



Enacting teacher evaluation in Norwegian compulsory education: teachers' perceptions of possibilities and constraints

Silje Kristin Gloppen¹ 

Received: 9 January 2023 / Accepted: 30 May 2023 / Published online: 28 June 2023
© The Author(s) 2023

Abstract

This article reports from a case study investigating enactment of teacher evaluation (TE) policies in two lower secondary schools in a large Norwegian municipality. The aim of the study was to explore teachers' perceptions of TE, and to unpack how new policy initiatives were shaped by translations, mediations, and negotiations in a national context characterised by relatively low-stakes accountability. The analysis was informed by municipal and local school documents, in situ observations, and semi-structured in-depth interviews with seven teachers, focusing on the interplay between material, interpretative, and discursive elements. Findings indicate that TE practices reflect traditional approaches to teachers' professional development based on classroom observations (COs) and peer counselling, marked by symmetrical relationships, reflexive collaboration, and collective knowledge-sharing. However, a non-hierarchical collective-oriented culture seems to be disrupted by result management and standards that define 'the good teacher' or 'the good lesson' combined with individual COs. Teachers raise concerns that market discourses which treat pupils as 'customers' and teachers as 'providers of results' may encourage strategic behaviours to comply with expectations and preserve one's good reputation. Implications for policy and practice are discussed with regard to how TE as a discursive strategy contributes to the 'making' of particular teacher subjectivities.

Keywords Teacher evaluation · Case study · High versus low stakes · Performance-based accountability · Discursive strategy

✉ Silje Kristin Gloppen
s.k.gloppen@ils.uio.no

¹ Faculty of Educational Sciences, Department of Teacher Education and School Research, University of Oslo, Oslo, Norway

1 Introduction

According to the existing literature, teacher evaluation (TE) practices serve the dual and often conflicting purposes of professional development (PD) and accountability (Berliner, 2018; Hazi, 2022). Presented as a formative tool for professional development, teachers in multiple contexts encounter various designs and uses of TE, most commonly as a summative approach to measuring and comparing teaching practice across time and space (see for example Hallinger et al., 2014; Lillejord et al., 2014). In recent decades, one of the most enduring arguments for TE contends that ‘raising teaching performance is perhaps the policy direction most likely to lead to substantial gains in student learning’ (OECD, 2005). The Organisation for Economic Cooperation and Development (OECD) has subsequently played an important role in the dissemination of research and meta-analyses of ‘teacher effectiveness’ (see for example Reynolds, 2007) to promote and underscore the assumption that transparency and accountability inherently contribute to PD and strengthen the legitimacy of the teaching profession (OECD, 2011, 2013, 2019).

The various designs and uses of TE are shaped by differing purposes and emphases in different accountability contexts. The integration of TE tools and frameworks into national and local governance influences how TE is perceived and enacted by school actors in the field (Reddy et al., 2018). As discussed in the next section, the literature on teachers’ perceptions of TE reveals a discrepancy between policy intentions and teachers’ enactment. In combination with marketization reforms, the push for accountability in high- or low-stakes evaluation frameworks has accelerated the debate around how the ‘terrors of performativity’ (Ball, 2003) might actually limit PD. Critical studies point to a paradigm shift in the construction of teacher professionalism as logics of accountability and ‘governing by numbers’ reshape the ways in which teachers are recognised and valued (Holloway & Brass, 2018; Lewis & Holloway, 2019). One key concern in this regard is that, rather than merely identifying a gap between current and desired performance, TE feedback should also specify development activities that are likely to improve teaching practice and, by extension, student performance.

The context of Norwegian compulsory education makes a particularly interesting case as it has been labelled in the literature as a low stakes accountability context (Camphuijsen, 2020). Despite OECD advice to incorporate TE into a national framework for quality assessment (OECD, 2011, 2013), Norway has remained resilient to include TE into national assessment policies. To date, little is known about how TE policies are enacted in Norwegian classrooms. To address this knowledge gap, the present case study examined how teachers at two urban lower secondary schools in a large Norwegian municipality made sense of and negotiated new TE policies, practices, and expectations in a municipality with relatively well developed frameworks for quality assessment and accountability. The general purpose was to explore and discuss how teachers expressed their agency in the ongoing debate around the discursive impacts of TE as a governance tool. The investigation and analysis were guided by the following research question:

How are teacher evaluation policies enacted by teachers at two different schools in a large Norwegian municipality? This qualitative case study was informed by enactment theory, which holds that national and local policy initiatives and legislation interact with complex discursive processes that are contextually mediated and institutionally rendered (Ball et al., 2012). Policies are not ‘done’ at a single point in time but as ongoing processes of ‘becoming’ (Ball et al., 2012). In that sense, local discourses as well as educators’ agency in enacting TE policies influences meaning-making processes that include translation, mediation, negotiation, acceptance, and refusal. Drawing on existing international research, I sought to capture, display, and discuss the processes of ‘policies becoming’ in national and local TE initiatives, with particular regard to how the participants experienced the emerging tensions between expectations, PD, and accountability. I frame conceptualisations of TE as a policy tool according to its dual purposes: PD and performance-based accountability.

The data sources included key municipal and school documents related to the two case schools, semi-structured in-depth interviews with seven teachers and observations of staff meetings and appraisal interviews following classroom observations (COs). The participating teachers’ perceptions and reflections on possibilities and constraints revealed emerging tensions within and between their perceptions and the underlying logics of PD and TE as ‘discursive strategy’ (Ball et al., 2012). The study elucidates interactions between TE and teacher subjectivities and identifies what is at stake in TE as a tool for teacher governance in a low-stakes context.

The rest of the paper is structured as follows. To begin, I contextualise the study by presenting a brief overview of international research on teachers’ perceptions of TE before situating the study in a Norwegian context. I go on to outline enactment theory and the constructs and perspectives that informed the study. I then elaborate the methods, cases, and data before presenting and discussing the findings through the lens of enactment and previous research. The paper concludes by highlighting some neglected aspects of TE in Norway which might contribute to inform debate and future policies, namely, how teachers’ lived experiences display connections between structural constraints inherent in TE and teachers’ possibilities for agency.

2 Research on teachers’ perceptions of TE

Much research is provided on various designs and outcomes of specific models for TE, while less attention is paid to how teachers experience and respond to new policy initiatives. Research suggests that impact and sustainability of TE systems are affected by practitioners’ perceptions, and that further research of their lived experiences with regard to TE is needed to inform policy debates (Pauffer & Sloat, 2020). For the purposes of the present study, I reviewed the international

research on teachers' perceptions of TE as a broader context for the Norwegian case. Moreover, I pertained to observe how context matters (Ball et al., 2012), especially with regard to the influence of stakes attached to TE procedures.

In the USA, which can be regarded as a high-stakes context, approaches to TE have adopted logics of performance-based accountability applied for summative assessment and managerial decision-making regarding promotion, tenure, and retention (Lavigne & Chamberlain, 2017). In most states, value-added models control for prior testing history and student- and school-level variables to measure the 'value' a teacher 'adds' to student achievement and growth during a school year. These practices have been criticised for not upholding scientific standards such as validity, reliability, bias, and fairness (Amrein-Beardsley, 2014) which have prompted teachers to initiate lawsuits for unfair and unjust dismissals (Paige & Amrein-Beardsley, 2020). From this context, the literature reports unintended practical consequences that include declining teacher job satisfaction, motivation, and self-efficacy (Cuevas et al., 2018), as well as a decline in trust, agency, engagement, efficacy, and emotional resilience among teachers (Wilcox & Lawson, 2018). Other studies highlight increased teacher stress, fear, anxiety, and burnout (Conley & Glasman, 2008; Day, 2002; Donaldson, 2016; Ford et al., 2017; Larsen, 2009). In Dunn (2018) analysis of 23 publicly available resignation letters from teachers across the USA, he contended that the extensive evidence and teacher stories of unintended consequences problematised TE as a product of neoliberal policies that prioritise superficial performativity rather than authentic professional growth (see also Ball, 2003; Bradford & Braaten, 2018; Head, 2019). Head (2019) explained teachers' responses as a reflection of what he characterises as the disempowerment and de-professionalization of US educators in recent decades, and that TE functions as a means of shifting responsibility and blame for system-level failures onto the individual teacher. In addition, questions are raised as to whether high stakes linked to TE may have unintended consequences as described above, rather than the intended improvement of pupils' learning outcomes, but also that such systems may contribute to increase the learning gap between pupils linked to attendance in more or less affluent schools (Lavigne, 2014).

