

## National testing data in Norwegian classrooms: a tool to improve pupil performance?

Tobias Werler & Margaret Klepstad Færevaaag

To cite this article: Tobias Werler & Margaret Klepstad Færevaaag (2017) National testing data in Norwegian classrooms: a tool to improve pupil performance?, Nordic Journal of Studies in Educational Policy, 3:1, 67-81, DOI: [10.1080/20020317.2017.1320188](https://doi.org/10.1080/20020317.2017.1320188)

To link to this article: <https://doi.org/10.1080/20020317.2017.1320188>



© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 12 May 2017.



Submit your article to this journal [↗](#)



Article views: 282



View related articles [↗](#)



View Crossmark data [↗](#)

## National testing data in Norwegian classrooms: a tool to improve pupil performance?

Tobias Werler and Margaret Klepstad Færevaaag

Teacher Education, Western Norway University of Applied Sciences (HVL), Bergen

### ABSTRACT

This paper considers teachers' use of data from national school tests. These national tests are part of the Norwegian top-down accountability school system. According to official regulations, teachers have to use the test results to improve learning outcomes even if the test system is not able to deliver necessary data. However, previous research has shown that teachers apply teaching-to-test strategies. The focus of this paper is twofold. First, we ask, 'How do teachers perceive and interpret the data from national tests?' Second, 'How do teachers view their actions related to the data from national tests?' We base our research on data from semi-structured 5th-grade-teacher interviews. The transcribed text is subject to qualitative content analysis. We find that teachers are in a state of data illiteracy towards complex Item Response Theory tests. Inspired by Bernstein's concept of the pedagogic device, we see that the test data rules both teacher work in the classroom as well as knowledge provided to the pupils. The national tests seem to undermine teachers' autonomy, restrict teachers' practice and reinforce the impact of unfair structures on pupils' learning.

### ARTICLE HISTORY

Received 15 August 2016  
Accepted 12 April 2017

### KEYWORDS

national tests; teacher work; accountability; Norway; Bernstein; pedagogic device

This article explores how teachers perceive pupil assessment data, and in particular how they reflect on being held accountable for pupil learning (improvement). This is highly relevant, since the use of data for pupils' learning (data literacy) has not been, and is still not, an issue in Norwegian teacher education (Ffl 2015; NRLU 2016a, 2016b; Werler & Volckmar, 2015). Until recently, data were primarily used at the system and policy level (Lawn, 2013) to guide policy decisions and evaluate education reforms (Grek, 2009; Meyer & Benavot, 2013; Prøitz, 2015; Takayama, 2008). However, in recent years there has been a tendency to argue for the use of assessment (of learning) data in classrooms (Thomas & Brady, 2005; Udir, 2014, p. 1; Wells, 2009). In this context, teachers are seemingly held accountable for pupils' learning (Mausethagen, 2013a, b). Such accountability strategies aim to link macrolevels (policy) and microlevels (classroom).

Our article consists of six parts. In the first three sections we briefly discuss the concepts of accountability and national testing, and provide insight into recent research on national testing in Norway. Second, we present what we have coined 'the accountability paradox', which forms the basis for our research questions. Next, we present our theoretical lens ('the pedagogic device', Bernstein, 2000). Then we present our research design and our data, and the last three sections contain our analysis and discussion of the findings.

### Accountability in Norway

In general, the increased emphasis on test data use in education is mainly based on the implementation of accountability policies in various countries (Schildkamp, Ehren, & Lai, 2012). The main objective of accountability policies is to compel teachers to change their classroom practice to achieve improved, measurable pupil learning outcomes. According to Gregory (2003), this can be achieved either by holding teachers responsible for something or by defining expectations for which teachers are answerable. Researching recent Norwegian education policy, Hatch (2013) argued that answerability and responsibility are two distinct but linked aspects of accountability. However, any form of enactment of accountability policy seeks to fulfil expectations set by education and non-education stakeholders (Romzeck & Dubnick, 1993).

In relation to Norwegian research on the introduction of such systems of accountability and competition, we see a threefold effect in Norway. First, the systems have created school markets (Ball, 2007; Elstad, 2009). Second, such systems define what counts as valuable school knowledge (Bachmann & Sivesind, 2012; Rizvi & Lingard, 2010). Third, these policies address inequality in educational outcomes by creating tighter links between the policy environment and instruction (Diamond, 2007; Hallett, 2010).

Engeland, Langfeldt, and Roald (2008) and Elstad (2009) demonstrated that the combination of competition and test system do not really create stakes for municipalities. However, the situation looks very different for teachers in the Greater Oslo Area. Malkenes (2014) reported that teachers experience high-stakes testing since their salaries have been made partially dependent on test results.

We understand these phenomena as results of the enactment of accountability policies (Elstad, Hopmann, & Langfeldt, 2008; Hopmann, 2008, 2013). Such policies bring to the fore a bureaucratic rational choice concept assuming that teachers will respond to accountability policies (Burch & Spillane, 2006; Diamond, 2007). Historically, teacher work has been based on trust in teachers' work quality and teacher autonomy (Werler, 2015). However, it seems that such reliance on trust, autonomy and pedagogic competence is contested by this new governance system (Evetts, 2008; Mausethagen & Granlund, 2012; Karseth & Engelsen, 2013). The accountability policies place greater emphasis on pupils' learning outcomes and focus on teacher accountability for performance (Ingersoll, 2003; Power, 1997; Svensson & Karlsson, 2008). In the following section, we outline how statistical data on pupil learning outcomes and standard-based tests are interlinked, and why teachers must make use of assessment data.

### Why must teachers apply test data?

Neo-institutionalists (Meyer & Rowan, 1977, 2006; Powell & DiMaggio, 1991) have shown that in the past, neither schools nor their instruction were tightly linked to public administration. Rowan (2006) and Fullan (1991) argued that teachers were motivated in their work due to a focus on maximising their own benefits, and claimed that such self-seeking practice prevented pupils from optimal performance and might even have put the nation's economy and welfare at risk. They concluded that tighter links between policy environment, administration and teaching would result in improved learning outcomes.

Accordingly, accountability policies and processes are linked (Thomson, Lingard, & Wrigley, 2012). Such enumerative assessment data create an aura of authenticity and provide arguments for accountability. It has also been argued that numerical data carry explanatory power (Lawn, 2013). In short, it is the narrative about the quality of quantitative data resulting from national tests that links accountability with classrooms. Further, the narrative builds on the underlying assumption that such data enable teachers to better target their teaching to improve pupils' learning via data-driven decision-making (Wayman, & Jimerson, 2014).

Next, we discuss the present state of research on national tests in Norway.

### National tests in Norway

Compared to other research objectives, research on national tests in Norway is rather limited. Existing research is concerned with explaining the changes in test results over time.

Skedsmo (2011) pointed out that the recent standard-based curriculum reform (K-06, 2006) led to a move from an input- to an output-orientated policy. Schools have to ensure that pupils achieve competence aims. In line with the central idea of education accountability (Müller & Hernández, 2010; Sahlberg, 2010), a test system that provides descriptive data on pupils' achievement of educational standards (for a broader discussion, see Linn, 2013) was introduced as far back as 2004.

National tests (Norw. *nasjonale prøver*) currently benchmark pupils' learning outcomes in cross-disciplinary skills in reading and mathematics, and basic skills in English in the 5th and 8th grade. They are not optional. Reading and mathematical competencies are tested in year 9. According to the Norwegian Directorate for Education and Training, the purpose of the tests is to provide 'information to pupils, teachers, school administrators, parents, school owners and the regional and national authorities' (Udir, 2010, p. 5) in order to improve pupils' learning outcomes. The authorities expect teachers to work with the test results as an integral part of their professional practice (Udir, 2014). Therefore, the database contains not only pooled data on school and class performance; any class teacher can also find the performance profile of individual pupils.

National tests measure the cognitive performance of pupils, thus following the tradition of psychometric analysis. The computer-based test system builds upon Item Response Theory (IRT) analysis (Udir, 2016). All the test items in the national tests have been developed at universities (Oslo, Bergen, Stavanger and Trondheim) (Udir, 2016). Typically, those in charge of test development belong to the academic research community and therefore pursue other interests than those of teachers. However, this weakens relationships to school practice.

