .....	82
<i>Appendix 2: Approval for research (NSD)</i> .....	88
<i>Appendix 3: Information letter (Phase 1)</i> .....	89
<i>Appendix 4: Information letter (Phase 2)</i> .....	90
<i>Appendix 5: Overview of participating schools and informants in Phase 1</i> .....	91
<i>Appendix 6: Interview data</i> .....	92
<i>Appendix 7: On-site observation data</i> .....	93
<i>Appendix 8: Interview guide</i> .....	94
<i>Appendix 9: Teacher questionnaire (Phase 3)</i> .....	95
<i>Appendix 10: Examples of the coding process in NVivo (Article I)</i> .....	96
<i>Appendix 11: Process of analysis (Articles I and II)</i> .....	97
<i>Appendix 12: E-mail to respondents</i> .....	98
<i>Appendix 13. Overview of the phases, articles, questions, and analytical concepts</i> .....	99
<i>Figures and tables</i> .....	100



## **Part II: The Articles**

### **THE THREE ARTICLES**

#### **Article I**

Gunnulfsen, A. E., & Møller, J. (2017). National testing: Gains or strains? School leaders' responses to policy demands. *Leadership and Policy in Schools, 16*(3), 455-474.

#### **Article II**

Gunnulfsen, A. E. (2017). School leaders' and teachers' work with national test results: Lost in translation? *Journal of Educational Change, 18*(4), 495-519.

#### **Article III**

Gunnulfsen, A. E., & Roe, A. (2017). Investigating teachers' and school principals' enactments of national testing policies: A Norwegian study. (Accepted for publication in *Journal of Educational Administration*).



# 1. Introduction

This thesis reports the results of the Norwegian Micro Policy Making (MPM) project, which focuses on how and why lower secondary school professionals (school principals, senior leaders, and teachers) make use of national test results. In Norway and internationally, the use of large-scale student data has been studied by focusing on politicians', local authorities', and educational professionals' understandings about how the results can be used to enhance the quality of education in schools (e.g., Coburn & Turner, 2011; Datnow & Hubbard, 2015; Prøitz, Mausethagen, & Skedsmo, 2017a; Skedsmo, 2011). There is, however, limited knowledge about how and why local school professionals deal with new national testing policy expectations in their “everyday” work, especially in Norway.

The MPM project is primarily a study of educational policy scholarship focusing on localized policy making. The project investigates micro-level enactment of national testing policies and the interaction between school principals, senior leaders, and teachers while using data from national test results in Norwegian lower secondary schools. The MPM project sheds light upon the Norwegian context, where testing policies were introduced fairly recently, and it moves beyond this context, positioning the study within the international body of literature regarding increased external control by use of large-scale data (e.g., Datnow & Hubbard, 2015; Grek, 2009; Ozga, 2008).

In the MPM project, the concept of data use and school professionals' enactment of testing policies is defined by Coburn and Turner (2011) as individuals' interaction with test scores, grades, and other assessment tools to meet central policy demands. In the project, data use and micro policy making incorporate interactions between school principals, senior leaders, and teachers and how those individuals deal with, negotiate, talk about, and make sense of national student test results.

## 1.1 Aims and research questions

The overall aims of the MPM project are to develop new insights into national testing practices and regulations in Norway and to determine how school principals, senior leaders, and teachers deal with relatively new policy expectations, particularly their intentions for making use of national test results to enhance individual student learning and school quality. More specifically, the project aims to identify and discuss school professionals' micro policy making in an educational system that is rooted in social democratic norms and values of trust

and equity (Møller et al., 2007; Ottesen & Møller, 2016). This educational context is different from that of other countries with longer traditions of education regulation based on large-scale student data, such as England and North America. Through social interaction, local school actors can be understood as contributors to policy making and enactment (Bowe, Ball, & Gold, 1992). The main research question for the MPM project is as follows: *How and why do school professionals in a Norwegian social democratic context make use of national test results?*

Four overarching research questions have guided the study and will be addressed in the final discussion in Chapter 6. They should be seen in relation to the research questions underlying each article, which are presented in Table 4A. The overarching questions are as follows:

1. *How and why do school principals and senior leaders engage with policy initiatives related to national testing?*
2. *What are constraining and enabling factors in how local schools deal with the intention of national testing policy?*
3. *How and why do teachers engage with policy regarding the use of national test results, how do they perceive the principal's facilitation of this work, and what are their attitudes towards this work?*
4. *What kind of educational leadership is being enacted in a social democratic school context when there is a top-down demand to raise test scores?*

To explore the research questions of the MPM project, I conducted three separate studies, which are presented in three articles. The first study is concerned with school principals and senior leaders, their responses to how and why national testing policies are enacted in general, and their stories about making use of national test results. The second concerns school principals', senior leaders', and teachers' discursive talk about and enactment of national testing policies as well as their use of national test results. The third study explores teachers' experiences with the use of national test results and how school principals facilitate this work. Together, the three studies elucidate the formal school leader perspective (school principals and senior leaders) and the teacher perspective. The studies examine the relationship between school principals and teachers first by examining the responses of school principals and senior leaders reported in Article I, second by analyzing the on-site observations and informant-conducted video recordings of school principals, senior leaders and teachers reported in Article II, and third by analyzing the correlation between teacher

responses and attitudes towards the school's work with test results *and* principals' facilitation of this work reported in Article III.

An exploratory case study methodology (Yin, 1994) with a mixed methods approach (Greene, 2007) is used to produce and analyze sets of interview data, on-site observation data, informant-conducted video data, and survey data, all of which enables the researcher to closely investigate micro policy making in lower secondary schools. The approach is an exploratory sequential, multiphase design (Creswell, 2013; Creswell & Clark, 2011). The rich data enabled observation of the language of policy, how policy rhetoric and discourses occur, and how power and negotiation (bargaining and talk) play a key role in policy. Additionally, the data enabled analysis of teachers' attitudes and experiences regarding national testing policies as well as school principals' role in facilitating micro policy making concerning national testing policies.

Conceptually, the project opened up the possibility of employing theoretical perspectives to illuminate different configurations of micro policy making. Through such an approach, the empirical data thus functioned as a significant mirror in which the theoretical framework could be reflected. The concept of "making use of" national test results can be explained as the way in which school principals, senior leaders, and teachers jointly construct discourses about, negotiate, and deal with central governing intentions and demands for national testing policies. Analysis of this micro policy making is made possible by using theoretical perspectives on educational policy within a micro-political approach (Ball, 2012), including perspectives on micro policy enactment, policy actors, and policy work (Ball, et al., 2012) as well as on crafting policy coherence (Seashore Louis & Robinson, 2012). In the MPM project, I consider data use and testing to be the core idea of policy, and the relatively new testing regulations in Norway are chosen as the focus of the project. Articles–I–III show that national test data is used in different ways and considered to be an accepted part of local school practices. More importantly, they show that school professionals are trusted to know other data and types of action can affect the school's work with student learning and school quality.

I will continue this introduction by positioning the MPM project in the context of international policy and research before moving briefly into the Norwegian testing policy context. Finally, I outline the sections of the extended abstract. The research questions of the three articles are also presented in chapter 4 (Table 4A), including an overview of the data and descriptions of the actors and findings.

## 1.2 Positioning of the study

The rationale for this project is grounded in policy makers' transnational emphasis on student achievement and large-scale student testing during the last two decades, which is closely connected to policy makers' desire to develop and control policy demands by central governments (Ozga & Lingard, 2007). This transnational emphasis generates demand for primary research about how micro policy making actually occurs, especially in Norway's comprehensive education system, which is strongly rooted in ideologies that emphasize aspects of equity closely linked to social democratic values (Møller et al., 2007; Ottesen & Møller, 2016).

Educational reform, such as the recent introduction of national testing in Norway, not only emphasizes national and local needs but also can be viewed as part of a trend adopted and influenced by large-scale international comparison tests of the Organization for Economic Co-operation and Development (OECD). The idea of improving schools' quality by use of large-scale student data has become what Røvik (2014) defines as a global phenomenon or a travelling reform concept. This "travelling belief" in large-scale student data has produced new concepts and ideas connected to standards and school improvements (Sundberg & Wahlström, 2012).

Research has shown that reforms are enacted in different contexts where different actors interpret and understand reforms and policy intentions according to what is considered to be understandable and applicable (Karseth, Møller, & Aasen, 2013). Institutional work has been done at different levels, as demonstrated by a Norwegian analysis of what goes on at the central level after a new reform is developed (Karseth & Møller, 2014). In the MPM project, I investigate these "travelling beliefs" and institutional work at the micro policy level, focusing on what is seemingly missing in the field: micro policy making by school professionals in a national educational context in countries where policy beliefs regarding improvement of school quality by use of large-scale student data are uncommon and infrequently represented.

Research has highlighted the important role of teachers in the use of test data within a local school setting (e.g., Coburn & Turner, 2011; Datnow & Hubbard, 2015). Other studies have focused only on school principals' perspectives (Day, 2003; Leithwood & Hallinger, 2002; Møller, 2009). The MPM project describes and discusses leadership as it occurs within the talk, roles, and practices illuminated in the situations under study involving school principals, senior leaders, and teachers. As such, the MPM project implies that "school

leadership” is not necessarily associated with a particular position and may come from school principals, teachers, and others (Møller et al., 2007). By focusing on processes of social influence and activity as well as influences such as use of power, leadership is presented as an organizational value (Ogawa & Bossert, 1995).

Few studies have investigated how school principals, senior leaders, and teachers *interact* when dealing with and responding to national educational policy expectations. There also seems to be a need for more knowledge about local school-level practices when school professionals create and conduct testing policy discourses as they deal with large-scale test results to improve student learning, classroom practices, and school quality. Studying how school professionals mutually respond to policy initiatives such as national testing might enhance knowledge about how policy expectations are enacted in schools because the use of large-scale test results creates new representations in school principals’, senior leaders’, and teachers’ practices (Honig & Hatch, 2004). The notion of policy expectations is exemplified by Norwegian authorities’ demands that schools are expected to make use of national test results in a specific way to enhance individual student learning and school quality (Norwegian Directorate for Education and Training, 2011).

The Norwegian authorities’ expectations regarding use of national test results are closely connected to the concept of accountability policies, the underlying assumption of which is that the results of standardized tests will be used by schools to make better decisions about classroom practices. Accordingly, accountability policies can be called “regulation by results” (Maroy & Voisin, 2015, p. 39) or “governing by numbers” (Grek, 2009). Accountability is a multi-layered concept tied to the central rationality of trust, and practices of accountability can reveal very different social relationships with regulation (e.g., against deliberation) as well as express different evaluative criteria for external and internal goods (Ranson, 2003).

The different relations and purposes of accountability create a rationale to more closely study micro policy making using Ranson’s (2003, p. 461) distinction between the “external goods of effectiveness,” such as wealth, status, and power, and the “internal goods of excellence,” such as the virtues of justice, courage, and friendship. This distinction is especially important in the Norwegian social democratic educational system, which has a long tradition of including public trust, civic service, social justice, and equity as central purposes of education. This ideological dimension of the purpose of education has been a compelling subject of research. Offering a comprehensive education system that provides equal opportunities for everyone is a high priority in the social democratic system (Karlsen,

2000). The idea implies that all students, regardless of gender, residence, and socio-economic background, have the right to free, equitable education that is adapted to their abilities and interests. In conjunction with this ideological base, until the 1990s, the national curriculum was considered the most important tool for governing the work of Norwegian schools (Bachmann, Sivesind, Afsar, & Hopmann, 2004; Hopmann, 2003; Sivesind, Bachmann, & Afsar, 2003).

### **1.3 The Norwegian testing policy context**

National student testing was introduced in Norway in 2004 mainly due to the below-average results of the 2001 Program for International Student Assessment (PISA) (Bergesen, 2006; Lie et al., 2005). With the average PISA-results and the introduction of the new Norwegian curriculum in 2006 (Norwegian Ministry of Education and Research, 2013), policy makers decided that the reform should be followed up by a National Quality Assessment System (NQAS), of which national student testing was a part (Møller, Ottesen, & Hertzberg, 2010; Skedsmo, 2009).

For the two first years, researchers, school leaders, and teachers criticized the quality of the national tests due to the lack of information about the tests as well as the increased workload, and a majority of the students in higher secondary schools boycotted the tests. Moreover, teachers were concerned about the possible consequences of the use of league tables. Consequently, the tests were stopped in 2006. One year was spent improving them, and in 2007, the tests were reintroduced (Lie et al., 2005) and were met with far less criticism. The number of basic skills tested was reduced from four to three and tests were conducted in spring rather than fall in order to more effectively use the results for formative student assessment.

The Norwegian national tests measure students' performance in the basic skills: reading, numeracy, and English<sup>1</sup> reading. In the Norwegian context, the government acknowledges two functions and intentions of national test results: (1) *improvement of school quality* with accumulated data on the system level and (2) formative assessment to *enhance individual student learning at the local level* (Norwegian Directorate for Education and Training, 2011). Reading and numeracy are basic skills that are part of all subjects at all levels. Hence, teachers must work to strengthen students' basic skills in all text-based subjects. Norwegian national and local policy makers and administrators gradually push for

---

<sup>1</sup> English involves only English as a second language; it is not a part of this study.



increased test scores and argue that large-scale test data are useful for guiding school improvement (Mausethagen, 2013). It is emphasized that large-scale testing is intended to provide instructional diagnosis, identify needs, and ensure accountability (Møller & Skedsmo, 2013). The tests are compulsory for all students in grades 5, 8, and 9.

## **1.4 Clarifications and limitations of the empirical case**

Reporting the MPM project in this thesis provides opportunities to go deeper than the separate articles into new perspectives and discussions significantly related to the main aim. However, it might also have some limitations, and therefore it is necessary to make a few clarifications. First, the concept of “a school” cannot solely represent “micro policy making.” In this thesis, school principals, senior leaders, and teachers represent the school and the use of national test results represents the micro policy making perspective. Moreover, leadership as it occurs within the talk, roles, and practices illuminated in the situations under study is analyzed. The thesis does not focus on how personal conditions, such as gender, age, and work experiences, may influence the empirical case, even though they are always important factors, nor on students, parents, or local authorities, even though these and other school actors also might influence schools’ micro policy work.

The micro policy perspective is used analytically to distinguish between broader policy developments and debates (macro perspective) and face-to-face processes that take place at the local school (micro perspective). The empirical case does not concentrate on issues related to school development or school change or pay direct attention to best practices. The analysis focuses on situations in which basic skills and national tests were the main issue, such as the talk in meetings between school professionals and the interviews with school principals and senior leaders. This is done partly because of the large amount of data and partly because of the main aim of the MPM project, which is to analyze and understand the situations in which school professionals deal with and talk about national test results.

## **1.5 Outline the extended abstract**

This thesis is divided into six chapters aiming to contextualize, exemplify, and discuss the overall MPM project. The study has already been overviewed and positioned nationally as well as internationally. *Chapter 2* further positions the study by presenting the state of the research field. *Chapter 3* outlines the theoretical and analytical framework central to the MPM project, which primarily draws upon on the policy enactment theory developed by Ball

(2012) and key analytical concepts regarding leadership, crafting of policy coherence, coupling mechanisms, accountability policies, policy discourse, policy actors, and policy work. *Chapter 4* presents the research design and argues for the relevance of the mixed methods approach. The methodological choices are thoroughly discussed, and the research questions, participants, data, and analysis are explained. Finally, research credibility, including reliability, validity, and ethical concerns, are addressed. *Chapter 5* summarizes the three articles reported in this thesis, including their main findings and discussions. In *Chapter 6*, the overall study findings, contributions, and possible implications are discussed.

## 2. State of the field

This chapter offers an overview of the field of national and international research on large-scale student testing and schools' micro policy making using national test results in a context that emphasizes globalization, assessment policies, and accountability perspectives. The studies I have reviewed use different terms to report research on large-scale student testing, including *large-scale data use*, *data use*, and *testing*. While the term *national test results* is used in the MPM project, this review includes the terms used in each cited study. In addition, the term *school professionals* is used to refer to principals, senior leaders, and teachers. *Leadership* is used as in the cited works, and educational leadership is considered to be directly linked to educational purposes and practices and trained and accredited educational professionals working to develop the school's pedagogy and curriculum and thus improve student learning and achievement (Gunter, 2016).

### 2.1 Literature search

A two-fold literature review is conducted to situate the findings presented in this thesis and their contribution to the field of research on both the Norwegian and international context. The literature review is a conceptual summary of the characteristics of the field, and it is not exhaustive (Maxwell, 2006). According to Grant and Booth's (2009) definition, this review can be defined as both an *overview* and a *rapid review*, which is an assessment of what is already known about accountability perspectives, testing policies, and local schools' data use (p. 95). Websites that could be accessed with the University Library of Oslo, such as ORIA, The Institute of Education Science, ERIC, and Google Scholar, were searched in the autumn of 2016 for key words such as the following: "national testing" OR "data use," "national testing" OR "data use" AND "school leadership," "accountability policies" AND "data use" OR "national test results," "teachers' experiences" AND "national testing," "policy enactment" AND "national tests" OR "data use."

This relatively wide search resulted in more than 60 international English-language articles, most of which were published in North America and England ("international" here is used to refer to all studies that are not Norwegian). The first search included articles that addressed national testing policies, accountability policies, teacher practices, and school leadership. Next, the search was specifically delimited to include only studies with the terms "data use" and "accountability" that were published in the last ten years, partially for pragmatic reasons and partially because national testing was substantively established in the

Norwegian educational context in 2006. This last specification yielded about 30 articles in the review. Moreover, educational leadership handbooks were reviewed to identify perspectives on data use (e.g., English, 2011; Townsend & MacBeath, 2011; Waite & Bogotch, 2017). To shed light on the state of transnational educational policy scholarship and policy analysis, books about the globalization of education policy were reviewed (e.g., Ball, 2008; Henry; Lindgard, Rizvi, & Taylor, 2001; Rizvi & Lindgard, 2010). Finally, throughout the nearly four years of this project, specific journals specializing in educational policy and leadership were reviewed (e.g., *Journal of Educational Change*, *Leadership and Policy in Schools*, *Journal of Educational Administration*). For a more detailed overview of the article search, see Appendix 1.

The Norwegian search yielded no empirical studies that specifically focused on school principals', senior leaders', and teachers' micro policy work and data use. Exclusion criteria were a time limit of about ten years and the Norwegian words for “national tests,” “assessment,” “leadership,” and “governance.” About half of the studies about the Norwegian context were written in Norwegian, and the rest were written in English and published internationally. Most of the ten studies targeted the national, municipal, and school system level, mainly focusing on the new governance practices of accountability policies (e.g., Prøitz et al., 2017a; Skedsmo, 2009). Other Norwegian studies discussed the pedagogical content of the basic skills measured by the national tests (e.g., Berge, 2005). This literature review takes up these perspectives in the Norwegian context, and the international studies more broadly cover local school practices, data use, and accountability policies. This chapter presents and discusses the Norwegian studies before introducing the international studies that investigate schools' policy enactment and data use. Next, what characterizes the research is discussed by presenting research on data use from assessment and accountability perspectives. Finally, how this thesis contributes to the current literature and practices in the Norwegian context is addressed.

## **2.2 New governing dimensions of education in Norway**

Due to the recent introduction of national testing in Norway, existing research on data use is dominated by conceptual studies of accountability policies, curricula, and governance, primarily focusing on national testing, which is part of the NQAS (e.g., Elstad & Sivesind, 2010; Skedsmo, 2009). Moreover, discussions on the didactical and pedagogical issues of national testing of the basic skills that comprise all school subjects have also had a dominant

status in Norwegian articles published in the last decade (e.g., Berge, 2005; Hertzberg & Roe, 2016; Roe, 2010). One Norwegian study used empirical data to examine the tensions that arose when national testing was discussed between teachers, school leaders, and superintendents (Mausethagen, 2013). There are, however, few studies on micro policy making at the local school level that focus on national tests as a tool for enhancing individual student learning and teacher practice. Similarly, there is little research focusing on *how* school principals, senior leaders, and teachers collaboratively deal with policy intentions to make use of national testing results. In the following, I describe and discuss existing national and international studies to situate the MPM project more clearly within these focus areas.

### **2.2.1 National testing and policy rhetoric**

In the past decade, numerous conceptual research contributions have reported that the first mediocre PISA results released in 2001 and 2004 were crucial for curriculum reform and greater emphasis on student achievement in Norway (e.g., Elstad & Sivesind, 2010; Karseth & Sivesind, 2010; Møller & Skedsmo, 2013). The state secretary from 2001 to 2005, Bergesen (2006), claimed that the Norwegian public and politicians had to agree that the Norwegian school system was significantly flawed in terms of how it taught students knowledge and skills.

Based on her study of educational policies in Norway, Skedsmo (2009) found that system-level monitoring and controlling of schools is de-emphasized in the national policy rhetoric in favor of a discourse that focuses on learning and development (Skedsmo, 2009). Rhetorical focus is placed on the key actors, such as school principals and teachers, who are expected to use information from the test results to improve their practice in ways that enhance student outcomes, particularly national test results. After the recent Norwegian reform (Norwegian Ministry of Education and Research, 2013), comprehensive document analysis has shown that educational leadership and accountability have become dominant discursive themes in Norwegian education (Møller & Skedsmo, 2013). The following presents research that deals with how central educational policies for national testing and accountability measures are perceived and applied by school professionals in the Norwegian context and discusses how the MPM project contributes to existing knowledge.

### **2.2.2 Tensions around practices related to national tests in Norway**

In a robust longitudinal empirical study of the implementation of the recent Norwegian education reform, Møller, Prøitz, Rye, and Aasen (2013) find that national authorities have had a hierarchical and top-down understanding of how to implement the new reform, even though school leader and teacher autonomy was an important part of the policy discourse when the reform was introduced. Five years after introduction of the reform, they found that schools experienced a decrease in autonomy and lacked the ability to make use of and interpret the results of national tests (Møller, et al., 2013). Nevertheless, school leaders and teachers seem to have increased their knowledge of the reform. According to the principals in the study, increased access to, for example, national student test results is used to prioritize work to achieve quality development of schools, and the emphasis on large-scale student data has been important for clarifying the rationale of the reform (Møller et al., 2013). However, the understanding of the improvements produced by the reform seems to depend on the professional reflections and discussions that take place among teachers and principals when they interact with national test results.

In another robust empirical study, researchers found that there is a tendency for schools to work to improve subsequent test results. Practices of “result meetings” were identified as new organizational routines between the municipality and local school level as well as within schools. This study also found that principals found dialogue-based tools to contribute favorably to improvements in student achievement (Skedsmo & Mausethagen, 2016).

So far, studies have not taken into account *how* school professionals deal with and talk about intentions to use test results in their daily work or *why* it matters for the purpose of education in a social democratic educational context.

In their comprehensive evaluation report of surveys and interviews with actors at all levels of the educational system in Norway, Seland, Vibe, and Hovdhaugen (2013) found that principals viewed national tests as an effective tool to improve school quality and individual student learning. Teachers, however, were more reserved in their praise for national tests, claiming that they did not have time to use individual student test results and that they did not gain any new information about the students’ level of knowledge from the test results. They experienced tension between the tests as a control tool and as a tool for development. Teachers that worked in teams to analyze the test results were generally more satisfied with the test as a tool for development, while those that were responsibility for interpreting national tests saw few opportunities for further use of the test results in their teaching practice (Seland et al., 2013).

By observing and interviewing teachers and principals during a school year, Mausethagen (2013) recognized tensions between what is seen as “internal” (i.e., teachers’ everyday work) and what is “external” (i.e., policies and practices outside the main frame of teaching). Even though national testing was mainly seen as external to teachers’ work, teachers were involved in reshaping their professional discourse to ensure the relevance of their work and maintain its legitimacy. The use of national test results seemed to be viewed quite positively by local authorities and school leaders, but less positively by teachers, who found the tests to be too dominating in terms of their directions for planning and conducting their teaching practice. The use of the results was dependent on where in the system the actors were placed, and results were used for pedagogical and diagnostic purposes less often than for governing and controlling purposes (Mausethagen, 2013). So far, these findings and tensions maintain the need for closer investigation of micro policy making in schools and how and why school professionals locally make use of national test results in an educational context. To analyze and discuss how and why this occurs, it is necessary to understand data use as part of the “travelling belief” about large-scale student data that is linked to student learning and school quality.

### **2.3 Global governing and local pressure**

In a recent review of 129 articles on data use in education, Prøitz, Mausethagen, and Skedsmo (2017b) found that most such articles primarily investigate structures and systems around data use. Some focused on implementation and effectiveness, and others focused on reflections about developments in data use in education. Their review showed that data use is a growing research issue and has been largely debated internationally.