Similar patterns have been reported in other countries, where TE policies are less developed and the stakes are lower than in the USA. According to one Israeli study, teachers perceived that TE contributed to a shift from internal responsibility to external accountability for judgement and control (Avidov-Ungar, 2018). In Japan, Katsuno (2010) found that performativity was enacted through the micro-power relations within schools, potentially endangering staff relationships. In South Korea, Ha and Sung (2011) investigated contextual features and teachers' resistance to TE. They reported that teachers generally expressed negative opinions about TE programme implementation and that school cultural factors such as seniority and concerns about job security did not facilitate the original policy intention. Similarly, Chen and Teo (2020) reported that Chinese teachers favoured low-stakes assessments, which they believed were more indicative of learning, teaching, examination, and school accountability. In Portugal, Flores (2018) found that teachers' perceptions reflected uncertainty and scepticism owing to lack of recognition, a bureaucratic and summative structure and conditions that did not facilitate recommended practices. Flores also reported that TE provoked new debates within the teaching

profession. In a more recent study (2018), she identified persisting tensions and problems related to key features of the TE model, including a lack of recognition of those conducting appraisals, leading to tensions among staff members and deterioration of the school climate. Noting Portuguese teachers' resistance to COs and the conflicting roles of department heads, De Lima and Silva (2018) identified a need for leadership preparation and training to enhance observations and engage teachers in joint reflection on how to create opportunities for PD. On a more positive note, Delvaux et al. (2013) performed a multilevel analysis of Dutch teachers' perceptions of a TE system that emphasised the role of feedback in PD. According to their findings, teachers felt that the evaluation system had only limited effects on their PD, but teachers with less than 5 years experience reported that constructive feedback contributed positively to PD.

In general, the above research points to a dissonance between policy intentions and teachers' perceptions of TE, with a range of unintended consequences, across, and especially from within high stakes contexts. It seems important, then, to investigate these tensions in greater depth, in order to broaden the perspectives on enactment within differing contexts and how TE policies 'become' along lines of high- or low-stakes accountability.

3 The Norwegian context

Norway's social democratic governance principles and norms make it a particularly interesting context in which to explore TE. Specifically, the policy context is regarded as 'low-stakes', with less visible sanctions or rewards (Camphuijsen, 2020) and no current national framework for TE. Ensuring equal education for all is a central political goal, which means that all pupils are to be enrolled in regular classes and that teaching must be adapted to individual skills and needs. Despite proposals to implement TE as part of a national framework for quality assessment (NKVS)¹ (OECD, 2011, 2013; NOU, 2019:3), municipalities remain responsible for quality assessment in local schools and for the professional development of teachers and school leaders in primary and lower secondary education (Skedsmo, 2018). As part of the 'school-based assessment' of education quality, responsibility for TE is often delegated to principals, who are expected to support teachers by providing feedback from pupils and observations (Lejonberg et al., 2018).

The OECD is widely acknowledged as an important driver of national policy reform and efforts to improve education quality (Skedsmo, 2018). In particular, the 'PISA shock' of 2001 increased the political focus on enhanced student outcomes as an indicator of school quality. Based on organisational analyses, standpoint analyses, student surveys, and national testing, tools were subsequently developed to support

¹ The National Quality Assessment System for Schools (NKVS) was established in 2004 as a key element of the reform and new National Curriculum named 'Kunnskapsløftet 2006'. The framework is a public system for assessing the quality of Norwegian schools, based on results from multiple sources of information including various tests (including national tests), user surveys, and school inspections.

quality assurance and local school-based assessment. From 2000 onwards, local routines were developed for quality assessment and accountability, including results meetings, risk analysis, contracts, concern meetings, and performance appraisals (Skedsmo, 2018).

Despite the relative lack of TE frameworks and material result-based consequences, Norwegian teachers are indirectly assessed through national and local systems for quality assessment using a range of evaluation tools. The extent to which information from these tools is embedded in accountability practices varies according to local government policies. However, research suggests that national and local tools and school benchmarking have rearticulated and accelerated public and policy discussions around teacher professionalism (Gloppen & Novak, 2023; Mausethagen, 2013; Skedsmo, 2018), as performance indicators and national and international tests have increased the pressure on local actors to align with PBA drivers and expectations (Camphuijsen et al., 2020; Mausethagen et al., 2021). As public debate and discourse centres increasingly on ‘best practice’ or ‘evidence-based practices’ and ‘what works’, new TE tools and practices are emerging in Norwegian classrooms; see for example Kvernbekk (2011) for a detailed discussion of the concept of ‘evidence’ in evidence-based practice.

Building on international research on TE and teacher resistance, Elstad et al. (2015) investigated Norwegian teachers’ perceptions of TE-related stress. They found that stress and resistance were positively associated with the perceived control function of TE rather than actual practices. In a subsequent study of student evaluations of high school teachers, the same authors reported a positive association between teachers’ perceived usefulness and (1) developmental purposes, (2) how school leaders communicated, and (3) teachers’ acknowledgement of students’ ability to evaluate teaching (Elstad et al., 2017). The authors called for further research into TE and contextual factors in Norwegian education, but to date, these issues remain relatively underexplored (Elstad et al., 2015; Lillejord et al., 2014), and few studies have investigated the role of teachers’ enactment of TE in education governance. To address this lack of research on TE in Norwegian compulsory education, and in particular on teachers’ perceptions and concerns, the present study employs the enactment perspective (Ball et al., 2012) as elaborated below.

4 Analytical and methodological framework

The study draws on Ball et al. (2012) framework for examining ‘how schools do policy’—that is, how education policies are enacted in schools by micro policy makers—in this case, teachers. Rather than addressing how measures are implemented from policy level downwards to fields of practice, the enactment perspective examines how policy agendas interact with local practices (Ball et al., 2012), focusing on how policies ‘become’ through reflexive social practices and processes—how actors’ make sense of, interpret, mediate, negotiate, accept, or reject new initiatives (Ball et al., 2012, pp. 1–4). In different countries, various contextual features and adaptations shape the processes of translation, interpretation, mediation, and policy adaptation. From the perspective

of the state or the relevant institution, policies succeed to the extent that they gain acceptance and are internalised into new practices, based on a congruence between the values that underpin reforms and the historical and cultural context (Clarke & Newman, 2009; Møller & Skedsmo, 2013; Ozga & Jones, 2006). In the same way, institutional factors at county, municipal, and/or school levels affect how policies are enacted in schools (Ball et al., 2012). Different schools serve different groups of pupils and parents and often attract or retain teachers or school leaders with disparate capacities, potentials, and limits (Ball et al., 2012). For that reason, taking context seriously means reflecting on the nature of the situated and material contexts, professional cultures, and external environments in which policies are enacted (Ball et al., 2012).

To capture the constituent facets of policy processes, Ball et al. (2012, p. 15) emphasised three levels of analysis: the material, the interpretative, and the discursive. Material contexts can create ‘different practical possibilities and constraints for policy enactment’ (Ball et al., 2012, p. 10). In the present study, the material context is represented by municipal and local documents as manifestations of policy intentions and expectations. One municipal document captured the overall governance context in which the two case schools are situated, and two local school documents embodied the respective expectations and practices at the two schools. In addition, distinct features of the two case schools reflected local contextual factors like demographics, pupil intake, school leadership, organisational routines, and culture and staff composition in terms of age, experience, and gender. Other important material features include financial matters such as budgeting and funding, but these were not explored in the present study.

Regarding the *interpretive* level of analysis, I sought to identify how actors understand, co-produce, rewrite, and/or reject the content of written texts and local prescriptions for TE through interpretation and mediation. Although principals serve as mediating links between teachers and policy expectations at the municipal level, they are not directly represented here. However, their mediating roles are indirectly represented in how teachers make sense of the directive, prescriptive, and discursive elements of TE as communicated in interactions between teachers and their superiors.

Actors’ recognition of socially situated and significant identities emerges through language and interaction but is also shaped by values, objects, tools, and technologies (Gee, 2015). The discursive level of analysis explores how policy agendas—in this case, the evaluation of teachers’ work—are continuously translated into localised versions, shaped by teachers’ and school leaders’ professional discourses and normative belief systems. Policy initiatives often prompt disruptive transformation and may be inflected by competing sets of values and ethics (Ball et al., 2012). The present study explored disruptions or transformations that might be attributed to TE, as expressed in teachers’ reflections on the values that guide their professional conduct. The interplay between ‘possible’ strategies embedded in TE, actors’ attitudes, experiences, knowledge, and preferences and local contextual conditions inform the various roles of policy actors as enthusiasts, narrators, translators, and critics (Ball et al., 2012).

5 Method

Framed by this enactment perspective, the present research was conceived as a descriptive multiple case study (Yin, 2012) of two local schools, with teachers as the units of analysis. The case study approach explores a target phenomenon in its real-world context (Yin, 2012) and is considered appropriate for studies of enactment because human interactions generate such rich data. In the present case, teachers' enactments of TE were explored 'to uncover patterns, determine meanings, construct conclusions, and build theory' (Patton & Appelbaum, 2003, p. 67). The aim was to generalise *within* rather than *from* the cases in order to develop 'thick' and adequate descriptions, interpretations, and explanations (Yin, 2012). To that end, the choice of cases, informants, and method of data collection were based on evidence from multiple sources (as outlined in the Introduction and detailed below). The general strategy for data analysis relied on the theoretical propositions of enactment theory (and important concepts from previous research as outlined) to build explanations based on cross-case synthesis (Yin, 2012).

