



Coping with performance expectations: towards a deeper understanding of variation in school principals' responses to accountability demands

Marjolein K. Camphuijsen^{1,2} 

Received: 22 May 2020 / Accepted: 9 November 2020 / Published online: 20 November 2020

© Springer Nature B.V. 2020

Abstract

In recent decades, performance-based accountability (PBA) has become an increasingly popular policy instrument to ensure educational actors are responsive to and assume responsibility for achieving centrally defined learning goals. Nonetheless, studies report mixed results with regard to the impact of PBA on schools' internal affairs and instructional practices. With the aim of contributing to the understanding of the social mechanisms and processes that induce particular school responses, this paper reports on a study that examines how Norwegian principals perceive, interpret, and translate accountability demands. The analysis is guided by the policy enactment perspective and the sociological concept of "reactivity", and relies on 23 in-depth interviews with primary school principals in nine urban municipalities in Norway. The findings highlight three distinct response patterns in how principals perceive, interpret, and translate PBA demands: alignment, balancing multiple purposes, and symbolic responses. The study simultaneously shows how different manifestations of two social mechanisms form important explanatory factors to understand principals' varying responses, while it is highlighted how the mechanisms are more likely to operate under particular conditions, which relate both to principals' trajectories and views on education, and to school-specific characteristics and the local accountability regime. The study contributes to the accountability literature by showing how, even in the relative absence of material consequences and low levels of marketization, standardized testing and PBA can drive behavioral change, by reframing norms of good educational practice and by affecting how educators make sense of core aspects of their work.

Keywords Performance-based accountability · Standardized testing · Policy enactment · Reactivity · School principals · Norway

✉ Marjolein K. Camphuijsen
m.k.camphuijsen@vu.nl

¹ Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

² Universitat Autònoma de Barcelona, Bellaterra, Spain

1 Introduction

In recent decades, metrics and indicators to evaluate the performance of individuals and organizations have proliferated in different policy domains (Mennicken and Espeland 2019). The spread of neo-liberalism and growing popularity of New Public Management (NPM) have contributed to numbers escaping “from the relatively restricted toolbox in which they were kept” and moving towards more accountability uses (Piattoeva and Boden 2020, p.4). In the education sector, a growing number of countries has adopted standardized tests to measure the performance of schools and teachers, and to hold educational actors accountable for learning objectives determined at the central level, usually emphasizing core subjects or basic skills (Ball et al. 2017; Verger et al. 2019). Following the increasing use of performance measures for accountability purposes, scholars have referred to an important shift in the governance of education, which goes from professional to performance-based accountability systems (Lingard et al. 2017). Key objectives of performance-based accountability (hereafter PBA) are to ensure that educators are responsive to and assume responsibility for achieving centrally defined learning goals, and to promote data-driven decision-making. Whereas generally aimed at regulating actors’ behavior, the specific institutional arrangements and instruments of PBA systems can differ according to various dimensions, including the type and alignment of accountability tools, the nature of accountability consequences, and the conceptions of actors that inform tool selection (Maroy 2015; Maroy and Pons 2019).

Despite the increasing popularity of PBA worldwide, studies report mixed results with regard to the impact of PBA on school organization and pedagogical practices. Whereas some educators seem to adapt their practices to meet accountability expectations, others respond by ignoring, avoiding, resisting, or re-writing policy demands, or by relying on quick and visible solutions, rather than on thorough and long-term changes (e.g. Barbana et al. 2019; Candido 2020; Diamond and Spillane 2004; Falabella 2014; Maroy and Pons 2019; Verger et al. 2020). In line with recent scholarship on policy enactment (Ball et al. 2012), these studies demonstrate how, rather than a linear and top-down process, putting policy into practice forms a creative, complex, and sometimes constrained social process. To understand variation in policy responses, it is key to examine how actors perceive, interpret, and translate policy demands in various ways, while being attentive to how this sense-making process is guided by educators’ pre-existing knowledge, beliefs, and practices (Coburn 2001, 2004; Spillane and Jennings 1997), as well as enabled or constrained by contextual factors (Braun et al. 2011).

While research interest in PBA reforms has sparked in recent decades, in particular “hard” or “strict” accountability systems have been studied extensively, usually in contexts also characterized by high levels of marketization, such as England, the USA, and Chile. “Hard” accountability systems tend to rely on external and closely aligned policy tools as well as high-stakes consequences, following the conception of a utilitarian and strategic actor (Maroy 2015). In contrast, other accountability approaches, including “soft” and “reflexive” systems, have remained under-researched (Maroy and Pons 2019). These approaches, informed by the conception of an actor that is socially embedded and reflexive, attempt to target actors’ internal feelings of responsibility and reflection. Nonetheless, whereas these approaches aim to instigate

change “from the inside” through influencing actors’ dispositions, some systems combine internal measures with external devices and moderate to significant accountability consequences (Maroy 2015). As a result of the predominant focus of accountability research on “hard” systems, a limited understanding persists of how educators perceive and respond to PBA demands in other contexts, as well as of how potential variation in actors’ responses can be explained. With the aim of contributing to this understanding, this paper reports on a study that examines how Norwegian principals reflect on and respond to PBA demands. More specifically, the study addresses the following research questions: (1) In what ways do Norwegian principals perceive, interpret, and translate PBA demands?; (2) what are the response patterns employed by principals to address PBA demands?; and (3) what are the social mechanisms and contextual conditions (local accountability regimes, school-specific factors, and personal trajectories) that explain the response patterns and the differences among them?

Norway forms a particularly interesting context for this object of study as the Norwegian approach to PBA differs in important regards from “hard” accountability systems, while at the same time significant local variation is found in accountability regimes. With recent policy documents placing strong emphasis on learning and basic skills as main missions for schools (Larsen et al. 2020), national tests and value-added models are increasingly used to hold teachers, school leaders, and local authorities accountable for students’ learning outcomes and acquisition of basic skills. Nonetheless, the system remains characterized by a relative absence of material consequences (such as financial rewards or sanctions), as well as low levels of marketization. Rather, the Norwegian approach to PBA combines administrative control devices with institutional regulations aimed at encouraging reflection, self-evaluation, and organizational learning, so to ensure educational actors adapt their practices in line with the competency aims formulated in the national curriculum and use achievement data for school improvement purposes. At the same time, despite these generic features at the central level, significant local variation exists in how accountability plays out in practice, following municipal discretion in terms of accountability tools and consequences, as well as local variation in the role played by external audiences, or “third-party” account-holders. With regard to the latter, local differences exist in school choice regulations (NSD 2016) and the level of activity of local media outlets in reporting on test performance. Such local variability makes Norway an excellent case to advance the understanding of how different accountability configurations and local policy contexts mediate policy enactment processes and policy outcomes.

This study focuses specifically on principals, as principals play a key role in the enactment of schools’ accountability (Coburn 2004; Diamond and Spillane 2004). While often juggling multiple, and sometimes conflicting, accountability demands from different audiences (Pollock and Winton 2016), principals act as key “managers in the middle” or policy brokers. The ways in which principals reflect on and respond to new policy demands are crucial, not in the least because principals’ reflections and actions have the potential of mediating teachers’ experiences and responses (Diamond and Spillane 2004; Spillane et al. 2002). Principals furthermore form a particularly interesting group of school-level actors as they, including in Norway, have often been specifically targeted by NPM reforms, expected to act as the vehicles of modernization of education in schools (Møller and Skedsmo 2013).

The analysis of this paper relies primarily on qualitative data derived from 23 in-depth interviews with primary school principals in nine urban municipalities in Norway, characterized by diverging local accountability regimes. In order to examine and explain different response patterns employed in reaction to PBA demands, this study relies on the sociological concept of “reactivity”, understood as the way “individuals alter their behavior in reaction to being evaluated, observed, or measured” (Espeland and Sauder 2007, p.6). More specifically, the study relies on two social mechanisms identified by Espeland and Sauder (2007) to understand the reflexive interactions between actors and measures, to undertake an attempt to explain *why* principals respond in particular ways to standardized testing and PBA.

The paper is structured as follows. The next section presents a review of previous research on how educators respond to PBA demands. Based on this review, the gap in the existing literature is identified, which this paper tries to address by relying on the “reactivity” framework, outlined in the subsequent section. Thereafter, the Norwegian educational context is briefly explained, which is followed by the study’s methodology. Subsequently, the study’s findings are presented in the form of three predominant responses: (a) alignment; (b) balancing multiple purposes; and (c) symbolic responses. The final section discusses the main results and concludes by arguing that even in the relative absence of material consequences and low levels of marketization, standardized testing and PBA can drive behavioral change, by reframing norms of good educational practice, and by affecting how educators make sense of core aspects of their work.