Evaluation of the national tests has found that the further removed people are from actual teaching, the more they support the system, and vice versa (Allerup, Kovac, Kvåle, Langeldt & Skov, 2009). Tveit (2014), investigating the entire assessment system of Norwegian schooling, argued that the national tests have contributed to 'holding municipalities and schools accountable for their pupils' results' (p. 232). Seland, Vibe, and Hovdhaugen (2013) emphasised that such tests are valued as a tool for improvement

efforts by school leaders. Furthermore, 35% of teachers they interviewed expressed that they practised test-relevant tasks throughout the school year, while 61% admitted that they practised test-relevant tasks shortly before the pupils took their tests (p. 107).

In the following, we discuss, based on current research, how teachers cope with test data. We also discuss research highlighting typical issues and problems linked to teachers' work with test systems.

### National test data and teachers' work

There is hardly a teacher in Norway who is unfamiliar with the terminology of formative assessment and assessment for learning (Black & William, 1998). Whilst both concepts have to some extent been part of the teacher education curriculum for several decades, this is not true for pedagogical data literacy. Pedagogical data literacy is framed as the ability to transform information (assessment, school climate, behavioural, snapshot and longitudinal, etc.) into actionable teaching concepts (Mandinach, Firedman & Gummer, 2015). Following Pierce and Chick (2011), it seems reasonable to assume that Norwegian teachers can read values and understand features such as scales or graphs, and interpret specific data points within graphs or tables. Yet it is rather unlikely that teachers are able to compare, contrast and critique multiple datasets, or that they have knowledge of the school contextual factors (e.g. pupil demographics and local events) that gave rise to the data. Irrespective of whether they have this ability, they are confronted with IRT-based assessment data that are subject to debate and at the same time lauded as providing meaning and facts (Desrosières, 1998). Moreover, the Directorate of Education and Training has admitted that the tests are unable to detect causes for achieved results since they are 'one-dimensional constructions' (Udir 2016, p. 9). Thus, teachers have to 'interpret' the test results (Udir 2016, p. 8), since such test systems cannot provide diagnostics for groups or individual pupils. Against this backdrop, research has shown that a major cause of pupils' learning outcomes is parents' level of education (Grøgaard, Helland, & Lauglo, 2008) – a factor that teachers cannot change.

This creates a paradoxical situation. First, teachers are held accountable for results they can influence only slightly, since parents' level of education is the most important factor. Furthermore, it is difficult for teachers to improve pupils' learning outcomes because they do not know which variables they can (or should) change due to the limited data provided by the tests and the teachers' limited data literacy. Based on these observations, it is reasonable to argue that teachers have to guess what causes poor test results if they wish to improve pupils' learning

outcomes. Guidelines from the Norwegian authorities also recommend this strategy (Udir, 2014, p. 6). In light of their commitment to pupils (responsibility), it is reasonable to argue that such guesswork would be experienced by teachers as somewhat unprofessional. In order to help pupils, they are likely to develop evasive strategies.

This paradox is also reflected in some empirical data. Chavannes, Engesveen and Strand (2011, p. 36) found that school owners, as well as school leaders, have developed structures they judge as valuable to improve results (concentrated teacher resources, staff training, provision of materials to improve teaching and learning). Beyond that, they found that the main strategy used by schools is discussing factors that may explain the test (Chavannes, Engesveen & Strand, 2011, p. 39). Waters (2013) revealed that there is a negative correlation between test results and schools' internal use of management by objectives. Overall, Isaksen and Hjelm Solli (2014), investigating school owners', school leaders' and teachers' work with test results, found that routines and plans for follow-up initiatives were missing. Uncertainty concerning how to use the results was also evident (Isaksen & Hjelm Solli, 2014, p. 42). Johansen (2015) found that teachers use test results for ability streaming. Evaluation of the national testing system revealed that teachers are frustrated with the information outcome from the tests, and are not prepared for providing feedback to pupils (Seland et al., 2013, p. 101).

Still, even international research has indicated that teachers struggle with using data to inform their own practice. Teachers are struggling with data systems, time use and lack of knowledge about how to use data to improve instruction (Anderson et al., 2010; Goertz, Olah, & Riggan, 2010; Valli & Buese, 2007; Wayman et al., 2012). Overall, research has shown that teachers are insufficiently prepared to effectively integrate assessment results into their practice (DeLuca & Bellara, 2013; Wayman, & Jimerson, 2014). It is, however, striking that existing research has not sufficiently investigated how teachers use test data to help pupils to improve their learning results.

### The research problem

Based on the above observations, we wished to learn how teachers experience this situation in which they are held accountable for results they can to little extent influence, while also having limited access to information about variables they could change. In this context, we operationalise the research problem by asking two questions: First, how do teachers perceive and interpret the data from national tests? Second, how do teachers view their actions related to the data from national tests? To answer the first question, we study teachers' experience test

data in a low-stakes system in which teachers feel responsible for their pupils (Hatch, 2013). By proposing the second research question, we aim to understand not only how teachers **enact policy**; applying the concept of the pedagogic device (Bernstein, 2000) will also help us to uncover the **internal grammar of the test system**. The answers to both questions will provide insight into how teachers cope with education accountability in order to avoid a deadlock that could possibly put their professionalism at risk.

In the following section, we outline our theoretical lens – the Bernsteinian concept of the pedagogic device. The concept of the pedagogic device allows us to identify how national tests function as a relay for policy dominance over teacher autonomy in Norwegian classrooms, since it is able to show how knowledge about test results is transformed into pedagogic actions.

### Analytical lens: a Bernsteinian reading of national tests

National tests are part of a wider policy design connecting test scores with teachers' accountability (Hatch, 2013). We see the policy as 'a multidimensional and value-laden state activity that exists in context' (Fitz, Davies, & Evans, 2005, p. 34). Policy is not a text or a document alone; rather, it is a process of organising specific rationalities (Ball, 2008), and merging different values and contingencies with a specific context (Maguire, Ball, & Braun, 2010). In this process, we find interpretation of interpretations (Rizvi & Kemmis, 1987) translating texts into contextualised action on both administrative and social levels. This has been termed 'policy enactment' (Ball, Maguire, & Braun, 2012, p. 3).

One finds, at the end of this translation chain, that teachers are enacting policy directives in classrooms through a series of mediations (Ball, Maguire, & Braun, 2012). Such enactment is expressed in teachers' work with national exams, i.e. its practical application in the classroom and the local evaluation of results, as well as by the reflection and work towards pupils' performance improvement.

According to Singh (2015), it is possible to characterise national exams as a cultural relay between macro and micro level. The tests, as well as corresponding instructions for how to use them, convey knowledge from mid-policy actors, such as the Norwegian Directorate for Education and Training (NDET/Udir), to teachers working (giving lessons, evaluation, counselling, etc.) in classrooms. Bernstein operationalises such enactment of policy using the concept of the pedagogic device (Bernstein, 2000; Bernstein & Solomon, 1999). The device is not of a technological nature; it refers to processes of applying rules to control the awareness

of actors. According to Bernstein's model (2000, p. 37), distributive rules are fundamental; they serve the production of knowledge. Recontextualising rules transform such knowledge, and in turn produce evaluative rules.

*Distributive rules* regulate the power relationships between social groups (Singh, 2002) by distributing different forms of knowledge. Wong and Apple (2003) stated, more precisely, that such rules facilitate social order through knowledge distribution and the formation of social group consciousness. Au (2008) pointed out that distributive rules not only deliver curriculum standards but also favour certain types of knowledge, e.g. via the implementation of test systems. The *recontextualising rules* are dependent on the distributive rules (Singh, 2002). Through recontextualisation, the test discourse is moved from its original site of production (universities, public administration) to another site (schools). Since the test knowledge is created at universities, it is not identical to school knowledge. It is important to note that recontextualisation of (truth-based) test knowledge turns it into pedagogic discourse (Udir, 2010, p. 5). On the third level in this hierarchy are *evaluative rules*, which constitute specific pedagogic practices (Singh, 2002). In broad terms, these rules dictate what teachers will recognise as valid modes of teaching. It has been shown that in light of the purpose of the test system, teachers tend to be primarily concerned with pupils' acquisition of curricular content (Au, 2008).