The case schools were purposively sampled on the basis of their approximate similarity in terms of size and teacher density. Grades or marks are not applied in Norwegian primary education, but are introduced as the pupils enter lower secondary level. The two lower secondary schools were intentionally selected as they offered the possibility for investigating how and to what extent pupils' grades were incorporated into TE practices. In addition, to ensure that they shared the same official ambitions and expectations, the schools were sampled from the same municipality. To generate richer data and to support the exploration of contextual factors' influence on TE enactment, I chose two schools that served distinct populations in terms of expected results and socio-economic status (SES). The schools' principals were contacted by email, and they helped to recruit participants by informing teachers about the study and encouraging them to participate. To facilitate exploration of similarities and differences in perception and personal characteristics, I sought to recruit teachers who differed in age, gender, and professional experience.

In spring 2018, I conducted on-site observations and seven semi-structured interviews, each lasting approximately 1 h (Appendixes 16.1. and 16.2.). The interviews were recorded and transcribed verbatim and were imported to NVivo (QSR) for thematic coding and analysis, along with relevant municipal and local documents and observation notes. Coding and categorization were guided by themes deduced from the analytical framework and reviewed literature in a protocol (Appendix 16.3.) that specified the focal topics for the observations and informed the content and structure of the interview guide (Appendix 16.4.). Interview questions and observations especially sought to capture material elements (practices, standards or criteria, use of results, and knowledge resources applied), interpretative elements (the dual purposes for TE described in the introduction), and discursive elements (such as for example notions of teacher professionalism). The interviews and observations pertained to how actors experienced and reflected on the possibilities and constraints of TE, and emergent themes and new perspectives were identified through qualitative content analysis of the data (Mayring, 2015). The participating teachers' experiences were coded accordingly and compared across the two case schools.

During investigations at both sites, I remained alert to any findings that *contrasted* with the relevant existing literature. Additionally, I attended specifically to *diversity* and the *variations* within and between the school contexts, informant responses, and observation situations, which enabled me to explore and broaden my understanding of the target phenomena. The case schools and study informants were assigned pseudonyms to preserve their anonymity.

6 Case descriptions

6.1 The municipality

Situated within the same municipality in a large city, the two schools aspired to set ‘ambitious’ goals for their pupils, and municipal strategy documents and local quality assurance routines specified relatively well-developed structures for quality assessment. Formulated as general expectations, the relevant policy assigned responsibility to principals for ensuring teachers’ participation in regular COs and performance appraisals by superiors as part of PD. A detailed observation form described ‘factors that are crucial for good quality in teaching’, emphasising results as ‘effects’. As part of an enhanced teacher education programme, a framework was established to appraise the performance of newly qualified teachers, including at least six observations during their first year. An agreement between first-year teachers and their assigned mentors—usually experienced teachers—specified joint responsibilities and guidance routines, along with expectations regarding roles and obligations. All teachers were expected to display ‘engagement, desire to set goals for professional development and openness to feedback’. Supervisors were expected to set aside time and to be available for support and advice, as well as preparing plans and cycles for COs and mentoring throughout the school year. The overall aim of this framework was to ensure that newly qualified teachers gained a systemic understanding of the school and an increased sense of security, self-insight, and self-confidence in their new role. Schools were also assigned a regional director, who was responsible for organising result meetings and ensuring that national and municipal expectations were met.

6.2 School 1: ‘Hill’

Located on the west side of the city, Hill employs about fifty people, and its 550–600 pupils come mainly from higher SES backgrounds. The school’s management includes a principal, an inspector with administrative functions and three mid-level team leaders for grades 8, 9, and 10. At the time of data collection, the former principal had recently resigned, and the new principal had only been at the school for a few months. During this period of transition, a team leader was assigned responsibility for several school functions.

6.3 School 2: 'Valley'

Located on the east side of the city, Valley employs about forty people and has 400 pupils. A relatively large proportion of these students come from low-income families and minority backgrounds. The school is organised in the same way as Hill, with a principal, administrative employees (assistant principal and inspector) and three mid-level team leaders in grades 8, 9, and 10. In addition, the school has established a team to address psychosocial issues, and a coordinator has been appointed to organise COs and teacher mentoring (Table 1).

All of the participants had completed formal teacher education.

7 Findings

The findings presented here focus in particular on the *material* and *interpretative* elements of policy (as outlined at p. 7). Material elements include descriptions in municipal and school documents referring to intended and expected TE practices. Other material elements referred here include in situ observations of teachers' interactions with peers and superiors as part of the schools' organisational routines for TE. The interpretative area is addressed through the interview data on teachers' enactment of TE and in situ observations capturing contextual influences on the interpretation of local practices within the two school cultures. These findings ground the subsequent discussion of TE as a *discursive strategy*.

7.1 Material elements: school policies and practices

Based on municipal guidelines and internal processes involving teachers, 'Valley' had developed an all-in-one booklet containing rigorous descriptions of everyday school routines, including rules, laws, and regulations, organisational structure, and meeting arrangements at all levels. The document assigned 2 h per week to plenary meetings for 'teachers' learning and development', in which researchers, external experts introduced by the municipality, and individual teachers would share their knowledge and teaching experiences. The document also detailed expectations regarding the teacher's role. A section entitled 'The good lesson' described the structure and content of a good teaching session, with examples specifying the number of minutes per section.

TE procedures at 'Valley' mainly involved regular COs (announced or unannounced) by the supervision coordinator or team leaders and occasionally by the principal. These observations were recorded for subsequent joint analysis and reflection, often using iPads or GoPro cameras. Observations were typically 30 min long, with a further 30 min of post-observation discussion. The teacher and supervisor were expected to discuss and agree on the scope of the observation beforehand.

Table 1 Overview of participating schools and informants

| School | Employees (approximate numbers) | Pupils (approximate numbers) | Location/demographics | Documents | Teachers | Age | Years of experience as a teacher |
|--------|---------------------------------|------------------------------|---|---------------------------------|----------------|-----|---|
| Hill | 50 | 550–600 | Relatively high socio-economic status (SES) | Municipal | Ada (female) | 67 | 43 years as a teacher |
| Valley | 400 | 400 | Relatively low socio-economic status (SES) | Municipal and locally developed | Hilde (female) | 25 | First year as a teacher |
| | | | | | Berit (female) | 28 | Second year as a teacher |
| | | | | | Emilio (male) | 28 | 3 years, plus 10 years of leisure club experience and as a substitute teacher |
| | | | | | Nils (male) | 38 | 19 years as a teacher |
| | | | | | Nora (female) | 46 | 8 years, plus 1 year as a teacher specialist |
| | | | | | Omera (female) | 27 | 1 year as a teacher, plus 2 years as teacher assistant |

Team leaders also conducted an annual review,² which would sometimes address pupil performance. As part of the school's efforts to build individual pupils' character, strengths, and motivation to participate in the school community, Valley placed increasing emphasis on assessment for learning and was moving towards a grade-free approach to student assessment.

Valley's local document communicated expectations of an organisational focus on building character and a positive identity through explicit routines and standards for strengthening teacher-pupil relationships. To that end, two class teachers were assigned to each group of pupils to ensure adequate capacity. Teachers were expected to participate in pupil activities during leisure time and to send positive messages to pupils and their parents by SMS. Teachers and the leadership group were also expected to greet pupils with a handshake at the main entrances each morning and teachers again when their pupils entered the classrooms.

In contrast, Hill had no such document specifying local guidelines. COs conducted by senior personnel (principals or team leaders) had been in place for 4 or 5 years, and team leaders were responsible for the performance appraisals of all teachers in their class teams. TE routines were guided by the central municipal documents referred to above and by the provisions of the mentoring programme for first-year teachers. In the 2 years prior to this study, teachers had been encouraged to initiate and conduct COs, performance appraisals, or peer counselling to ensure a common focus ahead of classroom observations.

COs (planned or unexpected) for experienced teachers were conducted at least once a year by team leaders and occasionally by the principal. In the case of novice teachers, COs were conducted at least twice a year, although municipal documents prescribed six observations annually. According to the participating newly qualified teachers, the reduction from six observations, COs, per year to two reflected an ongoing lack of time and resources. Before each CO, the teacher and supervisor would have a short informal conversation to discuss the main focus and scope of the observation and agree a perspective on growth. An observation commonly lasted for 30 min, with approximately 30 min of conversation afterwards. One of the newly qualified teachers also referred to a specific scheme prescribed by her leader as part of a mentoring programme for first-year teachers, where school management brought subject teachers together to exchange their experiences following peer counselling sessions.