2 Literature review: school actors’ responses to PBA

In recent years, a growing body of studies has examined how educators respond to accountability approaches characterized by the ambition to elicit change “from the inside” by influencing actors’ dispositions, as well as a relative absence of material consequences. A key finding of these studies is that policy enactment processes can differ significantly from policy intentions and even contradict key assumptions of the action theory underpinning such PBA reforms. For example, based on a study conducted in three schools in French Belgium, Barbana et al. (2019) showed how the clash between the accountability instruments and educators’ own views on instruction and student assessment discouraged many teachers from adopting the anticipated “reflexive attitude” and from making substantive changes to their classroom practices. At the same time, the authors found that a minority of teachers expressed a more positive attitude towards the instruments and used them to reflect on and to a certain extent modify their practices¹ (Barbana et al. 2019). Similar findings were reported in the Brazilian context, where Candido (2020, p.22) found that a number of educators chose to adapt their discourses and practices to testing and accountability policies, while other school actors found ways to “rewrite the rules of the ‘game’ to fit their own interests”. Also in Norway, studies report mixed results with regard to the impact of standardized testing and PBA demands on educators’ practices. Whereas some studies report how school leaders employ symbolic responses to policy demands emphasizing

¹ The authors classified the three schools as “atypical”, as they were theoretically selected for their “a priori distance towards the worldviews embedded into the accountability instruments” (Barbana et al. 2019, p.11).

test scores (Gunnulfson and Møller 2017), other studies show how national testing and PBA have an important impact on instructional strategies and schools' internal affairs (Elstad 2009; Seland et al. 2013; Skedsmo 2018).

A second key contribution of this growing body of literature entails the documentation of side-effects, formerly in particular associated with “hard” accountability approaches (e.g. see Au 2007; Mittleman and Jennings 2018). For example, in both Germany and Israel, where accountability systems were deliberately designed without attaching high-stakes consequences so to avoid the emergence of side-effects, scholars report effects such as teaching to the test, educational triage, and curriculum narrowing (Feniger et al. 2015; Thiel et al. 2017). Rather than attributing side-effects to the stakes of accountability, Thiel et al. (2017) suggest that side-effects might form systematic problems of accountability in education, while Feniger et al. (2015, p.3) point towards the “power of numbers”, arguing that “the use of external standardized tests, in itself, causes a shift in the way actors in the educational field think and speak about education”.

So far, a limited understanding prevails of how to interpret and explain the complex, creative, and sometimes unanticipated responses adopted in these accountability contexts. That is, little remains known of *why* and *under what circumstances* educators may adopt particular responses. By identifying the social mechanisms that induce particular response patterns, and by establishing the conditions under which they operate, a deeper understanding can be gained of “why we observe what we observe” (Hedström and Swedberg 1998, p.9). With the aim of contributing to this understanding, this study relies on the sociological concept of “reactivity”, and more specifically, the framework developed by Espeland and Sauder (2007), which identifies two social mechanisms that induce reactivity.

2.1 Reactivity as an analytical device to interpret and explain responses to PBA

In recent decades, awareness has grown that, because people are “reflexive beings who continually monitor and interpret the world and adjust their actions accordingly”, social measures such as standardized tests are “reactive” (Espeland and Sauder 2007, p.2). While some see it as a methodological problem that people adapt their actions in response to being measured (Campbell 1957), others consider reactivity a promise and a vehicle for inducing behavioral changes in desired ways. Considering that PBA systems tend to rely on the latter understanding of reactivity, it is key to examine the reflexive interactions between educators and PBA instruments, in order to gain a deeper understanding of the reactions employed by key actors, as well as the effects they give rise to.

To do so, this study relies on the framework developed by Espeland and Sauder (2007). Based on a large-scale study on law school rankings, they identified two mechanisms that produce reactivity to social measures: self-fulfilling prophecies and commensuration.² Rather than restricting the definition of a self-fulfilling prophecy to false beliefs (Merton 1968), Espeland and Sauder (2007, p.11) refer to “processes by

² While Espeland and Sauder (2007) present self-fulfilling prophecies and commensuration as two distinctive mechanisms of reactivity, they simultaneously emphasize how the two mechanisms can interact in the production of behavioral change.

which reactions to social measures confirm the expectations or predictions that are embedded in measures or which increase the validity of the measure by encouraging behavior that conforms to it". Key in this regard is the understanding of social measures, designed to evaluate the performance of individuals or organizations, as carrying tacit assumptions about what constitutes "quality", "excellence", or "success", thereby reframing or constructing new norms of what is considered relevant, valuable, and desirable. By encouraging actors to see themselves and behave according to the norms of good practice embedded in measures, thereby reinforcing their validity, social measures can create self-fulfilling prophecies (Espeland and Sauder 2007). In this light, various scholars have argued how, in the education sector, the increasing use of performance metrics has reoriented the purposes of schooling and redefined the education profession (Ball 2003), while simultaneously showing how educators can come to internalize or embody new definitions of "quality", "excellence", and "success", fostering efforts of norm compliance (Courtney 2014). Another way in which social measures can operate as a self-fulfilling prophecy relates to the effects of measurement on perceptions and actions of external audiences (Espeland and Sauder 2007). In particular when the precise, quantitative distinctions between individuals and institutions are increasingly perceived as relevant and "natural", even statistically insignificant differences can have real consequences for measured objects. That is, when external audiences act upon such differences, for example by raising their voice or choosing another provider, differences that initially largely resulted from measurement noise can become real and strengthen over time.

Commensuration, the second social mechanism identified by Espeland and Sauder (2007), entails "the comparison of different entities according to a common metric" (Espeland and Stevens 1998, p.313). Prices constitute an example of commensuration, which have become a highly naturalized form of comparing the value of disparate goods or services. Standardized test scores are another example of commensuration, which also enable the formal comparison of disparate entities, such as schools that are located in different parts of the country, with diverging histories, cultures, and student populations. While self-fulfilling prophecies induce behavioral changes as a result of actors adapting their actions in response to altered expectations, commensuration shapes behavior by changing "what we pay attention to, which things are connected to other things, and how we express sameness and difference" (Espeland and Sauder 2007, p.16). One way in which commensuration affects sense-making is by simplifying and de-contextualizing information, while organizing what remains into numbers that often appear rational, objective, and robust, and are easy to interpret and quick to compare and disseminate. The more such numbers become taken-for-granted ways to evaluate and compare goods or entities, the more attention risks being diverted from other ways of expressing difference. For example, the more standardized test scores or rankings are perceived and acted upon as proxies of school quality, the more attention shifts away from other differences between institutions, in particular differences that are hard to quantify. Another feature of commensuration that affects sense-making relates to the creation of precise and hierarchical relationships between measured objects, which enables the possibility to compare oneself to others, as well as previous versions of oneself, thereby affecting how entities make sense of one another and themselves, and changing how one determines and assesses "progress". In recent years, it has been suggested that the constitutive power of commensuration can, in part, be attributed to

the relation between data and affect (Sellar 2015). As argued by Sellar (2015, p.135), commensuration can shape actors' experiences and behavior as a result of "emotional or felt effects that data and associated judgements have on those whose practices are made commensurate in order to be compared and evaluated, sanctioned and rewarded". In this regard, scholars have highlighted how performance data, in particular when used to compare and judge individuals or institutions, can engender feelings such as pride, shame, and envy (Ball 2003), and as such influence sense-making processes.

2.2 Measurement, transparency, and accountability in the Norwegian context

In Norway, the increasing presence and regulatory power of external actors concerned with measurement, observation, and evaluation have altered historical self-regulatory dynamics of the education profession (Skedsmo and Mausethagen 2016). With the turn of the millennium, growing concerns about below-expected learning outcomes of Norwegian students in basic skills such as reading contributed to the increasing call for external assessment of student performance as well as external control of educators' competence and results (Møller and Skedsmo 2013). In 2004, a national quality assessment system was adopted, which includes national standardized tests. The tests are administrated at the beginning of the 5th, 8th, and 9th grades and measure students' acquisition of basic skills in reading and numeracy as well as their performance in English, reflecting competency aims as formulated in the national curriculum at the end of grades 4 and 7. Value-added models have been published at the school level since 2016, in response to calls for more accurate measurement of schools' contribution to student learning.