Next, we briefly present both the research design and the applied method of analysis (qualitative content analysis [QCA]). We then present the categorised data, before moving on to the analysis showing how test data function as a pedagogic device. In the final section, we discuss our findings in relation to the manifold challenges to teacher professionalism.

### Design and method

To operationalise the research problem, we carried out semi-structured interviews in which we invited teachers to talk about their thoughts and experiences regarding the national test paradox. We focused mainly on three phases: We asked the teachers (1) about processes regarding the time immediately after pupils completed the tests; (2) to elaborate on their reflections after being informed about the results; and (3) to talk about how they use the provided data to help pupils. By asking questions related to the first topic, we wanted to learn about teachers' cognitive work in a state of uncertainty, knowing that future interpretation of test results will be the subject of public opinion. For the second topic, we tried to gather informants' general response and attitudes towards the national tests. The last topic was

developed mainly to collect data on how the teachers act and cope with pedagogically paradoxical situations.

The transcripts are based on data collected from individual semi-structured interviews ( $n = 18$ ). The informants represent six different schools (*barneskole*, grades 1–7), and all interviews were carried out weeks before the 2016 tests. Sites selected for data collection represent various rural, suburban and urban schools. All informants are experienced teachers (women, 35–62 years) who have arranged national tests in the subject Norwegian several times. All teachers are trained (via a general teacher training programme, three to four years) and have taken part in further education. We chose teachers working in the 5th grade, since Seland, Vibe and Hovdhaugen (2013, p. 101) identified 5th-grade teachers as the being least content with the current situation. Differences in age, period of training or extent of experience were of minor importance for the analysis. Retrospective questioning was conducted in order to capture teachers' reflections and thoughts about past actions. We took into consideration that retrospective narration will inevitably lead to some blurring of factual information, and were conscious of this effect throughout the interviews and the analytical work. Since the units of analysis are teachers' reflections, we treat the voices of the teachers as a shared voice, even if this means that nuances contained in the data set will not be shown. Following Yin's methodological approach (Yin, 2003), the case study allows for greater understanding of teachers' actions.

In the analytical work, QCA (Kohlbacher, 2006; Mayring, 2002, 2015) was used as a method for systematically understanding the text. Applying QCA means looking for themes, meanings and context in order to build a picture of teachers' 'emplaced everyday experiences', as well as to gain insight into how they 'understand and frame [such] experiences' (Wiles, Rosenberg, & Kearns, 2005, pp. 97–98). Kohlbacher (2006) emphasised that QCA not only takes a holistic approach, but also covers the complexity of the social situations. In our case, the empirical material builds on transcripts that were used to identify deductive categories of meaning. The deductive categories for the interview guide (and analysis) were generated based on the findings of Seland, Vibe, and Hovdhaugen (2013). The category system represents the latent meaning of the analysed material. The system functions as a starting point for interpretation of the text, and is the heart of the analysis. We identified the following relevant topics: having completed the tests, and improving pupil achievement (Seland, Vibe, & Hovdhaugen, 2013, pp. 101–130).

Informed by Mayring (2015), a clear meaning component analysis was chosen as coding unit for the first cycle of the coding process (the entire material). Weft QDA software was used for coding and analysis. Since the text of the empirical material consisted of interview transcripts, we used word groups or statements that could consist of several coherent sentences as coding units. As a coding rule for the material, we decided, in accordance with the deductive concept of the research project, to follow the three central topics of the semi-structured interviews.

The structure of our analysis was operationalised based on the aforementioned categories. Accordingly, we developed a categorisation matrix to review the transcripts and code the data according to the categories (Elo & Kyngäs, 2008). The data were subsequently classified into much smaller content categories. In practice, we analysed the empirical material first by coding all of the teachers' interview data using the topic code 'period after pupils have completed tests' (1). Next, we coded for 'thoughts and ideas about pupils' results' (2). We then coded for 'using data for improving pupils' learning development' (3). Using this matrix allowed us to distil teachers' individual responses down to crucial elements. We have chosen to use direct quotes to illustrate important features.

Since we considered only 5th-grade teachers working in primary schools, our findings are limited to that group of teachers. Furthermore, we have to take into consideration the fact that the teachers condensed their retrospective reporting about past events and practice. The teachers might also have influenced the findings due to possible hidden agendas, despite the great care taken in categorising the data.

In the next section, we present the categorised empirical material and a stepwise analysis. With regard to the empirical material, we use letter numbers (according to the data file) to indicate statements made by informants. This methodology does not allow for tracing of individual informants. We use this mode of presentation since it is the collective statements, rather than single informants, that are of importance here.

## Findings

In the following section, we present prominent issues related to our data. These issues include how teachers approach the test results, how they cope with them and how they work in relation to the test data.

### *Acting without knowing*

In the interviews, teachers were asked to talk about their experiences and thoughts immediately after their class had completed the national tests. The

teachers were not asked to indicate the length of this period. This retrospective questioning was used in order to uncover teachers' understanding of the work-related value of the national tests.

### **Parents as stakeholders**

The major topic raised by the teachers was their future communication of the results from the national exams to pupils' parents. Even if the teachers did not yet know the individual or class results, their first thoughts concerned parent evenings [283–331] and parent–teacher conferences [14,545–14,563]. Both arguments are characterised by reflections on the upcoming presentation of the test results to a public audience [39,893–39,974]. Furthermore, they told us that parents had generally received quite positive feedback prior to the time of publication of the test results. As such, they expressed concerns that the results might not match earlier communication about pupils' learning results [40,114–40,220]. The concern expressed by the teachers reveal that they experience / parents as perceiving the results from national tests as very important information, and that parents look forward to finding out the results. The teachers seem to believe that parents have confidence in the test results representing the 'truth' about the performance of the class their child is attending.

### **Teacher insecurity**

Teachers reported several emotional responses related to the tests. Primarily, they described the period after completion of the national tests as stressful [14,293–14,318], and that they experienced periods of hectic activity [14,421–14,436]. Again, they reported stress related to assumptions that pupils 'did badly' [450–494] when they did not yet know the results. While this response is surprising, teachers talked about their desire to know the results [52,394–52,418], [39,575–39,595], [52,010–52,060].

### **The school level**

The teachers' initial reflections about this period seem to be influenced by school-based processes related to the national tests. In contrast to their own experiences, the teachers reported that 'nothing happened' at the school level right after the tests [22,918–22,943], [46,737–46,858]. The teachers reported that the schools would generally 'put the results aside' [46,737–46,858] and 'just carry on' [51,844–51,892]. It was also pointed out that the headmaster suggested making use of the tests: 'the headmaster said that I could find individual pupil reports online' and 'use them in conversation with the pupil' [23,206–23,359]. Interestingly, the teachers did not mention that headmasters gave advice on how results could or should be used. The teachers

indicated in the interviews that they did not know how to use the pupil profiles.

### **Proficiency of experience**

Finally, the teachers reported that right after the tests they conducted informal conversations with other teachers [14,506–14,528] or the head of department [14,875–14,896] at their school on how they thought pupils performed. The dialogues focused mainly on the topic of confirmation of teachers' experience-based everyday theories about 'their' classes' levels of performance. Teachers expressed collective doubts about the accuracy of their non-data-based judgements of their classes' performance. Accordingly, they stated that they were concerned about whether the results 'show the same picture as the one we see' [52,010–52,060]. They also discussed specific test items they thought were in need of further explanation [252–281].

### **Getting to know the results**

A second cluster of questions posed to the teachers concerned their reflections on the publication of results. In this section, we tried to unearth the informants' general responses and attitudes towards the national exams. The dominant response from the teachers concerned their everyday theories. The teachers reported that the test results confirmed their implicit knowledge about the performance levels of their classes. A second observation was that teachers talked about the discrepancy between test results and teachers' individual performance judgements of individual pupils.

### **Between confirmation and surprise**

The teachers mentioned that the test results provided did not contain 'big surprises' [47,726–47,828] at the level of whole classes [48,024–48,099]. Another teacher said that 'I thought it went as expected' [1349–1416]. Typically, the teachers mentioned 'that they (the teachers!) know where their pupils are at' [57,395–57,466]. In other words, the teachers maintained that their everyday theories about pupils' performance levels coincide with measurement results. While teachers expressed experiences of confirmation regarding the class as a whole, we also found statements of surprise. This is particularly true for individual pupil results. Here, the teachers talked about positive and negative deviance from expected test results [1083–1170], [15,142–15,297]. A characteristic statement is: 'I got some positive and some negative surprises' [42,421–42,459]. A suggested reason for this is that teacher(s) 'possibly did not know well enough what they (the pupils) could (achieve)' [1172–1218].