A large number of researchers (e.g., Ball, Maguire, & Braun, 2012; Grek, 2009; Hall, 2013; Hardy, 2014; Mintrop & Sunderman, 2009; Pollock & Winton, 2016) have empirically and conceptually contributed to an understanding of how policy demands characterized by high-stakes standardized testing put strong pressure on schools, and how school principals have to manage the pressure and demands of various accountability systems in their work. This pressure needs to be understood in relation to local school professionals’ opportunities within their micro policy making practices. Other researchers have contributed to the understanding of how a range of intended and unintended consequences might occur when local school actors, especially teachers, individually deal with policy initiatives related to

student assessment and large-scale student test data (e.g., Collins, Reiss, & Stobart, 2010; Harris & Brown, 2009; Mausethagen, 2013).

On one hand, many peer-reviewed empirical studies suggest that focusing on data about student performance encourages collaboration among teachers in order to improve practices as well as defend their practices to important stakeholders (e.g., Datnow & Hubbard, 2015). On the other hand, a substantial body of literature suggests that data use in certain accountability contexts changes teachers' perspectives away from a wide-ranging approach, narrows the curriculum, and stimulates teaching and learning strategies that promise to quickly raise test scores (e.g., Collin et al., 2010; Valli & Buese, 2007). Overall, these studies show that data provided by standardized tests embody particular representations of students' learning outcomes that enable users to see some aspects related to teaching and learning processes while other aspects are constrained.

### **2.3.1 Large-scale assessment in education**

There is growing evidence from extensive empirical research that increased use of large-scale and high-stakes testing as a tool for student assessment is restricting students' indefinable learning, such as creative action and understanding of innovation (Hargreaves & Shirley, 2008; Sahlberg, 2010). The concept of assessment for learning is defined as any assessment for which the first priority is to promote pupils' learning. According to the theoretical analysis of Black, Harrison, Lee, Marshall, and Wiliam (2004), assessment for learning differs from assessment designed primarily for accountability or ranking purposes. In their extensive review of assessment literature, Black and Wiliam (1998) found a clear distinction between large-scale student testing and classroom assessment. Assessments come in all shapes and sizes, ranging from international monitoring use to work with individual pupils in the classroom. The focus of this thesis is studies investigating large-scale assessment and schools' use of large-scale student test results. Studies on assessment in general are included only when they shed light on large-scale data use.

Based on a review of 40 articles published from 1993 to 2003 focusing on how different types of assessment are implemented in classrooms, Broadfoot and Black (2004) claimed that formative assessment is in conflict with the requirements of simultaneous summative large-scale assessments. To overcome what the authors described as "the severe limitations of external testing" (p. 17), they argued that teachers need to use their own knowledge about their students as a source of data for the purposes of certification and



accountability. Additionally, national testing and large-scale student test data must be called into question at a deeper level with investigation and “interpretation at micro as well as at macro levels” (p. 14).

Crooks, Kane, and Cohen (1996) and Firestone and Rosenblum (1988) make it clear that the strongly emphasized “empowerment” agenda of educational reforms cannot be achieved alongside corrective use of high-stakes testing to raise “standards.” They claim that although it may be theoretically possible to unite formative and summative assessment for the purposes of student growth and empowerment, doing so would require a substantial shift in the prevailing political and policy priorities of many countries.

In their multi-method investigation of leadership at the school, district, and state levels, Anderson, Leithwood, and Strauss (2010) concluded that the school principal plays an important role in establishing the purposes and practices of data use.

In Halverson et al.’s (2007) year-long study, no schools demonstrated the capacity to provide systematic feedback on student learning due to data-driven decision making across instructional programs. The researchers also claimed that school principals need to build new structures of formative feedback to generate information for teachers to adjust their instructional practices, concluding that school principals need to “reframe the traditional data-use practices of schools in terms of external accountability” (p. 6).

### **2.3.2 Large-scale data use for improving schools and student learning**

In his study of data use to improve schools, Stobart (2008) argued that while policy makers may welcome high-stakes large-scale data testing programs, the response from those tested may be very different. On the one hand, Hardy (2014) found in his extensive interviews with school professionals in Australia that teachers, to some degree, seem to use the demands for increased test scores to inform learning and improve practice. On the other hand, after his six-month study of school professionals in England, Hall (2013) claimed that teachers’ identities are jeopardized when school principals and administrators make an effort to implement assessment practices within accountability policies such as large-scale testing. According to Baker and LeTendre (2005), it has not been determined whether large-scale student testing on a national level creates its own dynamic by introducing new national reforms that in turn require more testing and assessment to determine their effects. In their analysis of data from the Third International Mathematics and Science Study (TIMSS), which was conducted from 1994 to 1999, the authors concluded that when large-scale data use is characterized by

systematic and sustained reflection on multiple forms, it might create processes of continuous improvement (Baker & LeTendre, 2005).

Data use can become infused into the structure and culture of an organization, as shown by Park and Datnow (2009), who found in their multiple case study of four urban school systems in the U.S. that school leaders created an ethos by primarily focusing on learning and continuous improvement rather than blame. They argued that when leaders strategically construct diagnostic, motivating, and prognostic frames to promote a culture of data use for continuous improvement, data use is more successful as a tool for school improvement and student learning. Their findings suggest that a framing of staff empowerment, learning ethos, and building human capacity is an important leadership tactic that must be carefully considered when policy expectations are introduced. The push for data-driven decision making, which in the MPM project is defined as the policy expectations of data use, is a challenge for school leaders in terms of linking the practices and cultures of their schools to new kinds of data provided by accountability systems. Conversely, Park and Datnow (2009) argued that when data use is characterized by sporadic examination of national test results, it is not likely to have the desired results. In other words, data-driven decisions can take many different forms depending on what kind of data is used, for what purposes, and by whom.

Within this context, large-scale student data has become a major component of reform agendas. Although the current review revealed some analytic and conceptual articles that aim to understand the phenomena of large-scale data use (e.g., Coburn & Turner, 2011; Lachat & Smith, 2005; Young, 2006), there was a predominance of critical studies of large-scale testing and data use as tools for improving schools and student learning, even though some of these studies concluded with suggestions on how to understand, promote, and support data use. The most critical studies were focused on the U.K. and Nordic contexts (e.g., Hall, Gunter, & Bragg, 2013; Hopmann, 2008; Møller et al., 2013). Within North America, the majority of critical studies on data use practices were related to high-stakes accountability perspectives (e.g., Anderson et al., 2010; Elmore, 2005; Halverson et al., 2007; Marsh, Bertrand, & Huguet, 2015).

According to Earl and Katz (2002), data has the potential to be a very powerful and useful mechanism for helping schools change in productive ways. They admit, however, that data use is not straightforward and that large-scale data use can be a deeply analytic and emotional process that is closely tied to the context and people involved as well as to the available data (Earl & Katz, 2002). The potential utility of data is to some extent supported

by the teacher perspective, as stated by Abrams, Varier, and Jackson (2016), who suggest that, under applicable circumstances, such as valid test items, good support, and time to discuss the results with others, benchmark student testing programs have the potential to provide meaningful formative assessment. Regarding the notion of “valid” test items, validation in large-scale data use is defined as the possibility to identify the results at a local, detailed, and small scale for each student. Although we know a lot about the possibilities and limitations of large-scale data use for improving schools and student learning, there is still a need for further investigation about how data use policies are actually dealt with in local schools.

### **2.3.3 Performance of accountability**

Building on his groundbreaking and comprehensive analysis of historical and comparative research, Hopmann (2008) argued that the experience of accountability and testing policies depends on deeply engrained “constitutional mind-sets” and how diverse cultures conceptualize the relation between the public and its institutions. In another conceptual article, Elmore (2005) claimed that policy makers as well as professional reformers often think the reforms they sponsor are unique, which leads to serious and adverse consequences for schools, and that “most external accountability systems embody primitive and unspecified theories of school improvement”. (p. 138). According to Elmore (2005), the key issues of large-scale data use for improving schools and student learning as a part of the reform and accountability system seem to be repeatedly associated with the conceptualizations and misconceptions of diverse cultures.

Furthermore, research has demonstrated that methods of improving teaching and learning through accountability policies have had limited effects. For example, in a study of 100 U.S. teachers, Berryhill, Linney, and Fromewick (2009) found that accountability policies lacked support for teacher burnout and self-efficacy. Ball (2003) extensively documented what he calls the unintended and dysfunctional results of the culture of accountability and performativity in the English educational system, writing about the “confusing interplay” of trust/distrust inside the discourses of contemporary public sector development and that there sometimes is talk of “reform fatigue.” Ranson (2003) described the serious effects of the new contexts of accountability on teachers’ practices and the consequential diminishing of pedagogies and educational goals in England.

In their book *Globalizing Education Policy*, Rizvi and Lingard (2010) claim that educational systems and governments around the world have become increasingly unable or unwilling to pay for educational development and have therefore enacted market-based reforms. This has led to an “almost universal and global shift from social democratic to neo-liberal orientations in thinking about educational purposes and governance, resulting in policies of greater demand for accountability” (p. 3). The authors note that policy has become even more important for running educational systems in the context of globalization and accountability. Policy making can hence be understood as a fundamentally political process that involves major trade-offs between values. In particular, public education policies have to simultaneously deal with a range of values, such as equality, excellence, autonomy, accountability, and efficiency. Micro policy making as a kind of trade-off between values is a central issue in the MPM project.

#### **2.4 Contribution to existing research and practice**

This review has discussed characteristics of policy research and identified crucial gaps that should be addressed in further studies of micro policy making in schools. In sum, it has demonstrated a need for empirical studies on micro policy making and clarification of central policy intentions regarding the use of national test results for school development and student learning. The review has been important for identifying the need for studies that include school professionals, including school principals, senior leaders, and teachers, within one study about data use. Even several studies were relevant to the current study in terms of findings and conceptualizations of data use in schools, the MPM project adds to the field by illuminating the dynamics between local practices and ideological contexts. As such, the MPM project aims to fill the research gap regarding how and why school principals, senior leaders, and teachers deal with, negotiate, and make sense of national testing policies. As both Ball (2012) and Rizvi and Lingard (2010) have argued, much is known about data use and large-scale assessments as travelling beliefs and contested practices. However, this knowledge does not include perspectives on micro policy making in schools. Based on the review, I will argue that there is a need for a new way of thinking about education policy as localized policy making, which leads to the theoretical perspectives introduced in Chapter 4.

### **3. Micro policy making in schools**

National testing policies can be understood as situated within overlapping systems of institutional boundaries (Maroy, 2012; Olsen, 2005). In these overlapping systems, policy actors interact at the macro-, meso-, and micro-levels, such as international and national governmental entities, or at the municipal or local school levels. National testing policies are textual phenomena that constrain and open possibilities regarding material related to policy, and the responses to these texts have “real” consequences. These consequences are experienced within the context of the practice to which the policy refers and is addressed (Bowe et al., 1992). The MPM project is focused on the micro policy level. The school is the subject of the policy under study, and school professionals are considered policy actors. The phenomenon under study is the micro policy making of national policies.

The MPM project is delimited to examination of lower secondary school principals’, senior leaders’, and teachers’ enactment of national testing policies. As an educational policy study, the project opens possibilities for investigating national testing policy intentions for data use and enables investigation and discussion of the micro policy work between school professionals. The project is challenging as it must pay attention to the broader social and civic context and not treat school professionals’ use of national test results as existing in an institutional vacuum. Thus, the study includes an understanding of the wider institutional structures, roles, functions, and norms that enable or constrain micro policy making. The following theoretical grounding is intended to illuminate the main research question—how and why do school professionals in a Norwegian social democratic context make use of national test results—by combining different theoretical perspectives. The following sets up a conceptual framework to further support and design the study by presenting perspectives on micro policy enactment, policy actors, discursive roles, micro policy talk, and crafting policy coherence, all of which are relevant for a critical discussion of the MPM project and its contribution to the field.

#### **3.1 Micro policy enactment**

The MPM project is centrally concerned with what Ball et al. (2012) describe as the “diverse and complex ways in which a set of education policies are made sense of, mediated and struggled over, and sometimes ignored, or, in other words, *enacted* in schools” (p. 3). More precisely, the project is focused on how school principals, senior leaders, and teachers make use of the results of national tests, which are an example of education policies. Enactments

are collective and collaborative, and they occur in the interaction and inter-connection between diverse actors, texts, talk, and technologies that constitute ongoing responses to policy. Moreover, enactments are “negotiations and translations of policies which go on at these points of connections over time and space” (p. 3). Policy is not “done” at one point in time; it is always a process of “becoming,” changing from the outside in and the inside out. Hence, policy enactment involves reviewing and revising policy as well as sometimes instructing or simply forgetting (cf. Ball et al., 2012, pp. 3–5). Ultimately, school principals’, senior leaders’, and teachers’ enactment of national testing policies might involve dealing with, making sense of, translating, and negotiating the use of national test results for instructing new teaching practice or generally ignoring the test results.

Conceptually, I follow Ball et al.’s (2012) call to take local contexts seriously in an effort to understand why policies are enacted differently in seemingly similar schools. Too often, research on policy implementation has taken for granted the meaning of policy itself. The MPM project is about the practices of policy, rather than policy “implementation” in any sense. Policy texts simply cannot be implemented straightforwardly, but must be translated from text to action, put into practice, and considered in relation to history and context (Ball et al., 2012). This project builds on Bowe et al.’s (1992) alternative framework for epistemological and political assumptions about “implementation” research and change in schools. When school professionals read policy documents on national testing, they learn *what*, *why*, and *when* new policies are meant to take place. However, these texts rarely tell school leaders or teachers exactly *how* work with national test results should happen.

The larger institutional context in this study is placed within an understanding of how education system is globally and nationally governed and valued and how education policy is “done” within the new institutionalist perspective. The spaces in which policy is thought and talked about have changed, and the characteristics of power within education policy are shifting in, for example, the relatively new national testing system in Norway. According to Ball (2012), the perspectives of power and conflict underlie almost all aspects of the study of values, interests, and change in organizational practices, and these dimensions involve a set of concerns that he calls “the micro-politics of the school” (p. 18). The concept of micro policy enactment in the MPM project is comprehensive, but limited regarding the interests of school professionals dealing with, negotiating, and practicing data use. It is not assumed that conflict is always destructive; it might also be a constructive part of the many “little-p policies that are formed and enacted within localities and organizations” (Ball, 2008, p 8). Accordingly, in micro-political analysis, the informal is assumed to have greater significance,

and social relations are usually the most important means for policy interpretation, policy translation, decision-making, and information exchange. In short, micro policy enactment can be called the “doing of policy” (Ball et al., 2012, p. 95). The following provides a framework for micro policy enactment, which in the MPM project is understood and analyzed as *micro policy making* in schools.

### **3.2 Micro policy actors**

Within the processes of policy discourse, individuals and groups of “authoritative policy interpreters” play an important role in “driving the way in which policies are selected and understood”. (Ball, et al., 2012, p. 8). The MPM project identifies groups or individuals with important roles in thinking and talking about policies in local schools as policy actors playing different discursive roles in their “policy work” (Ball et al., 2012). Accordingly, school professionals hold different roles regarding students’ parents, the media, and the community, and the roles and way in which they talk may occur in various locations (e.g., classroom, teacher break room, parent-teacher meeting, principal’s office). As such, enactment is likely to diverge because policy is usually dealt with in different arenas as well, and studies of micro policy making need to be conducted correspondingly, such as in formal and informal meetings between school professionals in different arenas.

An important aspect of the MPM project is what Ball (2008) refers to as “the language of policy; policy rhetoric and discourses” (p. 6). Inspired by the typology of “policy actors/positions which are involved in making meaning of and constructing responses to policy” (Ball et al., 2012, p. 49), a central step of the MPM project is to identify the different types of roles embedded in school principals’, senior leaders’, and teachers’ policy translation work (Table 3A). Ball et al. (2012) noted that the positions of *policy actors* (left column) are not necessarily attached to specific individuals and people may move between these roles in different aspects of their *policy work* (right column) and in their interactions with colleagues.

**Table 3A.** *Overview of policy actors and their policy work*<sup>2</sup>

<i>Policy actors</i>	<i>Policy work</i>
Narrators	Interpretation, selection, and enforcement of meanings
Entrepreneurs	Advocacy, creativity, and integration
Transactors	Accounting, reporting, support, and facilitation
Enthusiasts	Investment, creativity, satisfaction, and career-building
Translators	Production of texts, artifacts, and events
Critics	Monitoring management and providing counter-discourses
Receivers	Coping, defending, and dependency

According to Ball (2008), policy spaces are diversified and the territory and roles of influence are expanded. There are more diverse and a greater number of participants in the policy process than before, and sometimes it is difficult to know which voices count most or where and how key decisions are made. Policy discourse flows through new places such as think tanks and social media, gathering support and credibility as it does. The era of new governance does not merely represent a segment of party politics or ideology, but is part of a broad global shift in public service discourses involving language, ideas, organizations, and practices (p. 224).

The rules and thoughts (and talk) of school professionals about how to make use of national test results is a part of a “new legitimation” movement in the Norwegian education system. This also means that shared beliefs and cognitions play a vital role in use of test results, and that bargaining, conflict, and power are ever-present realities in the school (cf. Powell & DiMaggio, 1991, p. 27). Therefore, issues of change, power, and efficiency need to be addressed when studying how schools make use of national test results. One must bear in mind that perspectives on the global shift and performativity are vital for discussing the main research question: how and why do school professionals in a Norwegian social democratic context make use of national test results?

Ball (1994) has defined policy discourses as “ways of talking about and conceptualizing policy” (p. 109). Policy discourse can comprise both creation and consideration of texts because texts are related to action and enactment (Gale, 1999). According to Spillane, Halverson, and Diamond (2004), local discursive practices are situated in areas of tension between political and professional discourse, and even though such tensions can create discomfort, they must be negotiated to create organizational legitimacy. According to Ball (2008), policy discourses organize their own specific

---

<sup>2</sup> Modified from *How Schools Do Policy: Policy Enactments in Secondary Schools* (p. 49) by S. J. Ball, M. Maguire, and A. Braun, 2012, London: Routledge. Copyright 2012 by [Copyright holder]. Reprinted with permission.



rationalities, making particular sets of ideas obvious, common sense, and “true.” Discourses “mobilize truth claims and constitute, rather than simply reflect, social reality” (Ball, 2008, p. 7). Moreover, discourses “produce social positions from which people are ‘invited’ to speak, listen, act, read work, think, feel, behave and value” (Gee, Hull, & Lankshear, 1996, p. 10). In this thesis, micro policy work in schools and the policy enactment in local schools are recognized as deeply connected to policy discourse. In meetings between school leaders and teachers, invitations to speak, listen, act, read, value, and think are both formalized and, to a great extent, informal.

### **3.3 Micro policy talk and discursive roles**

The MPM project concentrates on the concept of discourse in connection to perspectives on policy talk between school professionals in the local school context and how this micro level discourse might be used to prioritize certain ideas and speakers and exclude others. According to Ball (2008), policies represent specific rules about truth and value, and the ways in which policies are worded and spoken about are part of the creation of their conditions of acceptance and enactment. The policy vocabularies construct what is challenging, needed, or not needed within, for example, national testing policies.

Ball (2012), focuses on the notion of autonomy, claiming that by granting autonomy over a specific curricular or classroom issue, the head (school principal) may effectively exclude teachers from talk about a range of other issues that affect the organization as a whole. By definition, autonomy might even limit the range of concerns over which the teacher can exercise influence. Following this claim, Ball (2012, p. 124) presents four types of participation and “talk” that have a role in schools’ decision-making, and in each case, leadership gives rise to and is perpetuated by a particular form of “political talk” in the organization, or what can be called “micro policy types of talk”. The four types of talk are authoritarian (preventing public expression), managerial (formal meetings), interpersonal (informal chats), and adversarial (public meetings and open debates). The responses to opposition of these types of political talk are, respectively, to stifle, to channel and delay, to fragment and compromise, and to confront. With regard to school principals, senior leaders’, and teachers’ micro policy talk about national test results, these four types of talk might occur in various settings and situations. In any case, the school principal has the power to either suppress or confront the issues being discussed or the issues on which he/she wants to focus.

Regarding the pressure to “deliver” and perform, Ball et al. (2012) claim that the word “focus” is interesting in many ways. In application of and discussion about the word “focus,”

it evokes the idea of bringing a lens to a point of concentration. It also suggests precise, organized, and effective action regarding teachers, school leaders and schools, pedagogies, procedures, and performance. The notions of leadership and different types of talk are imperative issues with regard to conformity and performativity in micro policy work and enactment of national testing policies. Leadership and talk is also relevant in combination with the theoretical perspectives of accountability, power, and coherence. I have chosen to apply these perspectives to illuminate the power relations in school professionals' micro policy making regarding national testing.

### **3.4 Crafting policy coherence**

It has been stated that a successful way to be effective in organizational settings is to cultivate and use one's social and political skills and to build on one's ability to influence, convince, and control others (Ferris et al., 2002; Mintzberg, 1983). What is of importance regarding the world of power and influence in micro policy making in schools is the degree to which various regulatory political and accountability tools are aligned, both amongst themselves and with different levels and actors of the educational system (Blase & Anderson, 1995). Thus, there can be more or less alignment of different policy factors that might influence the political skills of school professionals (the curriculum, evaluation, performance, skills, or educational "best practices"). If so, these factors will be able to serve as reference points in any form of accountability or evaluation of standards for "steering" central educational policies. If there is consistency in the values underlying policy reforms, it is more likely that the reforms will be adopted (Christensen & Lægreid, 2007; Ozga & Jones, 2006). For example, it is more likely that intentions regarding national testing reform and expedient use of test results will succeed if all involved parties work to remove tensions by identifying power structures and the underlying values associated with them.

Through the alignment process of translation for crafting coherence, policies are repeatedly interpreted and assembled by local actors (Ball et al., 2012). Thus, the local context in which the ideas are received, translated, and adapted into new practices must be considered (Newman & Clarke, 2009). In the MPM project, the crafting of coherence begins in the interactions between policy makers' policy initiatives and school professionals' policy practices (Seashore Louis & Robinson, 2012). According to Honig and Hatch (2004), coherence "depends on how policy actors make sense of policy demands and on the extent to which external demands fit a particular school's culture, political interests, and conceptions of professionalism and on-going operations" (p. 18). However, Coburn (2004) has found that

making sense of policy demands occurs with different coupling mechanisms, namely rejection, decoupling (symbolic responses with no effect), parallel structures (balancing of different priorities), assimilation (fit with the pre-existing understanding), and accommodation (substantial changes in the pre-existing understanding). Hence, the unavoidable variety in context means that the crafting of policy coherence will vary with different policy actors, such as school principals, senior leaders, and teachers. Inside-out or bottom-up approaches to coherence may generate local ownership in practice, but they may also fail to facilitate loyalty to the reform policies (Coburn, 2004).

Honig and Hatch (2004) argue for a method that treats coherence as a process “requiring school and school district central office leaders to work in partnership to constantly ‘craft’ or ‘negotiate’ the fit between external demands and schools’ own goals and strategies” (p. 17). This collaborative perspective suggests that neither local authorities nor school professionals can act alone to craft coherence in school improvement efforts. Rather, participants must work together to learn about, refine, and support the school’s improvement goals.

The MPM project also sees micro policy making in schools through the lens of a regime in which numbers are a “resource through which surveillance can be exercised” (Ozga, 2008, p. 264). Hence, national test results function as policy for “a resource of comparison” (p. 267) for improvements in quality and efficiency by making nations, schools, and students “legible” (p. 268). Numbers are becoming a greater part of policy discourse and the micro-political world within schools. According to Ball (2008), there is a paradox regarding market-based reforms because this monitoring discourse rests upon the promise of greater autonomy within educational systems, although it actually provides the state with new modes of governing, reducing autonomy, and shaping both individuals and individual conduct. Hence, a focus on autonomy and different types of relationships is also relevant when studying the crafting of policy coherence in micro policy making between school professionals when they collaboratively talk about national test results.

Different perspectives on accountability and responsibility generate tensions related to power and control in local school practices, and the differences create tensions within the conception and crafting of coherence regarding the policy intention for using national test results to improve school quality and individual student learning. There are both competing notions of and different approaches to accountability (cf. Pollock & Winton, 2015; Ranson, 2003). In the MPM project, the distinction between external and internal accountability as an analytical tool is emphasized (Abelman, Elmore, Even, Kenyon, & Marshall, 1999; Elmore,

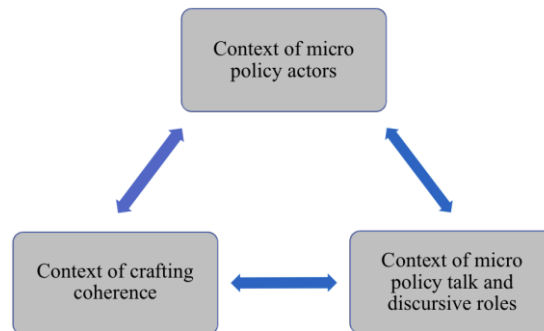
2005). External accountability implies performance-based accountability or “policies that evaluate, reward, and sanction schools on the basis of measured student performance” (Elmore, 2005, p. 134). In contrast, internal accountability is guided by individual responsibility or collective expectations, relying on internal normative structures that might be relatively immune to external influences (Elmore, 2005). O’Day (2002) claims that the combination of external and internal accountability addresses an underlying problem because external accountability might prevent the internal need to foster interaction concerning “common work and the sense of shared purpose” (p. 24). The tensions here lie in the contrasts between internal work with collective responsibilities, shared purpose, and values and external demand to improve the school’s test results and the crafting of coherence in this “two-dimensional landscape.”