The main rationales behind national testing and value-added models are to assess whether schools succeed in teaching pupils centrally defined learning objectives as well as to foster data-driven decision-making by providing teachers, school leaders, and local authorities with student performance data. Following the conception of an actor guided by social obligations, Norway's PBA system relies on a combination of external control devices as well as institutional arrangements that encourage self-evaluation processes and target internal feelings of responsibility and reflection. More specifically, a main external pressure mechanism forms the publication of test results on the government website "the School Portal" (skoleporten.no), where school results are presented in the form of comparisons to the municipal, county, and national average, following a benchmarking logic (Skedsmo 2018). In this regard, the media form an important "third-party" account-holder, by regularly reporting on municipal and school performance, often in the form of performative-oriented rankings and with a focus on narratives of success and failure (Elstad 2009). Here it is important to emphasize that the extent to which the publication of results plays out as a high-stakes mechanism for schools depends for an important part on the degree of school choice families enjoy.

Administrative supervision conforms to a second external control mechanism. Primary and lower-secondary schools report primarily to the municipal superintendent, who monitors and controls schools' results on various quality assessment measures, including standardized tests. Municipal authorities simultaneously play an important role in encouraging reflection, self-evaluation, and organizational learning on the basis of performance data, by supporting and following up schools' routines for analyzing and using assessment results to foster school improvement. As highlighted in the

introduction, significant local variability exists in municipal routines surrounding administrative supervision, support, and follow-up (e.g. Seland et al. 2013; Skedsmo 2018). In addition, municipalities have adopted different practices around the publication of test results on municipal websites and school choice regulations. With regard to the latter, while school choice in compulsory education is generally restricted in Norway, data from the Norwegian Centre for Research Data (NSD) reveal that 56 Norwegian municipalities (i.e. 19.5%) allow for “freer user choice in the area of compulsory education” (NSD 2016, p.256).³ Still, while such municipal regulations are likely to award families some degree of school choice, priority remains given to students residing in the school’s catchment area.

Whereas initially strongly resisted (Tveit 2014), recent studies indicate that over time, standardized tests have become more broadly accepted. In particular school leaders see benefit in having access to performance data (Seland et al. 2013), while teachers continue to struggle with how to integrate test data in their daily practices (Gunnulfson 2017; Mausesthaugen 2013; Mausesthaugen et al. 2017). Nonetheless, most schools have established systems surrounding how to use achievement data for school improvement purposes, in line with policy expectations (Seland et al. 2013).

3 Data and methodology

This paper relies primarily on qualitative data derived from 23 in-depth interviews with principals of primary schools located in nine Norwegian municipalities. The municipalities are located in eight different counties, dispersed across all regions of the country. Interviews were considered a particularly suitable method to gain a deeper understanding of principals’ worldviews, motivations, and professional trajectories, as well as perceptions, interpretations, and translations of PBA demands. Recognizing that principals’ perspectives are influenced by local policy contexts and school-specific factors (Braun et al. 2011), I sampled schools with the aim of guaranteeing variability in important factors, while ensuring comparability in others. With regard to the latter, I decided to focus specifically on public schools, which enroll 96% of the country’s student population at the level of compulsory education (Statistics Norway 2019). Additionally, all selected schools are located in urban municipalities. Compared to their rural counterparts, urban municipalities in Norway tend to possess greater institutional capacity to assist schools in policy enactment. As this study had a specific interest in the role of local authorities in supervising and supporting quality improvement efforts, institutional capacity was considered an important requirement.

For the selection of schools, I first classified Norway’s urban municipalities according to two criteria, with the aim of assuring variability in local policy contexts. The first criterion entailed the level of “strictness” of the local accountability regime (i.e. whether municipalities had a strong or weak performance orientation, as well as the type and level of alignment of the local accountability instruments). To determine local

³ The data are based on “Survey on Municipal Organization 2016, Municipality File”. The survey was financed by the Ministry of Local Government and Modernisation (KMD). The data are provided by Norwegian Institute for Urban and Regional Research (NIBR) and prepared and made available by NSD—Norwegian Centre for Research Data. Neither NIBR, KMD, nor NSD are responsible for the analyses/interpretation of the data presented here.

accountability regimes, previous research, local policy documents, and survey data were used. The latter data were collected during the school year of 2018–2019, in the context of the research project. The second criterion referred to the level of involvement of ‘third-party’ account-holders (i.e. parents and the media). This criterion was made up of two combined variables: first, whether municipalities employed freer or restricted regulations around school choice, and second, the level of activity of the local media in reporting on national testing. While I used NSD data (2016) to secure local variability in regulations around school choice, I relied on a unique database on local media coverage of national testing (2004–2018) to determine the level of activity of the local press. This database reveals important local differences in the extent to which schools are named, blamed, and praised for performance in local and regional newspapers.

Following this classification of municipalities, primary schools were selected in different local policy contexts. For the selection of schools, secondary data provided by the Norwegian Directorate for Education and Training, as well as survey data derived from the administration of the electronic questionnaire in the context of the research project were used. More specifically, the sampling criteria for the selection of schools included (a) school performance, (b) reported performative pressure levels, (c) reported reputation, and (d) reported pressure to maintain enrolment perceived by principals. Ideally, a proxy of the schools’ social composition would have been used as a fifth sampling criterion, as previous research has shown that student composition can influence how PBA policies, as well as school choice regulations, are experienced and enacted at the school level, in part because composition tends to correlate with school performance (e.g. Keddie 2013). Nonetheless, in the absence of available secondary data on Norwegian schools’ social composition, this was not possible. However, the interviews with principals, who described the dominant socio-economic status (SES) of their school’s student population, as well as the percentage of minority-language students, confirmed that the sample is characterized by significant variability in the schools’ social composition, as shown in Table 1. The sample of principals, moreover, is characterized by variety in gender (12 female principals and 11 male principals) and years of experience working as a school leader (ranging from four to 26 years). All principals worked as teachers before taking on the position of school leader, as is common in Norway, while 19 of them had obtained or were in the process of obtaining formal education in school leadership.

Data collection was informed by the policy enactment perspective (Ball et al. 2012; Braun et al. 2011), which recognizes that, rather than a straightforward, linear, and mechanical process, responding to policy demands forms a dynamic, non-linear, and negotiated process. Key actors at different levels (e.g. municipality, school, and classroom) are involved in interpreting and translating abstract policy ideas in complex and creative ways, enabled and constrained by local contexts and school-specific factors. As such, the interviews focused on gaining an understanding of how principals perceived, interpreted, and translated policy demands emphasizing data-driven decision-making and PBA, as well as the mediating role of personal and contextual factors. Specifically, the individual interviews, which were conducted between May 2019 and March 2020, followed a semi-structured interview script, which addressed (a) biographical information, (b) school characteristics and school context, (c) interpretations of testing, transparency, and accountability demands, (d) performative pressure, (e) pedagogical practices and data-use, and (f) administrative accountability. All interviews, which lasted on average 50 min, were taped and transcribed verbatim.

Table 1 Overview of participating schools

School ID	Student composition		Performance
	Predominant SES status	Percentage of minority-language students	
School 1	Low	> 75%	Low
School 2	Low	> 75%	Low
School 3	Low	50–75%	Low
School 4	Low	25–50%	High
School 5	Low	25–50%	Low
School 6	Average-low	> 75%	High
School 7	Average-low	25–50%	Medium
School 8	Average	50–75%	Low
School 9	Mixed	50–75%	Medium
School 10	Mixed	25–50%	Medium
School 11	Mixed	< 25%	Medium
School 12	Mixed	< 25%	Low
School 13	Mixed	< 25%	Medium
School 14	Mixed	< 25%	Medium
School 15	Average-high	< 25%	Medium
School 16	Average-high	< 25%	Low
School 17	Average-high	< 25%	Medium
School 18	Average-high	< 25%	High
School 19	High	< 25%	High
School 20	High	< 25%	High
School 21	High	< 25%	Medium
School 22	High	< 25%	High
School 23	High	< 25%	High

The data analysis consisted three phases. During the first phase, I performed a reading of all “raw” interview transcripts, in order to obtain a holistic view of the themes emerging during the interviews. During the second phase, a codebook was developed, which emerged in part from this first reading, as well as built on the heuristic distinction between policy interpretation and translation developed by Ball et al. (2012). The coding of all interviews, for which qualitative data analysis software was used (Atlas.ti), and the subsequent data analysis, allowed for the identification of three distinct response patterns in how principals perceive, interpret, and translate policy demands emphasizing data-use and PBA. The following overarching codes were particularly significant in this regard: (a) importance awarded to test results; (b) perception of PBA and data-use; (c) experience of performative pressure; (d) strategies to secure achievement of basic skills. The final phase aimed at making sense of and explaining the three response patterns, for which I relied on the reactivity framework. During this final reading, each response pattern was examined in further depth, and it was analyzed how different manifestations of the social mechanisms identified by Espeland and Sauder (2007) served as explanatory factors in interpreting the responses.