Over the previous decade, Hill had focused strongly on assessment for learning. After national testing was introduced in 2004, inter-level result meetings (e.g., municipality and principals; principals and teachers) became an important means of addressing progression. As Hill served a relatively high SES population, the municipality and regional directors expressed high expectations regarding the school's performance and results. As well as COs, team leaders conducted

² According to the The Norwegian Agency for Public and Financial Management (DFØ), the purpose of the annual review or *medarbeidersamtalen* is to consider the employee's own development requests in light of the company's current and future competence plans and tasks. The interview is additional to regular feedback on work performance.

employee interviews once a year, and result meetings followed the scheduled publication of results, especially in the case of national test results published by the municipality. As test results were increasingly regarded as indicators of teaching quality, employee interviews also began to focus on how individual teachers contributed to student results. According to one senior teacher (Ada), a more successful teacher was occasionally assigned to a low-performing class to ensure better national test results. However, the greater emphasis on performance outcomes also increased students' mental health issues related to stress and high achievement pressure, and this had become a matter of general concern over the preceding decade.

7.2 Interpretative elements 1: Hill teachers' enactments of TE

The analysis identified similarities and differences in the interpretation of TE policies at the two case schools as experienced by teachers. One of the main differences relates to the focus on results. At Hill, all three of the interviewed teachers were critical of the tendency to link TE to pupils' results, where individual teachers were held accountable for national and local standardised test results. Hilde considered it inappropriate to use grades as a basis for assessing teachers, as this criterion does not necessarily reflect the teacher's practice. She also expressed concern that this shift in focus might downplay the influence of contextual factors on pupils' grades.

Berit, a first-year teacher, also felt that tests and grades were an unsatisfactory measure of teacher performance: 'It's scary to look at pupils' results and say something about the teacher being good or bad. (...) I would not thrive in a context where results determined a [teacher's] pay rise'. She perceived tensions between the policy's intent and the reality as experienced by teachers: 'I think it is very easy to assume that you will get good grades if you work systematically. But it is a dangerous way of thinking to say that "in this society, the goal for everyone is to perform equally". That will never happen.'

Berit's words indicate a perceived gap between policy expectations that exceed classroom realities and the teacher's ability to influence her own working conditions. She also expressed a concern that the strong focus on results might make it more difficult for teachers to counteract or prevent mental health problems among their pupils.

So I think it's important to be able to distinguish between student performance and teaching practices. Otherwise, there will be an even bigger drive for results, making students excessively tired and leaving a lot of teachers feeling that they have to prove they are good enough. I think this is bad practice for any profession. In the case of doctors, for example, it would be wonderful if every surgical patient recovered, but that will never happen. I believe you cannot measure success in terms of the optimal [outcome].

Rather than using grades and test scores to evaluate teachers, Berit noted that performance appraisals by superiors help to make the teacher's responsibility

to improve more visible and transparent. She suggested that if the leadership team exhibited sensitivity and competence, TE might be appropriate in some cases to encourage teachers who were not suited to the profession to seek other career paths. Elaborating on contextual influences, Berit also reflected on the tensions caused by external accountability and expectations, especially among newly qualified teachers.

I think some teachers whose classes are troubled may feel that it is, to some extent, their fault ... even though I know it's not true. Poorly composed student groups are very demanding because there is so much unrest, and it is easy to feel 'I have to fix this'. Actually, the management should fix it, along with parents and teachers ... The class teacher is the one who really should fix the least (...) there is a lot of help available, but it's probably a bit scary to say 'I cannot manage my class, and it may not even be my fault'. Because you want to perform, and as a new employee you... at least you have a lot... You do not have a permanent job, right?

Berit's words reflect a strong professional ethos and an implicit expectation; the autonomous professional teacher is expected to solve all of the class's behavioural problems—to 'fix this', regardless of contextual factors and class composition. At the same time, she depicts the insecurity of the first-year teacher and how concerns about tenure and being held to account for 'performing well' can limit one's motivation to seek advice from peers or superiors on encountering difficulties.

Reflecting on the mentor's role, Hilde welcomed critical and challenging questions. However, she said, '...it's about *how* you ask those questions.' She linked the mentor's competence to how they relate to and understand classroom complexity. As a first-year teacher, she emphasised the need for concrete and direct advice, suggestions or support.

Reflecting on the various TE tools, Ada, an experienced senior teacher, was quite clear about the role of students in evaluating teaching.

I think it's a life-threatening way to go! (...) Because I think – at least if it's a question of wage differences – that's the extreme version. I don't think that's a good idea. I think teachers will become – or, at least, it may lead to teachers becoming – very concerned about 'pleasing' their students and making them happy. And of course they should be happy, but what they should be happy about is that you are delivering quality teaching – not that you are trying to gain a good reputation. So, I don't know... On the other hand, the students – in a way – are our customers, so... it's not unreasonable that they should have... an opportunity to say something. And to assess...

Ada's words raise at least three important issues regarding her understanding and conceptual interpretation of TE. The first of these is the severity of her opinions about certain approaches to evaluation and how they can be perceived as a threat—in this case, formulated as an existential threat. The second issue

relates to how student evaluations of teachers might affect salaries, triggering strategic behaviours and practices for ‘pleasing their students and making them happy’. Ada’s concern is that such practices might reorient teaching towards one’s ‘good reputation’ and ‘keeping the students happy’ rather than ensuring the quality of instruction.

A third issue concerns her framing of students as *customers*. Given the relatively low marketisation of Norwegian education, this perspective is especially interesting, as it suggests that the rhetoric of schools as providers within a market has become internalised, discursively constructing the teacher as a ‘provider’ of satisfied customers who have a say in the teacher’s instructional practice. Ada went on to say that she had been using a questionnaire ahead of interviews with pupils to ask ‘what they think I can do for them and what I could do better’. Berit also described how she self-evaluated in light of pupils’ responses: ‘They [the pupils] are my “little cases” – they just do not know it themselves’. These insights reveal the teachers’ willingness to take account of pupils’ perspectives in their efforts to improve teaching practices. However, as Ada emphasised, pupils’ evaluations of teaching must be open, transparent, and dialogue-based if it is to be useful.

Ada seemed to enjoy TE as collaborative PD when aligned with teachers’ professional culture of collaboration and knowledge sharing. ‘And I think meetings of maybe ten or twelve teachers can provide a lot of useful ... ideas [and] colleagues’ reflections ... about what works and what doesn’t.’ Ada’s view implies a TE approach that allows teachers a relatively high degree of freedom and autonomy, with less control by their superiors. Overall, she saw great value in collaborative and peer-oriented TE when time and resources permitted.

There must be no hurry. It must not be a case of peeking in, where someone stops by and looks at you and takes a little note, and then you sit together for half an hour and drink coffee and ‘get this over with’. That’s too rushed, but it’s typical of a teacher’s everyday life ...

7.3 Interpretative elements 2: Valley teachers’ enactments of TE

In general, the participating teachers at Valley seemed to interpret TE as a tool for PD rather than as a means of holding them accountable for results. In contrast to Hill, the teachers at Valley did not attend result meetings. However, national tests were discussed at school level and occasionally at teacher level or in annual reviews. Emilio described how the topic of national test results prompted an emotional response related to professional self-esteem when confronted with results that fell short of expectations.

You do not want to see that they [pupils] have declined because it suggests you have done something... you feel you have done something wrong, even if the management does not point it out. But it’s a matter of your self-esteem – your own view of how it’s going.

Emilio's perception of pressure highlights the impacts at both organisational and individual levels. Teams and teachers with poorer results become a topic of staff discussion, which increases the pressure on individual teachers to improve their performance. Nils, a senior teacher, reported positive experiences of COs and felt that they motivated him to try harder – 'a bit like a sports performance (...) So it's a bit like – how to put it – (...) being able to look yourself in the eye and be proud of what you have done.' Like Emilio, Nils' take on performance assessment illustrates how TE functions at an emotional level; being evaluated is linked to and 'works on' his professional self-esteem.

During his 3 years as a qualified teacher, Emilio said he had had no negative experiences of TE. On the contrary, he felt that his teaching practice had been improved by collaborative activities when sufficient time was allocated. Feedback from observers and post-observation discussions, especially with peers, was fruitful, as open questions enhanced reflection and required him to justify his decisions.

I wish I had more time to discuss and explore these issues ... That would have been very exciting. When we allocate plenty of time for team activities – for example, when I showed clips from my own teaching, with ... almost an hour to discuss them – that has proved very useful because you have a lot of time I like to go through one topic after another, and it has been a lot of fun, really.

Emilio's positive experiences of TE relate to his perception of how school management sought to build both individual and organisational trust. He felt that because teachers were allowed to discuss and disagree, a culture of openness and trust 'creates a sense of security.' He perceived this as an emphasis on growth rather than deficits, where TE allowed for creativity and experimentation without any anxiety about making mistakes. He concluded that 'If they imposed more controls, I think it would be a negative factor for the teaching community itself.'