In the Norwegian context, national testing policies have two objectives: first, to monitor and control students’ competence in basic skills such as reading and numeracy at a national level, and second, to use the test results to enhance individual students’ learning of reading and numeracy. Policy coherence is defined as the extent to which school professionals experience and craft consistency and agreement between national and externally initiated policy reforms with the values and demands that underlie them. In the Norwegian case, crafting policy coherence involves aligning the two different objectives of national testing policies and the underlying values and norms in local school practices by continuously dealing with policy demands in their discussions about national test results. It also depends on how school principals talk about the school’s results with senior leaders and teachers and what effort is made to reduce tensions and conflicts between internal work with collective responsibilities and the external requirements for developing school quality. Policy coherence requires one to accept the school as a real world of micro-political power and influence, bargaining, and negotiation (Blase & Anderson, 1995). Hence, the perspectives of power and influence on the crafting of coherence serve as tools to analyze the research question of this study.

### **3.5 Characterizing micro policy making in schools**

This conceptualizing chapter is concluded by introducing the notion of a continuous micro policy making cycle intended to recontextualize the contexts within which policy work occurs in schools. Inspired by and building on Bowe et al.’s (1992, p. 20) and their model for contextualizing policy making with the three contexts of policy making: a) *text production*, b) *practice*, and c) *influence*, a model for policy enactment in schools in which text production,

practice, and influence are represented in all three micro level policy contexts is created. The model of micro policy making in schools is represented in Figure 3A. Three primary micro policy making contexts are presented, bearing in mind that each context consists of a number of conceptual perspectives regarding how and why school professionals make use of national test results in a Norwegian social democratic context, such as leadership, accountability, performativity, and quality.



**Figure 3A.** *Contexts of micro policy making in schools*

Figure 3A represents the continuous micro policy making process of schools and is not meant to be read in a particular way; the three contexts are intertwined and related in a cyclical manner. The first context, the *context of micro policy actors*, is where policy enactment, including meaning making and construction of responses to policy, occurs and actors as narrators, entrepreneurs, and critics play a role. The second context, the *context of policy talk and discursive roles* is where policy actors in various settings and situations participate and where perspectives of leadership and power give rise to a particular form of “political talk”: the ways in which policies are worded and spoken about (Ball, 2012). The third context, the *context of crafting coherence*, is where actors, talk, and roles contribute to work with consistent underlying values, such as work to remove and/or identify tension, and ensure fit with the school’s culture, political interests, and ongoing practices (Honig & Hatch, 2004).

The contextualization of micro policy making in schools here highlights how a mixed methods approach has made it possible to investigate and analyze these complex contexts, which is further discussed in Chapter 4.

## 4. Methodology and data

The three empirical studies that were conducted and presented in Articles I–III employ different data sources and approaches but all intend to answer the main research question: *How and why do school professionals in a Norwegian social democratic context make use of national test results?* The overall research design and conceptual approach serve as a starting point for investigating how school professionals deal with national testing policies in public lower secondary schools in Norway. Table 4A presents an overview of the research questions, empirical data, and main findings of the three articles.

**Table 4A.** *Overview of research questions, empirical data, and main findings of the three articles*

Studies	Research questions	Empirical data	Main findings
<b>Article I</b> National Testing: Gains or Strains? School Leaders' Responses to Policy Demands.	(a) How do school leaders make sense of and respond to external demands related to a new policy context that emphasizes national test results for reading and numeracy? (b) How do they use national test results in their local work with school and student development?	Individual interviews of 4 school principals and 8 senior leaders.	Test data has become a tool for symbolic responses and helps school professionals focus on low-performing students. Internal accountability seems to trump external accountability.
<b>Article II</b> School Leaders' and Teachers' Work with National Test Results: Lost in Translation?	(a) How do school leaders and teachers position themselves in relation to each other in their interpretation and translation of national testing policies? (b) In what ways does school leaders' and teachers' language reflect their enactment of national testing policies?	Participant-conducted video recordings and on-site observations of leader team and teacher team meetings.	While responding to test data, school principals serve more as narrators and enthusiasts, senior leaders and teacher team leaders serve more as the messengers and enforcers, and teachers serve more as critics and the preventers of "overburdening."
<b>Article III</b> How Teachers Use National Test Results.	(a) How do teachers perceive principals' work to facilitate national test results at the local school level? (b) How do teachers and teacher teams use national test results, and what attitudes do teachers have towards national tests? (c) What relationship exists between principals' facilitating work with national test results and teachers' practices and attitudes?	Teacher questionnaire (N=176).	Teachers responded differently to the school principal's role in facilitating work with national test results. These differences correlated strongly with the teachers' practices and attitudes regarding their own work and the teacher teams' work with national test results.

This chapter begins by presenting the overall methodology and then discusses the purpose of using a mixed methods approach combining qualitative and quantitative data collection and analysis (Sections 4.1 and 4.2). Next, the design of the MPM project (Section 4.2.1), participants (4.3), data (4.4), and analyses (4.5) are described. Finally, the research's credibility in terms of reflexivity is addressed, including the researcher's own role, validity and analytical generalization, and ethical considerations (4.6).

## 4.1 Methodology

The MPM project is policy research seeking to contribute to critical policy studies by investigating perspectives on micro policy enactment, policy actors, discursive roles, micro policy talk, and crafting of policy coherence. In this case, data are interpreted as they integrate and compare with Norwegian policy demands for national testing policies and the study of how school professionals deal with these demands. The project is an exploratory instrumental multiple case study with a sequential multiphase design (Johnson & Christensen, 2014; Creswell, 2013; Creswell & Clark, 2011; Yin, 1994). The central components of a case study are questions about *how* and *why*. An exploratory and instrumental case study examines cases to generate ideas and learn about a phenomenon in a more general sense (e.g., school professionals locally dealing with national policies in general rather than using national test results in particular schools or in a particular municipality). An advantage of a multiple case study is its ability to more broadly analyze similarities and differences. A disadvantage is that studying multiple cases may lessen the depth of the analysis. This trade-off between breadth and depth is common in case study research (Johnson & Christensen, 2014).

The MPM project uses Bowe et al.'s (1992) model to "approach policy as discourse" (p. 13), which is constituted of possibilities and impossibilities, tied to knowledge on the one hand (analysis of problems and identification of solutions) and practice on the other (specifications of methods for achieving goals). The concerns of the project have been to explore micro policy making in terms of value dispute processes and micro policy discourses in lower secondary schools in one municipality as well as to portray and analyze the processes of active interpretation and meaning making that relate to policy texts and demands in practice. In part, this involves identification of resistance, adjustment, deception, and conformity within and between arenas of practice and maneuvering of conflicts and disparities between continuous discourses in these arenas (Bowe et al., 1992). Furthermore, it is important to acknowledge that policy intentions may contain ambiguities, uncertainties, and errors that provide and/or prevent particular opportunities for school professionals concerning micro policy making in schools. The national policy demand to make use of national test results is an ambiguity mainly because the notions of "results" and "make use of" are complex and may be understood differently in various contexts.

## **4.2 Mixed methods approach**

Mixed methods inquiry is “the planned and intentional incorporation of multiple mental models with their diverse constituent methodological stances, epistemological understandings, disciplinary perspectives, and habits of mind and experiences – into the same inquiry purposes of generatively engaging with difference toward better understanding of the phenomena being studied” (Greene, 2007, p. 30). In the MPM project, this inquiry embraces the difference between principals’ and senior leaders’ stories about how to use national test results and the teachers’ responses to the same issue in order to better understand the phenomena of national testing policies. In line with Greene’s (2007) definition, a mixed methods approach was chosen for this research to study the multiple ways of knowing how school professionals make use of national test results for a deeper and more enriched analysis. This study’s point of departure is that although philosophical paradigms may be incommensurable (Kuhn, 1962), most mental models, such as a “mixed way of thinking model,” are inherently dialogic. Most social research “seek[s] connection, conversation and understanding - one with the other” (Greene, 2007, p. 53). Moreover, mixed methods research is intended to pursue paradoxes, contradictions, and inconsistencies in which the priority [between the ways of knowing] is likely to be equal (Maxwell, 2012).

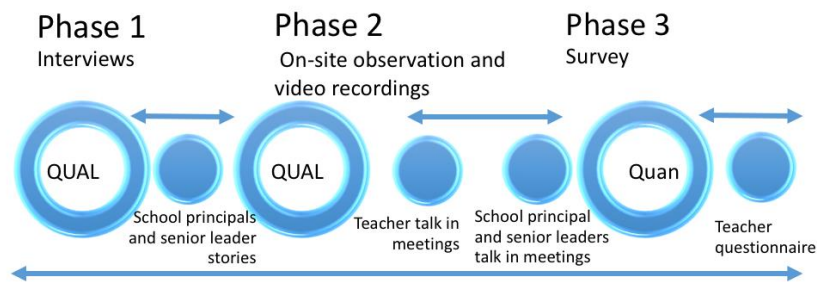
Here, aspects of Greene’s (2007) dialectic stance are used to understand that this specific stance is a sociopolitical argument as much as a methodological issue. The overall purpose of Greene’s (2007) stance is “meaningful engagement with difference” by gaining and discussing different kinds of information about aspects of the phenomenon under study (Greene, 2007, p. 14). In the MPM project, this means meaningful engagement with different kinds of information regarding how school professionals make use of national test results. The multiphase mixed methods approach is often guided by specific theoretical perspectives (see also the substantive stance, Bickman, 1987; Greene, 2007). In this study, policy sociology (e.g., Ball, 2008) and perspectives on policy enactment, policy actors, discursive roles, and crafting policy coherence have inspired the methodological approach in relation to the dialectic stance. The methods of the MPM project will be further discussed throughout this chapter.

### **4.2.1 The multiphase design**

Figure 4A illustrates the design of the three phases, what kind of data was collected in each phase, and how they influenced each other regarding the main aim to study how school



principals, senior leaders, and teachers make use of national test results. This approach is recognized as an exploratory sequential, multiphase design (Creswell, 2013; Creswell & Clark, 2011).



**Figure 4A.** *The research design to achieve the overall aim of investigating school principals’, senior leaders’, and teachers’ use of national test results*

As shown in Figure 4A, this PhD project has a qualitative dominant design (Creswell & Clark, 2011) as two of the data collection phases were qualitative (school leader interviews and on-site observations/video recordings from meetings). The approach is described as a sequential exploratory design in mixed methods literature, with one form (e.g., quantitative data) building on the other (e.g., qualitative data) in order to identify important and, so far, unknown variables between the data (Creswell & Clark, 2011; Johnson & Christensen, 2014; see also Greene, 2007, p. 131). The survey in Phase 3 was constructed on tentative conclusions, patterns, and generalizations based on the qualitative data obtained in Phase 1 and 2.

*Phase 1: School principals’ and senior leaders’ perspectives.* When designing this phase, the overall aim was to investigate school principals’ and senior leaders’ perspectives on how teachers and school leaders made use of national test results for reading and numeracy. This phase included individual interviews with both principals and senior leaders, all of whom had formal leadership responsibility. The interviews were semi-structured and conversational (Brinkmann & Kvale, 2014; Creswell, 2013) and lasted 1–1.5 hours.

*Phase 2: School leaders’ and teachers’ perspectives.* Phase 2 was intended to scrutinize school principals’, senior leaders’, and teachers’ perspectives when jointly constructing discourses of, negotiating and dealing with national testing policies, bearing in mind the concepts and findings of Phase 1. It includes on-site observations and participant-conducted video recordings of leader team and teacher team meetings in order to analyze and contrast the findings of Phases 1 and 2. Because school leaders’ *stories* about the school’s translation and negotiation of expectations for national testing policies were analyzed and

compared in Phase 1, the scope in Phase 2 was narrowed to focus on school leaders' and teachers' discourses and negotiation regarding *collaborative talk* in meetings.

*Phase 3: Teachers' perspectives.* Since dealing with policy expectations occurs at different levels of the school organization (Ball et al., 2012), it was important to compare the perspectives of formal school leaders (school principals and senior leaders), and the collaborative talk of school leaders and teachers from a teacher perspective. Since the teachers were part of the school principals' and senior leaders' stories about how national policy expectations were dealt with and negotiated, Phase 3 was designed as a quantitative descriptive analysis (Ary, Jacobs, Sorenson, & Razavieh, 1990) of how teachers experience their own work with national test results as well as teacher teams' work, and the principal's facilitation of work with national test results. Based on data obtained from Phases 1 and 2, constructs of thematic areas were developed to create the items in the teacher survey. The aim was to provide a different perspective of school principals' and school teachers' work with national test results.

Data obtained from the three phases are mostly reported separately; Article I reports the data from Phase 1, Article II reports the data from Phase 2, and Article III reports the data from Phase 3 but clearly builds upon Phases 1 and 2. Appendix 13 provides an overview of the phases and articles, including methods, research questions, participants, data, analytical concepts, mixed methods credibility, and the main findings.

### **4.3 Participants**

According to Ball et al. (2012), the notion of a school's enactment of policy intentions includes a variety of policy actors (see Section 3.2). A "school" cannot enact policy demands per se. Policy actors might be local authorities, superintendents, parents, media, students, teachers, and school leaders (Ball, 2012; Ball et al., 2012). Since this is a study of micro policy making in schools, it was essential to include school principals, senior leaders, and teachers in this research. A limitation of the study is its exclusion of other policy actors, such as students and parents. The samples include school principals, senior leaders, and teachers from lower secondary schools in a municipality with average to high national test scores.

The selection criteria included socioeconomic diversity, school size, and test scores. The starting point was selection of a municipality. The 428 municipalities in Norway are responsible for providing 10 years of compulsory education at the primary and lower secondary school levels. These municipalities vary in size as well as average income. Based

on national statistics concerning socioeconomic status, school size, and national test results, Adde was selected. It is a medium to large-sized municipality that, during the last three years, received a national and local reputation for its systematic work on national tests and average to high national test scores. Adde lies in the central eastern part of Norway. The following presents a sampling of each of the three phases, which have a multilevel, nested, and parallel relationship and are three of four major types of criteria for mixed methods research (Johnson & Christensen, 2014, p. 272).

*Sample 1:* Many Norwegian schools are small. To protect respondents' anonymity, schools with more than 300 students were included since one selection criterion was based on Adde's average to high national test scores. Selection of large schools makes it more difficult to expose the pupils' identity. Six schools in Adde matched the criterion. These schools were contacted by email, and four schools immediately expressed willingness to participate and thus were chosen. These schools were named Blue Mountain, Yellow Stone, Brown Hill, and Red Field. Their national test scores varied from average to high compared to the national average during the last three years. Formal leaders (4 school principals and 8 senior leaders) were chosen from these four lower secondary schools. A list of the schools with their alias names and school professionals' titles is provided in Appendix 5.

*Sample 2:* In order to obtain different perspectives and investigate the relation between school principals, senior leaders, and teachers, Brown Hill and Blue Mountain were invited for an in-depth study of leader team and teacher team meetings. This selection was primarily based on knowledge about the two schools obtained from Phase 1 and their different levels of national test results. The two schools were seemingly similar with regard to organizational structure and meetings and were part of the data in Phase 1. However, Blue Mountain self-reported having a more progressive pedagogical practice than Brown Hill, which had a reputation of being more traditional.

*Sample 3:* To obtain information about how teachers in lower secondary schools in Adde experienced and perceived their principals with regard to national testing policies and to compare differences in the responses of school principals and senior leaders to national test results, on-site observations, and participant-conducted video recordings, a teacher survey was prepared. Questionnaires were sent out to all lower secondary teachers in Adde ( $N=426$ ) by using the e-mail address on each school's website. Based on the website information, it was possible to determine that teachers of practical aesthetic subjects (e.g., arts and crafts, music, physical science, and food and health), assistant teachers with no formal teacher

education or responsibility for teaching, senior leaders<sup>3</sup>, and teachers whose emails were posted on the website but who were no longer employed at the school were ineligible. The sampling procedure resulted in 176 teacher responses. Accordingly, in Phase 3, the teacher sample consists of 176 teachers from 12 lower secondary schools in Adde.

## 4.4 Data

The following sections are meant to serve as a supplement to the accounts provided in each of the three articles.

### 4.4.1 Phase 1: Interviews

In Phase 1, data was collected through individual semi-structured interviews with principals and senior leaders in four different schools. The data consisted of 16 hours of audiotaped interviews, resulting in 185 pages of written transcripts. The interviews substantially contributed to the school principals' and senior leaders' answers about how and why they work with the results, when they work with the results, and who works with the results. Further, the interviews elucidated their thoughts about school quality, student results, and school results in general as well as the school's capacity to enhance student achievement.

According to Kvale and Brinkmann (2009), it can be claimed that qualitative interview research is an ability learned through the process of conducting interviews. Even when using an audio recorder in a mostly formal setting, the interviews typically functioned as conversations about issues. This experience is recognized in Ary et al.'s (1990) description of conducting a qualitative interview, which highlights the need to ask the questions in such a way as to obtain valid responses and record responses accurately and completely. A semi-structured interview guide (Appendix 8) was developed in order to contribute to comparable data from all 12 individual responses. The data was entered into NVivo 10 using free nodes from the inductive and interpretive readings. First, the transcripts were coded openly according to questions in the interviews that addressed the ways in which school leaders described micro policy making regarding national test results. The concepts they emphasized were *test results*, *practice*, *student learning*, and *school quality*. Second, the frequencies of different words associated with individual students or a school-wide perspective were

---

<sup>3</sup> Senior leader: In this study, the title "senior leader" is given to people with formal responsibilities associated with the former assistant principal role. Senior leaders are responsible for managing subject matter and human resources for their grade levels, including economic responsibility for parts of the school's budget.

examined. The data sets obtained from the interviews are accounted for in Appendix 6, and the process of analysis is described in Appendices 11 and 13.

#### **4.4.2 Phase 2: Observation**

In Phase 2, data consisted mainly of on-site observations and participant-conducted video recordings of meetings. However, a range of sources, like emails, school plans, and informal conversations with teachers, deputies, and principals before and after meetings functioned as background data. At Brown Hill, on-site observation of meetings was arranged beforehand, and all observations occurred at leader team and teacher team meetings for grade 9. Blue Mountain offered the opportunity to participate in any meeting at any time. The observed meetings were primarily concerned with work to improve student learning, plans for the next teaching period and the whole school year, and work with national test results.

The observation data marks a shift between Phase 1 and 2 from collecting self-reporting data to integrating these data with observations. Johnson and Turner (2003) pointed out that “observation is an important method, because people do not always do what they say they do” (p. 312). Additionally, according to Achinstein (1968), observations are influenced by the observer’s knowledge and intentions, and thus the researcher must be aware that observations are influenced by previous knowledge about the phenomenon in question as well as the danger of reporting what was observed. The positioning and point of view chosen by the researcher are significant, and as an observer, the researcher must attempt to take a non-participatory role and distance him- or herself from the observations. Here, the observations are defined as “observer-as-participant” observations, which means that the researcher takes on the role of the observer rather than the role of a participant (Johnson & Christensen, 2014). Audio recording was not conducted during the observations, which may limit the overall study. However, all observed meetings were carefully documented with notes written directly on a laptop computer or in a notebook, either verbatim or directly after the meetings. In addition, photographs were taken when somebody in the meetings wrote on the whiteboard or used visual tools like PowerPoint, and the researcher’s immediate thoughts about the observations were recorded while driving home from the sites, to prevent loss of important details (Johnson & Christensen, 2014).

Phase 2 emphasizes video data from the teacher team and leader team meetings at both schools as these data provide the most intensive view of the topic under study. Video footage was gathered in the fall, when the schools were conducting the national tests and

planning for the rest of the school year. This footage showed how school professionals talk about work with national test results for reading and numeracy and instances in which micro policy making was prominent. The senior leaders and teacher team leaders took responsibility for conducting video recordings of all the meetings. Blue Mountain primarily recorded leader team and teacher team meetings from the first week in August throughout October. Brown Hill recorded all the meetings for which national testing was on the agenda in the same period. Blue Mountain produced more than two-thirds of the video data, but an equal number of statements from both schools were chosen. These data build on the data obtained in Phase 1, which influenced the choice of codes and categories, such as the terms “dealing with,” “negotiation,” and “making sense of national testing policies and basic skills.” The analytical concepts are described in Appendix 13.

**Table 4B.** *Overview of the observation and video recordings of leader team and teacher team meetings in Phase 2*

School	Step 1: Observation and informal conversations (Spring 2015)	Step 2: Video-recordings “respondent-conducted” (Fall 2015)	Contextual information
<b>Blue Mountain</b>	<b>15 hours</b> 3 leader team meetings 2 teacher team meetings 2 plenary meetings	<b>32 hours</b> 8 leader team meetings 8 teacher team meetings	<i>Public online information about national testing. School results published online. School’s online documents about focus areas and local plans. Information on wall boards. Emails and the school’s webpage.</i>
<b>Brown Hill</b>	<b>5 hours</b> 2 leader team meetings 1 teacher team meeting	<b>8 hours</b> 2 leader team meetings 2 teacher team meetings	

Table 4B provides an overview of the data obtained in the first and second steps of Phase 2. The two schools video recorded all teacher and leader meetings, most of which included national testing and basic skills on the agenda. Together, the integration of data within and across Phases 1 and 2 clarified the school principals’, senior leaders’, and teachers’ interactions, placement of responsibility, and management of, negotiation of, talk about, and making sense of national student test results. The next phase integrates these qualitative data with a quantitative approach for the purpose of initiation, to evoke divergence or fresh insight and new perspectives, and to assess various facets of the same complex phenomenon (Greene, 2007, p. 103).

### **4.4.3 Phase 3: Teacher survey**

The questionnaire introduced a shift from Phase 1 and 2, which merged different types of qualitative data from four schools in a single municipality-based sample, to Phase 3, which built exclusively on quantitative survey data from all lower secondary teachers in the same municipality. I used a self-constructed questionnaire inspired by thematic areas developed from the two first phases (see Appendix 13). The survey was technically produced by following a web-based template from the University of Oslo. A total of 426 questionnaires were sent to all lower secondary teachers in the 14 lower secondary schools in the municipality, and 176 teachers conducted the survey. Based on the information from Phases 1 and 2 (e.g., that arts and crafts, physical science, and domestic science teachers were not involved with national testing), there is a high probability that approximately 300 respondents who received the survey were eligible. The response rate is hence estimated to be approximately 50 %. This may be a limitation of the study as the teachers who were expected to be involved did not recognize their eligibility. The survey was piloted with four colleagues at the University of Oslo and four lower secondary teachers without connection to the selected municipality. Their comments resulted in minor changes, mostly deletion of items. The final survey also collected information about which subjects the teachers taught and how long they had been working as teachers, and at the end, there was a field for voluntary supplementary comments.

The core 27 survey items had five-point Likert-type scales. The items represented a range of variables, which were directed towards the concepts of “leadership,” “expectations,” “experiences,” and “attitudes.” According to Ary et al. (1990), descriptive research asks questions about the nature, incidence, or distribution of variables and/or the relationships among these variables. In this case, items about the teachers’ practices with regard to the principal’s facilitation of work with national test results were created and defined within the nature and incidence of micro policy making concerning national tests. The items and questions are outlined in the teacher questionnaire in Appendix 9.

## **4.5 Analyses**

Mixed methods data analysis consists of analytic techniques applied to both quantitative and qualitative data as well as to a combination of the two forms of data in a multiphase project (Creswell & Clark, 2011). Hence, the data analyses in the MPM project involved looking across quantitative and qualitative data in order to assess how the information addressed the

research questions. The data analysis is described in detail in the three articles, and the following summary aims to show how the mixed methods multiphase design influenced the analyses. Drawing on the contextual overview of the field of national and international research on large-scale student testing and schools' micro policy making concerning national test results presented in Chapter 2 and the theoretical framing presented in Chapter 3, the analytical concepts that are essential to the analyses conducted in the research presented in Articles I, II, and III are described.

### 4.5.1 Analytical concepts

I used five overarching analytical concepts across and within the three phases: *policy enactment*, *policy actors*, *discursive roles*, *crafting coherence*, and *micro-policy practice*. As illustrated in Table 4C, I designed the first two phases to investigate school principals', senior leaders', and teachers' micro policy making concerning national testing policies, while the third phase was intended to capture teachers' experiences and attitudes with regard to the same issue.