It was moreover examined how the different manifestations of the two mechanisms interacted with contextual and personal factors in inducing each particular response pattern.

4 Findings

The presentation of the study's findings is structured according to the three main response patterns articulated by principals. However, before describing the different response patterns, it is important to highlight that the analysis identified a number of similarities in principals' interpretations and translations of policy demands emphasizing data-driven decision-making and PBA. It appeared that all principals, including those most critical of standardized testing and PBA, consider it an important school mission to make sure that students achieve the basic skills established in the national curriculum. Similarly, all principals report to have established routines for preparing students for the tests,⁴ as well as routines for analyzing test results. The most common use of test scores reported by principals is to identify students in need of support and follow-up, which all mention to do.

Despite these similarities, principals' perceptions and interpretations differ, with regard to both their general conceptions of the path to school improvement and their perceptions and interpretations of standardized testing and PBA. With regard to the latter, it was found that principals' perceptions diverge regarding whether they perceive national tests as valid, fair, and useful measurements, as well as how much importance they ascribe to the achievement of basic skills compared to other educational goals and purposes (i.e. how much emphasis they place on academic achievement). The analysis reveals that these different perceptions and interpretations guide principals' translations, generating three distinct response patterns, which I named (a) alignment, (b) balancing multiple purposes, and (c) symbolic responses.

The first pattern is articulated by principals who perceive the national tests as valid measures of crucial competencies and aim at obtaining the best possible test results by adopting top-down, performance-oriented management practices and data-driven decision-making. In other words, they align their practices to accountability expectations. The second pattern is employed by principals who also perceive the national tests as valid measures, but who, in contrast to principals in the first response category, reject a prioritization of the tested cognitive competencies over non-cognitive competencies. Moreover, rather than the main source of information, the tests form one out of various information sources that guide school development. The third pattern is articulated by principals who question the validity of the national tests and strongly emphasize non-cognitive competencies as key educational purposes. They respond predominantly symbolically to the expectation to raise test scores and employ data-driven decision-making. The three response patterns are summarized in Table 2 and explained in further depth below. For each pattern, it is highlighted how principals interpret and translate

⁴ The main rationale behind preparing students for the test is to make sure students are familiar with the test format and test situation. Such preparations are recommended by the Norwegian Directorate for Education and Training, and are perceived particularly important for primary schools, where students have little to no experience with taking tests.

Table 2 Overview of three response patterns

Response pattern	Interpretation	Translation	Mechanism of change
<p>Perception of national tests and PBA</p> <p>General conception of path to school improvement</p>			
<p>1. <i>Alignment</i> Strong commitment to PBA mandate and action theory</p>	<p>Tests are valid measures of education quality</p> <p>Tests measure crucial competencies</p> <p>Tests provide basis for school management and improvement of school practice</p> <p>Transparency and accountability for test results are positive and natural</p>	<p>Ambition to obtain highest possible academic performance defines the school vision and organizational goals</p> <p>Data analysis is highly routinized and serves to diagnose the schools' situation and change practices according to "what works"</p>	<p>Self-fulfilling prophecy: expectations embedded in standardized testing and PBA system are internalized and attempted to be realized</p>
<p>2. <i>Balancing multiple purposes</i> Partial/selective commitment to PBA mandate and action theory</p>	<p>Tests are valid measures of education quality</p> <p>Tests measure important competencies, but non-tested competencies are equally important</p> <p>Tests represent one out of various information sources that inform school development</p> <p>Concerns about harmful effects of transparency</p>	<p>A balanced focus on academic and social development defines the school vision and organizational goals</p> <p>Data analysis is highly routinized and serves as a source for collective reflection and decision-making</p>	<p>Commensuration: standardized testing creates a precise hierarchy of schools, which shapes schools' self-perceptions and guides decision-making</p>
<p>3. <i>Symbolic responses</i></p>	<p>Concerns about tests' validity</p>	<p>A balanced focus on academic and social development defines the</p>	<p>Commensuration: test scores are increasingly perceived as prime indicators to evaluate, judge, and compare school quality</p>

Table 2 (continued)

Response pattern	Interpretation	General conception of path to school improvement	Translation	Mechanism of change
<p>Perception of national tests and PBA</p>	<p>Tests measure important competencies, but non-tested competencies are equally important</p> <p>Tests serve as external control and provide information of limited use for school development</p> <p>Concerns about harmful effects of transparency and accountability</p>	<p>practices based on educators' professional knowledge and judgement</p>	<p>school vision and organizational goals</p> <p>Data analysis forms a less routinized and more isolated practice, with a limited impact on school development work</p>	<p>Self-fulfilling prophecy: this perception can generate pressure from external audiences to conform to the definitions of good practice embedded in standardized testing and PBA</p>
<p>Weak commitment to PBA mandate and action theory</p>	<p>Tests measure important competencies, but non-tested competencies are equally important</p> <p>Tests serve as external control and provide information of limited use for school development</p> <p>Concerns about harmful effects of transparency and accountability</p>	<p>practices based on educators' professional knowledge and judgement</p>	<p>school vision and organizational goals</p> <p>Data analysis forms a less routinized and more isolated practice, with a limited impact on school development work</p>	<p>Self-fulfilling prophecy: this perception can generate pressure from external audiences to conform to the definitions of good practice embedded in standardized testing and PBA</p>

policy demands, following the heuristic distinction between interpretation and translation developed by Ball et al. (2012). Interpretation refers to the initial reading of policy texts, a process during which actors construct meaning of policy ideas and attempt to make sense of policy demands. Translation relates to the language of practice and the decisions made by actors regarding how to put into practice abstract policy ideas (Ball et al. 2012). In addition, it is discussed how different manifestations of the social mechanisms outlined in the theoretical framework interact with contextual and personal factors in inducing the response patterns.

4.1 Alignment: achievement of basic skills as core school mission

4.1.1 Interpretation

The first type of response is articulated by principals who strongly align to the PBA mandate and action theory. Principals in this response category perceive basic skills as crucial competencies that students need in order to obtain further education and improve their life chances. One principal captured the dominant view by arguing that “if children cannot read, write and calculate, then they do not manage their life” (P1).⁵ Inspired by rhetoric often promoted by research on school effectiveness and visible learning, principals express a strong belief in the ability of educators to impact student learning, regardless of a student’s family background. Three of the four principals who more explicitly adopt this approach work at schools that perform (significantly) below municipal and national averages (P1; P2; P12). In two of these cases, the schools’ student populations are characterized by a high proportion of students with a minority background and low SES status. Despite working with more disadvantaged student populations, principals are strongly motivated by a “no-excuses motto” and work hard to raise expectations of students among teachers. The fourth principal in this response category works at a school that performs at the municipal and national average, with a student population characterized by a higher number of students with more advantaged backgrounds (P15).

All four principals describe the fact of having to take responsibility and being held accountable for students’ learning outcomes as “natural” (P12) and “positive” (P1). They generally support the publication of results, referring to transparency as an important trigger of school improvement.⁶ In a similar vein, the principals tend to speak positively about administrative control exercised by the municipal superintendent to secure obtainment of basic skills:

So, I am happy that [municipality X] is a very demanding municipality. It does not hold back. They are forward-leaning in almost everything they do, [they] make demands on their schools. I was told that I got three years to get good results at this school. My boss said it clearly. And that is how I like it. (P1).

⁵ To assure the anonymity of the research participants, each interviewee is referred to by a numerical code. The letter “P” stands for principal, while the number in the coding refers to the school ID (see Table 1). All interview quotes have been translated from Norwegian to English by the author(s).

⁶ Principals acknowledge that, if misused, the publication of results can have negative side-effects. Nonetheless, not publishing test results is still perceived as more problematic.

Even though principals highlight the importance of administrative pressure to ensure school actors take responsibility for academic results, they simultaneously explain that the pressure they themselves experience is for either an important part or fully self-imposed:

It is completely clear to them [local authorities] that we have a worse starting point and that things take more time. They have an understanding for that. We are not being compared to [neighborhood X, a very affluent neighborhood], even though we like to compare ourselves to [neighborhood X]. Our students will be just as successful as those in [neighborhood X]. This is actually our motto, to “out-compete” those in [neighborhood X]. That is the goal. [...]. There is nobody who gets angry because there are bad results, but we often become very disappointed ourselves. (P2).