### *Teachers' use of data*

An interesting observation is that teachers invited to share their general thoughts and ideas about the results 'jump[ed] to conclusions regarding actions' to be taken. They primarily expressed a sense of competition. Typically, teachers expressed that they learned about reading performance and 'whether we are close to results we [the school] want to achieve compared to other schools' [23,694–23,906]. Teachers also claimed that 'the results cannot help them at all,' [56,710–56,761] and that they need other tests to help them understand their situation [52,664–52,767]. A point the teachers typically made was that the results could potentially be abused by the municipal administration [56,762–56,844]. Others came to the immediate conclusion that they have to make use of teaching-to-test methods, based on repetition and practising tasks from the tests [16,861–16,921], [32,187–32,234]. One of the teachers said that she 'learn[s] what I have to practise even more' with the pupils [16,798–16,835].

### *Working with a paradox*

As already mentioned, teachers are expected to improve and develop the conditions for pupils' learning based on data from the national tests. In order to gain a better understanding of the teachers' general thoughts about and experiences with the national tests, we asked them to talk about the primary baseline for their work. Their answers fall into two categories: one relating to the enactment of governance policies (i.e. accountability) and management of expectations, and the other to the ability to offer help and support to pupils.

### *Governance, expectations and accountability*

In general, the teachers clearly stated that national tests are primarily a policy tool for 'improving Norwegian test results' [16,508–16,606], implicitly referring to the recent wave of large-scale assessment tests on which Norway scored at a level similar to other industrialised nations. Nevertheless, the teachers seem to have incorporated testing – conceptualised as non-diagnostic benchmarking – as a beneficial concept [40,312–40,422], even if they were unable to elaborate on what, exactly, they consider positive. It is possible that our informants tried to demonstrate loyalty towards the current governance system.

In relation to their pupils' results, the teachers were fairly satisfied [14,439–14,484]. Although not the official intention of the tests, they pay attention to other schools and their results. The teachers indicated that they talk about their school's results in order to rank and compare results with other schools or municipalities [1888–1993]. However, as soon as

teachers start to compare their pupils' results with the results of others, they feel that they are being held accountable. The teachers generally perceive that the results are not good enough, regardless of what the results actually are. They even go further, stating that they themselves are not good enough and that they do not do enough to help their pupils [14,740–14,763].

The teachers seem to assume that national tests are a positive, but they are at that the same time uncertain of how to help pupils improve and develop their learning, other than trying to change their approach to instruction. Commenting on activities linked to development work, teachers expressed that they accept that they cannot actually do much based on the data provided ('we don't get so much done anyway' [2272–2371]), and they confessed that they do not know how to help their pupils [14,740–14,763]. Furthermore, they expressed that they intentionally limit their activities. They said that 'we are just ourselves and do not have access to additional resources' [2414–2475]. Concerning efforts to make changes to instruction, they mentioned that they focus even more on developing pupils' reading proficiency [42,971–43,066], [43,100–43,257], [48,273–48,346].

### *Development and causes*

Overall, teachers are struggling with trying to meet the demands for supporting pupils' learning based on results from the national tests. Since teachers are mainly driven by altruistic motives (Ffl 2011, p. 41), it is no surprise that they assume the role of advocates for their pupils. They appear to try to understand the causes for the test results, and the test items on which a class performed badly, in order to find data-based evidence for why the results are the way they are [24,402–24,599]. This search for causes is mainly related to classes of test items [24,402–24,599]. Hence, one teacher pointed out that many failed to find data-based causes that could be used as starting points for helping the pupils [60,034–60,088]. It is worth mentioning that the teachers did not raise doubts about the tests or the test procedures. In their search for a cause, they take the pupils' perspective; they ask whether pupils simply 'had a bad day' [2060–2148] or 'have problems with interpreting tasks' [3147–3231]. Another informant wondered whether the tasks are just too demanding and complex [9785–10,151].

Throughout the interviews, the teachers talked about how to explain pupils' results causally [24,402–24,599]. They do so by guessing at underlying causes for the results. At the same time, they contest the validity of the pupils' results [9785–10,151]. From the teachers' perspective, this implies a further need for in-depth testing [9785–10,151].



They indicated that they are struggling with results and their pupil-related meaning [24,402–24,599]. According to the interview data, teachers do not in general know what challenges individual pupils' are faced with [59,947–60,033], [60,034–60,088].

### Analysis – the grammar of the national tests

When reviewing our data, we see that national tests are part of teaching practice. This allows us to see the national tests, including related data, as a pedagogic device (Bernstein, 2000; Singh, Thomas, & Harris, 2013). Our findings indicate that teachers actually do work with the tests; they make creative interpretations of a policy tool (Ball et al., 2012) in order to deliver education. In the following, we wish to show how the grammar of the test system functions.

#### Distributing knowledge

The national tests operate on an epistemological level; they distribute 'test knowledge'. As with other tests, national tests codify disciplinary knowledge created by scientific research at universities. Such expert knowledge is encoded in highly complex symbolic forms in the tests, i.e. the measured knowledge as well as the test theory. When teachers work with the tests (preparing instruction, arranging the tests, evaluating the results), they have to decode the test knowledge in order to access the tests from the outside. However, teachers are specialists in neither subject matter domains nor in test theory. This contradictory situation leads to several different responses amongst teachers.

Teachers' work depends to some extent on public opinion. Respondents in our study are sceptical of public opinion, and worried that any negative parent feedback on pupils' performance may influence (local) policymakers (school administration) to change course. However, the teachers seem to accept that parents have faith in information from testing metrics. They construct parents as powerful stakeholders, since they anticipate knowledge about parents' views on pupils' performance even before parents have had the chance to make statements about the issue.

These assumptions of anticipatory obedience are reasonable, since teachers start developing strategies that might help them justify the results when they have to present them to parents. In fact, teachers make unjustified guesses about the results based on their teaching experiences with the pupils. However, they seem to make those guesses under the influence of ideas about public opinion suggesting that their work is of poor quality. As their emotional responses indicate, they are concerned that the publication of test results will undermine the image of pupils'

performance they have previously presented to parents. Before the national tests in grade 5, teacher reports are the only source of information available to parents about their children's school performance; grades are not given until grade 8.

Distributive rules of test designers and policy-makers determine who has the power to decide. That causes a situation in which actors who are physically and practically distant from classrooms (Apple, 1995; McNeil, 2000) decide what counts as legitimate pedagogic discourse. As such, public administration shapes certain pedagogic orientations of teachers, who have to work with (i.e. enact) knowledge about the tests and tested knowledge. The unspecified dissonance between experience-based knowledge and test data functions as a driving force, making teachers comply and work with one-sided teaching strategies in order to improve test results (Seland, Vibe, & Hovdhaugen, 2013). Furthermore, we see an inclination amongst teachers towards regarding tested knowledge as legitimate, while untested knowledge is viewed as illegitimate, for the pedagogic discourse. Above all, those rules question teachers' proficiency and experience.

#### Recontextualisation of distributive knowledge

When teachers talked about publication of the test results, they framed the expression of their experiences by mentioning that the tests offer no new or relevant information. The results are no big surprise, the teachers said. In other words, the teachers' experience-based knowledge about pupil performance is confirmed and contested at the same time. Taking into consideration that the test results deliver very detailed data, the teachers indicated indirectly that they do not wish to deal more extensively with the results. Such argumentation indicates that teachers recontextualise distributive knowledge.

A possible explanation for this is that their (professional) intuition, based on daily work with pupils, provides enough information about pupils' achievement. Framing these claims from a traditional view on professionalism (e.g. Abbott, 1988; Brint, 1994; Larson, 2012; Lortie, 1975), this seems reasonable, since it indicates that teachers are able to make valid and valuable judgements about the quality of their professional work. This claim is supported by the fact that teachers present immediate actions to be taken. Further, teachers argue that there is no need for further in-depth analysis, since they have already made their 'reliable judgements' (teacher beliefs) in advance.