**Table 4C.** Overview of areas of analytical focus in each phase and data sources used to investigate school principals', senior leaders', and teachers' use of national test results

	Policy enactment	Policy actors, discursive roles	Crafting coherence and micro-policy practice
<b>Phase 1</b>	School principals and senior leader interviews School documents	School principals and senior leader interviews School documents	School principals and senior leader interviews School documents
<b>Phase 2</b>	Field notes School documents On-site observations and participant-conducted video recordings	On-site observations and participant-conducted video recordings Field notes School documents	On-site observations and participant-conducted video recordings Field notes School documents
<b>Phase 3</b>		Survey data	Survey data

The three areas of analytical focus build on the main research question: how and why do school professionals in a Norwegian social democratic context make use of national test results? These areas of analytical focus are based on internal interpretation, which occurs through a process influenced by the social and structural conditions of schools as workplaces (Coburn, 2003; Spillane, 1999; Spillane, 2012). Moreover, the process of putting policies into practice is always situated in context (Ball et al., 2012). The analyses are conducted in light of the theoretical concepts presented in Chapter 3, and the data under analysis are the school leader interviews conducted in Phase 1; the on-site observations, participant-conducted



video-data, field notes, and school documents (background data) obtained in Phase 2, and the survey data obtained in Phase 3.

#### **4.5.2 Analytical process**

To expand the methodological clarity of the MPM project, examples of how the analyses were conducted for various types of data are presented.

*School principal and senior leader interviews:* Content analysis of the transcriptions of the audio-taped interviews was conducted. To identify micro policy making, sensitizing concepts in their stories and how they described their understanding of how to make use of the test results were searched. One example of such a concept is “Well, this [the results] is not very interesting because we already have this information.” Hence, examples of sensitizing concepts include (a) conformity/non-conformity, (b) delaying mechanisms, (c) resistance, and (d) negotiation. The numerous readings of the transcribed interviews were intended to describe variations in responses regarding the use of national test results in a local school. Coding and analysis of the interviews can be characterized as data-driven because codes were developed through multiple readings of the material (Appendix 11). The emerging approach of the process of analysis concentrated on interpretation of the sense-making performed by the participating school professionals and how they dealt with, negotiated, and talked about their work with national test results. I conducted the next step in a more deductive manner, combining the distinction between external and internal accountability (Elmore, 2005) with the five coupling mechanisms Coburn (2004) identified as analytical tools.

Throughout the interviews, when an open-ended question about the school’s work with national test results was asked, it was necessary to elaborate upon the meaning of, for example, *teachers, school, work with, and results*. The explicitness of the answers very much depended on the ability to discuss the meaning of the words in questions such as the following: *Tell me about your school’s work with national test results and the teachers involvement*. This question immediately revealed the lack of information about the concepts. What did “involvement” actually mean? Having the control and flexibility to clarify meaning is one of the advantages of interviews (Ary et al., 1990). Data were entered in NVivo 10 using free nodes from the inductive and interpretive readings. First, the transcripts were coded openly according to the questions in the interviews that addressed the ways in which school professionals described their enactment and sense-making of national test results. The

concepts they emphasized were *test results, practice, student learning, and school quality*. Second, the incidence of words connected to individual students or whole-school perspectives was examined (Appendix 10). I discussed the analysis and preliminary interpretations with other researchers as part of the continuous validation process (Johnson & Christensen, 2014).

*On-site observations:* The field notes, pictures, and audio recordings were analyzed after the visits, including background documents from the on-site observations and created codes and categories for observation of the video data.

*Video recordings:* Video data were analyzed by observing the videos and reading the field notes and transcripts numerous times. The first viewing revealed the time at which the main topics (e.g., student learning, national test results, basic skills) were discussed and segments were marked as they occurred in the video. During the second and third viewings, the chosen segments were transcribed. Then, data were organized in a variety of matrices and memos and three readings of the transcripts were performed. The first reading was intended to identify how school principals, senior leaders, and teachers discussed and planned for activities related to national test results and work with basic skills. The second reading was performed to investigate how discursive talk was linked to variations of enactment. The third reading was intended to examine variations in individual and collective talk in leader team and teacher team meetings about how test results were used in local work to improve student learning. In addition, it revealed how teachers positioned themselves in relation to school leaders in the talk. During all readings of the transcripts and thinking about the data (Maxwell, 2012), memos were written and coding categories were developed on numerous small notes. Matrices were created to find patterns connecting the memos. In summary, the selected excerpts represented what Fairclough (1992) described as tension points indicated by language-in-use. This selection criterion enabled more comprehensive analysis of situated meanings (Gee, 2014). The main difference between the interviews in Phase 1 and the observations and video data in Phase 2 was the explicit way in which the informants talked about the teachers' practice, the school's vision, and reflections about policy expectations regarding national testing.

*School documents (background data):* The school documents consisted of the plans and information sheets that teachers and school leaders handed out or discussed during the meetings, information and histories hanging on the school walls, and information that was published on the schools' websites. Even the researcher's own pictures (photos) of the teacher team leaders writing on the white board functioned as school documents. These

documents functioned mainly as contextual information, but also as a guideline for asking clarifying questions during part 1 of Phase 2 (observation). Comparing the plans and documents with statements, information, and talk in meetings contributed to validation of the main data in Phases 1 and 2.

*Survey data:* The analysis in Phase 3 was performed using the data program Statistical Package for the Social Sciences (SPSS). The Cronbach's alpha (coefficient alpha) method, which estimates consistency in the data (Cronbach, 1951), were performed. Validation of the 27 items as one construct in terms of reliability showed a Cronbach's alpha of  $\alpha=.92$ . For each of the four thematic areas, Cronbach's alpha was  $\alpha=.83$  for construct 1 (facilitating principals),  $\alpha=.70$  for construct 2 (teachers' individual work with national test results),  $\alpha=.83$  for construct 3 (the team's work with national test results), and  $\alpha=.71$  for construct 4 (teachers' attitudes towards national test results). In the survey, the teachers responded to items concerning the analytical concepts of crafting coherence and micro-policy practice on a five-point Likert scale. The scale ranged from "not to any extent" to "to a very great extent" for the items concerning the three first thematic areas, which were concerned with experiences and practices. On the items that concerned the fourth construct (attitudes, perceptions, and viewpoints), the five-point scale ranged from "totally agree" to "totally disagree." The first concept (principals' facilitating) was measured with seven items. The second concept (teachers' work) was measured with six items. The third concept (teacher teams' work) was measured with five items, and the fourth concept (teachers' attitudes) was measured with nine items. The estimates formed the basis for the descriptions used in the analyses in Phase 3, including the methodological considerations related to the conduct of a descriptive, exploratory quantitative study (Creswell & Clark, 2011). In line with Ary et al. (1990), the overall aim of this phase of the study is to select applicable measurement procedures driven by the research questions.

The main aims of collecting these data were to further identify, elaborate, and discuss possible differences regarding the concept of micro policy making concerning national testing policies, which were revealed in the school principal and senior leader stories in Phase 1, and to identify the roles and practices of school principals, senior leaders, and teachers as policy actors, which were revealed in the observations and video data in Phase 2. SPSS was used to analyze the results and Cronbach's alpha was used to measure internal consistency (i.e., the closeness of the relation between each set of items within the concept). Thirdly, descriptive statistics were used to measure the distribution of responses for each item (Johnson & Christensen, 2014), and finally, bivariate correlations were conducted to detect

covariance between the teachers' experiences with their principals and the teachers' own practices and attitudes related to national tests.

## **4.6 Research credibility**

Research credibility emphasizes awareness of justifiable research and involves establishing that the results are credible or believable from the perspectives of participants in the study. Hence, credibility, which is synonymous with validity in quantitative research, is judged by respondents' recognition of consistency; in other words, participants confirm the findings (Kumar, 2011, p. 185). The following sections discuss the reliability, validity, and generalizability of this mixed methods research before addressing its ethical concerns.

### **4.6.1 Reliability**

Reliability is concerned with consistency or regularity, as emphasized in Silverman's (2013) definition: "reliability refers to the degree of consistency with which instances are assigned to the same category by different observers or by the same observer on different occasions" (p. 302). Still, research involving people can never be fully repeated. For example, the talk in a teacher team meeting can never be identically reconstructed. The three different measures of reliability used in this thesis are *reliability of results* (i.e., replication over time showing similar results), *intra- and inter-rater reliability* (i.e., consistent coding over time), and *measurement reliability* as a psychometric property (i.e., test-retest reliability).

*Reliability of results* is determined by the consistency of information from the school principals and senior leaders about the school's and teachers' use of national test results in Phase 1 and then again by information from the on-site observations and participant-conducted video recordings in Phase 2. Information about the school leaders from the interviews was used as part of the coding process in the analysis in Phase 2. The consistency of this information throughout the phases suggests that the results regarding the schools' use of national test results are reliable. Moreover, the multiphase design enabled reuse of the categories and coding from the interview protocol, observation protocol, and memos in the two first phases in Phase 3 when designing the questionnaire. The items in the questionnaire were constructed in order to collect data from teachers' stories and on-site observations in Phases 1 and 2. Using similar codes and categories made the comparison across the interviews, on-site observations, and survey data more reliable, which is important because

comparing data across time, situations, and perspectives is important in mixed methods studies (Creswell & Clark, 2011).

Conclusions and interpretations drawn from separate quantitative and qualitative strands and across those strands are regarded as “meta-inferences” (Teddlie & Tashakkori, 2009, p. 300). The codes and categories are presented in Appendices 9 and 10 and are represented in the items in the questionnaire presented in Appendix 9. Ary et al. (1990) state that observation of social phenomena, such as teacher meetings in schools, may produce changes that might not have occurred otherwise. A researcher may think that X causes Y, when in fact observation of X may cause Y. This is exemplified by the well-known Hawthorne effect, when workers’ productivity changed because they knew they were singled out for investigation (Ary et al., 1990). The investigator is a human being, and his or her presence in a situation may change the behavior of other human beings. Ary et al. (1990) suggest using hidden cameras and tape recorders to minimize this interaction in some cases. In this case, the two schools investigated in Phase 2 conducted the video recordings by themselves. This meant that the senior leaders were the one to push the “record” button before the start of every meeting. However, although the researcher was not physically present, knowing that the video recorder was running might have affected the school professionals’ actions. This could, in turn, have influenced the consistency of the results.

*Inter- and intra-rater reliability* was found in all the three phases. During analysis of the interviews, on-site observations, and video recordings, the data from Phases 1 and 2 were coded and reanalyzed three times as described in Articles I and II. Comparisons of the coding into categories indicated satisfactory connection. Article III reported analysis using SPSS. To ensure consistency in coding over time, all analyses were conducted several times by the researcher and the second author of Article III. *Measurement reliability* was found in the reliability measures of teachers’ responses in Article III. The survey was piloted with a representative group of lower secondary teachers and a group of teacher/lecturer colleagues. Their responses helped add distinctions to the questions and categories in the questionnaire and improved consistency in the survey. During analysis of the results, Cronbach’s alpha was used to validate the thematic areas and estimate consistency within each of them.

#### **4.6.2 Validity**

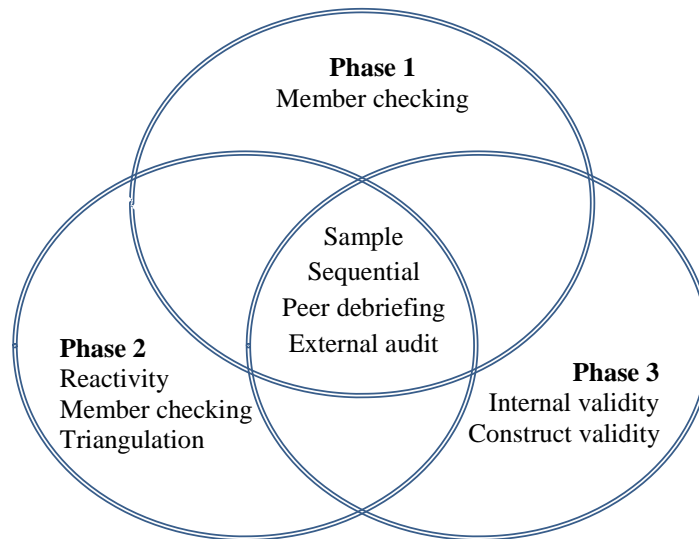
The mixed strategies for interviews, on-site observations, video recordings, and surveys might have reduced the risk that the conclusions would reflect the biases of only one method

and enabled a more secure understanding of the issue being investigated (Maxwell, 2012). Validity is, however, emphasized in different ways in qualitative and quantitative research. Equal emphasis is given to whether the interview as an example of a qualitative strategy and a questionnaire as a quantitative strategy measure what they are supposed to measure. Some variables that influence the validity of a questionnaire include how important the topic is for the respondents and whether the questionnaire protects respondents' anonymity (Ary et al., 1990). Quantitative research also considers issues of reliability, which means that participants' scores are consistent and stable over time. In qualitative research, there is more emphasis on validity than reliability to determine whether the accounts provided by the researcher and the participants can be trusted (Creswell & Clark, 2011; Lincoln & Guba, 1985).

Working with data concerning naturally occurring interactions, such as teacher team and leader team meetings, affords some advantages in regard to validity. Relying on video data and transcripts of this data can reduce the risk of individual bias in the analysis (Atkinson & Heritage, 1984). Other people can examine the episodes and judge the interpretations presented here. In other words, the video data makes visible the grounds for the analysis and the generated findings. Since only some episodes from the data corpus on the two cases in Phase 2 were selected, in order to give a detailed account, a major portion of the video data is not included in the article. This aspect of the detailed study that relies on transcripts of discursive talk can be described as a weakness. However, episodes were selected for their relation to particular themes that are relevant for illuminating discursive talk about national testing policies. The readers can make their own judgments about the trustworthiness of the interpretations of the selected themes and episodes presented here (Fairclough, 1992).

Validation is present in the entire research process of the MPM project. A variety of steps have been taken to reduce the threat of *multiple validities* (Johnson & Christensen, 2014). The first threat is the weakness of self-reported data. Since the interviews were all individually self-reported, a mixed methods approach was chosen to integrate several data sources across the three phases in an attempt to compensate for this limitation. The use of a mixed methods approach contributes to validity as each phase influences the design of the next (sequential validity) and multiple data sources are integrated throughout the phases (Creswell & Clark, 2011). Figure 4B illustrates this relationship, with the overlapping validity procedures placed in the center of the figure. In addition, sample integration, sequential validity, peer debriefing, and external auditing was performed in all three phases.

Additional validity procedures were addressed in different phases; member checking was performed in Phases 1 and 2, and internal validity and construct validity was measured in Phase 3 (Johnson & Christensen, 2014). I describe the processes for each phase in the next section. External validity is presented as a generalization in Section 4.5.3.



**Figure 4B.** Multiple validities in and integrated across each of the three phases

Phase I used *member checking* (Creswell & Clark, 2011) to assess the questions in the semi-structured interview guide by asking the school principals and senior leaders after the interviews if the questions were constructed adequately. The respondents confirmed that the questions were sufficient for investigating schools' work with national test results, as they were told before the interview sessions. About half a year after the interviews, the researcher visited one of the schools to comment on her immediate analysis of the school leaders' stories about how teachers made use of national test results, which was confirmed. Secondly, the researcher had the opportunity to visit one of the lower secondary schools to validate her analysis of the video recordings and on-site observations in Phase 2. The teachers, senior leaders, and principals were asked to comment on the preliminary analysis of the observed meetings.

*Reactivity* was evident once in Phase 1 and once in Phase 2. It was evident in Phase 1 during the school leader interviews and in Phase 2 during the on-site observations and video recording sequences. Participants that self-report their participation in interviews or conduct video recordings of themselves might be affected by the Hawthorne effect (Ary et al., 1990). In other words, school leaders and teachers may behave or perform better or differently than

when not interviewed, observed, or being recorded. However, during the meetings, interviews, and video sequences, it took only a few minutes before the school leaders and teachers seemed to forget that an observer or a camera was present. Particularly in the school leader interviews, it was possible for the responses to be compromised by the Hawthorne effect (i.e., whether interviewees adjusted their answers to match what they thought the researcher expected) and thus disrupt the validity of the data. However, in these interviews, the principals and senior leaders built on their individual self-confidence, professional experiences, and knowledge to confirm their understandings of work with national test results, student learning, and school quality. To strengthen the validity of the interviews, the transcripts of video recordings were combined with the interview transcripts and field notes from on-site observations. *Triangulation* occurred in Phase 2 during on-site observations and video recordings that occurred in the same school year with the same informants. There was no difference between the two different types of data (in situ and remote) obtained when the researcher was present and not present (Creswell & Clark, 2011).

*Internal validity* represents the extent to which the researcher can identify a cause and effect relationship among variables in the study. Correct cause and effect inferences can be made only if threats such as participant attrition, selection bias, and unfolding of participants are accounted for (Ary et al., 1990). The internal validity of the findings of the teacher survey in Phase 3 was measured. Teachers' responses to their own and the teacher teams' use of national test results, their attitudes towards the use of national test results, and their experiences with the principal's facilitation of work with test results were analyzed. Care was taken to control for a number of potential causes or explanations and ensure internal validity. First, for analysis of the results, SPSS and Cronbach's alpha were used to validate the thematic areas and estimate the consistency within each of them. Second, descriptive statistics were used to measure the distribution of responses for each item. Third, bivariate correlations were conducted to detect covariance between the teachers' experiences with their principals and the teachers' own practices and attitudes related to national tests. Finally, the schools under study were differentiated. The 27 items in the survey were validated as one construct in terms of reliability, with a Cronbach's alpha of 0.92. Second, each of the four thematic areas was validated by using Cronbach's alpha, which was 0.83 for principals' facilitation of work with national test results, 0.70 for teachers' individual use of national test results, 0.83 for the teacher team's use of national test results, and 0.71 for teachers' attitudes towards the use of national test results. Some absent variables, such as teachers' personal and psychological lives and situations, might represent important variables for reporting attitudes



and perceptions, which in turn might influence internal validity. Nonetheless, effort was made to ensure valid inferences based on findings, theory, and prior research.

*Construct validity* refers to the extent to which a higher-order construct, such as attitude, is accurately represented in the study (Johnson & Christensen, 2014; see also Ary et al., 1990). Abstract concepts like “teacher attitudes,” “teacher practices,” “principal facilitation” or “making use of national test results” are difficult to define accurately. Hence, construct validity was pursued, especially in the quantitative Phase 3, when the scale items were established based on the findings and themes of Phases 1 and 2 and when the questionnaire was piloted. For piloting, a panel of professors at the researcher’s university and former teacher colleagues were used to ensure the content validity of the survey items. The preliminary findings of Phases 1 and 2 in the thematic areas were used when creating and validating the items and the survey in Phase 3 (Ary et al., 1990).

*Sample integration* refers to the connections between samples in qualitative and quantitative research (Johnson & Christensen, 2014). Section 4.3 described how the teacher and leader teams in Phase 2 were sampled from the school leader sample and teacher sample in Phase 1. Combining the samples from Phase 1 and Phase 2 with all the teachers in Phase 3 increased the breadth and depth of the study. Breadth was increased by analyzing the 176 teachers’ responses regarding their practices and attitudes to work with national test results, and depth was increased by analyzing the video data and interviews with the school principals and senior leaders with regard to their attitudes and practices concerning work with national test results. Combining two qualitative teacher samples and one quantitative sample increased the depth of understanding of teacher practices and provided generalizable data concerning the responses of the population of 176 teachers in one municipality (Article III). Finally, the schools and teachers in Phases 1 and 2 were included in the sample in Phase 3.

The *sequential approach* and the sequential timing of the study improved the validity of the multi-phase design when the qualitative data was collected before quantitative data and when the three phases were integrated and built upon each other (i.e., an exploratory design) (Creswell & Clark, 2011). The ways in which Phase 1 influenced the design of Phase 2 and Phases 1 and 2 influenced each other and the design of Phase 3 are simplified in Table 4D.

**Table 4D.** *Validation of the sequential approach*

	<b>Phase 1</b>	<b>Phase 2</b>	<b>Phase 3</b>
<b>Credibility of mixed methods approach</b>	<p><i>Reliability:</i> Codes and categories of the findings in Phase 1 were used to analyze data in Phase 2.</p> <p><i>Sequential validity:</i> The design of Phase 1 influenced the design of Phase 2 as the self-reported data were combined with video data and on-site observations in Phase 1 and 2, as well as in the creation of the survey.</p>	<p><i>Reliability:</i> The codes and categories from Phase 1 were reused when analyzing the video recording data and field notes from on-site observations in Phase 2.</p> <p><i>Sequential validity:</i> The concept of leadership roles and perspectives on power found in Phase 1 were used in the analyses in Phase 2.</p> <p><i>Sample integration:</i> Participants were sampled from Phase 1. The perspectives of, for example, policy actors, policy roles, accountability, and power relations from Phase 1 were combined with the analyses in Phase 2 and Phase 3.</p>	<p><i>Sequential validity:</i> The findings from Phases 1 and 2 influenced the items and thematic areas involved in the planning, analyses, and contribution of Phase 3.</p> <p><i>Sample integration:</i> Schools and teachers in Phases 1 and 2 were included in Phase 3.</p>

The validity procedures described were intended to meet the *multiple validities* requirement for legitimization of mixed methods research.

### **4.6.3 Ethical considerations**

All the participants in this research gave their voluntary consent to participate after being informed that they could withdraw at any time (Appendices 3, 4 and 12). The phases of data collection and analyses were conducted in line with the ethical guidelines proposed by the Norwegian Social Science Data Services (NSD), which assessed and accepted the data collection situations that needed approval (Appendix 2).

In *Phase 1*, school leaders were asked to participate by e-mail (Appendix 12), and some were followed up with by telephone. This aligns with the ethical notion of avoiding gatekeepers (Ryen, 2011). All the participants were given written information about their right to withdraw at any time due to their right to freedom and self-determination (Busher & James, 2012). *Phase 2* revealed some advantages and disadvantages associated with using video recording as a data collection method. In the MPM project, the informants conducted video recordings themselves. Cameras were permanently stationed in the principals' office, where the leader team meetings were conducted, and in the 9th grade teacher room, where the teacher team meetings were held. The senior leaders and team leaders were responsible for pushing the record button before the start of every meeting. It is important to review how video recording and self-recording (as well as the method of recording) enhance the credibility of a study, record the environment in which the observation takes place, determine the data analysis that will be used, and affect the possible reasons for non-participation and

ethical issues. This also goes for on-site observations. As pointed out by Johnson and Christensen (2014), participants might perform unnaturally while being observed or video recorded. Additionally, even though the teachers, senior leaders, and school principals gave their voluntary consent, the use of video data raises critical questions about anonymity because of the way in which it displays the location, school, teachers, and students (Flewitt, 2006). During Phase 2, participants were assured that no clips would be used to present the findings and that the clips would be transcribed in order to analyze collaborative talk using a discourse-oriented approach.

In *Phase 3*, teachers were assured that the survey data would be treated anonymously. Their e-mail addresses were collected from the schools' official websites, and to further ensure their anonymity, the schools' names were coded as numbers. To ensure validity, the participants were given the opportunity to comment on the preliminary findings and the research process as a whole. According to Guba and Lincoln (1989), hiding the inquirer's intent can hinder uncovering of the findings. Relatively close personal interactions between researcher and informants might also produce problems regarding confidentiality and anonymity. When appropriate, drafts of the articles were presented to some informants. The researcher was invited to one of the schools to present the preliminary findings and went to one other school to present drafts of articles. At that point, I had written all three articles, with one published and the other two in review. The articles were displayed with parts of the transcripts and analysis so that the informants could better grasp what kind of research was being carried out. In addition, they could see the transcripts translated into English and obtain an idea of how their interview responses and collaborative talk would be presented in the articles.

## **4.7 Short summary**

This chapter has presented and discussed the key methodological considerations of this thesis, which is intended to be a reasonably reliable, valid, and ethically sound study (Johnson & Christensen, 2014). The following chapter presents and discusses the main findings and the study's research contribution.

## 5. Summary of the articles

The aim of the MPM project is to investigate how and why school professionals in a Norwegian social democratic context make use of national test results. This overarching aim was investigated through three sub-studies published in separate articles. The following will briefly summarize the articles, concentrating on the main findings presented in each. The findings will be further discussed in Chapter 6.

### 5.1 Article I

Gunnulfsen, A. E., & Møller, J. (2017). National testing: Gains or strains? School leaders' responses to policy demands. *Leadership and Policy in Schools, 16*(3), 455-474.