We probably do it to ourselves, I think. [...]. I do not think there are many principals who have no ambition to do well, so you put some pressure on yourself to do well and try to find the methods to boost results. After all, I have to admit, if I had bad results, I would feel that I have not delivered, I would go home with stomach ache. It touches me. (P15).

The pressure that principals impose on themselves, as exemplified by the quotes above, seems to derive for an important part from an internal sense of duty to bring about high academic achievement, and not to let their students down by allowing for sub-optimal performance. As test performance is more and more perceived as a proxy of the extent to which principals succeed in giving their students a chance to get ahead in life, principals increasingly measure their professional success against external metrics and value themselves according to the progress their school makes in moving up in the performance hierarchy. This exemplifies how principals can come to internalize norms embedded in standardized testing and the PBA system as to what counts as relevant purposes of education. In the words of one principal:

[The tests] give us a good clue as to what teaching should contain. Not that we should teach to the test, but it does give us some guidelines regarding what to emphasize, what is important for kids to learn. (P12).

4.1.2 Translation

Principals' actions, in turn, are steered by the belief that aiming at the best possible test results forms a key way to secure educational “success”. In an attempt to do so, principals adopt top-down, performance-oriented management practices and data-driven decision-making. More specifically, academic achievement is formulated as a core organizational goal, sometimes in the form of specific performance targets (e.g. the ambition to perform at or above the municipal average), which guide the schools' organizational, instructional, and pedagogical approaches. In particular in the case of the two lowest-performing schools (P1; P2), the perceived importance of a strong focus on academic achievement seems to contribute to the degradation of other organizational goals and identities. For example, after being newly appointed at a school that for years

had performed significantly below the national average, the principal of school 1 decided to build-down the former school vision, which strongly emphasized the relevance for students' multicultural identity. In contrast, the new vision that was created more strongly emphasizes academic achievement, a decision with significant consequences for how the school is run:

We have laid far below the national average. So, my ambition was to turn the school around. [To] think less about the multicultural and more about the academic content. We have been working a lot on classroom structure ... I have replaced half the school staff. [I have replaced] the entire management. So here have been major upheavals. (P1)

The four principals explain that at their schools, data analysis efforts are highly routinized and test data are used for multiple purposes, including when distributing resources, assessing and comparing teachers' performance, moving around staff members, and deciding on the school's focus areas and improvement plans. The four principals, moreover, show strong commitment to the use of achievement data to identify learning gaps and find out "what works", i.e. what methods and practices result in higher student learning. As explained by one principal, this entails constant reflections on questions such as "could we have done this differently, could we have done better, could we have worked differently so that more students would have understood the tasks?" (P2). In addition to identifying methods and practices that "work", data analysis efforts focus on assessing what particular content areas should receive more attention, so to adjust teaching accordingly.

4.1.3 Mechanism of change

As exemplified above, the four principals seem to have accepted and internalized the conception of education as embedded in standardized testing and the PBA system, which stipulates that the key towards educational "success" or "excellence" lies in securing optimal performance in tested skills. In their efforts, the four principals largely conform to these assumptions and increasingly attempt to become like the definition of "a successful school" embedded in standardized testing and PBA, sometimes at the expense of other educational purposes and identities. The above forms an example of how standardized testing and PBA, by imposing a particular definition of education and encouraging actors to see themselves and behave according to the norms of good practice, and can create self-fulfilling prophecies (cf. Espeland and Sauder 2007).

Various factors are likely to promote principals' internalization of external performance criteria, including the local accountability regime. The four principals work in three municipalities, all of which are characterized by a strong performance orientation. In all three cases, principals report that the municipal superintendent aims at belonging to the top-performing municipalities in Norway, even though two municipalities have a relatively high number of disadvantaged students. To secure this, the three municipalities have a long history of PBA demands and rely on an extensive toolbox that both controls schools and serves to promote organizational learning and reflection. The extensive use of PBA tools, as well as their strong alignment (cf. Maroy 2015), can contribute to a process of "socialization". That is, a climate, where implicit and explicit

expectations emphasize a focus on improving learning outcomes and push schools to conform to expectations surrounding datafication, can come to shape educators' identities and discipline school practices.

Still, whether principals working in such local contexts actually come to perceive test scores as proxies of professional success seems to depend on their views on education, which are shaped by their professional trajectories. In this regard, the analysis hints at the importance of the educational institution where principals obtained formal education in school leadership. It appeared that principals who had obtained their school leadership degree at the Norwegian Business School more often tended to evaluate their professional "success" according to performance metrics, compared to principals who studied at a pedagogical institute of a public university. From the interview data, it remains however unclear whether principals already held particular views on education before they applied to the Norwegian Business School, or whether they developed these views and perceptions during their leadership training or even later on.

4.2 Balancing multiple purposes: performance data to inform decision-making

4.2.1 Interpretation

The second response pattern is employed by principals who only partially embrace the PBA mandate and action theory. Principals in this response category ascribe significant importance to the achievement of basic skills, while simultaneously placing important emphasis on non-cognitive skills and social competencies as key educational purposes. A prominent sentiment among these principals is that the Norwegian school has never been only about learning, but equally about "Bildung". The principals who more explicitly adopt this approach work at schools that vary significantly from one another, both in terms of performance and social composition.⁷ What binds them together, more than an overlap in school-specific characteristics is principals' belief in the validity of the national tests, as well as their positive attitude towards the use of achievement data for school improvement purposes. In this regard, principals argue that the quality of the national tests has improved over time⁸ and explain that the tests respond to an important need for measures to assess whether schools "succeed with teaching" (P20).

National tests, I think, are a tool. After all, we need some assessments that can help us along the way. I think national tests help us to see how we are doing. And the tests have become much better. [...]. We have stopping points once a year, where we look at the academic results, but not in relation to whether the individual student has done well, but more in terms of how we do as a school. [...]. So, we use the results to reflect; "is this where we want to be"? "Do we feel satisfied with this?" (P10).

⁷ This response pattern was found to be the most commonly applied pattern. Principals who more explicitly adopt this approach include P4; P6; P7; P8; P9; P10; P13; P17; P18; P19; P20; P21; and P23.

⁸ Principals are in particular positive about recent changes to the test format, which allow for comparisons over time.

The tests are so good, of such high quality, that I have no problem to account for them. The texts are good, the questions are good, and they really reveal what is important. Also, it is very nice to have a benchmark "Do we reach what we aim for?" "Do we manage to have a positive development?" [...]. It gives us an indication of whether we manage to work systematically enough. (P8).

In other words, the tests are said to offer a “meta-view” (P8), which provides feedback to schools and principals about “where we are” (P20). Principals regard this feedback as an important source of information to foster collective reflection and school improvement.

The principals furthermore report that there are particular “expectations outside of the house on how to perform on the national tests” (P10), expressed by local authorities and politicians, and by some parents.⁹ Generally, principals explain that, even though they are regularly followed up for results and asked to explain and justify performance in particular by the municipal superintendent, such encounters are not experienced as threatening. Rather, such meetings are often described as arenas to gain ideas, get advice, and request support. Nonetheless, the construction of performance rankings, as sometimes presented during collective meetings with the municipal superintendent or in the media, seems to spark some emotional response by a number of principals. That is, such rankings are mentioned to elicit feelings of pride and shame, contributing to pressure and a desire to score well. Moreover, in particular negative publicity following low performance is said to affect some teachers’ and students’ confidence and motivation. The publication of results is therefore often referred to as doing more harm than good.

4.2.2 Translation

When looking more closely at the language of practice, it appears that principals’ belief in the validity of the tests as well as their desire to use test data as a pedagogical and organizational tool results in highly routinized data analysis efforts. Interestingly, even though principals express a critical attitude towards the construction of performance rankings, in particular by media outlets, the analysis highlights how, for many principals, the performance of their school compared to others, as well as to their own previous performance, forms an important reference point to reflect on and reassure their school is “on track”:

The most important thing about the national tests is internal use, it is an internal medicine. And in that context, it is ok to compare yourself to other schools and

⁹ The interview data reveal large differences in the extent to which parents are perceived as putting pressure on schools and educators to obtain high test scores. Many principals argue that parents are generally more concerned with whether their child has a good time at school, than with academic results. If they do express an interest in the latter, this tends to reflect a desire to find out how their own child performs, not necessarily the school as a whole. Nonetheless, at some schools, principals argue that parents show increasing interest in school performance, and raise their voice in case of below-expected performance. In extreme cases, test scores are used when “shopping” for schools (P7), even though opportunities to do so remain limited. That is, even in municipalities that employ freer regulations around school choice, priority remains given to students residing in the school’s catchment area.

other municipalities to see "are we in the right place, are our students weaker than others?". This can provide knowledge about the need to put in extra support in some areas. (P17).