Au (2008) pointed out that such recontextualisation communicates knowledge containing a theory of instruction as well. Even if teachers to some extent tend to 'teach-to-test' at a modest level, such practice

indicates the potency of the test knowledge when it comes to controlling schools and teachers' curricula. That teachers tend to adopt their pedagogies to meet the test-defined knowledge structures, as illustrated by Ball (Ball, 2003).

Teachers obviously tend to be convinced of the efficacy of their work and reject the opportunity to perform deeper analyses of the test results. Those findings can be seen as indicating that teachers implicitly argue for everyday theories to be maintained and not replaced by research-based data produced outside the local school, which they do not understand. Teachers' recontextualisation of distributive knowledge indicates that their discretion-based judgement and experience-based knowledge is devalued by powerful stakeholders.

### **Evaluative rules in the classroom**

In order to comply with their accountability, teachers tend to regulate the selection of content, the form of its transmission, as well as pupils' social conduct. In other words, national tests combined with the demand to improve pupils' learning outcomes function as physical manifestations of the evaluative rules in the classroom.

Even if the Norwegian test system is characterised as a low-stakes system, teachers feel as though they are held accountable. Although they argued that test results of their classes are not good enough, they also indicated that their quality of work is not good enough and that they do not do enough to help the pupils. Moreover, they indicated that they try to understand the test items in which their classes performed badly in order to find data-based evidence for the results. In their search for a cause or explanation, they take the pupils' perspective by guessing at underlying causes for the results. Such measures indicate that teachers are struggling with the meaning of the data, and that they do not learn about pupils' individual challenges. Interestingly, teachers seem not to cast doubts on the tests or the test procedures themselves.

The procedures connected to the outlined use of national tests and their results seem to have an impact on teachers' selection of content, on how they give lessons and on how they distribute knowledge to groups of pupils. Thus, the teachers' awareness of the tests has the power to define how they specify 'suitable contents under proper time and context' (Wong & Apple, 2003, p. 85). As such, the tests and the data produced function as a data-determined manifestation of power over classroom practice by non-professionals (stakeholders, politicians, administrators). In light of the above observations regarding accountability, one can say that the national tests constitute a policy tool in the classroom. Overall,

the analysis reveals that tests work as a symbolic ruler controlling teachers' autonomy.

In the following, we sum up our arguments to answer the research questions. First, we asked how grade 5 teachers perceive and interpret data from national tests. Based on the analysis, we infer that the experience of those teachers falls into four partially overlapping areas: power, expertise, professionalism and accountability. Regarding the power dimension of test data, teachers feel helpless facing powerful stakeholders (parents). Teachers experience that they have been pushed into a powerless situation, where authority over their work is given to parents. It might be possible that teachers ask whether parents are the new experts of pupils' learning. Teachers also feel that the data provided by the test system steers their focus. However, at the same time they are not convinced about the importance of the tests for pupils' success in life. In other words, teachers feel that the part of their work that is not test related becomes insignificant. Second, teachers position themselves as 'lost in translation' when it comes to reclaiming expertise. They clearly have a feeling of being non-experts with regard to data interpretation and use. In the current accountability system, they experience a devaluation of their expertise by non-school agents. They argue that those non-school agents (test designers at universities; local school boards) are setting the agenda for what is acknowledged as valuable test knowledge. Somewhat speculatively, one may argue that teachers as non-experts question the legitimacy of the tests and test data. Third, in relation to expertise, teachers experience that their work with the test data is non-professional. Hence, they see their work as incompatible with their moral code of conduct as defined by the teachers' union (Union of Education Norway, 2012). They feel that 'data-driven' work has nothing to do with their pupils.

Fourth, teachers experience that they must assume responsibility for results that they can only slightly influence compared to the impact of parents' socio-economic situation. In particular, the fact that they take the core idea of testing into the classroom points to the power of implicit social control in schools, which is virtually invisible to outsiders. Seen from an accountability point of view, teachers are caught between the contradictory demands of school administrations and the needs of the pupils for whom they are responsible.

Our second research question was concerned with how teachers enact the accountability policy. We have been mainly interested in how they think about their actions related to data from the national tests. Applying our theoretical lens (the pedagogic device) enabled us to identify three enactment processes used by the teachers. As non-experts concerning test

theory, teachers accept the validity of the test data and distribute it to a public audience (parents), and we see that they think their work is construed as being of low quality among public stakeholders. A second trend identified in our material is that teachers are working with recontextualisation of knowledge provided to the public audience to defend their pupil expertise. As a consequence, they think that understanding the data is unnecessary. At the same time, they think that their expertise is devalued. Finally, we see that teachers take the evaluation rules of the test system into the classroom. On the one hand, teachers think that they have to select curriculum content according to the tests' knowledge domains. On the other, they feel that they are held accountable for the results achieved and therefore have to guess what the causes for these test results may be.

### **Discussion: lessons learned**

As mentioned earlier, we see the national tests as a pedagogic device. The tests are, as such, an ensemble of rules enacting policy as teaching practice. National tests stand out as public communication and rule teachers' work (Bernstein, 2000, p. 26). Because these rules are hierarchically ordered and mandatory, they recontextualise classroom practice, as well as teachers' autonomy. It is mainly the mandatory aspect of the national tests that defines what has to be regarded as 'important knowledge' (Bernstein, 2000, p. 31). In the following, we discuss how teachers think about their accountability in a setting where they have little influence on key factors of their work and limited data literacy. Finally, we identify how national tests function as a relay for policy dominance over teacher autonomy.

### **Accountability & data literacy**

National tests stand out as an evaluation policy that is not an integral part of the triad of curriculum development, enacted pedagogy and a school-based system of evaluation serving the local development of a community school. The national tests have been made part of teachers' work. Historically, such evaluation work was a natural element of teachers' professional actions. As the interviews reveal, teachers have not chosen to perform national tests on the basis of a professional need. The limited use of results from the national tests points to what Bernstein calls a 'meaning gap' (Bernstein, 2000, p. 30). Furthermore, teachers suffer from not having the level of data literacy needed to help them to understand the significance of the results. However, the meaning gap gives teachers room for manoeuvre: teachers expressed that they ignore the test data or apply further diagnostic test

systems that might possibly help them to act responsibly towards their pupils. Nevertheless, we also see some contradictions. The teachers exhibited a positive attitude towards the tests. It is thus reasonable to argue that teachers can read the results as an indication of whether they emphasise the 'correct' knowledge. By looking into the test data, they may discover to what extent their classroom practice complies with current educational policy. In other words, for the teachers, the national tests function as a loyalty indicator for them. This might possibly be a result of a lack of confidence that they have the psychometric and statistical knowledge needed to interpret the test results, which is not and has never been part of Norwegian teacher education.

### **Dominance over teacher professionalism**

In the following, we discuss the topics of autonomy, restricted practice, and time allocation with regard to the limitation of professionalism.

### **Contested teacher autonomy**

The presence of national tests in classrooms seems, through the tests' foundation in computer capacity and 'datafication', to have contributed to an epistemological shift from concerns of causality and understanding to concerns of correlation (Mayer-Schonberger & Cukier, 2013, pp. 61–67). Our study, as well as several other studies (see above), allow for the conclusion that national tests create an illusion (Ball, 2013, pp. 66–68) to policymakers (both national and local) and headmasters that it is possible for teachers to fulfil expectations defined by others. As we explain in the following, the findings in our study indicate that national tests lead to what we call 'relative teacher professionalism'. Such professionalism is characterised by centralised decision making about teachers' work. In our case, we see that teachers have to make use of data resulting from the national tests. It is no longer a matter of professional judgement. Further, teachers command on limited authority to find solutions, since major decisions are made by those who are far from classrooms. However, even teachers limit their authority by applying teaching to the test strategies or by guessing. We cannot see teachers developing scientifically based efforts to understand and interpret test data. They solely apply embodied tacit knowledge.

The varying responses of the teachers involved in our study regarding the national tests show that the very existence of the test results creates a situation of inherent conflict or struggle. Test results give rise to actors (groups) inside and outside the educational profession defending or criticising the tests and the results (for a broader discussion, see Aasebøe, 2015, pp. 60–61). According to Bernstein (2000), it is the

device itself (i.e. the test system) that creates an arena of struggle. In other words, the test system partly transmits teachers' professional power and control to actors outside the profession.