Nora, a senior teacher at Valley, shared Emilio's views about the school and the leadership team, and she enjoyed the knowledge-sharing culture and the focus on professional development. She spoke of high expectations—her own and those of the school's management and the municipal authorities—and said she had an 'intense desire for her pupils to succeed'. She felt that expectations had shifted in recent years from the issue of higher grades to a broader responsibility for pupils' success beyond the classroom.

More is expected of us as teachers now, requiring us to adopt a more inclusive approach that offers students different types of assessment and feedback.

However, rather than being motivated by payment for results, Nora was quite clear that pupils' success was a matter of professional responsibility. As a senior teacher, she echoed Ball et al.'s description of senior colleagues as 'shields' and policy translators for younger and less experienced teachers (2012, p. 63). She expressed concerns about demanding expectations and workloads, especially for newly qualified teachers: 'There is a limit to how much responsibility you can place ... on them

because I find that many people quit very quickly. There is a very large turnover—maybe ten or twelve who quit each year ... maybe more.’ She also described how COs that did not go well were sometimes a topic of further discussion.

...it may be that people experience being observed as very negative – almost a bit humiliating ... because of the form of observation, which I know others have experienced ... where the leader, for example, brings a cell phone into the classroom where he is due to observe and then sits there, half scrolling and half observing, and the subsequent conversation is badly planned. So, I know that many – or at least some – have experienced this as demotivating and almost as a rebuke.

Nora displayed an awareness of the possible unintended consequences of COs for individual teachers and underscored the importance of competent and sensitive observers in building confidence and trust. ‘Because being an observer is not something you should take lightly. I think is extremely—to be observed can be painful ... maybe not painful, but it’s a bit like—you are vulnerable in such situations, and the observer must be aware of this.’

For Omera, a first-year teacher, her first performance appraisal following a CO (which included a video recording) focused on the issue of mastery and how pupils responded to her various teaching practices. The assessment coordinator asked open questions like ‘What do you think yourself?’ and ‘How can you hold their attention in this situation?’ The exchanges between Omera and the appraiser seemed relaxed and informal in tone. During the interview, Omera reflected on the value she found in seeing the class and her approach to instruction from a different perspective, which made her more aware of ‘the little things’ she often tended to forget.

Omera said that her COs to date were always unannounced, which meant there was no conversation beforehand. She described her first CO as follows.

I was not told about it [the unannounced CO], and I don’t think my colleagues are told about them either, unless of course they request it. And then it’s like, when the bell rings, I go to the classroom, and the observer is standing outside and then joins the students. So, if I’m a little unprepared for that particular class, I’m pretty much panicking at that point. (...) My heart rate gets a little higher, and I feel I have to perform a bit better when someone comes from management. (...) And then afterwards, they just leave, and they don’t talk to me at all from the time they come in until they leave...And then it’s up to me to seek out that person afterwards to ask for their feedback.

Omera’s story highlights a discrepancy between the expectations communicated in the municipal document and the actual practice of TE. Her account of the experience points to the absence of formative PD inputs, including a lack of agreement about the CO’s purpose and no planned or systematic post-observation performance appraisal. She seemed to find the situation somewhat humiliating, not least because she had to seek out the observer to ask for feedback.

Omera also spoke of a situation during a CO regarding ‘the reading quarter’ (*lesekvarten*). She said that she and her pupils found this reading time important

and valuable: ‘These pupils never read at home—this is their only chance to actually read!’ However, as the assessor did not see the pupils ‘producing’ anything, he asked her to change her ways, and she tried to encourage the pupils to write as they read.

But I found that this might not work so well, and I’m a little nervous about being observed in the reading quarter again because right now they just sit and read (laughter). I’ve sort of... reverted [to old practices]. So it’s a bit like ‘Okay, if someone comes from the management during the reading quarter, I will quickly go up to the board and write two questions!’

Nora also described her experiences of municipal efforts to improve teaching and learning through detailed direction.

I tend to get really annoyed about directives from the district leader that say things like ‘Now you should do everything like this or like that in your classroom.’ Because I think he [the municipal director] knows nothing about what I do... and... I do what I do because it works. Yes, I can get very annoyed (...), but it’s been a long time since that happened. I remember the last time, when we received a whole cardboard box with these laminated foils to hang on the walls all around the classroom so that the pupils could learn some formulas and rules ... I was really annoyed about that ... it was just disruptive.

When asked what she did next, Nora replied, laughing, ‘I pretended that nothing had happened!’ Omera’s and Nora’s responses illustrate how attempts to control classroom practices without considering the local context and knowledge can provoke rejection by teachers of all ages and experience as interference with professional autonomy and discretion. However, Omera’s concern about her next CO and subsequent response suggest compliance with expected practices, regardless of her own professional judgment, perhaps because she was a first-year teacher.

8 Discussion

The informants’ interpretations confirm the need for a hermeneutic approach when investigating policy enactment, which entails both material and discursive elements. This section summarises the similarities and differences in material and interpretative elements before discussing how TE is continuously shaped and negotiated by teachers’ (and school leaders’) professional and normative belief systems within local and national discourses.

8.1 Material elements

Although the case schools were located in the same municipality and therefore subject to the same overall expectations and performance management structures, contextual features seem to have contributed to differences in local policy enactment.

First, the analysis revealed material differences related to demographics and pupil intake. At Hill, a majority of pupils came from an ‘ethnic Norwegian’ background (i.e. born in Norway of Norwegian parents). As these were generally high SES families, high expectations at municipal level tended to increase pupils’ stress levels. On the other side of the city, Valley served pupils with parents from immigrant and relatively low SES backgrounds. This meant that explicit expectations regarding results were lower, but there were higher expectations in relation to social inclusion and motivating pupils to complete their education.

In terms of teacher composition, Hill’s staff came mainly from ethnic Norwegian backgrounds and included both experienced and younger and newly qualified teachers. At Valley, a majority of teachers were younger and came from more mixed ethnic backgrounds. Many held masters’ degrees, and some exhibited high local competence and social skills in extracurricular and leisure activities. The two schools shared the same organisational structure (one principal, one assistant principal and three team leaders for grades 8–10). Hill had plenary teacher meetings, focusing occasionally on PD. In addition, there were regular result meetings at group and individual levels. At Valley, individual-level result meetings were rare. Instead, Valley reinforced and integrated TE into continuous PD arrangements that included weekly plenary meetings targeted at teacher development, regular COs conducted by the principal, team leader, or assessment coordinator.

At Hill, municipal documents were adopted more or less directly, which meant that expectations regarding TE and teaching practices were less detailed. The scope for professional autonomy and a somewhat looser approach to teacher follow-up may owe in part to a relatively well-resourced student population who can accommodate the traditional norms of schooling. On the other hand, Valley’s comprehensive local document communicated detailed expectations regarding ‘the good lesson’, as well as a school ethos that extends the teacher’s role beyond the classroom.

These features contribute to some interesting differences between the schools in terms of the fourth material element of culture. Teachers and leaders at Hill appeared ambitious but were also aware of the unintended consequences of high pressure, pupils’ stress, and anxiety. At Hill, TE followed municipal expectations, but was not seen as a key tool for continuous PD. The management seemed more distant, and TE was less integrated into knowledge sharing or overarching perspectives and targets in the same way as at Valley. For that reason, Hill’s approach to TE did not activate a collective or individual commitment to PD as a tool for improvement or the ‘common good’. In contrast, the culture at Valley was ambitious, innovative, and progressive, focusing more on collective than hierarchical processes. Teachers at Valley referred to leaders’ explicitly ‘open door’ approach, with a low bar for discussing issues and scope for experimentation or trial and error. TE was integrated into continuous PD and linked to the school’s ‘larger project’: to pursue equity, equality, and high ambitions despite poorer conditions for learning prevalent among many of the pupils. The principal’s and team leaders’ practices of meeting and greeting pupils at the main entrance each morning reflected a commitment to being ‘on the same page’ as teachers in the shared effort to build a positive school ethos.

At both schools, TE practices took the form of individual COs followed by formative performance appraisals and/or post conferences involving peers or superiors.

However, the differences between the two student populations seemed to influence the relationship between policy and practice. At Valley, there was closer teacher follow-up, and the school's expanded practices reflected its cultural features. Additionally, Valley was experimenting with the use of tablets and GoPro cameras for COs, providing common ground for subsequent discussion between appraiser and appraised in post-CO conferences and in the weekly knowledge-sharing meetings.

In summary, the above material elements help to account for teachers' interpretations of TE in light of the dual purposes referred to in the Introduction: TE for PD and TE as accountability, as elaborated below.