Based on collective reflections on comparisons of schools' relative performances, principals determine whether there exists a need to adapt organizational, pedagogical, or instructional practices. For example, reflection on test performance can contribute to the decision to revisit the distribution of resources, or to offer teachers to take part in professional development courses. Moreover, if it appears that tested students do not manage specific tasks, particular content can be given more emphasis, intensive courses for low-performing students are sometimes set up, or adaptations to teaching methods are considered. The principals explain that some of the implemented changes only account for the grade-level of the tested students, while other changes are more systematic and imply changes at non-tested grade-levels. From this, it appears that, depending on the perceived need to implement changes, the focus on test performance can have an important impact on the school's core activities and practices. Nonetheless, change initiatives remain based on dialogue and collective agreement between school leaders and teachers, and test data form one out of various sources of information that guide principals' decision-making processes. Moreover, in contrast to principals in the first response category, the principals in this response category reject a prioritization of academic performance over other competencies and continue to place significant emphasis on non-cognitive skills and social competencies in their development projects and focus areas, even in the case of low academic performance.

4.2.3 Mechanism of change

In contrast to principals in the first response category, principals in the second category only partially agree with the conception of "a successful school" embedded in standardized testing and PBA, and continue to balance multiple educational purposes. Instead, the new social relations constructed by commensuration form a more central explanatory factor to understand principals' responses. That is, standardized testing, as an example of commensuration, creates a precise hierarchy of schools, which seems to affect how principals evaluate how their school is faring, as they increasingly rely on performance comparisons to assess "progress" and assure their school is "on track". As such, the perceived need to undertake action is increasingly shaped by a school's relative performance, where previous performance, or the performance of other schools or the municipality as a whole, serves as reference points.

At first glance, it appears that performance comparisons contribute in particular at low-performing schools to a perceived need to adapt practices so to move up in the performance hierarchy. It is also the principals of these schools that express particular discomfort with the construction of performance rankings by some municipal superintendents or media outlets, which further contribute to a desire to score well, so to avoid public humiliation. Nonetheless, upon deeper examination, it appears that also principals of average-performing schools or even those at the top of the performance hierarchy can perceive a need to adapt their organizational, instructional, or pedagogical practices as a result of performance comparisons. That is, many principals seem to evaluate their performance in relation to performance expectations, taking their student

population into account. As such, also average- and high-performing schools can perceive a need to undertake action, for example when their relative performance drops or when they perform below what can be expected from their student body.

4.3 Symbolic responses: value-driven education and a focus on the non-measurable

4.3.1 Interpretation

The third response pattern is employed by principals who show weak commitment to the PBA mandate and action theory. These principals tend to more actively question the validity of the national tests,¹⁰ while simultaneously argue that the tests form a narrow measure of what the schools' priorities are. Principals' concerns about the tests' validity reflect more general skepticism regarding the often-proclaimed superiority of standardized, quantitative data over particularistic ways of knowing. Instead, principals emphasize the importance of professional judgement and knowledge in fostering school improvement. Moreover, rather than a predominant focus on performance and academic excellence, principals support a broader approach to learning and emphasize the humanistic aims of education, arguing that "if we get bad at that, it is dramatic, both democratic and socially" (P4). Three of the six principals who more explicitly adopt this approach work at schools that perform below municipal and national averages (P3, P5, P16), while the other three principals work at schools that perform at or above average (P11, P14, P22). A common sentiment expressed by all six principals is that, rather than imposing pressure on oneself to obtain high academic results, principals feel pressured to "make sure all students feel safe and have a good time at school" (P3).

As the principals in this response category generally reject the norms embedded in the tests and PBA system as to what constitutes quality education, they express particular concern about how public and political debates on school quality have narrowed to a discussion of test scores:

[The tests] show a small part of the picture. I feel that maybe it has become a little too much, that everything is measured only by them. It seems like the only thing people talk about, in a way, how the school performs at national tests. [...]. Clearly it is important that pupils achieve basic skills, but there are many other things that are important here as well. [...]. I think that the curriculum has swung and that it gets far too much attention compared to everything else that we have to do as well. (P16).

The six principals report that in particular local politicians, the media, and some parents increasingly rely on test scores to evaluate and compare schools, sometimes drawing quick conclusions based on what principals refer to as uncritical readings of narrow, quantitative data. As a result, principals worry that even small and insignificant differences in performance can have real consequences for schools, when external

¹⁰ Principals express concerns about the validity of the national tests on multiple grounds. For example, one principal explains that it is not realistic to demand from 5th-graders to sit still for 90 min to conduct a test. As students struggle to do so, the tests are perceived to measure 'who has best prepared their students to sit as quietly as possible' (P5), rather than what students have learned or can do. Other principals argue the tests predominantly reflect students' backgrounds, rather than the quality of the school's teaching practices.

actors will use this as a basis to question and interfere with the school's educational project and pedagogical approach:

We, according to the value-added model, are such a school: we perform well, but according to the student base we have, we should be able to perform better, if you read those numbers a bit uncritically. And then I got a bit worried about it because we see that they take them... I got kind of nervous because politicians read numbers very easily and just look at some tables, and think "OK, but then we do this and that". But when Statistics Norway explains how to read the value-added model, they provide a report of over 100 pages. (P22).

When we have good results, they are out to tell us that we should have had better results, with this local neighborhood. Then I think, "No, we cannot start with those discussions". But of course, I am concerned that we have to deliver good enough. We have to deliver such good results on national tests that we do not get pressured, right? Because the moment we start delivering really poorly, then we will, then we will start to get critical ..., then our entire pedagogy will be put under pressure. (P11).

4.3.2 Translation

Even though the fear of being called out or told what to do by external actors motivates principals to act, they strongly disapprove of what they refer to as short-sighted solutions to quickly raise test scores, such as spending more time on particular test content when students score low on specific tasks. As argued by one principal, "it might be that it is not more mathematics that you should have when you are bad at math. It might be that you need more arts and crafts, or more physical education" (P11). Principals acknowledge that schools adopting such "quick fixes" are likely to see an increase in their test scores in the short-term, but at the same time emphasize that in the long-run "they fall like a rock, because they have not worked on what underpins the results" (P11). Rather than prioritizing academic performance over other educational purposes, the vision statements of the six schools highlight a similar focus on inclusion, solidarity, and creativity as core values, while practical esthetical subjects occupy a central place in the schools' development projects. Moreover, in contrast to principals in the first response category, the notion of setting specific performance targets is dismissed, and principals are critical of efforts to look for straightforward solutions to complex learning problems. Rather, they argue, focus should lie on continuous and steady improvement of classroom practices.

While the use of national test data is not completely rejected, the tests are generally regarded as offering too narrow, too limited, and too unreliable information to be used when making important school decisions. Rather, principals report that the national tests are predominantly used to identify students in need of additional support and follow-up. In contrast to the other schools, data analysis efforts seem less routinized and appear to form a more isolated practice. In some cases, principals themselves do not take part in data analysis meetings but leave these efforts to a team of teachers. Whereas some of the principals appear to perceive some value in using test data to identify learning gaps, for others it seems primarily a way to comply with institutional

expectations surrounding data-use. While they, in line with policy expectations, have established systems for analyzing test scores, they continue to express much greater faith in and report to rely on other assessment measures as well as teachers' judgement and knowledge to foster school development.

Despite that such predominant symbolic responses allow principals to keep running their schools as they see most appropriate, in particular at those schools where performance swings from year to year, or at schools that perform below what can be expected from their student population, the perceived threat of increased pressure from local politicians, as well as complains by parents, forms an important source of concern, which sometimes forces principals to respond in ways that challenge their own principle beliefs and values:

When the Knowledge Promotion¹¹ and all this came, with more focus on results and stronger steering, we noticed that with the new grade-levels, we began to retreat. We did not dare to be as progressive [as before]. We did not get any better results, on the contrary, but we were tricked into it. [...]. We did the same things, but we did a little less, because we had to do a little more of this. So then you end up training kids on what they are not good at, instead of cultivating what they are good at, and then you try to do the rest afterwards. (P11).