Teachers' reflections about and professional response to rules made by others (that they have to carry out the tests) not only create a new power context for teachers' work and pupils' learning. Powerful authorities (e.g. the government, NDET and local policy actors) increasingly define the scope of teachers' pedagogic actions. Teachers' work with the national tests not only indicates policy control over curriculum knowledge, but changes how teachers conceptualise and plan their classroom practice, in terms of what they deem valuable to teach. The test system may create a situation in which teachers lose authority over subject matter expertise.

Furthermore, we see some conflict in the teachers' professional self-understanding that might undermine their professional power. This conflict has its origins in what teachers value as reliable data. On the one hand, teachers evaluate national tests as a professionally non-reliable source of information. On the other, they do not criticise its psychometric approach. The resulting contradiction suggests that teachers value their experience, as well as their beliefs, as sources of reliable data.

### **Restricted teacher practice**

In light of the professionalisation efforts, the way teachers use data suggests that their relative pedagogic autonomy is contested. The tests and the data function as a data-determined manifestation of power over classroom practice by non-professionals (stakeholders, politicians, administrators). The teachers' statements nevertheless show that they have accepted the testing system. The missing open protest indicates a general pedagogical shift towards acceptance of test contents and test logics. As our study reveals, teachers use test data not as instruments to help individual pupils to fulfil their potential; instead, they primarily alter their teaching practice to help pupils improve their results on future tests. This is made evident by the improvement strategies teachers have chosen: they assign tasks that pupils were previously struggling with. Since teachers do not know the causes of pupils' test results, they can only hope that repetition of tasks pupils were struggling with, or repetitive and intensified practice of procedures they have to know, will make those pupils 'get the point'. The teachers' responses to the issue indicate that teachers' data use builds on the assumption that a 'more-of-the-same' strategy will improve individual pupils' learning results.

### **Time allocation and unjustness**

The pedagogic device, i.e. the national tests, limits teachers' professional space since they are forced to take an unequivocal stand. Whether teachers are criticising or defending the test system, their autonomy is relative. Teachers' autonomy is constrained by the tests. As a rule, it privileges certain kinds of testable knowledge. Independently of whether teachers resist or accept the test system, they have to allocate some time to the issue, and that time can no longer be allocated to pupils who need specific support. This creates an unjust situation. In other words, the pedagogic device of 'national testing' demands (implicitly) of teachers the use of various forms of time allocation, which intensifies unjust processes of schooling (Au, 2008). Consequently, the national tests and teachers' work with or struggle against them create conditions for society's reproduction (Bernstein, 2000, p. 53). Hence, teachers' work with or against the tests manifests existing social structures and limits a society's capacity to change. In particular, teachers' work with the test system regulates pupils' identities, as well as pupils' educational success.

### **Conclusions**

The implementation of the national test system as part of teacher accountability in Norway has, over time, created a situation in which teachers experience and interpret data from highly complex IRT tests in many ways, since teachers have limited data literacy. Interestingly, the teachers in this study did not address their data illiteracy in the interviews. Rather, they brought the results into collegial discussion. The teachers also did not discuss whether the tests are linked to the curriculum they have to teach. Rather, they discussed 'what those students brought to schools' (Popham, 2007, p. 167). Furthermore, they assume only partial responsibility when communicating the results to parents. In this state of data illiteracy, they engage in pseudo-professional argumentation to justify or explain results. Obviously, they build on the assumption that most of the parents are not qualified to judge either the results or the teachers' explanations. However, they express responsibility for the pupils' learning. Based on the fact that national tests are of a low-stakes nature, teachers ignore the accountability paradox we outlined earlier. It is possible to ask whether they demonstrate resistance to current policy by their ignorance. We wonder whether the tests have the power to confuse teachers about the knowledge they should focus on. Irrespective of the test results, teachers do not know how they can help pupils achieve better test results. Our research suggests that one should strengthen efforts to

develop a test system that makes sense for both teachers and pupils.

As we see it, national tests in classrooms stand out in the empirical material as being reductive and decontextualised. We cannot see how national test data can help teachers improve classroom practice and thereby facilitate holistic learning for all pupils. Beyond that, we learned that further research should focus on how national tests contribute to even sharper reproduction of a society's inequality.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

- Aasebø, T. (2015). Teachers' discourse about the Norwegian national tests. In S. St. Hillen & A. Aprea (Eds.), *Instrumentalism in education – where is Bildung left?* (pp. 59–71). Münster: Waxmann.
- Abbott, A. (1988). *The system of professions. An essay on the division of expert labor*. Chicago, IL: The University Chicago Press.
- Allerup, P., Kovac, V., Kvåle, G., Langfeldt, G., & Skov, P. (2009). *Evaluering av det Nasjonale kvalitetsvurderings-systemet for grunnsopplæringen*. Kristiansand: Agderforskning.
- 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. doi:10.1080/15700761003731492
- Apple, M. W. (1995). *Education and power*. New York, NY: Routledge.
- Au, W. (2008). Devising inequality: A Bernsteinian analysis of high-stakes testing and social reproduction in education. *British Journal of Sociology of Education*, 29(6), 639–651. doi:10.1080/01425690802423312
- Bachmann, K., & Sivesind, K. (2012). Kunnskapsløftet som reformprogram: Fra betingelser til forventninger. In T. Englund, E. Forsberg, & D. Sundberg (Eds.), *Vad räknas som kunskap?: Läroplansteoretiska utsikter och inblickar i lärarutbildning och skola* (pp. 242–260). Stockholm: Liber.
- Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, 18(2), 215–228. doi:10.1080/0268093022000043065
- Ball, S. J. (2007). *Education plc: Understanding private sector participation in public sector education*. London: Routledge.
- Ball, S. J. (2008). *The education debate*. Bristol: The Policy Press.
- Ball, S. J. (2013). *Global education inc.: New policy networks and the neo-liberal imaginary*. London: Routledge.
- Ball, S. J., Maguire, M., & Braun, A. (2012). *How schools do policy. Policy enactments in secondary schools*. London: Routledge.
- Bernstein, B. (2000). The pedagogic device. In B. Bernstein (ed.), *Pedagogy, symbolic control, and identity: Theory, research, critique* (pp. 25–39). Lanham, MD: Rowman and Littlefield Publishers.
- Bernstein, B., & Solomon, J. (1999). 'Pedagogy, identity and the construction of a theory of symbolic control': Basil Bernstein questioned by Joseph Solomon. *British Journal of Sociology of Education*, 20(2), 265–279. doi:10.1080/0142569995443
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5, 7–74. doi:10.1080/0969595980050102
- Brint, S. (1994). *In an age of experts. The changing role of professionals in politics and public life*. Princeton, NJ: Princeton University Press.
- Burch, P., & Spillane, J. P. (2006). The institutional environment and instructional practice: Changing patterns of guidance and control in public education. In H. D. Meyer & B. Rowan (Eds.), *The new institutionalism in education* (pp. 87–102). Albany, NY: State University of New York Press.
- Chavannes, I., Engesveen, H., & Strand, E. (2011). *Nasjonale prøver som grunnlag for skoleutvikling og kontroll* (Master thesis). University of Oslo. Oslo.
- DeLuca, C., & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, 64, 356–372. doi:10.1177/0022487113488144
- Desrosières, A. (1998). *The politics of large numbers: A history of statistical reasoning*. Cambridge, MA: Harvard University Press.
- Diamond, J. B. (2007). Where the rubber meets the road: Rethinking the connection between high-stakes testing policy and classroom instruction. *Sociology of Education*, 80, 285–313. doi:10.1177/003804070708000401
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115. doi:10.1111/j.1365-2648.2007.04569.x
- 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., Hopmann, S., & Langfeldt, G. (2008). *Ansvarlighet i skolen. politiske spørsmål og pedagogiske svar: resultater fra forskningsprosjektet [Achieving school accountability in practice]*. Oslo: Cappelen akademisk forlag.
- Engeland, Ø., Langfeldt, G., & Roald, K. (2008). Kommunalt handlingsrom. Hvordan forholder norske kommuner seg til ansvarsstyring i skolen? In G. Langfeldt, E. Elstad, & S. Hopmann (Eds.), *Ansvarlighet i skolen. Politiske spørsmål og pedagogiske svar. Resultater fra forskningsprosjektet [Achieving school accountability in practice]* (pp. 178–203). Oslo: Cappelen akademisk forlag.
- Evetts, J. (2008). The management of professionalism: A contemporary paradox? In S. Gewirtz, P. Mahony, I. Hextall, & A. Cribb (Eds.), *Changing teacher professionalism: International trends, challenges and ways forward* (pp. 19–30). London: Routledge.
- Fitz, J, Davies, B, & Evans, J. (2005). *Education policy and social reproduction: class inscription & symbolic control*. London: Routledge.
- Følgjegruppa for lærarutdanningsreforma (Ffl) / Munthe, E., Henriksen, M., Hjetland, H., Hustad, B., Isaksen, T., Jahr, H., ... Werler, T. (2011). Frå allmennlærer til grunnskulelærer. Innføring og oppstart av nye grunnskulelærerutdanningar. Rapport nr. 1 frå Følgjegruppa til Kunnskapsdepartementet. Stavanger: Universitetet i Stavanger.
- Følgjegruppa for lærarutdanningsreforma (Ffl) / Munthe, E., Henriksen, M., Hjetland, H., Hustad, B., Isaksen, T., Jahr, H., ... Werler, T. (2015). Grunnskulelærerutdanningane etter fem år. Status, utfordringar og vegar vidare. (Teacher Education after five years. Status, challenges and the way ahead) Rapport nr. 5 frå Følgjegruppa til