As the title suggests, this article addressed school leaders' responses to the school's use of national test results as an example of policy enactment within different systems of accountability. The aim of this article was to report a study conducted in a Norwegian context that is regarded as a part of a low-stakes accountability system. The following research questions guided the study: (a) *How do school leaders make sense of and respond to external demands related to a new policy context that emphasizes results based on national tests in reading and numeracy?* (b) *How do they use national test results in their local work with school and student development?* Data consisted of individual semi-structured interviews with four principals and eight senior leaders from four different schools in one municipality. Three themes emerged from the analysis of the interview data: (a) test data seemed to be working as a tool for symbolic responses, (b) the school leaders focused on low-performing students when talking about data use, and (c) internal accountability trumped external accountability. Based on the analysis, the main argument of this study is that the school principals and senior leaders were more concerned about teacher practices than report meetings at the superintendent level.

### 5.2 Article II

Gunnulfsen, A. E. (2017). School leaders' and teachers' work with national test results: Lost in translation? *Journal of Educational Change, 18*(4), 495-519.

The aim of this article was to explore how school professionals dealt with new policy expectations emphasizing schools' use of national test results. The data consisted of approximately 60 hours of participant-conducted video recordings and on-site observations of leadership and teacher team meetings. In combination with a discourse analytical approach, perspectives on policy enactment served as analytical concepts. The research questions were as follows: (a) *How do school leaders and teachers position themselves in relation to each other in their interpretation and translation of national testing policies?* (b) *In what ways does school leaders' and teachers' language reflect their enactment of national testing policies?* Three main discursive themes emerged: (a) school principals functioned mostly as narrators and enthusiasts, (b) senior leaders and teacher team leaders served as messengers and enforcers, and (c) teachers served as critics and preventers of "overburdening." Based on the analysis, it is argued that schools' local work with national test results can develop in different directions with school professionals holding different discursive roles.

### **5.3 Article III**

Gunnulfson, A. E., & Roe, A. (2017). Investigating teachers' and school principals' enactments of national testing policies: A Norwegian study *Journal of Educational Administration*. (Accepted for publication in *Journal of Educational Administration*).

The aim of this article was to investigate Norwegian lower secondary school teachers' use of and attitudes towards national test results and perceptions of the school principal's facilitation this work. The article addressed the following questions: (a) *How do teachers perceive principals' facilitation of work with national test results at the local school level?* (b) *How do teachers and teacher teams use the national test results, and what attitudes do teachers have towards national tests?* (c) *What is the relationship between principals' facilitation of work with national test results and teachers' practice and attitudes?* This study was based on survey data from lower secondary teachers ( $N=176$ ) in the municipality of Adde. The following main findings were illuminated: (a) teachers engaged in different practices and the principal had high confidence, (b) the school principal's "clearness" matters, and (c) national test results matter less at the individual teacher level. To a large extent, teachers and teacher teams use national test results differently. Based on the analysis, it is argued that teachers and teacher teams perceived a lack of information and understanding of how to use the results in their direct work to enhance individual student learning.

## 6. Discussion

To reiterate, the overall research question for the MPM project is as follows: *How and why do school principals and senior leaders in a Norwegian social democratic context make use of national test results?* In addition, four overarching research questions guided the project: (1) *How and why do school principals and senior leaders engage with policy initiatives related to national testing?* (2) *What are the constraining and enabling factors in local school work to deal with the policy intention of national testing?* (3) *How and why do teachers engage with policy regarding the use of national test results, how do they perceive the principal's facilitation of this work, and what are their attitudes towards this work?* (4) *What kind of educational leadership is being enacted in a social democratic school context when there is a top-down demand to raise test scores?* These four questions will be discussed in the following sections.

### 6.1 School professionals' engagement with policy initiatives

RQ: How and why do school principals and senior leaders engage with policy initiatives related to national testing?

All three articles of the MPM project provide insight into how school professionals engage with national test results. Article I addresses school principals' and senior leaders' reports of their own engagement with the results; Article II addresses engagement in collaborative talk between school principals, senior leaders, and teachers about this work; and Article III addresses school principals' engagement from teachers' perspective, focusing on how teachers perceive their school principal's facilitation of work with national test results. When seen together, the separate articles provide findings of similarities and variations to be further discussed. First, intentions and translations within micro policy making concerning national policy intention are addressed, and second, the symbols and tensions appearing in school professionals' work with the policy intention for national testing are discussed.

#### 6.1.1 Intentions and translations

Honig and Hatch (2004) state that when organizational actors use new information, they incorporate it or deliberately decide not to incorporate it into organizational rules or policy.

This process of interpretation involves deciding whether and how to incorporate the information into organizational policy (Weick, 1995). Interpretation is essential because, typically, numerous policy responses or non-responses may “fit” a given situation (Yanow, 1996). The findings of the MPM project illustrate that the national policy intention of large-scale student testing seems to be, more or less intentionally, lost in both interpretation and translation as the schools ensure fit to the given situation and prevent incorporation of the intention.

The most prominent and overarching finding regarding school principals’ and senior leaders’ policy engagement was their disbelief that the information from the national test results could be beneficial for developing teachers’ instructional practice. However, the survey results presented in Article III illuminated that school principals do facilitate work with national test results, even when they are doubtful of the usability of the results. The overall results illuminate the “double” policy intentions of national test results: that they should serve (1) as tools to improve school quality by gathering accumulated data at a system level and (2) as formative assessment tools to enhance individual student learning at the local level (Norwegian Directorate for Education and Training, 2011). These intentions were perceived differently by different school professionals. In particular, the second intention was not expressed by school professionals.

The identified differences and ambiguities are supported by Pettersson, Popkewitz, and Lindblad (2017), who described large-scale assessment-based activities (e.g., the use of national test results) as grey zone activities that exist between policy, research, and practice. Pettersson et al. (2017) state that the seemingly technical appearance of the statistics of international large-scale assessment creates a situation in which numbers become a codification and standardization of reality and planning. The apparent rigor and uniformity are transported across time and space so as to not require intimate knowledge and personal trust. The new construction of references to define “reality” can be linked to national testing policies in which numerical data provides more than an “objective way” of seeing reality, instead “instituting” reality by creating a “common cognitive space” that can be both observed and described through data (Borer & Lawn, 2013). This might explain why school professionals had difficulty recognizing their own engagement with national test results; the common cognitive space is enlarged by the national testing policy, resulting in a loss of intimate knowledge and personal trust. The findings of the MPM project hence align with the perspectives of “new cognitive spaces and realities” and “grey zone” activities because the most important intention of Norwegian national testing policy (i.e., to enhance student

learning) might be fulfilled by the talk about or effort to not talk about “large numbers.” Then, what is left might be pressure to “deliver” and perform (cf. Ball et al., 2012), which in turn could lead to a situation in which the school professionals’ micro policy talk about national test results mostly involves defending and crafting stories about the school’s “well-being” (Valli & Buese, 2007). The MPM project partly confirms these aspects when focusing on teachers’ perspective but goes further by arguing that school professionals’ engagement can be regarded as micro policy making for practices between and beyond national policy intention.

### **6.1.2 Symbolic responses and tensions**

The level of engagement with policy intentions was most prominent in the first sub-study (Article I), when the school principals and senior leaders were asked about the policy demand to report the results. The responses were more concerned with teacher practices than report meetings at the superintendent level. Hence, the overall findings suggested that the school leaders faced no difficulty navigating the tension between the aspects of external accountability and internal responsibility. Moreover, the key aspect of capacity building for individual student learning was decoupled from the policy demand to make use of national test results in this regard. Accountability for national test results did matter to them, but not in a direct way (e.g., making substantial changes in instructional practices). A possible explanation is that Norwegian school leaders do not yet run any risk with this approach, even though the national policy has changed to more strongly emphasize outcomes during the last ten years.

Additionally, findings in Article I and II revealed that school professionals had a key focus on low-performing students, which allowed them the opportunity to combine ongoing practices (intensive reading and numeracy courses for low-performing students) with testing policy engagement. This revealed that the responses to work with national test results seemed to be mostly symbolic and the attention paid to developing student learning and school quality were not directly coupled to responses to work with national test results. This finding aligns with the findings of Coburn (2004). However, in Article III there was a strong correlation between how teachers perceived the principal’s facilitation of work with national test results and basic skills and the teachers’ attitudes. This correlation suggests that the notion of making use of national test results was present in the school principals’ engagement. The school principals’ facilitative work and engagement was, to some extent,



confirmed in the stories and talk presented in the two first articles, but mostly in the ways that they made efforts to align the school's current practices with subject matter and basic skills (i.e., reading and numeracy).

All three articles confirmed the dominance of internal accountability aspects, especially in Article I, where internal responsibility trumped external accountability. The narratives that the school principals and senior leaders put forth in the interviews emphasized, criticized, and questioned the formative assessment intention of national testing. The reported narratives in Article I and collaborative talk in Article II were dominated by methods of enhancing student learning other than use of national test results. Overall, the most striking findings regarding tension were identified in the responses and observations regarding how senior leaders functioned as enforcers and messengers, while the principals' initiatives and engagement was more in the background. Unlike earlier studies (e.g., Coburn & Turner, 2011; Mausethagen, 2013; Møller & Skedsmo, 2013), the MPM project highlighted how tensions lie in the contrasts between the discursive roles of internal work with collective responsibilities, shared purpose, and values, and the external demand to improve the school's test results.

## **6.2 Micro policy making - constraining and enabling factors**

RQ: What are the constraining and enabling factors affecting local schools' work to meet the policy intention of national testing?

The findings reported in the three articles demonstrate differences in the schools' enactment of national testing policies. As established in Articles I and II, the interview responses and on-site observations and video recordings illuminated differences in power relations and in the values and internal beliefs represented in the discursive talk. The teachers' survey responses reported in Article III illustrated variances in the relationship between the principal's clear (or indirect) facilitation of the work with national test results and teachers' practices and attitudes. Hence, the findings illuminated dissimilarities in micro policy making both *between* the schools and *within* single schools in the municipality. These disparities are regarded as constraining and enabling factors, particularly those concerning the perspectives of values and power at play when dealing with national testing policies.

### **6.2.1 Values at play**

The overall findings show that micro policy talk of national testing among school leaders and teachers emphasized practicalities. The attention paid to how to arrange the tests represents a kind of surrender to the testing policy practice, but not to the testing policy intention. In such circumstances, it is more comfortable and less confrontational to focus on practicalities as a form of compromise. To balance comfortable practices, the emphasis on low-performing students in the talk about the use of test results seems to have a mitigating function, perhaps because the comprehensive education system in Norway is strongly rooted in ideologies and norms and emphasizes aspects of equity that are linked to social democratic values (Møller et al., 2007; Ottesen & Møller, 2016). By “gaming the system’ with symbolic responses and non-conforming practices, values can be regarded as both enabling and constraining factors. They are enabling as low-performing students are safeguarded but constraining due to the national policy intention of using the test results as tools to improve school quality and as formative assessment tools to enhance individual student learning.

Both national and local context is important for understanding why policies are dealt with differently in seemingly similar schools (Ball et al., 2012). Findings in the three articles revealed that schools differ in their stories and collaborative talk as much as in their engagement with local educational authorities. These alterations were supported in the survey in Article III, where variances between the schools relating to the thematic areas of principals’ facilitation and teachers’ attitudes and practices were significant. Yet, these results do not explain the cause for the differences, merely confirming that school context and leadership matter in the micro policy making of national testing policies, which again is confirmed by the results of Articles I and II.

Policy makers across the globe frequently understand large-scale testing as something that is intended to provide instructional diagnosis, to identify needs, and to ensure accountability (Abu-Alhija, 2007). This is also the case in Norway, where national and local policy makers justify national testing as a tool for improving schools. The findings of the three articles illuminated student assessment as an area of focus for school development determined at the superintendent level. However, none of the three articles revealed schools’ work with national test results as a tool in general work involving formative student assessment. An interpretation of this finding is that talk about student learning is focused on the value of internal responsibility and the school’s established practices of assessment for learning as part of daily student–teacher relations (shared beliefs played a vital role). Work

with national test results has hence become a parallel practice, disconnected from what appeared to be the schools' more relational practice for assessment of student learning. The question that must be asked is whether large-scale testing and formative student assessment are consistent values. As such, the findings of the MPM project add different perspectives with regard to the values at play.

According to Prøitz et al. (2017b), both national and international educational contexts seem to heavily involve data use. The MPM project also aligns with the notion that everything is becoming data, but it goes beyond earlier findings by applying the following perspectives: the value of *time* and the *use of time* in schools to develop student learning. As shown in Articles I and II, the value of time was brought to the forefront when the questions of how and why the schools worked with national test results were asked. Especially in Article II, the video recordings illuminated that the teachers served as defenders of time. The different discursive roles in the schools—the principals' roles as narrators and enthusiasts and teachers' role as defenders of time to do what they defined as “more important things” than work with national test results—were prominent and representative of the values at play with the actors. This discursive talk can be interpreted in the context of “micro power talk” or “micro policy talk,” and the perspective of time can be understood in the context of “micro policy time,” which can be related to theoretical perspectives on education policy, power relations, and teachers' work.

### **6.2.2 Power at play**

Central concerns of the MPM project are power and conflict, which underlie almost all aspects of studying values, interests, and change in organizational practices (Ball, 2012). All members of an organization can influence others using resources provided to them in their role; thus, school leadership is not necessarily synonymous with a position and may be associated with school principals, teachers, or others (Foster, 1986; Ogawa & Bossert, 1995). Hence, practices are permeated by power seen in a relational and situated way rather than as a top-down and linear phenomenon (Fairclough, 1992). The MPM project demonstrated how language and discursive talk were used and illuminated the power exercised through chains of events.

It is in the collaborative spaces of meetings with colleagues where school professionals make decisions about policy and how to engage with it. The “power talk” (defending or initiating talk) found in Article II and positioning identified in the

conversations occur with teachers, principals, and senior leaders. Yet, exercise of power based on the analysis of interviews reported in Article I appears to be more “powerful” in the school principal narratives than when observing talk in meetings between school principals, senior leaders, and teachers. In Article II, the exercise of power through school principal’s talk somehow transforms or transfers to the teachers with help from the senior leaders’ discursive role as messengers between the principal and teachers. For example, the issue of time use equally competed with the principals’ and teachers’ understanding of whether work with national test results was equivalent to other tasks on the agenda. The principal initiated the issue of time use, but the power of the initiative faded or transformed in the teacher team meetings with two short questions from one teacher: “Do we need to do this?” and “Who has decided it?” After this loss of power, the senior leaders had to leave the meeting with unfinished business and once again ask the principal what she had actually decided. This finding highlights that leadership as a relational process always involves some sort of influence and exercise of power (cf. Gronn, 2010; Liljenberg, 2015; Sørhaug, 1996). The MPM project has illuminated that power from below seems to be most prominent and that power and trust as relational processes are at play in schools’ micro policy making.

Ball et al. (2012) asserted that policies do not actually inform school actors about what to do; in fact, they barely dictate practice. The findings of the MPM project illustrate school professionals’ quest to determine the school’s data use and how the formal and informal power in play exist within this context. The notion of informal power is illustrated in Article II. The findings based on the survey data in Article III illuminate aspects of principals’ formal power because they directly present teachers’ perceptions of their principal. The quantitative results reported in Article III are not sufficient for making any claims about or analyzing the aspects of power relations or values in play. Nevertheless, these findings connect power perspectives to the theory of coherence, in which “coherence depends on the extent to which external demands fit a particular school’s culture, political interests, aspirations, conceptions of professionalism and on-going operations” (cf. Honig & Hatch, 2004, p. 18). Senior leaders seemed to be losing their power as formal leaders in their effort to contribute to the crafting of coherence, and the principals’ distinct initiatives and messages crumbled away both within and between the different arenas of official meetings. These observations highlight the importance of informal roles, discursive talk, and power relations. The MPM project enables questioning of new formal roles and leadership structures produced by “distributing leadership” in school organizations.

## 6.3 Teacher engagement and principal facilitation

RQ: How and why do teachers engage with policy regarding the use of national test results, how do they perceive the principal's facilitation of this work, and what are their attitudes towards this work?

This question is of particular interest when assuming that national testing and accountability practices exert pressure on schools (e.g., Ball et al., 2012; Hall et al., 2013; Hardy, 2014; Mintrop & Sunderman, 2009). As a result of this pressure, school principals might feel that they need to focus on national testing as a contribution to school development and student learning. In spite of the close connections between teacher engagement with national test results and principals' facilitative work, the MPM project also illuminated great differences within and between schools regarding this matter.

### 6.3.1 Teacher attitudes and engagement

Blase and Anderson (1995) defined the real world of schools as a political world of power and influence, bargaining, and negotiation. While micro politics mostly involve conflict and competition, it also involves cooperation and building of support to achieve goals. Hence, micro policy making in schools concerns attitudes and collaborative engagement towards a common purpose and crafting of policy coherence. Article II and III addressed the issues of collaboration, talk, and facilitation in the qualitative and quantitative data, whereas in Article I, the qualitative data addressed the school principals' and senior leaders' stories in a similar way.

The results of the MPM project indicated that when a principal less clearly facilitates use of national test results, the teachers report that the results have little or no value for improving individual students' learning, that the work needed to improve students' learning takes too much time away from what they define as more important tasks, and that they lack knowledge about how to use the results. In contrast, the more the principal serves as "a driving force," the less teachers have negative opinions about the national tests. However, a principal cannot establish a clear internal agenda alone, and a discretionary external policy will not in itself help a principal develop a clear focus. Hence, connecting the individual leadership perspective with Ball's (2012) micro-political perspective enables a more broad understanding of how school professionals make, more than enact, national policies locally.

The MPM project confirms that local schools' policy making seems to be strongly influenced by individual principals' actions (cf. Ball et al., 2012). Nevertheless, the findings of the MPM project go beyond former studies, showing how the senior leaders needed to serve as messengers in order to enforce the principals' intentions and demands.

The results of Article III illuminated a contradiction in the responses: more than 54.5 % of teachers reported that they recognized national test results as useful tools, but about half of the teachers also reported that they lacked knowledge about how to use the reading results (48 %) and numeracy results (40 %) in their subject(s). This may mean that they acknowledged that the results may be useful (*why*) but not *how*. Hence, these findings, which represent the lack of quality of quantitative data (national test results), defy the underlying (policy) assumption that such data enable teachers to better target their teaching to improve pupils' learning (Wayman & Jimerson, 2014). The MPM project presents important perspectives on teachers' attitudes and engagement and how power and control exist within the school context. The project also illuminates how new leadership perspectives and discursive role emerge and how the teachers deal with national testing policies as part of practices of "crafting the school's wellbeing."

Moreover, Articles I and II shed light on school principals' practices when initiating or instructing teachers to make use of test results in the context of their general assessment practice. But the findings also showed that the teachers refused to do so due to a lack of time, which revealed internal decoupling between the school principal and teachers. The quantitative data revealed that 11.3 % of teachers reported that they did not take responsibility for purposeful monitoring of students' results and 13.7 % reported that they did not take responsibility for following up on the school's national test results. However, none of the teachers reported that they *did not at all* work with national test results. In this regard, the MPM project has illuminated how teacher engagement and attitudes are significant factors affecting how schools deal with national testing policies.

### **6.3.2 The distinct principal**

In this study, I define a facilitative principal as a principal that discursively invests in work with policies. In addition, such a principal engages with direct initiatives and outspoken demands. The results reported in Article I and II revealed that the principals' ways of facilitating work with national test results were manifold. The third study addressed this variety through teachers' perceptions of their principals. More than 90% of the teachers self-

reported that they found their principals to be confident in their work with basic skills. However, this confidence might be due to indifference or negative attitudes. The MPM project illuminates how school principals, senior leaders, and teachers influence each other's work with new national policy expectations and answers Ball et al.'s (2012) call to take local contexts seriously in an effort to understand why policies are enacted differently in seemingly similar schools. The MPM project confirms that, too often, research on policy enactment has taken for granted the meaning of policy itself, but policy texts must be translated to action, put into practice, and seen in relation to history and context (Ball et al., 2012).

When school professionals read policy documents on national testing, they learn *what*, *why*, and *when* new policies are meant to take place, but these texts rarely tell school leaders or teachers exactly *how to* work with national test results. Interpretation is the initial reading of policy texts, while translation is closer to the language of practice. Ball (1994) defined policy discourses as “ways of talking about and conceptualizing policy” (p. 109). The discursive role of the principal as an enthusiast when talking about and conceptualizing policy can hence be interpreted as synonymous with the principal's distinct facilitative role. In Article II, the findings illustrated different approaches to being distinct and facilitative. As such, the MPM project adds to the knowledge about conceptualizing policy by showing that senior leaders' roles as messengers and enforcers appeared to be similar even though their principals have different ways of being distinct and facilitative. As reported in Article II, the principals led the conversations in leader team meetings and instructed the senior leaders to focus on certain topics. As such, the principals held the discursive roles of enthusiast and narrator, but the responsibility to inform the teachers was given to the senior leaders.

The overall findings show that the correlation between principals perceived as facilitative with clear expectations and dialogue and teachers engaged in systematic work with the test results was high. In contrast, if the principal is not clear about this work, the correlation to indifferent practice and attitudes towards work with national test results was high. Another example of enthusiastic and distinct talk can be found in the school leaders' stories of data use, which may be described as strategic behavior, or “window-dressing,” to comply with the superintendents' external demands. It may also be interpreted as a parallel approach in which school leaders try to balance different priorities, such as preexisting student assessment practice and the “new” practice of using national test results to enhance student learning. Intensive reading courses for low-performing students is one example of a preexisting practice that was connected to use of national test results as a “new” practice. The distinctiveness of a facilitating principal might hence be recognized within a leadership

practice that consciously attends to the distinction between “large talk about small numbers” (internal responsibility) and “small talk about the large numbers” (external accountability). In both, numbers constituted a resource through which surveillance and development could be exercised (cf. Ozga, 2008). As such, the findings regarding surveillance go beyond Ozga’s (2008) findings.

## 6.4 Leadership in a social democratic testing context

What kind of educational leadership is being enacted in a social democratic school context when there is a top-down demand to raise test scores?

According to Ball (2008), the era of new governance does not simply represent party politics or ideology but also is part of the broad global shift in public service discourses regarding language, ideas, organizations, and practices. Perspectives on this global shift are vital for investigating micro policy making in schools. The global shift represents a top-down or “outside-in” context and is likely to influence leadership. So far, literature on governing through data use has been situated within the English educational context, which features pressure on schools, school leaders, and teachers (Ozga, 2008). This differs from the Norwegian context, where professional discretion is still strong. However, schools cannot choose to opt out of the tests. Within this context of top-down demands, school professionals need to navigate their practice-related priorities during micro policy making in schools.

### 6.4.1 Leadership and performativity

In the MPM project, the policy intention for use of national test results has been viewed as part of a system of control and performance (e.g., Elstad, 2009; Hopmann, 2008). According to Ball (2008), performativity is “a culture or a system of accountability that employs judgements, comparisons and displays as means of control, attribution and change” (p. 57). In the MPM project, the findings regarding “policy talk” and “power talk” were analyzed in light of the new language being embraced in education that may be eroding a broader long-term discussion about the core purpose of education (Møller, 2017). One of the main tensions seems to be between discourses about competition that underpin new public management on the one hand and discourses rooted in social democratic ideologies linked to notions of equity, participation, and comprehensive education on the other hand (Skedsmo & Møller,



2013). The overall findings of Articles I and II showed that the school principals, senior leaders, and teachers did not make use of “performative” language in their narratives and talk about national testing policies or comply with the notion of competition regarding the use of national test results. The results in Article III showed that, overall, the teachers conformed with the conducting of the tests, but work with the results to enhance individual student learning was inconsistent between schools in the municipality. Hence, this finding may imply that the school professionals responded to the national testing policy demands in ways that aligned with their own values and beliefs and current work with school quality, teacher practices, and student learning.

These findings can be connected to Ball’s (2008) perspectives on a performativity system in which one feels constantly judged in different ways, by different means, and according to different criteria and in which the performances of individuals or organizations serve as measures of productivity or output or as displays of “quality.” The issue of who controls the field of judgement and what is judged is crucial. Yet, even when school professionals find themselves in a constant state of competition during micro policy making, four out of five teachers responded either neutrally or positively to the following statement: “Publication of the national test results has no value for improving the quality of our school.” This may indicate that publication of and competition regarding test results is not a big concern for teachers, who would rather focus on the school’s internal responsibility for improving school quality and individual student learning. This does not mean that the school principals, senior leaders, and teachers in the study were not interested in discussing quality; rather, they talked about it in other ways. However, these findings do not explain whether this “non-conforming” performativity and quality practice can be linked to the Norwegian social democratic education context.