4.3.3 Mechanism of change

In contrast to the principals in the first response category, who imposed pressure on themselves to obtain high test scores, the six principals in this response category predominantly experience external or socially imposed pressure to aim at high test performance. This pressure is exercised by local politicians, the media, and some parents, following the increasing perception that test scores form prime indicators to evaluate, judge, and compare school quality. This perception is influenced by features of commensuration. That is, by simplifying and condensing information, while making it seem more authoritative, forms of commensuration (i.e. standardized tests) attract attention (Espeland and Sauder 2007). The more attention is paid to test performance, the more differences between entities become predominantly expressed by the interval on the shared performance metric. Other ways of distinguishing between entities become, in turn, less salient. For example, as explained by the principals, media reports on test performance frequently use headings such as “the best schools”, while contextual information is often erased from such reports, and little attention is paid to whether data are fit for cross-comparisons. As such, the presentation of test scores appears robust and definite, and is increasingly used for making general claims about the overall quality of schools.

As outlined above, principals are in most cases able to keep running their schools as they see fit, regardless of pressure exercised by external audiences. With the exception of one, all principals work in municipalities characterized by a weak performance orientation and loosely coupled accountability tools. This is likely to provide them

¹¹ Name of the curriculum reform introduced in 2006, which transformed how Norwegian schools were governed.

greater leeway to respond predominantly symbolically to the expectation to raise test scores and use achievement data to foster school improvement. Nonetheless, regardless of the weaker performance orientation of municipal superintendents, pressure exercised by local politicians, the media, and parents can sometimes force principals to respond in ways that go against their own views on education. These cases highlight how the increasing perception of local politicians, the media, and some parents that test scores from relevant and robust indicators of school quality, shaped by features of commensuration, can reinforce self-fulfilling prophecies, by generating social pressure to conform to the definitions of good practice embedded in standardized testing and PBA.

5 Discussion and conclusion

This study has examined how primary school principals reflect on and respond to being measured, compared, and held accountable for school performance on national tests. The findings highlight three distinct response patterns in how principals perceive, interpret, and translate policy demands emphasizing data-driven decision-making and PBA. Principals who strongly align to the PBA mandate and “theory of action” attempt to secure optimal performance by adopting performance-oriented management practices and by using achievement data to plan school improvement. In contrast, principals who only partially embrace the PBA mandate and action theory seem to respond by balancing multiple purposes. That is, principals report to rely on performance data to identify learning gaps and reflect on pedagogical and organizational challenges, but reject a prioritization of academic performance over non-cognitive competencies. Finally, principals who experience a significant mismatch between their views on education and the central policy demands predominantly employ symbolic responses. While they comply with the institutional expectation to use test data, this appears to remain a more isolated practice from their core activities.

By examining the three response patterns through the lens of two social mechanisms identified by Espeland and Sauder (2007), this study has attempted to explain *why* principals respond in particular ways to standardized testing and PBA. The findings indicate that different manifestations of self-fulfilling prophecies and commensuration form important explanatory factors to understand how standardized testing and PBA can give rise to complex, creative and sometimes unanticipated responses. The study simultaneously highlights how the mechanisms are more likely to be activated under particular conditions, which relate to both principals’ trajectories and views on education, as well as to school-specific characteristics, such as the school’s relative performance level and parental expectations. Finally, the level of “strictness” of the local accountability regime, as well as the level of activity of the local press, is found to play an important role.

More specifically, the study’s findings show how standardized testing and PBA can operate as a self-fulfilling prophecy when principals come to see themselves and act according to the criteria of good practice embedded in the tests and PBA system. This implies that measures, such as standardized tests, increasingly create what they are meant to describe. It was found that the local accountability regime, and in particular the history of PBA demands and the degree of alignment between PBA tools, can favor principals’ internalization of performance metrics as proxies of professional success.

Nonetheless, whether principals actually do so seems to depend on their views on education, which are influenced by their professional trajectories. In contrast to previous research, which shows that the point of entry to the teaching profession affects principals' desire to be seen as successful according to system metrics (see Heffernan 2017), this study's findings hint at that for Norwegian principals the institution where they obtained formal education in school leadership, that being the Norwegian Business School or a pedagogical institute at a public university matters more than the entry point to the profession. Nonetheless, from the interview data, it is not possible to draw definite conclusions about which factors shaped principals' perceptions. Further research is needed to examine the potential relationship between principals' values and views on education and the type of educational institute where they obtained formal education in school leadership.

For those principals who to a lesser extent perceive test scores as proxies of professional success, social pressure can still contribute to them conforming to the expectations embedded in the tests. This form of pressure can grow when external audiences or society in general come to embrace the definition of education quality promoted by the tests and PBA system, a process which can be reinforced by features of commensuration. On the one hand, this seems to affect low-performing schools in particular, as these schools become increasingly perceived as "low-quality" schools and sometimes face significant administrative and social pressure to improve their performance, which in turn can reinforce self-fulfilling prophecies. Nonetheless, this study's findings indicate that average- and high-performing schools are not spared from this form of social pressure. With the recent publication of the value-added model, attention has shifted towards whether schools perform in line with what can be expected of them, considering their student body. That is, increasing use of the value-added model by external audiences such as local politicians and the media, following the perception that the value-added model forms a more accurate measure of the school's contribution, implies that high-performing but low-contributing schools are now subject to increased questioning.

The analysis, furthermore, shows that features of commensuration influence not only external audiences' perceptions of schools, and as such can impact principals' behavior, but also principals' self-perceptions. To assure schools are "on track", many principals monitor and compare their performance, where previous performance or the performance of other schools or the municipality serve as reference points (see also Skedsmo 2018). This information is then used to reflect on whether there exists a need to adapt organizational, instructional, or pedagogical approaches. Similar findings were reported by Feniger et al. (2015, p. 15), who showed that school comparisons based on test scores became "a major lens through which principals look at their own school and accordingly make decisions". While some principals interviewed for this study rejected that test scores reflect school quality, the majority of them expressed the belief that test data uncovered an important truth about the school situation and the effectiveness of the schools' teaching practices.

Principals' supportive attitude towards the use of performance data to reflect on and modify educational practices, which aligns to findings reported by Seland et al. (2013), may reflect technological advances, such as improvements made to the test format, as well as development of support systems in how to make productive use of achievement data. At the same time, the relatively broad support can imply a process of naturalization or institutionalization. Whereas initially strongly resisted, the tests seem to have become more accepted over time (see also Gunnulfsen and Møller 2017; Mausethagen

2013), in particular among primary school principals (Seland et al. 2013). Similarly, despite that principals differed in how much importance they ascribe to the achievement of basic skills compared to other educational goals and purposes, all proclaimed to perceive acquisition of basic skills as an important school mission, which resulted in at least a mild interest in how individual students performed at the national tests. Nonetheless, despite principals' willingness to take responsibility for centrally defined learning goals, and to use achievement data for pedagogical adaptation, most principals remain critical of some of PBA's policy tools, most notably the publication of results. While many argue to value transparency in terms of learning outcomes, the interpretation and use of test scores by external actors, specifically the media and some politicians, are perceived as doing more harm than good.

While the study's findings have provided insight into the reflexive interactions between actors and measures, it is important to highlight that the self-reported nature of the data implies that the findings, in particular with regard to classroom practices, should be regarded as beliefs and intentions, rather than as evidence of what is happening in classrooms (Creswell 2009). Similarly, even though some principals may have embraced the PBA mandate, they remain for an important part dependent on teachers' willingness and capacity to incorporate real changes in the classroom. Compared to school leaders, teachers have generally positioned themselves more critical to the usefulness of national test data (Seland et al. 2013; Skedsmo 2018), and studies have highlighted how teachers continue to struggle with how to respond to demands from national tests (Gunnulfsen 2017; Mausethagen 2013). Previous research has moreover indicated that willingness to use performance data does not guarantee productive data-use (Gunnulfsen 2017; Mausethagen et al. 2017). This highlights the need for future research on how teachers reflect on and respond to the different mediations of PBA demands employed by principals, possibly combining interviews with teachers with classroom observation.

To sum up, by examining how school principals in different local accountability regimes and at different schools perceive, interpret, and translate PBA demands, this study has contributed to opening up the "black box" of policy enactment in accountability contexts characterized by an ambition to elicit change "from the inside" by influencing actors' dispositions, as well as a relative absence of material consequences. Moreover, by adopting a mechanism-based approach, and by examining the conditions under which the mechanisms operate, this study has contributed to the understanding of how to interpret the variegated school responses adopted in these accountability contexts. The study's findings highlight how standardized testing and PBA, even in the absence of material consequences and low levels of marketization, can drive behavioral change, by establishing new norms of good educational practice, and by changing how educators make sense of core aspects of their work.