8.2 Interpretative elements

Teachers at both schools seemed to perceive TE as a useful tool for both PD and knowledge sharing, but this was seen to depend on competent appraisers and a sensitive approach. With timely and adequate follow-up, CO-based TE was seen to contribute to reflection and was considered helpful for improving teaching practices, especially among the younger teachers. As the main tool in current use, CO can be linked to earlier well-known Norwegian practices for teacher education and continuous professional development (see for example Dahl et al., 2016). As a tool for peer collaboration and joint PD, teachers clearly viewed TE as constructive, useful, valuable, and even fun. At both schools, principals and leaders seemed to be aware of the need to build and maintain relational trust. Post-CO performance appraisals were characterised by open questions, allowing space and time for mutual reflection on practices (perhaps more so at Valley).

Regarding the accountability function of TE, the associated tools were differently applied, but national tests, surveys, and exam results were mentioned by teachers at both schools. Teachers at Hill were especially critical of national test results meetings, which they interpreted as an indirect form of TE. While they felt that individual result meetings and employee appraisals were used to hold teachers to account, teachers at Valley felt that discussions of national test results appealed to their internal sense of accountability, promoting self-reflection, self-direction, and self-control. These efforts to self-regulate found further support in the local booklet's explicit and detailed 'standards for good teaching practices'.

8.3 Discursive elements

Ball (1990) warned that concepts such as 'empowerment' and 'teacher autonomy' could prompt more oppressive systems of control, with more sophisticated and less direct methods for keeping people 'in line'. More recent research on high- versus lowstakes accountability regimes refers to the transformative capacity of performance-based policies as 'technologies of control' far beyond the more visible provisions of high-stakes regimes (Hangartner, 2019). According to Hangartner (2019, p. 5), technologies like data-driven evaluation, feedback, leadership, and self-reflection use self-direction as a concealed form of regulation by harnessing the teachers' sense

of autonomy as self-control. By capturing ‘softer’ and more formative approaches to TE, the present study seems to corroborate that view. Norway is generally characterised as a low-stakes accountability context, and neither of the two case schools linked TE to explicit sanctions or bonuses. The findings illuminate how local interpretations of policy artefacts generate discourse around ‘the good teacher’ and how they are to be evaluated and to self-evaluate. Teachers’ emotional responses when confronted with test results seem to bear on their self-esteem, illustrating how results indirectly reinforce TE by engaging and motivating self-directed efforts to improve teaching in line with expectations or standards.

At Valley, a ‘moral ethos’ seemed to contribute to a sense of ‘being in this together’, with a collective commitment to TE as a means of enhancing both individual and collective practices ‘for the common good’. In this sense, the teachers experienced a moral obligation to engage with the wider purpose of schooling, sometimes exceeding the mandate and possibilities at hand. Nora referred to a relatively large turnover attributed to some of the tensions that arise when schools and teachers are expected to ‘compensate for society’ (Bernstein, 1970).

Ball et al. (2012) described junior teachers as ‘receivers’ of policy who exhibit high levels of compliance and ‘policy dependency’ (p. 63). Newly qualified teachers look for guidance and direction; rather than attempting to be creative, they focus on managing the classroom rather than attending to ‘the bigger picture’ (Ball et al., 2012, p. 63). In contrast, Berit was very aware of the ‘bigger picture’, or at least of how contextual factors, policy expectations and management’s facilitative role might influence teachers’ working conditions. The tensions she describes capture the intersection of two discursive notions of ‘the professional teacher’: the teacher as subject, internally motivated to stretch her professional responsibility, doing her best for the good of the pupils, and the teacher’s professional value in terms of her ability to meet expectations and deliver performance and results, regardless of contextual constraints over which she has no control. Hilde’s descriptions of how challenging class composition prompted teachers to work harder to ‘fix it’ rather than seeking advice or help underscore how discourses bear on trust and self-perception. Any breach of trust between leaders or peers may discourage newly qualified teachers from seeking developmental advice and guidance, with unintended consequences for professional and organisational development.

The concerns raised about the fairness of TE, especially when linked to student performance data, illustrate several discrepancies between policy expectations and teachers’ reality in the classroom. First, teachers at both schools noted how contextual factors can be downplayed as more emphasis is placed on individual teachers’ ability to improve by standards. Second, the focus on individual teachers’ performance and competition was seen as a conflict between professional values and policy expectations. According to Ada, linking TE to student performance data or anonymous student evaluations is ‘a life-threatening way to go’.

Ada also raised concerns about how the logics of market and accountability contribute to a change in expectations of the teacher as ‘provider’ and, in particular, how marketization can change teachers’ relationships with peers, superiors, pupils,

and their parents. Skewing professional responsibility towards accommodating the judgements of peers, superiors, or pupils could prompt strategic behaviours to protect one's own reputation. A study on Norwegian school actors' responses to the implementation of national testing in 2004 documented strategic reorientation of behaviours in line with new expectations (Camphuijsen et al., 2020). It seems important, then, to understand teachers' concerns about TE and the associated evaluation tools in terms of reputational issues and how discourses around quality and professionalism can influence both public perceptions and teachers' self-perceptions.

Despite these concerns, and consistent with findings from other contexts (see for example Flores, 2018), an increased focus on individual teachers' contributions to student outcomes appeared to encourage teachers to question and reflect on traditional norms and values. Teachers at both schools exhibited a strong internal sense of the professional teacher as attentive to contextual factors, pupils' individual needs, and social-democratic norms and values like professional collaboration, sharing, caring, compassion, and encouragement. Performance-based accountability was seen to represent a normative breach of the collectivist non-hierarchical traditions of teacher collaboration and the social-democratic values associated with being a teacher. This aligns with Hangartner (2019) account of widespread silent resilience to performative accountability governance in low-stakes contexts, as teachers reverted to previous practices in the face of attempts to control or interfere with teacher autonomy. This resistance is well illustrated by Nora's story about the district director's laminated foils and Omera's reversion to existing practices while struggling to comply.

According to Hangartner (2019), resistance to high-stakes, performative management may engender emancipation rather than a self-disciplining technology of control by prompting teachers to engage in collegial reflection and so reclaim their autonomy. In Norway, teachers' enactment of TE indicates a similar resilience to control, accountability, comparison, competition, and hierarchical structures involving asymmetrical power relations. Despite efforts to audit and control through TE, informants seemed to remain focus on maintaining a context of mutual trust by focusing on collaboration and knowledge sharing within an open, symmetrical culture that encourages professional discussions.

9 Concluding remarks

Policy enactment studies highlight the holistic interplay between policy intentions, prescribed tools for pursuing policy goals, the underlying logics of these tools, and their interaction with living human beings in local contexts—the micro-policy makers. In attempting to capture this interplay, the present study investigated how TE as a policy tool for enhancing teaching practices was enacted at two urban lower secondary schools in a large Norwegian municipality.

Concerning what might be at stake in a low-stakes context, senior and less experienced teachers expressed critical concerns about the conflicting uses of TE as a performance measure *and* a driver of professional development. In particular, teachers' reflections illuminate how TE seems to 'work on' both individual teachers,

organisational structures and cultures, functioning as a discursive policy tool. TE in combination with use of test results affected teachers' self-esteem and indirectly reinforced self-directed efforts to improve teaching in line with expectations or standards. TE practices informed by the discourse of marketisation were seen to reinforce a potential for strategic behaviour, as well as to undermine organisational trust and teachers' professional autonomy by allowing them less discretion to make professional judgements. On the other hand, TE was considered a source of professional support and guidance, and even perceived as *fun* when sufficient time was allocated and competent and attentive superiors or peers exhibited the necessary sensitivity to contextual factors inside and outside the classroom. TE was also seen to introduce new and fruitful arenas for collective professional development and discussion, both at local level and in the overarching national discourse of 'teacher professionalism'. Experienced and novice teachers made sense of, and engaged with TE, but also linked their experiences to the larger context of policy expectations in which their work is embedded. The findings thus exemplify 'the becoming' of policy tools informed by performance and accountability logics when shaped by contexts that may prompt resistance to such policy provisions in favour of a commitment to professional development.

Although limited by design in terms of rigour and generalisability, the study's main contribution is a new and deepened understanding of the connections between structural constraints inherent in TE and subsequent possibilities for teacher agency. In this case, exemplified by illuminating the complexity and nuanced lived experiences of Norwegian lower secondary teachers, and their convergent and divergent perceptions of the appropriateness of TE. The findings and discussion provide important insights necessary not only to better understand TE and its discursive impact on Norwegian teachers but also the importance of including teachers' perceptions to inform development of current and future policies in Norway and beyond. Further explorations are necessary to understand municipal actors' and school leaders' mediating roles when enacting TE policies and also what makes up the contextual features grounding the Norwegian resilience towards TE.