#### **6.4.2 Leadership and the micro politics of leading**

In the interviews, the principals seemed to be more concerned with reporting the school’s results in result meetings with the superintendent than clearly responding to the policy demands in the school. These double discursive roles as both a “foreign minister” and a “domestic minister” challenge the notion of crafting policy coherence as the two roles represent different levels of coherence: the macro and meso policy making levels and the micro policy enactment level. Micro policy work to craft coherence may be a symbolic response, but it is still a response to the national policy intention. The school professionals’

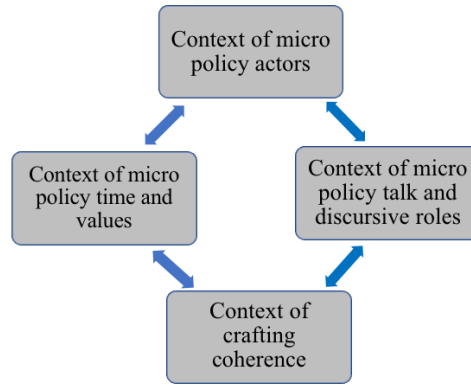
work to craft coherence can hence be regarded as part of “playing the game” and of gaming the system.

The analysis of the results reported in Article III can be easily considered a celebration of principals’ sustained hard work, and a strong correlation between principals’ facilitation and teachers’ attitudes and practices can be found. The MPM project may be criticized for focusing too much on personal capacities and hence complicating the notion that the school principal’s work is embedded in wider social structures of power. There is no doubt that policy makers add unnecessary pressure to the roles of school principals, but it is also known that the school principal’s success depends to a large degree on his or her relationship with teachers as well as parents and superintendents (cf. Møller, 2017).

In the MPM project, leadership and the micro politics of leading are illuminated by investigation of how and why school principals, senior leaders, and teachers deal with national testing policies, which can be interpreted as skills in “big” and “small” talk about “large” and “small” numbers. Small numbers can be described as the uncountable results related to school quality and student learning. Within the frame of the micro politics of leading, school professionals can be regarded as holding the discursive role of protector of the school’s “immune system”; they are trying to avoid taking too much “medicine” (not complying with policy intentions) after the “diagnosis” (test results) is given. Schools seem to stay “healthy” due to their resilience and internal work. Hence, school principals’ practices are cumulative parts of the educational political world, and principals need to be aware of the contexts of micro policy making, leading of politics, and micro politics of leading.

## **6.5 Short summary**

A summary of the analysis can be shown as a conceptual model of the contexts of micro policy making in schools (Figure 6A).



**Figure 6A.** *Contexts of micro policy making in schools in the MPM project*

Figure 6A shows how the outputs of the MPM project identify the contexts of micro policy making within a continuous cycle of how *micro policy actors*, *micro policy talk*, *micro policy roles*, *micro policy time*, and *micro policy values* interact in a process. Figure 6A hence represents the components of micro policy making in schools, in distinct opposition to micro policy taking of central, external policy demands.

## **7. Conclusion**

The main aim of the MPM project has been to investigate the micro policy making of school professionals in lower secondary schools in Norway, emphasizing new policy expectations for national testing. The main research question is as follows: How and why do school professionals in a Norwegian social democratic context make use of national test results? The findings of the articles and the discussion in the previous chapter contributed to increased understanding of micro policy making in lower secondary schools in the Norwegian context. Specific attention has been paid to how and why testing policies are dealt with in lower secondary schools. The mixed methods approach and empirical and theoretical triangulation have been useful for pursuing the research questions posed for the study. This final chapter offers a brief conclusion regarding the main contributions, limitations, and implications for future research.

### **7.1 Contributions**

The following summarizes the main contributions of the MPM project.

#### **7.1.1 Theoretical and conceptual contributions**

The MPM project contributes theoretically to policy scholarship by developing a model of micro policy making in schools with different contexts of micro policy time, values, actors, coherence, talk, and roles. (Figure 6A). School professionals' work with national test results and school improvement is closely connected to internal responsibility. Rewards and sanctions matter less than internal individual and collective expectations and values, and principals were widely considered to hold the discursive roles of narrator and enthusiast to encourage productive processes. Senior leaders were "squeezed in the middle," with some prominent teachers acting as critics of the new policy. Lack of time was a recurring theme.

In summary, this study contributes knowledge about how micro policy making may take unpredictable forms. The outcome—the finding that national test results are viewed as a tool for improving schools and student learning—does not necessarily align with local and national authorities' expectations. A strong correlation between how the principals facilitated work with national test results and the teachers' practices and attitudes was found. Another important finding is the variety of responses to national testing by the schools in the municipality. The project contributes to an understanding of the "inside-out" (cf. Coburn, 2004) and collective approach to crafting coherence, but it also strengthens the role of

individual principals, building upon other studies (cf. Anderson et al., 2010; Marsh et al., 2015; Park & Datnow, 2009).

Overall, this study shows that there is a need for cross-level dialogue between national and local policy makers, including local schools, about the long-term effects of national testing policies in order to ensure purposeful and appropriate education for all children rooted in social democratic norms and values of trust and equity. Cross-level dialogue about testing policies and large-scale assessment also involves supporting school professionals' obligations to the purpose of a school within a policy environment that increasingly requires performative accountability.

### **7.1.2 Methodological and empirical contributions**

In the MPM project, multiple ways to seek knowledge about the research questions and overall purpose of education have been used to engage with and embrace different types of information about how and why school principals, senior leaders, and teachers work with national test results. Multiple lenses have allowed for focus on research and theories that situate national policy intentions and schools' micro policy making as well as enabled analysis of power structures in order to address issues related to the core purpose of school. Greene's (2007) dialectic stance and her understanding that this stance is a sociopolitical argument as much as it is a methodological issue have been vital. The dialectic stance allowed the project to be designed in phases, with each phase influencing the next and the interrelation between school principals, senior leaders, and teachers identified within and across the schools in the municipality.

Including fewer dimensions in the research design would not have provided the same insight into the material, which is crucial not only for identifying the perspectives of school principals and senior leaders but also examining teachers' own perceptions of the school principals' as well as their own practices regarding data use. Research on data use in schools often uses fewer dimensions and typically focuses on principals, senior leaders, or teachers. Therefore, the main methodological contribution of this thesis is its mixed methods approach and identification of school principals', senior leaders' and teachers' perspectives on work with national test results.

Empirically, this thesis contributes new knowledge about micro policy work in local schools, including negotiation, "micro policy time," and "micro policy talk" about national policies in Norway. This new knowledge is of special concern since national testing is a

relatively new policy demand and the Norwegian social democratic education context is unlike education contexts in, for example, the U.S. or England. In addition, the findings are vital for developing a greater understanding of micro policy making concerning national testing for improving school quality and individual student learning, as specified in the national policy intention.

Why the principals acted and talked differently and why the teachers re-acted differently should be examined in further research. In addition, future research must identify and understand how the contexts of competition and performativity are dealt with at the policy maker level both locally, in municipalities, and nationally.

The findings of the MPM project have also revealed inconsistencies. The empirical contribution concerning the principal's discursive role and facilitating practices as a formal school leader seems to depend on individual (hero or villain) perspectives as well as collective perspectives with regard to the principal's relationship with teachers. Hence, the contribution relates to the complex perspective of leadership as not being either-or, but both a collective and individual practice. This project has allowed for critical analysis of empirical evidence within the wider power structure of micro policy making in schools, although the analysis is limited to a sample of principals, senior leaders, and teachers in one municipality.

## **7.2 Implications**

In the MPM project, it has been crucial to describe school professionals' policy making concerning national testing. Issues regarding educational policies and data use are often ideologically laden, and a descriptive view is necessary in terms of the empirical contribution. The project was intended to describe school professionals' micro policy making and not take an evaluative view in terms of whether Norwegian national testing policies would weaken or strengthen work to improve school quality and student learning. By taking a more descriptive view, the project explicitly aimed to "open up the black box" of school principals', senior leaders', and teachers' practices of micro policy making. Hence, the framework serves as a starting point for the first-order constructs (thematic areas) of the actors rather than "top-down" analysis of theory-defined concepts. Although the findings of the empirical study are only analytically generalizable (Kvale & Brinkmann, 2009), new insights from the study can be followed by suggestions for what could be improved or further considered by different actors, like many discourse analyses (Winther Jørgensen & Philips, 1999). Given that the MPM project addresses a topic that is highly relevant to both politicians

and professionals in Norway and elsewhere, it is essential to elucidate the implications of the study.

### **7.2.1 Implications for school professionals**

There is a need for school principals and school professionals to identify and create spaces and time for *micro policy talk* regarding national policy intentions. There is also a need for school professionals to recognize the increased demand for authorities to be responsible for providing high-quality education. Yet, the different ways of doing so should be publicly debated and contested. The findings reported in Article III and the high correlation between principals' facilitation and teachers' practice and attitudes might indicate that principals need to provide opportunities for senior leaders and teachers to collaboratively discuss school principals' and teachers' questions. Moreover, as indicated in Article II, senior leaders' roles as messengers and enforcers of policy intentions and the teachers' roles as critics and preventers of overburdening need to be openly and collaboratively discussed in light of the principal's role as narrator, enthusiast, and "foreign minister." Due to the ways in which different roles and power relations affect the school's micro policy making, it may be beneficial for every actor holding a role to be present in these discussions, bearing in mind the importance of democracy in schools and the core purpose of education. How the teachers perceive their principal and the senior leaders facilitating this work is also an issue to be illuminated in further discussions in order to reach a collaboratively unified understanding of the intentions, challenges, and limitations of educational policies. This also includes the necessity of discussing the "why" (or "why not") of large-scale assessment policies, not only the "how," "what," and "when."

### **7.2.2 Implications for policy makers**

The variety of micro policy making that occurs in schools provides knowledge about how policies can develop in directions other than those intended. The findings contrast policy makers' relatively strong belief in "governing by numbers." The findings are imperative for policy makers in terms of how, to what extent, and why/why not school professionals relate to accountability policies and external control. Based on the findings of this study, it can be argued that policy makers need to be aware of incoherence in the field in terms of values and beliefs regarding the core purpose and broader aims of education as well as the issue of increased attention to testing and performance. For example, the analyses have shown that

national testing is still a partly external issue for school professionals, even though the tests are integrated into school professionals' practices and partly accepted. Moreover, this study has confirmed the predominant conceptions that national student test results are not particularly useful for guiding school improvement or instruction at the classroom level.

This is not to say that the school professionals in this study did not “respond” to these incentives, because they did, the pressure connected to accountability may not be a useful response from a policy perspective, or the symbolic responses within the internal school perspective were entirely immune to the influence of external testing policy. Rather, the practice of conducting intensive reading and numeracy courses was both reconstructed and inspired by national test results, and it contributed to renewed commitment to previous practices. Yet, if large-scale assessment and national testing are not acknowledged by local school actors as tools for improving school quality and student learning, these are implications that politicians need to know. If use of large-scale tests to measure student learning outcomes increases, and if it is attached to risks for teachers and schools, politicians need to be aware of how adverse results can be achieved. Currently, this is not the situation in Norway, but nonetheless, it is important to be aware of these issues in order to develop a better “policy for practice” to ensure good education for all students. Well-informed policy expectations might encourage school professionals' micro policy making regarding the purposes and local practices concerning educational policies.

### **7.3 Limitations and need for further research**

Before concluding this thesis, possibilities for future research need to be addressed. However, to further develop this area of research, it is necessary to look back at the choices that were (and were not) made to improve the understanding of micro policy making in schools. To a certain extent, all three sub-studies have methodological shortcomings, and broadly put, there are limitations of both the theoretical and methodological choices that have been made throughout the research process. In the next section, emphasis is placed on the latter as the choices made regarding the theoretical and methodological frameworks are discussed in the previous sections.

Even though the study included on-site observations and video recordings of meeting practices to understand different aspects of translation and negotiation of policy expectations, more comprehensive observations of practices in other situations in local schools are needed, including situations in which students and parents are micro policy actors. This would enrich



the current research base on micro policy making in schools. Another issue to be addressed in further research is the extent to which the findings of this study are specific to the social democratic education context. For instance, studies of countries with stronger external accountability policies suggest that schools comply with and conform to increased external control of their work (Coburn, 2004; Hall et al., 2013; Hardy, 2014). Therefore, future studies should employ a strong comparative design and transnational analysis to draw attention to how crafting of coherence and micro-policy enactment play out in school leadership and teaching in local school contexts and diverse accountability contexts.

Moreover, it is worth further investigating the sense of control over student learning and school quality gained by use of large-scale test data, which local policy makers argue they need. Future research might seek to discover how and why micro policy making occurs, especially with regard to local and national ideas about accountability. The dealings, negotiations, power, and sense-making that take place between actors at different levels are important for developing a greater understanding of multifaceted and contested ideas about external and internal accountability. All in all, at least three more questions must be answered: What kind of (big) data might provide the knowledge that *schools* need in order to address relevant improvements in the quality of education? How, if at all, can large-scale data be used in schools? And for what purpose? These and other questions concerning micro policy making of education policies should continue to be answered by research using different (comparative) designs to enhance knowledge and understanding of these practices globally and locally.

# References

- Abrams, L., Varier, D., & Jackson, L. (2016). Unpacking instructional alignment: The influence of teachers' use of assessment data on instruction. *Perspectives in Education*, 34(4), 15-28.
- Abu-Alhija, F. N. (2007). Large-scale testing: Benefits and pitfalls. *Studies in Educational Evaluation*, 33(1), 50-68.
- Achinstein, P. (1968). *Concepts of science. A philosophical analysis*. Baltimore, MD: Johns Hopkins Press.
- Anderson, S., Leithwood, K., & Strauss, T. (2010). Leading data use in schools: Organizational conditions and practices at the school and district levels. *Leadership and Policy in Schools*, 9(3), 292-327.
- Ary, D., Jacobs, L. C., Sorenson, C., & Razavieh, A. (1990). *Introduction to research in education* (4th ed.). Orlando: Holt, Rinehart, and Winston Inc.
- Atkinson, J. M., & Heritage, J. (1984). *Structures of social action*. Cambridge: Cambridge University Press.
- Bachmann, K., Sivesind, K., Afsar, A., & Hopmann, S. (2004). *Hvordan formidles læreplanen? En komparativ evaluering av læreplanbaserte virkemidler—deres utforming, konsistens og betydning for læreres praksis*. [How is the curriculum communicated? A comparative study of curriculum-based tools – their design, consistency and significance for teacher practice]. Kristiansand: Høgskolen i Agder.
- Baker, D., & LeTendre, G. K. (2005). *National differences, global similarities: World culture and the future of schooling*. Redwood City: Stanford University Press.
- Ball, S. J. (1993). Education policy, power relations and teachers' work. *British Journal of Educational Studies*, 41(2), 106-121,
- Ball, S. J. (1994). *Education reform: A critical and post-structural approach*. Buckingham: Open University Press.
- Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, 18(2), 215-228.
- Ball, S. J. (2008). *The education debate*. Bristol: Policy Press.
- Ball, S. J. (2012). *The micro-politics of the school: Towards a theory of school organization*. London: Routledge.
- Ball, S. J., Maguire, M., & Braun, A. (2012). *How schools do policy: Policy enactments in secondary schools*. London: Routledge.

- Berge, K. L. (2005). Skrivning som grunnleggende ferdighet og som nasjonal prøve-ideologi og strategier. [Writing as a basic skill and national test-ideology and strategies]. In A. J. Aasen. & S. Nome (red.) *Det nye norskfaget [The new subject Norwegian]*, (pp. 161-188.) Bergen: Fagbokforlaget.
- Bergesen, H. O. (2006). *Kampen om kunnskapsskolen*. [The fight for a knowledge-based school]. Oslo: Universitetsforlaget.
- Berryhill, J., Linney, J. A., & Fromewick, J. (2009). The effects of education accountability on teachers: Are policies too stress provoking for their own good? *International Journal of Education Policy and Leadership*, 4(5), 1-14.
- Bickman, L. (1987). The functions of program theory. *New Directions for Evaluation*, 1987(33), 5-18.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8-21.
- Borer, V. L., & Lawn, M. (2013). Governing education systems by shaping data: From the past to the present, from national to international perspectives. *European Educational Research Journal*, 12(1), 48-52.
- Bowe, R., Ball, S. J., & Gold, A. (1992). *Reforming education and changing schools: Case studies in policy sociology* (Vol. 10). London: Routledge.
- Brinkmann, S., & Kvale, S. (2014). *InterViews - Learning the craft of qualitative research interviewing* (3rd ed.). Thousand Oaks: SAGE Publications.
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of Assessment in Education. *Assessment in Education: Principles, Policy & Practice*, 11(1), 7-26.
- Christensen, T., & Læg Reid, P. (2004). Governmental autonomisation and control: The Norwegian way. *Public Administration and Development*, 24(2), 129-135.
- Coburn, C. E. (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.
- Coburn, C. E., & Turner, E. O. (2011). The practice of data use: An introduction. *American Journal of Education*, 118(2), 99-111.
- Collins, S., Reiss, M., & Stobart, G. (2010). What happens when high-stakes testing stops?

- Teachers' perceptions of the impact of compulsory national testing in science of 11-year-olds in England and its abolition in Wales. *Assessment in Education: Principles, Policy & Practice*, 17(3), 273-286.
- Creswell, J. W. (2013). *Qualitative inquiry & research design. Choosing among five approaches* (3rd ed.). Los Angeles: SAGE Publications.
- Creswell, J. W., & Clark, V. L. P. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks: SAGE Publications.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy & Practice*, 3(3), 265-286.
- Datnow, A., & Hubbard, L. (2015). Teachers' use of data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record*, 117(4), 1-26.
- Earl, L., & Katz, S. (2002). Leading schools in a data-rich world. In K. A. Leithwood & P. Hallinger (Eds.), *Second international handbook of educational leadership and administration* (Vol. 8, pp. 1003-1022). New York: Springer Science & Business Media.
- Elmore, R. F. (2005). Accountable leadership. *The Educational Forum*, 69(2), 134-142.
- Elstad, E. (2009). Schools which are named, shamed and blamed by the media: School accountability in Norway. *Educational Assessment, Evaluation and Accountability*, 21(2), 173-189.
- Elstad, E., & Sivesind, K. (Eds.). (2010). *PISA: Sannheten om skolen?* [PISA: The Truth About School?]. Oslo: Universitetsforlaget.
- English, F. W. (Ed.). (2011). *The Sage handbook of educational leadership: Advances in theory, research, and practice*. London: SAGE Publications.
- Fairclough, N. (1992). *Discourse and social change*. Cambridge: Polity Press.
- Firestone, W. A., & Rosenblum, S. (1988). Building commitment in urban high schools. *Educational Evaluation and Policy Analysis*, 10(4), 285-299.
- Flewitt, R. (2006). Using video to investigate preschool classroom interaction: Education research assumptions and methodological practices. *Visual Communication*, 5(1), 25-50.
- Foster, W. (1986). *Paradigms and promises: New approaches to educational administration*. Buffalo, NY: Prometheus Books.

- Gale, T. (1999). Policy trajectories: Treading the discursive path of policy analysis. *Discourse: Studies in the Cultural Politics of Education*, 20(3), 393–407.
- Gee, J. P. (2014). *An introduction to discourse analysis: Theory and method*. London: Routledge.
- Gee, J. P., Hull, G., & Lankshear, C. (1996). *The new work order: Behind the language of new capitalism*. Boulder, CO: Westview Press.
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91-108.
- Greene, J. C. (2007). *Mixed methods in social inquiry* (Vol. 9). New York: John Wiley & Sons.
- Grek, S. (2009). Governing by numbers: The PISA “effect” in Europe. *Journal of Education Policy*, 24(1), 23-37.
- Gronn, P. (2010). Leadership: Its genealogy, configuration and trajectory. *Journal of Educational Administration and History*, 42(4), 405-435.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. London: SAGE Publications.
- Gunter, H. (2016). Intellectual histories of school leadership: Implications for professional preparation. *Acta Didactica Norge*, 10(4), 27-47.
- Hall, D., Gunter, H., & Bragg, J. (2013). Leadership, new public management and the re-modelling and regulation of teacher identities. *International Journal of Leadership in Education*, 16(2), 173-190.
- Hardy, I. (2014). A logic of appropriation: Enacting national testing (NAPLAN) in Australia. *Journal of Education Policy*, 29(1), 1-18.
- Hargreaves, A., & Shirley, D. (2008). Beyond standardization: Powerful new principles for improvement. *Phi Delta Kappan*, 90(2), 135-143.
- Harris, L. R., & Brown, G. T. (2009). The complexity of teachers’ conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, 16(3), 365-381.
- Henry, M., Lingard, R., Rizvi, F., & Taylor, S. (2001). *The OECD, globalisation and education policy*. West Yorkshire: Emerald Group Publishing Limited.
- Hertzberg, F., & Roe, A. (2016). Writing in the content areas: A Norwegian case study. *Reading and Writing*, 29(3), 555-576.
- Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher*, 33(8), 16-30.

- Hopmann, S. T. (2003). On the evaluation of curriculum reforms. *Journal of Curriculum Studies*, 35(4), 459-478.
- Hopmann, S. T. (2008). No child, no school, no state left behind: Schooling in the age of accountability. *Journal of Curriculum Studies*, 40(4), 417-456.
- Johnson, R. B., & Christensen, L. (2014). *Educational research: Quantitative, qualitative, and mixed approaches* (6th ed.). Thousand Oaks: SAGE Publications.
- Johnson, R. B., & Turner, L. (2003). Data collection strategies in mixed methods research. In A. Tashakkory & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral sciences* (1st ed., pp. 297-319). Thousand Oaks: SAGE Publications.
- Jørgensen, M. W., & Phillips, L. (1999). *Diskursanalyse som teori og metode*. Frederiksberg: Forlaget Samfundslitteratur.
- Karlsen, G. E. (2000). Decentralized centralism: Framework for a better understanding of governance in the field of education. *Journal of Education Policy*, 15(5), 525-538.
- Karseth, N., Møller, J., & Aasen, P. (2013). *Reformtakter. Om fornyelse og stabilitet i grunnsopplæringen*. [Reform beats. About renewal and stability in primary education]. Oslo: Universitetsforlaget.
- Karseth, B., & Møller, J. (2014). «Hit eit steg og dit eit steg»-Et institusjonelt blikk på reformarbeid i skolen. [“One step here and one step there” – An institutional view on work with reforming school]. *Norsk pedagogisk tidsskrift [Norwegian Pedagogical Journal]*, 98(06), 452-468.
- Karseth, B., & Sivesind, K. (2010). Conceptualising curriculum knowledge within and beyond the national context. *European Journal of Education*, 45(1), 103-120.
- Kumar, R. (2011). *Research methodology* (3rd ed.). London: SAGE Publications.
- Kvale, S., & Brinkmann S. (2009). *Interviews. Learning the craft of qualitative research interviewing*. London: SAGE Publications.
- Lachat, M. A., & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed at Risk*, 10(3), 333-349.
- Leithwood, K. A., & Hallinger, P. (Eds.). (2002). *Second international handbook of educational leadership and administration* (Vol. 8). Dordrecht: Springer Science & Business Media.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry* (Vol. 75). London: SAGE Publications.
- Maroy, C. (2012). Towards post bureaucratic modes of governance: A European perspective. In G. Steiner-Khamsi & F. Waldow (Eds.), *World yearbook of education*

2012. *Policy borrowing and lending in education* (pp. 62-79). London & New York: Routledge.
- Maroy, C., & Voisin, A. (2015). Comparing accountability policy tools and rationales: various ways, various effects? In H.-G. Kotthoff & E. Klerides (Eds.), *Governing educational spaces: Knowledge, teaching and learning in transition* (pp. 35-58). Rotterdam: Sense Publishers.
- Marsh, J. A., Bertrand, M., & Huguet, A. (2015). Using data to alter instructional practice. *Teachers College Record*, 117(4), 1-40.
- Mausethagen, S. (2013). Talking about the test. Boundary work in primary school teachers' interactions around national testing of student performance. *Teaching and Teacher Education*, 36, 132-142.
- Maxwell, J. A. (2006). Literature reviews of, and for, educational research: A commentary on Boote and Beile's "Scholars Before Researchers." *Educational Researcher*, 35(9), 28-31.
- Maxwell, J. A. (2012). *Qualitative research design: An interactive approach* (Vol. 41). London: SAGE Publications.
- Mintrop, H., & Sunderman, G. L. (2009). Predictable failure of federal sanctions-driven accountability for school improvement—and why we may retain it anyway. *Educational Researcher*, 38(5), 353-364.
- Møller, J. (2009). School leadership in an age of accountability: Tensions between managerial and professional accountability. *Journal of Educational Change*, 10(1), 37-46.
- Møller, J. (2017). Leading education beyond what works. *European Educational Research Journal*, 16(4), 375-385.
- Møller, J., Eggen, A., Fuglestad, O. L., Langfeldt, G., Presthus, A. M., Skrøvset, S., ... & Vedøy, G. (2007). Successful leadership based on democratic values. In Day, C., & Leithwood, K. (Eds.). *Successful principal leadership in times of change* (pp. 71-86). Rotterdam. Springer Netherlands.
- Møller, J., Ottesen, E., & Hertzberg, F. (2010). Møtet mellom skolens profesjonsforståelse og Kunnskapsløftet som styringsreform. [The meeting between the school's professional understanding and The Knowledge Promotion as governing reform]. *Acta Didactica Norge*, 4(1), 1-23.
- Møller, J., & Skedsmo, G. (2013). Modernising education: New Public Management reform