Acknowledgements I would like to thank all principals involved in the study for kindly giving their time to participate in an interview. I would also like to acknowledge and thank Clara Fontdevila, Antonina Levatino, Guri Skedsmo and Antoni Verger, as well as two anonymous reviewers. The paper has benefited greatly from their careful readings of earlier drafts.

Funding This work was supported by the European Research Council under Grant 680172.

Data availability Anonymized data can be made available to researchers upon request.

Compliance with ethical standards

Conflicts of interest The author declares no conflict of interest.

Code availability Atlas.ti coded. Coding protocol can be made available upon request.

References

- Au, W. (2007). High-stakes testing and curricular control: a qualitative meta-synthesis. *Educational Researcher*, 36(5), 258–267.
- Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, 18(2), 215–228.
- Ball, S. J., Maguire, M., & Braun, A. (2012). *How schools do policy: policy enactments in secondary schools*. London: Routledge.
- Ball, S. J., Junemann, C., & Santori, D. (2017). *Edu-net: globalisation and education policy mobility*. London: Routledge.
- Barbana, S., Dumay, X., & Dupriez, V. (2019). Local implementation of accountability instruments in the French-speaking community of Belgium. *European Educational Research Journal*, 19(2), 94–108.
- Braun, A., Ball, S. J., Maguire, M., & Hoskins, K. (2011). Taking context seriously: towards explaining policy enactments in the secondary school. *Discourse: Studies in the Cultural Politics of Education*, 32(4), 585–596.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4), 297–312.
- Candido, H. H. D. (2020). Datafication in schools: enactments of quality assurance and evaluation policies in Brazil. *International Studies in Sociology of Education*, 29(1–2), 1–32.
- Coburn, C. E. (2001). Collective sensemaking about reading: how teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145–170.
- Coburn, C. E. (2004). Beyond decoupling: rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, 77(3), 211–244.
- Courtney, S. J. (2014). Post-panopticism and school inspection in England. *British Journal of Sociology of Education*, 37(4), 623–642.
- Creswell, J. (2009). *Research design: qualitative, quantitative, and mixed methods approaches*. Los Angeles: SAGE Publications, Incorporated.
- Diamond, J., & Spillane, J. (2004). High-stakes accountability in urban elementary schools: challenging or reproducing inequality? *The Teachers College Record*, 106(6), 1145–1176.
- 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.
- Espelund, W. N., & Sauder, M. (2007). Rankings and reactivity: how public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1–40.
- Espelund, W. N., & Stevens, M. (1998). Commensuration as a social process. *Annual Review of Sociology*, 24, 312–343.
- Falabella, A. (2014). The performing school: the effects of market and accountability policies. *Education Policy Analysis Archives*, 22(70), 1–26.
- Feniger, Y., Israeli, M., & Yehuda, S. (2015). The power of numbers: the adoption and consequences of national low-stakes standardised tests in Israel. *Globalisation, Societies and Education*, 14(2), 183–202.
- Gunnulfson, A. E. (2017). School leaders' and teachers' work with national test results: lost in translation? *Journal of Educational Change*, 18(4), 495–519.
- Gunnulfson, 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.
- Hedström, P., & Swedberg, R. (1998). Social mechanisms: an introductory essay. In P. Hedström & R. Swedberg (Eds.), *Social mechanisms: an analytical approach to social theory* (pp. 1–31). Cambridge: Cambridge University Press.
- Heffernan, A. (2017). The accountability generation: exploring an emerging leadership paradigm for beginning principals. *Discourse: Studies in the Cultural Politics of Education*, 39(4), 509–520.

- Keddie, A. (2013). Thriving amid the performative demands of the contemporary audit culture: a matter of school context. *Journal of Education Policy*, 28(6), 750–766.
- Kommunal Organisering. (2016). Kommunefil. NSD – Norsk senter for forskningsdata AS, NSD, 2016. <https://doi.org/10.18712/NSD-NSD2406-V1>.
- Larsen, E., Møller, J., & Jensen, R. (2020). Constructions of professionalism and the democratic mandate in education: a discourse analysis of Norwegian public policy documents. *Journal of Education Policy*, 1–20. <https://doi.org/10.1080/02680939.2020.1774807>.
- Lingard, B., Sellar, S., & Lewis, S. (2017). Accountabilities in schools and school systems. In G. Noblit (Ed.), *Oxford research encyclopaedia of education* (pp. 1–28). New York: Oxford University Press.
- Maroy, C. (2015). Comparing accountability policy tools and rationales: various ways, various effects? In H. G. Kothhoff & L. Klerides (Eds.), *Governing educational spaces: knowledge, teaching, and learning in transition* (pp. 35–56). Rotterdam: Sense Publication.
- Maroy, C., & Pons, X. (2019). *Accountability policies in education. a comparative and multilevel analysis in France and Quebec*. Cham: Springer.
- 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.
- Mausethagen, S., Prøitz, T., & Skedsmo, G. (2017). Teachers' use of knowledge sources in 'result meetings': thin data and thick data use. *Teachers and Teaching*, 24(1), 37–49.
- Mennicken, A., & Espeland, W. N. (2019). What's new with numbers? Sociological approaches to the study of quantification. *Annual Review of Sociology*, 45(1), 223–245.
- Merton, R. K. (1968). *Social theory and social structure* (3rd ed.). Glencoe: Free Press.
- Mittleman, J., & Jennings, J. L. (2018). Accountability, achievement, and inequality in American public schools: a review of the literature. In B. Schneider (Ed.), *Handbook of the sociology of education in the 21st century* (pp. 475–492). Cham: Springer.
- Møller, J., & Skedsmo, G. (2013). Modernizing education – NPM reform in the Norwegian education system. *Journal of Educational Administration and History*, 45(4), 336–355.
- Piattoeva, N., & Boden, R. (2020). Escaping numbers? The ambiguities of the governance of education through data. *International Studies in Sociology of Education*, 29(1–2), 1–18.
- 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.
- Seland, I., Vibe, N., & Hovdhaugen, E. (2013). Evaluering av nasjonale prøver som system [Evaluation of national testing as a system]. Published online: <https://www.nifu.no/publications/1025043/> Accessed 1 Feb 2020.
- Sellar, S. (2015). A feel for numbers: affect, data and education policy. *Critical Studies in Education*, 56(1), 131–146.
- Skedsmo, G. (2018). Comparison and benchmarking as governing processes in Norwegian schools. In L. M. Carvalho, L. Levasseur, M. Liu, R. Normand, & D. A. Oliveira (Eds.), *Education policies and the restructuring of the educational profession* (pp. 137–158). Singapore: Springer.
- Skedsmo, G., & Mausethagen, S. (2016). Accountability policies and educational leadership – a Norwegian perspective. In J. Easley, P. Tulowitzki, & P. (Eds.), *Accountability and educational leadership – country perspectives* (pp. 205–223). London: Routledge.
- Spillane, J. P., & Jennings, N. (1997). Aligned instructional policy and ambitious pedagogy: exploring instructional reform from the classroom perspective. *The Teachers College Record*, 98(3), 449–481.
- Spillane, J. P., Diamond, J. B., Burch, P., Hallett, T., Jita, L., & Zoltners, J. (2002). Managing in the middle: school leaders and the enactment of accountability policy. *Educational Policy*, 16(5), 731–762.
- Statistics Norway. (2019). Pupils in primary and lower secondary school. Statistics Norway. <https://www.ssb.no/en/utdanning/statistikker/utgrs>. Accessed 1 Feb 2020.
- Thiel, C., Schweizer, S., & Bellmann, J. (2017). Rethinking side effects of accountability in education: insights from a multiple methods study in four German school systems. *Education Policy Analysis Archives*, 25(93), 1–32.
- Tveit, S. (2014). Educational assessment in Norway. *Assessment in Education: Principles, Policy and Practice*, 21(2), 221–237.
- Verger, A., Parcerisa, L., & Fontdevila, C. (2019). The growth and spread of national assessments and test-based accountabilities: a political sociology of global education reforms. *Educational Review*, 71(1), 5–30.
- Verger, A., Ferrer-Esteban, G., & Parcerisa, L. (2020). In and out of the 'pressure cooker': schools' varying responses to accountability and datafication. In S. Grek, C. Maroy, & A. Verger (Eds.), *World yearbook of education 2021*. London: Routledge.