- Kunnskapsdepartementet. Stavanger: Universitetet i Stavanger.
- Fullan, M. (1991). *The new meaning of educational change*. London: Cassell.
- Goertz, M. E., Nabors Oláh, L., & Riggan, M. (2010). From testing to teaching: The use of interim assessments in classroom instruction. CPRE Research Report (No. RR-65), Philadelphia, PA: Consortium for Policy Research in Education.
- Gregory, R. (2003). Accountability in modern government. In B. G. Peters & J. Pierre (Eds.), *Handbook of public administration* (pp. 557–568). London: Sage.
- Grek, S. (2009). Governing by numbers: The PISA 'Effect' in Europe. *Journal of Education Policy*, 24(1), 23–37. doi:10.1080/02680930802412669
- Grøgaard, J. B., Helland, H., & Lauglo, J. (2008). *Elevenes læringsutbytte: Hvor stor betydning har skolen? En analyse av ulikhet i elevers prestasjonsnivå i fjerde, syvende og tiende trinn i grunnskolen og i grunnskuret i videregående [Pupil learning outcomes: How much influence does the school have? An analysis of inequalities in pupils' achievement in fourth, seventh and tenth grade in basic school and in first grade in upper secondary school]*. Oslo: NIFU.
- Hallett, T. (2010). The myth incarnate: Recoupling processes, turmoil, and inhabited institutions in an urban elementary school. *American Sociological Review*, 75(1), 52–74. doi:10.1177/0003122409357044
- Hatch, T. (2013). Beneath the surface of accountability: Answerability, responsibility and capacity-building in recent education reforms in Norway. *Journal of Educational Change*, 14(2), 113–138. doi:10.1007/s10833-012-9206-1
- 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. doi:10.1080/00220270801989818
- Hopmann, S. T. (2013). The end of schooling as we know it? *Journal of Curriculum Studies*, 45(1), 1–3. doi:10.1080/00220272.2013.767570
- Ingersoll, R. M. (2003). *Who controls teachers' work? Power and accountability in America's schools*. Cambridge, MA: Harvard University Press.
- Isaksen, T., & Hjelm Solli, A. (2014). *Resultatene er kommet! Hva nå? Nasjonale prøver – fra fokus på resultat til utvikling av kvalitet [The results have come! What now? National tests – from focus on results to quality development]* (Master thesis). Universitetet i Tromsø. Fakultet for humaniora, samfunnsvitenskap og lærerutdanning, Tromsø.
- Johansen, N. (2015). *Kartlegging som en ressurs for elever med lese- og skrivevansker. En kvalitativ studie på mellomtrinnet [Mapping as a resource for students with reading and writing difficulties. A qualitative intermediate stage study]* (Master thesis). University College of Hedmark. Hamar.
- Karseth, B., & Engelsen, B.U. (2013). Læreplanen for Kunnskapsløftet: Velkjente tråkk og nye spor [The curriculum for the Knowledge Promotion: Well-known trails and new tracks]. In Karseth, B., Møller, J. & Aasen, P. Reformtakter. *Om fornyelse og stabilitet i grunnsopplæringen [Reform measures about renewal and stability in basic education]*. Oslo: Universitetsforlaget, 43–60.
- Kohlbacher, F. (2006). The use of qualitative content analysis in case study research [89 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 7(1). Retrieved from <http://nbn-resolving.de/urn:nbn:de:0114-fqs0601211>
- Larson, M. S. (2012). *Rise of professionalism. Monopolies of competence and sheltered markets*. New Brunswick, NJ: Transaction Publishers.
- Lawn, M. (2013). Introduction: The rise of data in education. In M. Lawn (Ed.) *The rise of data in education systems: Collection, visualization and use* (pp. 7–11). Oxford: Symposium Books. doi:10.1177/1753193412470225
- Linn, R. L. (2013). *Test-based accountability. The Gordon Commission on the future of assessment in education*. Retrieved May 5, 2016, from [http://www.gordoncommission.org/publications\\_reports/assessment\\_education.html](http://www.gordoncommission.org/publications_reports/assessment_education.html)
- Lortie, D. (1975). *Schoolteacher: A sociological study*. Chicago, IL: University of Chicago Press.
- Maguire, M., Ball, S. J., & Braun, A. (2010). Behaviour, classroom management and student 'control': Enacting policy in the english secondary school. *International Studies in Sociology of Education*, 20(2), 153–170. doi:10.1080/09620214.2010.503066
- Malkenes, S. (2014). *Bak fasaden i Oslo skolen [Behind the facade of Oslo school district]*. Oslo: Res Publica.
- Mandinach, E., Friedman, J., & Gummer, E. (2015). How can schools of education help to build educators' capacity to use data? A systemic view of the issue. *Teachers College Record*, 117, 1–50.
- Mausethagen, S. (2013a). A research review of the impact of accountability policies on teachers' workplace relations. *Educational Research Review*, 9, 16–33. doi:10.1016/j.edurev.2012.12.001
- Mausethagen, S. (2013b). *Reshaping teacher professionalism. An analysis of how teachers construct and negotiate professionalism under increasing accountability* (PhD thesis). Oslo: Oslo and Akershus University College of Applied Sciences.
- Mausethagen, S., & Granlund, L. (2012). Contested discourses of teacher professionalism: Current tensions between education policy and teachers' union. *Journal of Education Policy*, 27(6), 815–833. doi:10.1080/02680939.2012.672656
- Mayer-Schonberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think*. New York, NY: Houghton.
- Mayring, P. (2015). Qualitative content analysis: Theoretical background and procedures. In A. Bikner-Ahsbals, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education* (pp. 365–380). Dordrecht: Springer.
- Mayring, P. (2002). Qualitative content analysis – Research instrument or mode of interpretation? In M. Kiegelmann (ed.), *The role of the researcher in qualitative psychology* (pp. 139–148). Tuebingen: Verlag Ingeborg Huber.
- McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York, NY: Routledge.
- Meyer, H. D., & Benavot, A. (Eds.) 2013. *PISA, power and policy: The emergence of global educational governance*. Oxford: Symposium Books.
- Meyer, H. D., & Rowan, B. (2006). *The new institutionalism in education*. Albany, NY: State University of New York Press.
- Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83, 340–363. doi:10.1086/226550
- Müller, J., & Hernández, F. (2010). On the geography of accountability: Comparative analysis of teachers' experiences across seven European countries. *Journal of Educational Change*, 11, 307–322. doi:10.1007/s10833-009-9126-x