Appendix

Appendix 1 Interview data

| Nr. | Date | School | Informant | Duration (in minutes) |
|-----|------------|--------|-----------|-----------------------|
| 1 | 07.03.2018 | Hill | Ada | 01:00 |
| 2 | 02.03.2018 | Hill | Berit | 01:07 |
| 3 | 02.03.2018 | Hill | Hilde | 00:53 |
| 4 | 20.03.2018 | Valley | Emilio | 00:43 |
| 5 | 20.03.2018 | Valley | Nils | 00:29 |
| 6 | 21.03.2018 | Valley | Nora | 00:47 |
| 7 | 20.03.2018 | Valley | Omera | 00:45 |

Appendix 2. On-site observation data

| Nr. | Date | School | Occasion | Topic/scope, field notes | Duration |
|-----|------------|--------|---|--|-----------------|
| 1 | 30.10.2017 | Valley | Conversation with principal | Background information about the school: material aspects. Leadership perspectives on TE policy enactment: practices, conceptualisations, challenges, and possibilities | 00:41 |
| 2 | 08.11.2017 | Valley | Plenary teacher meeting | Enactment: culture, communication between levels | 00:37 |
| 3 | 25.01.2018 | Valley | Plenary teacher meeting: 'teachers' learning' | Nora's presentation on, and discussions of what motivates pupils' learning. | 00:42 |
| 4 | 28.02.2018 | Hill | Conversation with assistant principal | Background information about the school: material aspects. Leadership perspectives on TE policy enactment: practices, conceptualisations, challenges, and possibilities | 00:22 |
| 5 | 28.02.2018 | Hill | Classroom observation, Berit | Enactment: teaching practices, interplay between observer and the observed. | 00:30 |
| 6 | 01.03.2018 | Hill | Classroom observation and performance appraisal, Hilde | Enactment: teaching practices, interplay between observer and the observed. | 00:20 and 00:50 |
| 7 | 07.03.2018 | Hill | Observation of performance appraisal, Berit | Enactment: teaching practices, interplay between observer and the observed. | 00:30 |
| 8 | 07.03.2018 | Hill | Conversation with assistant principal | Leadership perspectives on TE policy enactment: practices, conceptualisations, challenges, and possibilities | 01:16 |
| 9 | 08.03.2018 | Valley | Conversation with principal | Leadership perspectives on TE policy enactment: practices, conceptualisations, challenges, and possibilities | 01:06 |
| 10 | 14.03.2018 | Valley | Conversation with assessment coordinator | TE policy enactment: practices, conceptualisations, challenges, and possibilities | 01:03 |
| 11 | 16.04.2018 | Valley | Classroom observation and observation of performance appraisal, Omera | Teacher's enactment: teaching practices, interplay between observer and the observed. | 00:15 01:23 |

Appendix 3. Example of protocol for thematic content analysis and coding in Nvivo

| Teachers | | | |
|----------------------------------|-------|------------|----|
| Name | Files | References | |
| Features of teacher knowledge | | 7 | 14 |
| Features of the school | | 8 | 30 |
| Culture for evaluation | | 1 | 2 |
| Organisational routines | | 7 | 34 |
| Perceived accountability | | 7 | 19 |
| Perceived control | | 5 | 14 |
| Perceived expectations | | 7 | 36 |
| Perceived pressure | | 5 | 23 |
| Data use | | 1 | 3 |
| Standards or Criteria | | 5 | 20 |
| Perceived changes | | 5 | 13 |
| Perceptions of school leadership | | 6 | 25 |
| Assessor qualifications | | 4 | 6 |
| Perceptions of TE practices | | 8 | 63 |
| Practices for TE | | 8 | 60 |
| Professional development | | 7 | 34 |
| Autonomy | | 1 | 5 |
| Perceived room for discussions | | 8 | 32 |
| Trust (2) | | 8 | 34 |
| Motivation for work | | 6 | 9 |
| Resistance | | 7 | 17 |
| Use of research in PD | | 3 | 6 |
| Values and Theories | | 6 | 20 |

Appendix 4. Interview guide

Introduction

The interview lasts about 1 h and will be recorded on an audio file. No one but me will be listening to this, and you will be anonymized when the interview is transcribed. You can choose to withdraw from the project at any time—your participation is 100% voluntary. Consent form.

Main questions and follow-up questions

Teacher background

| <i>Main question</i> | <i>Follow-up questions</i> |
|--|--|
| Can you tell me a bit about why you chose to become a teacher? | What is your educational background? Subject? Continuing education? How long have you been teaching? Have you taught elsewhere before, if so where? What do you teach currently? |
| If you were to give a brief presentation of what characterizes this school, what would it be like? | Why did you choose this school as your workplace? What are the differences and similarities from previous experiences? |

Practices for TE

| <i>Main question</i> | <i>Follow-up questions</i> |
|---|---|
| How do you carry out TE here at your school? | Tell me a bit about how this is carried out; How often is your teaching assessed? Which methods/tools are used? Who is responsible for implementing TE? How is the assessment followed up afterwards? Documentation? What sources of information do you use to support your work? |
| How would you describe the work you do before, during and after TE? | Are there guidelines for how you should work with teaching assessment, and if so: what are these like? Does the assessment have consequences for your teaching? Positive? Negative? |
| In which arenas are TE discussed? | Tell a little more about the various meetings; Joint meetings? Staff meetings? Who leads them? How is the purpose of the meeting communicated? Do you feel that the purpose(s) of the meeting(s) is (are) clear? |
| How do you work to develop your teaching skills? | How/where do you find sources of teaching methods that you believe produce good learning results? How do you feel that the school's management facilitates new ideas related to teaching practice? Examples: various teaching methods - actively using student results to improve teaching? |

Purposes of TE

| <i>Main question</i> | <i>Follow-up questions</i> |
|--|---|
| Can you say a little about what you experience as the results of TE? - What positive and/or negative consequences have you experienced? | How do you feel that different tools for TE are presented and introduced? How do you feel that the TE tools fulfill the purpose/goal, as you understand it/it through the sources you use? How do you experience that the tools are received and used by your colleagues? |
| What do you understand to be the purpose of TE? | What do you think about the tools/means/frameworks to achieve this purpose/goal? In what ways can you agree/disagree with the purpose of the assessment? How do you think that TE can contribute to achieving the school's, and your school's, overall goals? Personal goals? |
| What do you think about TE being used to hold teachers accountable for the quality and results of the school? | Is this something you know by personal experience? Who do you think is responsible for the students' at your school achieving good results? |
| To your knowledge, which sources/resources of information guide TE at your school? (Government guidelines/municipal guidelines/supervisors/research/principal contracts/individual contracts...etc.) | Which sources/documents of information do you apply as a teacher when working with TE? How do you feel that these sources contribute in an assessment situation? |
| As a teacher - how do you understand the term <i>professional development</i> ? | In what ways do you think TE contribute to professional development? Individually, in the teaching staff and/or the teaching profession as a whole? How does it contribute to strengthening the school's work with results follow-up? Quality development? |

...on the use of data/student performance results in TE

| <i>Main question</i> | <i>Follow-up questions</i> |
|---|--|
| Are student results or results from user surveys (students, staff, parents) used in the assessment of your work as a teacher? | What types of results are used/ weigh the most? How do you experience this – positive/negative experiences? |
| What expectations do you perceive to be prominent with regards to your work, right now? | Who provide expectations for teachers? Politicians? The municipality? Leaders? Associates? Parents? Have these changed over time? |
| What expectations do you have of yourself as a teacher? | How have these changed over time? |

...on autonomy, professionalism, organisational culture and trust

| <i>Main question</i> | <i>Follow-up questions</i> |
|---|---|
| What particular theories or values are important to you as a teacher? | How did you learn or become aware of these? |

| <i>Main question</i> | <i>Follow-up questions</i> |
|--|--|
| Do you have any examples of matters decided by the school management that have consequences for your teaching? | How do you regard this? Positive? Negative? Is there something common to the municipality that you must do? How do you see it? Of the team? What happens if you do not comply with what is decided? Do you feel that communication goes both ways - that you have the opportunity to communicate upwards/to the sides in the system? |
| What significance does TE have for your professional/pedagogical development? | In what ways are you motivated by assessment/feedback? |
| How do you feel about being assessed by your superior/colleague/supervisor/external expert? | What do you think about the professional competence of those involved in evaluating you/your teaching? What do you think about your relationship with those involved? Trust/lack of trust? |
| How do you collectively work with TE? | How do you discuss experiences from practices/results? How do you feel that the organization/culture facilitates a collaborative climate characterized by mutual trust? |

Thank you for your participation!

Funding Open access funding provided by University of Oslo (incl Oslo University Hospital)

Data availability The participants of this study did not give written consent for their data to be shared publicly. Supporting data is therefore not available.