- in the Norwegian education system. *Journal of Educational Administration and History*, 45(4), 336-353.
- Møller, J., Prøitz, T., Rye, E., & Aasen, P. (2013). Kunnskapsløftet som styringsreform. In B. Karseth, J. Møller & P. Aasen (Eds.). *Reformtakter. Om fornyelse og stabilitet i grunnsopplæringen*. (pp. 23-41). [The Knowledge Promotion as governing reform. Reform Beats. About renewal and stabilization in basic education]. Oslo: Universitetsforlaget.
- Norwegian Directorate for Education and Training. (2011). *Nasjonale prøver. Veiledning for skoleeiere og skoleledere* [National tests. A guide for superintendents and school leaders]. Oslo.
- Norwegian Ministry of Education and Research. (2013). *Læreplan for grunnskolen og videregående skole* [Curriculum for elementary and secondary school]. Retrieved 1. August, 2014 from <https://www.udir.no/laring-og-trivsel/lareplanverket/>
- O'Day, J. (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.
- Olsen, J. P. (2005). Maybe it is time to rediscover bureaucracy. *Journal of Public Administration Research and Theory*, 16(1), 1-24.
- Ottesen, E., & Møller, J. (2016). Organisational routines—the interplay of legal standards and professional discretion. *European Educational Research Journal*, 15(4), 428-446.
- Ozga, J. (2008). Governing knowledge: Research steering and research quality. *European Education Research Journal*, 7(3), 261-272.
- Ozga, J., & Jones, R. (2006). Travelling and embedded policy: The case of knowledge transfer. *Journal of Education Policy*, 21(1), 1-17.
- Ozga, J., & Lingard, B. (2007). Globalisation, education policy and politics. In Lingard, B & Ozga, J. (Eds). *The RoutledgeFalmer Reader in Education Policy and Politics* (pp. 65-82). Oxon: Routledge.
- Park, V., & Datnow, A. (2009). Co-constructing distributed leadership: District and school connections in data-driven decision-making. *School Leadership and Management*, 29(5), 477-494.
- Pettersson, D., Popkewitz, T. S., & Lindblad, S. (2017). In the grey zone: Large-scale



- assessment-based activities betwixt and between policy, research and practice. *Nordic Journal of Studies in Educational Policy*, 3(1), 29-41.
- Pollock, K., & Winton, S. (2016). Juggling multiple accountability systems: How three principals manage these tensions in Ontario, Canada. *Educational Assessment, Evaluation and Accountability*, 28(4), 323-345.
- Powell, W. W., & DiMaggio, P. J. (Eds.). (1991). *The new institutionalism in organizational analysis*. Chicago: University of Chicago Press.
- Ranson, S. (2003). Public accountability in the age of neo-liberal governance. *Journal of Education Policy*, 18(5), 459-480.
- Sørhaug, T. (1996). *Om ledelse: makt og tillit i moderne organisering*. [About leadership: Power and trust in modern organizing]. Oslo: Universitetsforlaget.
- Prøitz, T. S., Mausethagen, S., & Skedsmo, G. (2017a). Data use in education: Alluring attributes and productive processes. *Nordic Journal of Studies in Educational Policy*, 3(1), 1-5.
- Prøitz, T. S., Mausethagen, S., & Skedsmo, G. (2017b). Investigative modes in research on data use in education. *Nordic Journal of Studies in Educational Policy*, 3(1), 42-55.
- Rizvi, F., & Lingard, B. (2010). *Globalizing education policy*. London: Routledge.
- Roe, A. (2010). Nasjonale leseprøver—hva de måler og hvordan resultatene kan brukes. [National reading tests – what they measure and how they may be used]. *Bedre Skole*, 1, 41-47.
- Røvik, K. A. (2014). Translasjon—en alternativ doktrine for implementering [Translation—an alternative doctrine for implementation]. In Røvik, K. A., Furu, E. M., & Eilertsen, T. V. (Eds.). *Reformideer i norsk skole. Spredning, oversettelse og implementering*. [Reform ideas in Norwegian school. Distribution, translation and implementation] (pp. 403-418). Oslo: Cappelen Damm akademisk.
- Sahlberg, P. (2010). Rethinking accountability in a knowledge society. *Journal of Educational Change*, 11(1), 45-61.
- Seashore Louis, K., & Robinson, V. M. (2012). External mandates and instructional leadership: School leaders as mediating agents. *Journal of Educational Administration*, 50(5), 629-665.
- Seland, I., Vibe, N., & Hovdhaugen, E. (2013). *Evaluering av nasjonale prøver som system*. [Evaluation national tests as a system]. Oslo: Nordisk institutt for studier av innovasjon, forskning og utdanning.

- Silverman, K. (2013). *The threshold of the visible world*. London: Routledge.
- Sivesind, K., Bachmann, K. E., & Afsar, A. (2003). Nordiske læreplaner. [Nordic curriculum]. Oslo: Laerningscentret.
- Skedsmo, G. (2009). *School governing in transition. Perspectives, purposes and perceptions of evaluation policy* (Unpublished doctoral thesis). University of Oslo, Norway.
- Skedsmo, G. (2011). Formulation and realisation of evaluation policy: Inconsistencies and problematic issues. *Educational Assessment, Evaluation and Accountability*, 23(1), 5-20.
- Skedsmo, G. & Mausestagen, S. (2016): Accountability policies and educational leadership – a Norwegian perspective. In Easley, J. and Tulowitzki, P. (Eds.). *Accountability and Educational Leadership – country perspectives* (pp. 205-223). London: Routledge
- Spillane, J. P. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: The mediating role of teachers' zones of enactment. *Journal of Curriculum Studies*, 31(2), 143-175.
- Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education*, 118(2), 113-141.
- Spillane, J. P., Halverson, R., & Diamond, J. B. (2004). Towards a theory of leadership practice: A distributed perspective. *Journal of Curriculum Studies*, 36(1), 3-34.
- Sundberg, D., & Wahlström, N. (2012). Standards-based curricula in a denationalised conception of education: The case of Sweden. *European Educational Research Journal*, 11(3), 342-356.
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. London: SAGE Publications.
- Townsend, T., & MacBeath, J. (Eds.). (2011). *International handbook of leadership for learning* (Vol. 25). London: Springer Science & Business Media.
- Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal*, 44(3), 519-558.
- Waite, D., & Bogotch, I. (Eds.). (2017). *The Wiley international handbook of educational leadership*. New Jersey: John Wiley & Sons.
- Wayman, J. C., & Jimerson, J. B. (2014). Teacher needs for data-related professional learning. *Studies in Educational Evaluation*, 42, 25-34.

- Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3). London: SAGE Publications.
- Yanow, D. (1996). *How does a policy mean?: Interpreting policy and organizational actions*. Washington, DC: Georgetown University Press.
- Yin, R. (1994). *Case study research. Design and methods*. Thousand Oaks: SAGE Publications.
- Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education*, 112(4), 521-548.

# Appendices

## Appendix 1: Literature review (Alphabetical order)

Author/year	Purpose	Methodology	Main findings
Abrams, Varier, & Jackson (2016)	Examine teachers' data use for instruction to reveal the extent to which instruction is aligned with established content standards and assessment.	Focus group interviews with 60 U.S. upper elementary and middle school teachers from 45 schools.	Teachers aligned instruction and assessments with the state curriculum in order to improve student performance. While teachers found day-to-day informal assessments to be essential for shaping instruction, periodic formal assessments helped them monitor student progress. Teachers described the challenges associated with the state curriculum and the lack of infrastructure to support data use.
Anderson, Leithwood, & Strauss (2010)	Examine data use and organizational conditions influencing data use by principals and teachers and the relationship between data use and student performance.	Multi-method investigation of leadership at the school, district, and state levels.	Leadership among principals plays an important role in establishing the purpose and expectations of data use, access to expertise, and follow-up actions. Principals' and teachers' use of data is strongly shaped by district leaders in the context of state accountability systems. Statistical evidence linking patterns of data use to achievement test results was weak and limited to elementary schools.
Baker & LeTendre (2005)	Investigate the circulation of knowledge, education, and education policy.	Book of data from the Third International Mathematics and Science Study (TIMSS) conducted in 1994 and 1999 in the U.S.	Analysis of whether a global style of teaching (mathematics) exists. Mathematics teachers' working conditions are mostly affected by "global institutional forces in the core tasks," while national forces affect their work environment.
Ball, Maguire, & Braun (2012)	Aim to develop a theory of policy enactment and empirical critical exploration of policy enactment.	Two-and-a-half-year-long qualitative case study of lower secondary schools.	Policy enactment beyond implementation and the importance of context, discourses, representations, and translations.
Berge (2005)	Discuss how to understand the concept of basic skills and how the content of basic skills is represented in national tests.	Conceptual and theoretical work mainly about writing as a basic skill.	National tests lack value if they are not carefully monitored by schools and teachers. Thorough pedagogical monitoring of a local school will help the teachers to master use of national test results.
Bergesen (2006)	Discuss the intentions and content of the new Norwegian curriculum "The Knowledge Promotion."	Conceptual, historical, and theoretical analysis.	It is important to give schools the freedom to make organizational, methodological, and didactic choices. Thereafter, society needs to gain control and monitor the results of those choices through reports and tests.
Berryhill, Linney, & Fromewick (2009)	Investigate U.S. elementary school teachers' perceptions of their state's	Survey data of 100 teachers of grades 3–5 and interviews with	There is a path model relating a lack of policy support to teacher burnout via two mediators: role conflict and reduced self-efficacy. The results of interviews with a subset of teachers were

	accountability policy, particularly its effect on their job engagement.	20 teachers.	consistent with the model.
Black & Wiliam (1998)	Discuss the shift away from concentration on the restricted forms of test, which are only weakly linked to students' learning experiences.	Literature review of 250 articles.	Theoretical analysis of the nature of feedback and discussion of models for future formative assessment practices.
Black, Harrison, Lee, Marshall, & Wiliam (2004)	Find evidence about how to improve formative assessment. The King's-Medway-Oxfordshire Formative Assessment Project (KMOFAP).	Observation in classrooms, interviews with groups of teachers ( $N=24$ ). Interviews with principals and discussions with students.	Concerns focus on questioning, grading, peer- and self-assessment, and the importance of classroom dialogue. There is a need to review school policies.
Broadfoot & Black (2004)	Discuss the content and concept of assessment in the first ten years of <i>Assessment in Education</i> .	Review of the first ten years of the journal <i>Assessment in Education</i> .	The notion of formative assessment is in conflict with the requirements of simultaneous summative large-scale assessments.
Coburn & Turner (2011)	Provide a conceptual guide for how to study data use.	Conceptual essay and review article.	Using data requires the user to interpret the data and construct implications for the next steps. Data can be a source of power, particularly in the current accountability environment. Hence, data use also involves power relations.
Collins, Reiss, & Stobart (2010)	Determine what happens when high stakes testing stops using a comparative analysis of England and Wales.	Quantitative and qualitative data obtained from a telephone survey of 600 respondents and eight group interviews with 74 participants.	Practical science activities, including investigations, are becoming an important feature of science lessons in Y6 in Wales, and early indications suggest increased opportunities for investigative science in Y6 after abolition of high-stakes testing in Y6 in Wales.
Crooks, Kane, & Cohen (1996)	Show that the chain model can be a very useful tool for validating an existing assessment process.	Conceptual framework.	An eight-step model for validation. The purpose of the assessment must be clarified first by carefully considering the expected actions based on the assessment as well as desired impact of those actions and of the overall assessment process.
Datnow & Hubbard (2015)	Examine teachers' capacity for and beliefs about data use.	Review article.	Teachers have varied beliefs about data use, and some feel they lack the ability to use data to inform instruction. In order to be more successful, capacity building should directly address teachers' beliefs, and data use must be decoupled from external accountability demands and involve a variety of information about student learning.
Earl & Katz (2002)	Theorize about how leaders can become competent and confident in interpreting data.	Conceptual book chapter.	To be a leading school in a data-rich world, leaders must: <ul style="list-style-type: none"> <li>• develop a habit of inquiry,</li> <li>• become data literate, and</li> <li>• create a culture of inquiry.</li> </ul>
Elmore (2005)	Investigate (changes in)	Conceptual	A change in leadership demands is largely a

	leadership demands in American public schools.	article. The outlined framework is built on previous research funded by CPRE.	consequence of the introduction of performance-based accountability policies that evaluate, reward, and sanction schools based on measured student performance. Policy makers' misconceptions of accountability lead to serious and adverse consequences for schools.
Firestone & Roseblum (1988)	Understand teacher and student commitment.	Conceptual framework and empirical data from 10 urban high schools.	Teacher and student commitments are mutually reinforcing. Five school factors can increase commitments: relevance, respect, support, expectations, and influence.
Grek (2009)	Examine the Programme for International Student Assessment (PISA) as a major and influential component of educational work.	Text analysis and interview data from key policy actors in Brussels.	PISA, through its direct impact on national education systems in Europe and beyond, has become an influential tool in new political technology for governing the European education system based on statistics.
Hall (2013)	Offer an analytical account of how educational leadership has emerged within the context of new public management.	Qualitative review of five six-month to one-year case studies of secondary schools	Discourses of leadership have enabled schools and teachers to adapt to educational modernization in ways that have contested teacher professionalism.
Hall, Gunter, & Bragg (2013)	Examine the public sector's and new public management's effort to re-model teacher identities, construct normative discursive practices, and identify the accompanying potential for socio-ideological control of teachers.	Focus on one school of five secondary schools investigated in case studies. Interviews over a period of six months. Document analysis and Q-methodology for leadership practices.	NPM has created normative discursive leadership (as distributed leadership) practices that operate as a form of identity work in teachers' professional lives. Within NPM, the development of standardized testing systems is accompanied by major emphasis on pupil, teacher, and school performance in such tests.
Hardy (2014)	Explore how the strong policy push for improving test results in Queensland, Australia influenced school practice.	55 individual interviews with teachers and principals from three primary schools.	Policy enactment in schools is characterized by competing interests, involving not only interpretation, translation, and critique but also active appropriation of teachers' political concerns.
Hargreaves & Shirley (2008)	Investigate how the Organization for Economic Co-operation and Development is related to school leadership in Finland, the U.K., Canada, and the U.S.	Conceptual and empirical study. Interviews with school leaders from Finland.	A high-trust culture with the "peer factor" is better than a culture with the "fear-factor." Cooperation and responsibility matter. Data can inform and enhance teacher decisions but should never drive instruction.
Hertzberg & Roe (2016)	Investigate how teachers choose to work with content writing when they define the premises themselves.	Quantitative and qualitative data from a network of secondary schools with established cross-curricular	While teachers had to deal with differences in subject-specific text norms, they also discovered similarities across subject areas in which they could work collectively. The development of a broader instruction repertoire and more goal-related use of scaffolding strategies in the classroom, most of which was in accordance with

		school-based writing projects.	a process approach to writing, had a major effect on participants.
Hopmann (2008)	Provide a frame for discussing questions about the international rise of accountability as a key tool of social change.	Study drawing on historical and comparative research using PISA as an example.	Even though accountability is a global phenomenon, the ways and means of enacting and encountering accountability are not. How accountability is experienced depends on deeply engrained “constitutional mind-sets” (i.e., diverse cultures of conceptualizing the relation between the public and its institutions).
Karseth & Sivesind (2010)	Investigate the implications of a global curriculum language and how it can be dealt with in terms of content knowledge and reform.	Policy document analysis.	New language for change and worldwide expectations could harmonize curricula across national borders. The design of the reform is of core significance and is highly dependent on the way in which matters and meanings are conceptualized in professional semantics within and beyond a national context.
Lachat & Smith (2005)	Investigate how large-scale data can be used to examine progress and guide improvement, how factors and conditions either promote or prevent data use, and the policy and practice implications regarding effective data use for high school reform.	Case study focusing on data use in five low-performing urban high schools.	Several key factors have an impact on data use in the study sites: the quality and accuracy of available data, staff access to timely data, capacity for data disaggregation, collaborative use of data organized around a clear set of questions and leadership structures that support school-wide use of data.
Marsh, Bertrand, & Huguet (2015)	Examine critical aspects of teachers’ use of student learning data and the actions taken in response to data analysis in order to alter instruction practices.	Year-long comparative, multi-method study of six low-performing middle schools in four districts.	Dialogue and the dynamic relationship between two types of expertise may help explain the ways in which professional learning communities and coaches facilitate deep-level changes in pedagogy. In addition, school leadership and district-level context emerged as key factors.
Mausethagen (2013)	Examine tensions that occur in meetings when primary school teachers in Norway discuss national testing and how they are handled.	Observations of teacher and principal meetings and interviews of principals and teachers.	Tensions about national testing develop regarding issues of professional knowledge, curriculum, caring relationships, and formative assessment. National tests are mainly dealt with outside teachers’ work.
Mintrop & Sunderman (2009)	Investigate how high-stakes accountability policies and sanction systems do not work and why they are still being done.	Survey of teachers in impoverished school districts, building on a longitudinal six-state study of No Child Left Behind (NCLB).	There seem to be good ideas about how to improve the meaningfulness of accountability systems for educational practice, but they fail to address the underlying flaws of NCLB. The authors argue that an alternative to the current sanction system and broadening of the social welfare agenda are needed.
Møller & Skedsmo (2013)	Investigate the way in which ideas connected to NPM reforms were introduced and interpreted in the Norwegian education sector.	Policy document analysis.	Three areas of discursive struggle were identified. Ideologies about schooling, contested issues of teacher professionalism, and strategies for improving education are presented. NPM reforms sped up when Norway was listed among the “lower-performing” schools according to international tests. Leadership and accountability became the dominant themes.

Møller, Prøitz, Rye, & Aasen (2013)	Understand the new educational reform in Norway.	Document analysis, surveys, and interviews.	Norwegian national authorities have a hierarchical and top-down understanding of implementing the reform even though “steering from below” was an important part of the policy discourse when the reform was initiated.
Park & Datnow (2009)	Examine leadership practices in school systems that are implementing data-driven decision making.	Qualitative data from a multi-case study of four urban school systems in the U.S.	School leaders created an ethos of learning and continuous improvement rather than one of blame. Leaders distributed decision-making authority in a manner that empowered different staff members. Directed recourse for building human and social capacity.
Pollock & Winton (2016)	Explore how three principals from Ontario, Canada manage the tensions between multiple accountability systems in their work.	Qualitative data from interviews with principals. Student data from 14 schools.	Principals were able to find ways to be accountable to a variety of communities and ideals but could not ignore various accountability pressures. Success hinged on the ability to negotiate these pressures. This delicate balancing act requires school principals to develop the skills necessary to successfully juggle these competing demands.
Prøitz, Mausethagen, & Skedsmo (2017)	Report the findings of a literature review of research on data use in education.	Systematic mapping and review of articles published in English, German, and Scandinavian.	The characteristics of the total corpus of 129 articles on data use in education vary across different contexts, countries, and regions. In all contexts, the studies primarily investigate structures and systems concerning data use. Some mostly concern implementation and effectiveness, and others concern reflections on the developments of data use in education.
Ranson (2003)	Examine and discuss how accountability practices and options have emerged.	Conceptual analytical framework. Critique of the emerging public-private partnership (PPPs) regimes.	Accountability practices and dispositions have changed over time and generated perverse, unintended consequences. The notions of consumer choice and capital ownership have been strengthened over time, and public society has become an “audit and performative society.” There are possibilities for democratic accountability.
Roe (2010)	Discuss how national tests of reading can be measured and how the results can be used.	Conceptual and theoretical.	It is important for teachers in all subjects to monitor and follow up on the results of national tests because they measure basic reading skills in all subjects.
Sahlberg (2010)	Stress the distinction between intelligent and non-intelligent accountability policies and how they direct teachers and students to learning differently.	Conceptual article within the Finnish school system context.	There is no evidence that test-based accountability policies improve the quality and efficiency of public education. Increased high-stakes testing restricts students’ engagement in creative action and understanding of innovation. There is a need to increase networking, build trust, and strengthen collective responsibilities within and between schools.
Seland, Vibe, & Hovdhaugen (2013)	Evaluation report commissioned by the Norwegian national authorities.	Large-scale surveys and interviews with actors at all levels of the educational system.	Norwegian principals view national tests as a tool to improve school quality and individual student learning. The teachers, however, were more reserved. They did not have time to use the individual student test results and did not gain any new information about students’ knowledge.
Skedsmo & Mausethagen (2016)	Investigate the policy intentions of the new Norwegian evaluation tools and principals’	Policy document analysis and survey data	It is uncertain whether all policy intentions can be fulfilled or whether the tools might produce other effects.



	perceptions regarding use of these tools to improve student achievement.	about principals' perceptions.	
Skedsmo (2011)	Investigate the policy intentions of the evaluation tools and principals' perceptions regarding use of the tools to improve student achievement.	Document analysis and survey data (principals).	There are problematic aspects of the tools' modes of regulation and thereby the choice of tools. There are inconsistencies between the policy intentions and the use of tools in practice.
Spillane (2012)	Identify and discuss some conceptual and analytical tools for studying data in practice.	Theoretical and conceptual article including excerpts from field notes in particular schools.	The concept of organizational routines is framed. Researching data use should be partially about the study of practice in schools in order to understand how school staff use data and what sorts of data they use in their everyday work.
Stobart (2008)	Investigate how assessment shapes the way we see ourselves, how we learn and what we learn, and how it can be used to encourage effective learning and teaching. <i>Testing Times: The Uses and Abuses of Assessment.</i>	Theoretical and conceptual book.	While policy makers may welcome high-stakes large-scale data testing programs, the response of those tested may be very different. Tests of potential are really tests of attainment. Culture and context are emphasized.
Valli & Buese (2007)	Examine the impact of federal, state, and local policies on the roles of elementary school teachers in an era of high-stakes accountability.	Large mixed-methods, longitudinal study of fourth- and fifth-grade teachers.	Role expectations are increased, intensified, and expanded in four areas: instructional, institutional, collaborative, and learning. These changes had unanticipated, and often negative, consequences on teachers' relationships with students, pedagogy, and sense of professional well-being.
Young (2006)	Explore the influence of grade-level team norms and district and school leadership on teachers' data use.	Qualitative study of four schools in two different districts. Embedded system perspective.	Establishing a rationale for teachers to use particular data, model data use, and structure time to learn about using data is a deliberate agenda-setting activity. Varying degrees of loose coupling between the case study districts underscore how grade-level norms and agendas mediate teachers' collaborative use of data.

## Appendix 2: Approval for research (NSD)

Norsk samfunnsvitenskapelig datatjeneste AS  
NORWEGIAN SOCIAL SCIENCE DATA SERVICES



Harald Hårfagres gate 29  
N-5007 Bergen  
Norsay  
Tel: +47-55 58 21 17  
Fax: +47-55 58 96 50  
nsd@nsd.uib.no  
www.nsd.uib.no  
Org nr: 985 321 884

Ann Elisabeth Gunnulfsen  
Institutt for lærerutdanning og skoleforskning Universitetet i Oslo  
Postboks 1099 Blindern  
0317 OSLO

Vår dato: 24.10.2014

Vår ref: 39413 / 3 / JSL

Deres dato:

Deres ref:

### TILBAKEMELDING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 10.08.2014. Meldingen gjelder prosjektet:

39413 *School Leadership and the Enactment of Policy Demands A Micro Political Perspective (Skoleledelse og håndtering av politiske krav Et mikropolitisk perspektiv)*  
Behandlingsansvarlig *Universitetet i Oslo, ved institusjonens øverste leder*  
Daglig ansvarlig *Ann Elisabeth Gunnulfsen*

Personvernombudet har vurdert prosjektet og finner at behandlingen av personopplysninger er meldepliktig i henhold til personopplysningsloven § 31. Behandlingen tilfredsstiller kravene i personopplysningsloven.

Personvernombudets vurdering forutsetter at prosjektet gjennomføres i tråd med opplysningene gitt i melde skjemaet, korrespondanse med ombudet, ombudets kommentarer samt personopplysningsloven og helseregisterloven med forskrifter. Behandlingen av personopplysninger kan settes i gang.

Det gjøres oppmerksom på at det skal gis ny melding dersom behandlingen endres i forhold til de opplysninger som ligger til grunn for personvernombudets vurdering. Endringsmeldinger gis via et eget skjema, <http://www.nsd.uib.no/personvern/meldeplikt/skjema.html>. Det skal også gis melding etter tre år dersom prosjektet fortsatt pågår. Meldinger skal skje skriftlig til ombudet.

Personvernombudet har lagt ut opplysninger om prosjektet i en offentlig database, <http://pvo.nsd.no/prosjekt>.

Personvernombudet vil ved prosjektets avslutning, 01.07.2018, rette en henvendelse angående status for behandlingen av personopplysninger.