- Nasjonalt råd for lærerutdanning (NRLU). (2016a). *Nasjonale retningslinjer for femårig grunnskolelærerutdanning, trinn 1–7*. Oslo: Universitets- og høskolerådet. Retrieved December 1, 2016, from [http://www.uhr.no/documents/Godkjent\\_1\\_7\\_010916.pdf](http://www.uhr.no/documents/Godkjent_1_7_010916.pdf)
- Nasjonalt råd for lærerutdanning (NRLU). (2016b). *Nasjonale retningslinjer for femårig grunnskolelærerutdanning, trinn 5–10*. Oslo: Universitets- og høskolerådet. Retrieved December 1, 2016, from [http://www.uhr.no/documents/Godkjent\\_5\\_10\\_010916.pdf](http://www.uhr.no/documents/Godkjent_5_10_010916.pdf)
- Pierce, R., & Chick, H. (2011). Teachers' intentions to use national literacy and numeracy assessment data: A pilot study. *38*(4), 433–447. doi:10.1007/s13384-011-0040-x
- Popham, J. (2007). The no-win accountability game. In C. Glickman (ed.), *Letters to the next President. What we can do about the real crisis in public education* (pp. 166–173). New York, NY: Teachers College Press.
- Powell, W., & DiMaggio, P. J. (1991). *The new institutionalism in organizational analysis*. Chicago, IL: University of Chicago Press.
- Power, M. (1997). *The audit society: Rituals of verification*. Oxford: Oxford University Press.
- Prøitz, T. S. (2015). Uploading, downloading and uploading again—concepts for policy integration in education research. *Nordic Journal of Studies in Educational Policy*, *1*(1), 27015. doi:10.3402/nstep.v1.27015
- Rizvi, F., & Kemmis, S. (1987). *Dilemmas of reform. An overview of issues and achievements of the participation and equity program in victorian schools 1984–1986*. Geelong: Deakin Institute for Studies in Education.
- Rizvi, F., & Lingard, B. (2010). *Globalizing education policy*. London: Routledge.
- Romzeck, B., & Dubnick, M. (1993). Accountability and the centrality of expectations in american public administration. In J. Perry (ed.), *Research in public administration* (pp. 37–78). London: Jai Press.
- Rowan, B. (2006). The new institutionalism and the study of educational organizations: Changing ideas for changing times. In D. Meyer & B. Rowan (Eds.), *The new institutionalism in education* (pp. 15–32). Albany, NY: State University of New York Press.
- Sahlberg, P. (2010). Rethinking accountability in a knowledge society. *Journal of Educational Change*, *11*, 45–61. doi:10.1007/s10833-008-9098-2
- Schildkamp, K., Ehren, M., & Lai, M. (2012). Editorial article for the special issue on data-based decision making around the world: From policy to practice to results. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, *23*(2), 123–131. doi:10.1080/09243453.2011.652122
- Seland, I., Vibe, N., & Hovdhaugen, E. (2013). *Evaluering av nasjonale prøver som system [Evaluation of the national test system]*. Oslo: NIFU.
- Singh, P. (2002). Pedagogising knowledge: Bernstein's theory of the pedagogic device. *British Journal of Sociology of Education*, *23*(4), 571–582. doi:10.1080/0142569022000038422
- Singh, P. (2015). Pedagogic governance: Theorising with/after Bernstein. *British Journal of Sociology of Education*, *1*–20. doi:10.1080/01425692.2015.1081052
- Singh, P., Thomas, S., & Harris, J. (2013). Recontextualising policy discourses: A Bernsteinian perspective on policy interpretation, translation, enactment. *Journal of Education Policy*, *28*(4), 465–480. doi:10.1080/02680939.2013.770554
- Skedsmo, G. (2011). Formulation and realisation of evaluation policy: Inconcistencies and problematic issues. *Educational Assessment, Evaluation and Accountability*, *23*, 5–20. doi:10.1007/s11092-010-9110-2
- Svensson, L. G., & Karlsson, A. (2008). Profesjoner, kontroll og ansvar [Professions, control and responsibility]. In A. Molander & L. I. Terum (Eds.), *Profesjonsstudier [Studies of professions]* (pp. 261–274). Oslo: Universitetsforlag.
- Takayama, K. (2008). The politics of international league tables: PISA in Japan's achievement crisis debate. *Comparative Education*, *44*(4), 387–407. doi:10.1080/03050060802481413
- Thomas, J. Y., & Brady, K. P. (2005). Chapter 3: The elementary and secondary education act at 40: equity, accountability, and the evolving federal role in public education. *Review of Research in Education*, *29*(1), 51–67. doi:10.3102/0091732X029001051
- Thomson, P., Lingard, B., & Wrigley, T. (2012). Introduction: Ideas for changing educational systems, educational policy and schools. *Critical Studies in Education*, *53*(1), 1–7. doi:10.1080/17508487.2011.636451
- Tveit, S. (2014). Educational assessment in Norway. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 221–237.
- Union of Education Norway. (2012). *Professional ethics for the teaching profession*. Oslo: Utdanningsforbundet. Retrieved February 2, 2017, from [https://www.utdanningsforbundet.no/upload/1/L%C3%A6rerprof\\_etiske\\_plattform\\_a4\\_engelsk\\_31.10.12.pdf](https://www.utdanningsforbundet.no/upload/1/L%C3%A6rerprof_etiske_plattform_a4_engelsk_31.10.12.pdf)
- Utdanningsdirektoratet/ Udir. (2010). *Rammeverk for nasjonale prøver*. Oslo. Retrieved May 12, 2016, from [http://www.udir.no/Upload/Nasjonale\\_prover/2010/5/Rammeverk\\_NP\\_22122010.pdf?epslanguage=no](http://www.udir.no/Upload/Nasjonale_prover/2010/5/Rammeverk_NP_22122010.pdf?epslanguage=no)
- Utdanningsdirektoratet/ Udir. (2014). *Til lærere. Hvordan bruke nasjonale prøver som redskap for læring?* Oslo. Retrieved May 12, 2016, from <http://www.udir.no/PageFiles/84379/Larerbrosjyre-bokmal.pdf>
- Utdanningsdirektoratet/ Udir. (2016). *Metodegrunnlag for nasjonale prøver*. Oslo: Udir. Retrieved December 01, 2016, from <http://www.udir.no/globalassets/filer/vurdering/nasjonaleprover/metodegrunnlag-for-nasjonale-prover.pdf>
- 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. doi:10.3102/0002831207306859
- Waters, P. (2013). *Mål- og resultatstyring i grunnskolen. Bidrar mål- og resultatstyring til gode resultater? [New public management in primary school. Does new public management contribute to good results?]* (Master thesis). Høgskolen i Oslo og Akershus, Fakultet for samfunnsfag. Oslo.
- Wayman, J., & Jimerson, J. (2014). Teacher needs for data-related professional learning. *Studies in Educational Evaluation*, *42*, 25–34. doi: 10.1016/j.stueduc.2013.11.001
- Wayman, J. C., Jimerson, J. B., & Cho, V. (2012). Organizational considerations in establishing the data-informed district. school effectiveness and school improvement. *American Journal of Education*, *23*(2), 159–178. doi:10.1086/505058
- Wells, A. S. (2009). "Our children's burden": A history of federal education policies that ask (now require) our public schools to solve societal inequality. In M. A. Rebell & J. R. Wolff (Eds.), *NCLB at the crossroads: Re-examining the federal effort to close the achievement gap* (pp. 1–42). New York, NY: Teachers College Press.

- Werler, T. (2015). Commodification of teacher professionalism. *Policy Futures in Education*, 14(1), 60–76. doi:10.1177/1478210315612646
- Werler, T., & Volckmar, N. (2015). Norway. In W. Hörner, H. Döbert, L. Reuter, & B. Von Kopp (Eds.), *The education systems of Europe* (pp. 603–619). Dordrecht: Springer.
- Wiles, J. L., Rosenberg, M. W., & Kearns, R. A. (2005). Narrative analysis as a strategy for understanding interview talk in geographic research. *Area*, 37(1), 89–99. doi:10.1111/area.2005.37.issue-1
- Wong, T. H., & Apple, M. W. (2003). Rethinking the education–state formation connection: The state, cultural struggles, and changing the school. In M. Apple (ed.), *The state and the politics of knowledge* (pp. 81–108). London: Routledge.
- Yin, R. K. (2003). *Case study research, design and methods*. Thousand Oaks, CA: Sage.