Vennlig hilsen

Katrine Utaaker Segadal

Juni Skjold Lexau

Kontaktperson: Juni Skjold Lexau tlf: 55 58 36 01  
Vedlegg: Prosjektvurdering

## **Appendix 3: Information letter (Phase 1)**

(Author's translation from Norwegian)

### **Request for participation in a PhD project**

“School leadership and national testing”

#### **Background and purpose**

This PhD project is related to the field of educational leadership and performed at the Department of Teacher Education and School Research at the University of Oslo.

The aim is to investigate how school leaders in lower secondary schools (principals, senior leaders, teacher team leaders, and teacher leaders) use national test results as a means of developing school quality and student learning. The main research question of the study is as follows:

*How do schools (leaders and teachers) work with national test results for reading and numeracy?*

I am addressing your school based on my positive dialogue with the superintendent about my interest in schools within the municipality that are relevant for participation in my study. I emphasize that the study is voluntary. The selection criteria are lower secondary schools with more than 300 students. Additional criteria for the selection of two schools for on-site observation in the second phase of the study are the extent to which the schools have used national test results for reading and numeracy from 8<sup>th</sup> grade to 9<sup>th</sup> grade.

#### **What does participation in the study involve?**

Participation in the study involves interviewing the formal leaders of the school (principal, senior leader, teacher leader, teacher team leader, etc.). An audio recorder will be used in the interviews. Furthermore, participation in the second part of the study will involve observation and logging of formal meetings between teachers and school leaders with national tests on the agenda. All data will be anonymized. Finally, the study will involve a survey of all the teachers at the school. The data will be anonymized in the thesis. All questions will concern the main purpose of the study: school leadership and the use of national test results. Examination of the school's results for reading and numeracy in 8<sup>th</sup> and 9<sup>th</sup> grade for this year and three years back will be included in the study. Only the school leaders and teachers will be interviewed and observed. Parents and students are not a part of this study.

#### **What will happen to the information about you?**

All personal information will be treated confidentially. Only my main supervisor, Professor Jorunn Møller, and any computer supervisor will have access to the study. Otherwise, the names of individuals and schools will be anonymized in the transcription and analysis process. Participants will not be able to be recognized in publications. The project is scheduled to end on 1 July 2018. Recording material will be deleted. The results will be presented to each school if requested.

#### **Voluntary participation**

It is voluntary to participate in the study, and you can withdraw your consent without giving any reason at any time. If you withdraw, all information about you will be deleted. The study has been reported to the Norwegian Center for Research Data (NSD).

If you wish to participate or have questions about the study, please contact me: Ann Elisabeth Gunnulfsen, [a.e.gunnulfsen@ils.uio.no](mailto:a.e.gunnulfsen@ils.uio.no), telephone: +47 97759928.

Best regards  
Ann Elisabeth Gunnulfsen (sign)

## **Appendix 4: Information letter (Phase 2)**

(Author's translation from Norwegian)

### **To the school's leadership and teachers at the 9<sup>th</sup> grade level**

#### **Information about participation in video recording – research project**

##### **Background and purpose**

I am a PhD student at the Department of Teacher Education and School Research at the University of Oslo. I hereby inform you of your school's participation in Phase 2 of my PhD project. In this phase of the study, part of the method is video recordings of meetings. Earlier on-site observations and the school's educational plans will be included in the compiled material as background data. Data collection involves video recordings of leadership meetings and team meetings at the 9<sup>th</sup> grade level from August to September 2015. The study focuses on schools' (teachers' and school leaders') discussion about practice, with particular focus on basic skills and use of national test results. The third and last part of the study is a questionnaire aimed at a large number of teachers in the Eastern region. I will come back to this. Criteria for selecting your school in this second part of the study have been the extent of improvement in results on national testing in reading and numeracy the 8<sup>th</sup> to 9<sup>th</sup> grade, as well as a pre-admission for the conducting of video recordings.

##### **What does participation in the study involve?**

The video recordings will be conducted by one of the senior leaders. The researcher will not be present. All data will be anonymized by the researcher. The analyses are conducted based on organizational theories and the process of sense-making (Weick, 1995) and sociological theory on educational policy. Emphasis will be placed on analytical generalizability. This means that the informants shown on the video recordings represent a general analytical range of teachers and school leaders. Only school leaders and teachers are the subject of the study. Parents and students are not included.

##### **What will happen to the information about you?**

All personal information will be treated confidentially. Only my main supervisor, Professor Jorunn Møller, and any computer supervisor will have access to the study. Otherwise, the names of individuals and schools will be anonymized during the transcription and analysis processes. Participants will not be able to be recognized in publications. The project is scheduled to end on 1 July 2018. Recording material will be deleted. The results will be presented to each school if requested.

##### **Voluntary participation**

It is voluntary to participate in the study, and you can withdraw your consent without giving any reason at any time. If you withdraw, all information about you will be deleted. The study has been reported to the Norwegian Center for Research Data (NSD).

If you wish to attend or have questions about the study, please contact me: Ann Elisabeth Gunnulfsen, [a.e.gunnulfsen@ils.uio.no](mailto:a.e.gunnulfsen@ils.uio.no), telephone: +47 97759928.

Best regards  
Ann Elisabeth Gunnulfsen (sign)

## Appendix 5: Overview of participating schools and informants in Phase 1

School	Informants	General information	Average test score
<b>Blue Mountain</b>	Rebecca (principal) Rosa (deputy manager) Ronny (deputy manager) Reese (deputy manager)	400 students, 8th–10th grade	High
<b>Brown Hill</b>	Gina (principal) Gary (deputy manager) Greta (deputy manager)	350 students, 8th–10th grade	Medium/high
<b>Yellow Stone</b>	Brigitte (principal) Brita (deputy manager) Barry (deputy manager)	300 students, 8th–10th grade	Medium/high
<b>Red Field</b>	Hilary (principal) Henry (deputy manager)	320 students, 8th–10th grade	Medium

---

## Appendix 6: Interview data

Data collection in Phase 1: Individual interviews with principals and senior leaders

<b>Nr</b>	<b>Date</b>	<b>School</b>	<b>Informant</b>	<b>Title</b>	<b>Duration</b>
1	02.09.2014	Blue Mountain	Rebecca	Principal	1:12
2	09.09.2014	Blue Mountain	Rosa	Senior leader	1:06
3	09.09.2014	Blue Mountain	Ronny	Senior leader	1:09
4	09.09.2014	Blue Mountain	Reese	Senior leader	0:51
5	09.10.2014	Yellow Stone	Brigitte	Principal	1:08
6	09.10.2014	Yellow Stone	Brita	Senior leader	1:02
7	09.10.2014	Yellow Stone	Barry	Senior leader	0:51
8	20.10.2014	Brown Hill	Gina	Principal	1:05
9	20.10.2014	Brown Hill	Gary	Senior leader	0:48
10	20.10.2014	Brown Hill	Greta	Senior leader	1:06
11	18.12.2014	Red Field	Hillary	Principal	0:55
12	18.12.2014	Red Field	Henry	Senior leader	1:12

## Appendix 7: On-site observation data

Nr	Date	School	Meeting	Topic – categories – field notes	Duration
1	240215	Blue Mountain	Principal/ researcher	Information about the “extra” planning evening with the teachers.	13.00 – 13.30 <b>30 min</b>
2	240215	Blue Mountain	Plenary teacher meeting	Principal orientation. Students’ presentation of their work. General information of spring term.	13.30 -14.30 <b>60 min</b>
3	240215	Blue Mountain	9 <sup>th</sup> grade teacher meeting	Evaluation of an integrated subject project (9A).	14.30 -15.30 <b>60 min</b>
4	240215	Blue Mountain	Plenary teacher meeting	Principal’s model of cooperative learning in small groups.	16.00 – 16.30 <b>60 min</b>
5	240215	Blue Mountain	8 <sup>th</sup> grade teacher meeting	Interdisciplinary topics. Planning for spring term. Mostly administrative tasks, such as scheduling lesson plans (8A and B).	16.30 – 18.00 <b>90 min</b> 18.30 – 19.30 <b>60 min</b>
6	240215	Blue Mountain	Breaks and dinner	Informal discussions and social talk.	15.30 – 16.00 <b>30 min</b> 18.00 – 18.30 <b>30 min</b>
7	250215	Blue Mountain	8 <sup>th</sup> grade teacher meeting	First hour includes teacher groups A and B. Second hour includes only teacher group A.	13.30 – 15.30 <b>120 min</b>
8	100315	Blue Mountain	Pedagogical team (principal, deputies, and team leaders)	Critical discussion about an exposure draft for new national regulations for student assessment.	08.30 – 10.00 <b>90 min</b>
9	100315	Blue Mountain	Leader team meeting	Practical issues. Spring term planning. Exam planning. One short break.	10.30 – 13.00 <b>150 min</b>
10	100315	Blue Mountain	Plenary teacher meeting	Principal orientation. 9 <sup>th</sup> grade students’ presentation of their “student enterprises.” “Pedagogical Candies” (10 <sup>th</sup> grade) presentation of criteria for and results of making a film.	13.30 – 15.00 <b>90 min</b>
11	120315	Brown Hill	Leader meeting	Discussion of the research method/topic and identification of strategies to inform teachers about it.	12.00 – 13.00 <b>60 min</b>
12	120315	Brown Hill	Observation	Students work with iPads for writing and reading in various subjects. Two classes, one of which works outside.	13.00 – 14.00 <b>60 min</b>
13	020315	Brown Hill	Leader meeting	Discussion of NT results.	Cancelled
14	120315	Brown Hill	Leader meeting	Planning for plenary discussion of NT results.	12.00 – 13.00 <b>60 min</b>

## **Appendix 8: Interview guide**

(Author's translation from Norwegian)

### **QUESTIONS FOR THE SEMI-STRUCTURED INTERVIEW GUIDE**

The numbered questions are key questions. Bullet points under each main question served as support questions when necessary.

#### **1. Tell me about the school. (Starting issues)**

- What is the intake area and student and teacher demographics (ethnicity, gender, and class)?
- What kind of reputation does the school have?
- What is the students' learning environment?
- How would you describe the school climate and culture (safety, creativity, professionalism, critical thinking)?
- How would you define an effective/successful school?
- How would you describe your school in light of this definition?
- Which school development projects are you working on right now?

#### **2. What does the student performance look like today? (Exam + national tests)**

- Define student performance (knowledge aims/social objectives).
- Does the school perform as well as it should? Why or why not?
- Are there students who do not succeed as expected? If so, why?
- What is the school's capacity to work on improving student performance?

#### **3. Tell me about the school's work with national test results for literacy and numeracy.**

- Describe teachers' involvement.
- How are the tests conducted?
- How and when do you work with the results? Who is responsible? How are the staff involved? What conclusions/implications do you draw?
- What opportunities and challenges might exist regarding the use of national test results to strengthen individual students' learning?
- What do you consider to be the school's strengths related to work with national test results? What are the challenges?
- What do you think about the possibilities for improving the school's results in national tests?
- Do you think the national tests could contribute to the development of quality in teaching practice at (school name)?
- To what extent do you think you (and the school) are accountable for the national test results for reading and numeracy?
- What do you think is the role of basic reading and numeracy skills in the curriculum?

#### **4. How would you describe yourself as a leader?**

- What are your strengths/weaknesses?
- Describe any strategies you as a leader make use of regarding work with national test results for reading and numeracy.

#### **5. Is there anything else you would have wanted to talk about?**



## Appendix 9: Teacher questionnaire (Phase 3)

Thematic areas	Items (A five-point Likert-scale ranging from “Not at all” to “To a very great extent”).	Criteria for validation
<p><b>1. Principals’ facilitative work with national tests</b>  <i>To what extent do the following statements/allegations match the situation at your school?</i></p>	<ol style="list-style-type: none"> <li>1. The principal has confidence in my work with basic skills.</li> <li>2. The principal requests information about how I follow up on the national test results.</li> <li>3. The principal represents the school’s work with quality development in a good way in the wider community.</li> <li>4. The principal engages in dialogue with teachers on the curriculum content.</li> <li>5. The principal is clear about what is expected of my work with the national test results.</li> <li>6. The principal expects that teachers will work systematically with the national test results.</li> <li>7. National tests are a topic at the <i>appraisal meetings</i> with the principal.</li> </ol>	<p>Leadership            Experiences            Expectations</p>
<p><b>2. Teachers’ own work with national test results</b>  <i>To what extent do the following statements/allegations fit your situation?</i></p>	<ol style="list-style-type: none"> <li>1. I am involved in conducting national tests.</li> <li>2. I take responsibility for following up on the school’s national test results.</li> <li>3. I take responsibility for purposeful monitoring of the students in my course when the national test results are ready.</li> <li>4. I discover new aspects of my students’ competence in dealing with their national test results.</li> <li>5. I use assignments from the national reading test when I go through the reading results with my students.</li> <li>6. I use assignments from the national numeracy test when I go through the numeracy results with my students.</li> </ol>	<p>Practices            Experiences            Leadership</p>
<p><b>3. National tests and the teacher team and local school</b>  <i>To what extent does the following statement/allegation fit your teacher team/your school?</i></p>	<ol style="list-style-type: none"> <li>1. The entire teacher team takes responsibility for monitoring student results on national tests.</li> <li>2. The teacher team discusses the school’s teaching practice in light of the national test results.</li> <li>3. The teacher team has influence on the time spent working with students on national test results.</li> <li>4. The teacher team actively uses the national test results to improve students’ individual learning outcomes.</li> <li>5. Using the national test results is a part of the school’s plan for quality development.</li> </ol>	<p>Leadership            Practices            Experiences</p>
Construct	Items (A five-point Likert-scale ranging from “Strongly disagree” to “Strongly agree”).	Criteria for validation
<p><b>4. Perceptions/views about national tests</b>  <i>To what extent do you agree with the following statement?</i></p>	<ol style="list-style-type: none"> <li>1. The national test results have little or no value in terms of individual students’ learning and development.</li> <li>2. The national test results are useful tools for working with individual students’ learning and development.</li> <li>3. Work with national test results takes too much time away from other, more important tasks at school.</li> <li>4. I lack knowledge of how to use the national test results for reading in my subject(s).</li> <li>5. I lack knowledge of how to use the national test results for numeracy in my subject(s).</li> <li>6. National tests for reading have contributed to new knowledge about reading as a basic skill in my subject(s).</li> <li>7. National tests for numeracy have contributed to new knowledge about numeracy as a basic skill in my subject(s).</li> <li>8. The school’s overall national test results are very important for the quality of our school’s work.</li> <li>9. Publication of the national test results has no value for improving the quality of our school.</li> </ol>	<p>Attitudes            Expectations            Experiences            Leadership</p>

# Appendix 10: Examples of the coding process in NVivo (Article I)

## 1. Coding Query (free node: “quality,” “learning,” “national testing”)

The screenshot shows the NVivo interface with a coding query named "Nasjonale prøver - Coding by I". The left sidebar lists nodes such as "Ansvarliggjøring", "Arbeid med resultater på n", "Grunnleggende ferdigheter", "Holdninger", "Kvalitet", "Ledelse", "Lærers praksis", "Læring", "Læringsmiljø", "Metode", "Nasjonale prøver", "Organisering", "Resultater", "Samarbeid", "Satsingsområder", "Skolekultur", "Styring", "Tid", and "Vurdering". The main window displays a list of references with their coverage percentages. The selected reference (Reference 6) shows a text segment with highlighted words and phrases.

Name	Sources	References
Ansvarliggjøring	1	6
Arbeid med resultater på n	1	11
Grunnleggende ferdigheter	1	11
Holdninger	1	3
Kvalitet	1	22
Ledelse	1	27
Lærers praksis	1	19
Læring	1	29
Læringsmiljø	1	8
Metode	1	3
Nasjonale prøver	1	22
Organisering	1	19
Resultater	1	25
Samarbeid	1	21
Satsingsområder	1	3
Skolekultur	1	13
Styring	1	2
Tid	1	3
Vurdering	1	12

Reference 6 - 0.34% Coverage

...Prøvegruppene er vi svært lite interessert i. Fordi de er normert, og fordi de har flyttet seg

## 2. Example of word frequencies (free node: “practice”)

The screenshot shows the NVivo interface with a word cloud visualization. The left sidebar lists nodes such as "Kvalitet", "Læring", "Nasjonale prøver", "Praksis", "Resultater", and "Vurdering". The main window displays a word cloud with the word "ikke" being the most prominent. The word cloud is generated from a query named "Alle i fase 1".

Word Cloud Tools

WORD CLOUD

Alle i fase 1

Word Frequency Query Result

Word Frequency Criteria

Search in: All Sources, Selected Items..., Selected Folders...

Display words: 1000 most frequent

With minimum length: 4

Grouping: Exact matches (e.g. "talk"), With stemmed words (e.g. "talking"), With synonyms (e.g. "speak"), With specializations (e.g. "whisper"), With generalizations (e.g. "communicate")

## Appendix 11: Process of analysis (Articles I and II)

### Process of analysis (Phase 1 - Article I)

	<b>Aim</b>	<b>Tools of analysis</b>	<b>Empirical research questions</b>
<b>1st reading</b>	To identify how school leaders make sense of policy initiatives related to national test results.	Search for sensitizing concepts.	How do school leaders talk about work with national test results? How do school leaders make sense of their ability to enhance individual students' learning?
<b>2nd reading</b>	To investigate how the leadership team's work is linked to variations in accountability.	Internal versus external accountability. Enabling and constraining factors. School leaders' use of coupling mechanisms.	How do school leaders respond to being held responsible for national test results for literacy and numeracy? Which coupling mechanisms dominate?
<b>3rd reading</b>	To describe variations in response to the use of national test results in work with school improvement.	Variations in individual and collective work. The logic of conformity versus non-conformity.	How do school leaders explain variations? How do they define the school's capacity to improve using national test results?

### Process of analysis (Phase 2 - Article II)

	<b>Aim</b>	<b>Tools of analysis</b>	<b>Empirical research questions</b>
<b>1st reading</b>	To identify situations in which school leaders and teachers in the two schools discussed and planned for activities related to national test results as well as how they did so.	Word choice. When and how discussion about national testing occurred in the video material.	When and how do school leaders and teachers interpret and translate national testing policies? How do they talk about work with national test results?
<b>2nd reading</b>	To identify how the participants positioned themselves in relation each other in their interpretation and translation of national testing policies.	Modalities and binaries. How words are used in suggestions for planning of activities and placing responsibilities regarding improvement of student learning.	In what ways does school leaders' and teachers' language reflect interpretation and translation of national test results and policy?
<b>3rd reading</b>	To synthesize variations between the two schools' individual and collective talk occurring as responses to the use of test results in local work to improve student learning.	Comparison of moments of difficulty, as indicated by school leaders' and teachers' discussion about future plans and understanding of work with national test results for improving student learning.	How do school leaders and teachers position themselves in relation to each other in their interpretation and translation of national test results? What are the similarities and differences in school leaders' and teachers' collective talk?

## **Appendix 12: E-mail to respondents**

(Sent **November 4th, 2014** - my translation from Norwegian to English)

**To the principal and senior leaders at ..... school:**

Dear Principal and Senior Leader,

I am contacting you in connection to my PhD work, in which I focus on leadership and national testing in lower secondary schools. Attached, you will find the information letter with my request for your school's participation and the possibility of withdrawing at any time from the project. Initially, I want to conduct interviews with everybody holding leadership positions at the school (about one hour with each participant).

Gradually, it may be appropriate to conduct on-site observations of teacher meetings, team meetings, and leader meetings with a focus on how the school makes use of national test results. This will, however, be decided after the first phase, and I will only need two schools from the sample. I will come back to you regarding this question.

I have been in contact with the superintendent, who feels positively about the study. The sample of the study includes all lower secondary schools in ... municipality. It is important to emphasize that the participation is voluntary and that all data will be anonymized. Findings and other contributions will be offered to any school that participates. Please contact me if you have any questions.

Hoping for a positive reply.

Best regards,

Ann Elisabeth Gunnulfsen  
<http://www.uv.uio.no/ils/personer/vit/anngun/index.html>  
PhD-candidate, Educational Leadership  
University of Oslo

Mobile Phone: 97759928

## Appendix 13. Overview of the phases, articles, questions, and analytical concepts

	<b>Phase 1 – Article I</b> <b>Interviews with school leaders</b>	<b>Phase 2 – Article II</b> <b>Observation and video recordings of leader team and teacher meetings</b>	<b>Phase 3 – Article III</b> <b>Teacher questionnaire</b>
<b>Title of article</b>	National Testing: Gains or Strains? School Leaders' Responses to Policy Demands	School Leaders' and Teachers' Work with National Test Results: Lost in Translation?	Investigating teachers' and school principals' enactments of national testing policies: A Norwegian study
<b>Methods</b>	Qualitative	Qualitative	Quantitative
<b>Main (sub) research question(s)</b>	How do school leaders make sense of and respond to external demands related to a new policy context that emphasizes national test results for reading and numeracy?  How do they use national test results in their local work with school and student development?	How do school leaders and teachers position themselves in relation to each other in their interpretation and translation of national testing policies?  In what ways does school leaders' and teachers' language reflect their enactment of national testing policies?	How do teachers perceive principals' facilitative work with national test results at the local school level? How do teachers and teacher teams use national test results, and what are teachers' attitudes towards national tests?  What relationship exists between the principals' facilitative work with national test results and the teachers' practices and attitudes?
<b>Participants (Section 4.2)</b>	Sample 1 12 lower secondary school leaders (principals and senior leaders)	Samples 1 and 2 2 leader teams (7 leaders, including the 12 school leaders from Sample 1), 2 teacher teams (16 teachers)	Samples 2 and 3 176 teachers (including the 16 teachers from Sample 2)
<b>Data (Section 4.3)</b>	School principal and senior leader interviews	Video recordings and on-site observations of leader team and teacher team meetings	Teacher questionnaire
<b>Analytical concepts (Section 4.4)</b>	Policy enactment School leadership and policy interpretation Policy enactment Coupling mechanisms Accountability	School principals, senior leaders and teachers as policy actors Discursive roles in policy enactment Translation and negotiation of policy expectations School context	Teachers as policy actors Crafting coherence and micro politics in schools Relational and contextual aspects of micro-policy practices
<b>Mixed methods credibility (Section 4.5)</b>	<i>Reliability:</i> Codes and categories of the findings in Phase 1 were used for analysis of data in Phase 2. <i>Sequential validity:</i> The design of Phase 1 influenced the design of Phase 2 as the self-reported data were combined with video data and observations in Phase 2.	<i>Reliability:</i> The codes and categories from Phase 1 were re-used to analyze the video recording data and field notes from on-site observations. <i>Sequential validity:</i> The concept of leadership roles and perspectives of power from the findings in Phase 1 were used in the analyses in Phase 2. <i>Sample integration:</i> School leaders and teachers were sampled in Phase 1. Perspectives of leadership and power relations from Phase 1 were combined with analyses in Phases 2 and 3.	<i>Sequential validity:</i> Findings from Phases 1 and 2 influenced the items and thematic areas of the planning, analyses, and contributions of Phase 3. <i>Sample integration:</i> Schools and teachers in Phases 1 and 2 were included in the Phase 3.
<b>Main findings (Section 5.1)</b>	This article shows how test data has become a tool for symbolic responses and focuses on low-performing students. Internal accountability seems to trump external accountability.	This article shows that, while responding to test data, school leaders serve as narrators and enthusiasts, senior leaders and teacher team leaders serve as messengers and enforcers, and teachers serve as critics and the preventers of "overburdening."	This article shows 176 teachers' different perceptions of their principals' role in facilitative work with national test results. These differences correlated strongly with teachers' practices and attitudes regarding their own and the teacher teams' work with national test results.

## Figures and tables

### TABLES

<i>Table 3A:</i>	<i>Overview of policy actors and their policy work .....</i>	<i>22</i>
<i>Table 4A:</i>	<i>Overview of research questions, empirical data, and main findings of the three articles.....</i>	<i>28</i>
<i>Table 4B:</i>	<i>Overview of the observations and video recordings of leader team meetings and teacher team meetings in Phase 2.....</i>	<i>36</i>
<i>Table 4C:</i>	<i>Overview of areas of analytical focus in each phase and data sources used to investigate school principals', senior leaders', and teachers' use of national test results.....</i>	<i>38</i>
<i>Table 4D:</i>	<i>Validation of the sequential approach.....</i>	<i>48</i>

### FIGURES

<i>Figure 3A:</i>	<i>Conceptualizing micro policy making in schools.....</i>	<i>27</i>
<i>Figure 4A:</i>	<i>Research design to achieve the overall aim of investigating school principals', senior leaders', and teachers' use of national test results ...</i>	<i>31</i>
<i>Figure 4B:</i>	<i>Multiple validities in and integrated across each of the three phases.....</i>	<i>45</i>
<i>Figure 6A:</i>	<i>Micro policy making in schools included in the MPM project.....</i>	<i>65</i>

Part II:

The Articles