Declarations

Conflict of interest The author declares no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability*. Routledge.
- Avidov-Ungar, O. (2018). Teacher evaluation following reform: the Israeli perspective. *Quality Assurance in Education*, 26(4), 511–527. <https://doi.org/10.1108/QAE-01-2018-0009>
- Ball, S. J. (1990). *Politics and policy making in education*. Routledge.
- Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, 18, 215–228.
- Ball, S. J., Maguire, M., & Braun, A. (2012). *How schools do policy: Policy enactments in secondary schools*. Routledge.

- Berliner, D. C. (2018). Between Scylla and Charybdis: reflections on and problems associated with the evaluation of teachers in an era of metrification. *Education Policy Analysis Archives*, 26(54). <https://doi.org/10.14507/epaa.26.3820>
- Bernstein, B. (1970). Education cannot compensate for society. *New Society*, 15(387), 344–347.
- Bradford, C., & Braaten, M. (2018). Teacher evaluation and the demoralization of teachers. *Teaching and Teacher Education*, 75, 49–59. <https://doi.org/10.1016/j.tate.2018.05.017>
- Camphuijsen, M. K., Moller, J., & Skedsmo, G. (2020). Test-based accountability in the Norwegian context: exploring drivers, expectations and strategies. *Journal of Education Policy*, 36(5), 624–642. <https://doi.org/10.1080/02680939.2020.1739337>
- Camphuijsen, M. K. (2020). Coping with performance expectations: towards a deeper understanding of variation in school principals' responses to accountability demands. *Educational Assessment, Evaluation and Accountability*, 1–27.
- Chen, J., & Teo, T. (2020). Chinese school teachers' conceptions of high-stakes and low-stakes assessments: an invariance analysis. *Educational Studies*, 46(4), 458–475.
- Clarke, J., & Newman, J. (2009). Elusive publics: Knowledge, power and public service reform. *Changing Teacher Professionalism* (pp. 63–73). Routledge.
- Conley, S., & Glasman, N. S. (2008). Fear, the school organization, and teacher evaluation. *Educational Policy*, 22(1), 63–85. <https://doi.org/10.1177/0895904807311297>
- Cuevas, R., Ntoumanis, N., Fernandez-Bustos, J. G., & Bartholomew, K. (2018). Does teacher evaluation based on student performance predict motivation, well-being, and ill-being? *Journal of School Psychology*, 68, 154–162. <https://doi.org/10.1016/j.jsp.2018.03.005>
- Dahl, T., Askling, B., Heggen, K., Kulbrandstad, L. I., Lauvdal, T., Mausethagen, S., Qvortrup, L., Salvanes, K. G., Skagen, K., Skrøvset, S., & Thue, F. W. (2016). *On the Role of the Teacher: A Knowledge Base*. Fagbokforlaget.
- Day, C. (2002). School reform and transitions in teacher professionalism and identity. *International Journal of Educational Research*, 37(8), 677–692.
- De Lima, J. Á., & Silva, M. J. T. (2018). Resistance to classroom observation in the context of teacher evaluation: teachers' and department heads' experiences and perspectives. *Educational Assessment, Evaluation and Accountability*, 30(1), 7–26.
- Delvaux, E., Vanhoof, J., Tuytens, M., Vekeman, E., Devos, G., & Van Petegem, P. (2013). How may teacher evaluation have an impact on professional development? A multilevel analysis. *Teaching and teacher education*, 36, 1–11. <https://doi.org/10.1016/j.tate.2013.06.011>
- Donaldson, M. L. (2016). Teacher evaluation reform focus, feedback, and fear. *Educational Leadership*, 73(8), 72–76. <https://doi.org/10.1080/10665684.2016.1144831>
- Dunn, A. H. (2018). Leaving a profession after it's left you: teachers' public resignation letters as resistance amidst neoliberalism. *Teachers College Record*, 120(9), 1–34. <https://doi.org/10.1177/016146811812000906>
- Elstad, E., Lejonberg, E., & Christophersen, K. A. (2017). Student evaluation of high-school teaching: which factors are associated with teachers' perception of the usefulness of being evaluated? *Journal for Educational Research Online*, 9(1), 99–117.
- Elstad, E., Lejonberg, E., & Christophersen, K.-A. (2015). Teaching evaluation as a contested practice: teacher resistance to teaching evaluation schemes in Norway. *Education Inquiry*, 6(4), 375–399. <https://doi.org/10.3402/edui.v6.27850>
- Flores, M. A. (2018). Teacher evaluation in Portugal: persisting challenges and perceived effects. *Teachers and teaching*, 24(3), 223–245.
- Ford, T. G., Van Sickle, M. E., Clark, L. V., Fazio-Brunson, M., & Schween, D. C. (2017). Teacher self-efficacy, professional commitment, and high-stakes teacher evaluation policy in Louisiana. *Educational Policy*, 31(2), 202–248. <https://doi.org/10.1177/0895904815586855>
- Gee, J. P. (2015). Discourse, small d, big D. *The international encyclopedia of language and social interaction*, 3, 1–5.
- Gloppen, S. K., & Novak, J. (2023). Making teachers by policy? The case of teacher evaluation in Norwegian education, Scandinavian. *Journal of Educational Research*. <https://doi.org/10.1080/00313831.2023.2192747>
- Ha, B. W., & Sung, Y. K. (2011). Teacher reactions to the performance-based bonus program: How the expectancy theory works in the South Korean school culture. *Asia Pacific Education Review*, 12(1), 129–141.
- Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement: an analysis of the evidence. *Educational Assessment, Evaluation and Accountability*, 26(1), 5–28.

- Hangartner, J. (2019). Control of teachers under conditions of low-stakes accountability. *On Education. Journal for Research and Debate*, 2(5). https://doi.org/10.17899/on_ed.2019.5.4
- Hazi, H. M. (2022). Reconsidering the dual purposes of teacher evaluation. *Teachers and Teaching*, 28(7), 811–825. <https://doi.org/10.1080/13540602.2022.2103533>
- Head, J. C. (2019). From the bottom-up: New York City teacher evaluation and the narration of mediated institutional assault. *Critical Studies in Education*, 60(1), 37–54. <https://doi.org/10.1080/17508487.2016.1214159>
- Holloway, J., & Brass, J. (2018). Making accountable teachers: the terrors and pleasures of performativity. *Journal of Education Policy*, 33(3), 361–382. <https://doi.org/10.1080/02680939.2017.1372636>
- Katsuno, M. (2010). Teacher evaluation in Japanese schools: an examination from a micro-political or relational viewpoint. *Journal of Education Policy*, 25(3), 293–307.
- Kvernbekk, T. (2011). The concept of evidence in evidence-based practice. *Educational Theory*, 61(5), 515–532. <https://doi.org/10.1111/j.1741-5446.2011.00418.x>
- Larsen, M. A. (2009). Stressful, hectic, daunting: a critical policy study of the Ontario Teacher Performance Appraisal System. *Canadian Journal of Educational Administration and Policy*, 95.
- Lavigne, A. L. (2014). Exploring the intended and unintended consequences of high-stakes teacher evaluation on schools, teachers, and students. *Teachers College Record*, 116(1), 1–29. <https://doi.org/10.1177/016146811411600103>
- Lavigne, A. L., & Chamberlain, R. W. (2017). Teacher evaluation in Illinois: school leaders' perceptions and practices. *Educational Assessment, Evaluation and Accountability*, 29, 179–209. <https://doi.org/10.1007/s11092-016-9250-0>
- Lejonberg, E., Elstad, E., & Christophersen, K. A. (2018). Teaching evaluation: antecedents of teachers' perceived usefulness of follow-up sessions and perceived stress related to the evaluation process. *Teachers and Teaching*, 24(3), 281–296.
- Lewis, S., & Holloway, J. (2019). Datafying the teaching 'profession': Remaking the professional teacher in the image of data. *Cambridge Journal of Education*, 49(1), 35–51. <https://doi.org/10.1080/0305764X.2018.1441373>
- Lillejord, S., Børte, K., Ruud, E., Hauge, T. E., Hopfenbeck, T. N., Tolo, A., & Fisher Griffiths, P. (2014). *Forms of teacher assessment that can have a positive impact on the school's quality. A systematic knowledge overview*. The Knowledge Center for Education. Norwegian Research Council.
- Mausethagen, S., Prøitz, T. S., & Skedsmo, G. (2021). Redefining public values: data use and value dilemmas in education. *Education Inquiry*, 12(1), 1–16. <https://doi.org/10.1080/20004508.2020.1733744>
- Mausethagen, S. (2013). Accountable for what and to whom? Changing representations and new legitimization discourses among teachers under increased external control. *Journal of Educational Change*, 14(4), 423–444.
- Mayring, P. (2015). Qualitative content analysis: theoretical background and procedures. In A. Bikner-Ahsbals, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education. Advances in Mathematics Education* (pp. 365–380). Springer. https://doi.org/10.1007/978-94-017-9181-6_13
- Møller, J., & Skedsmo, G. (2013). Modernising education: New public management reform in the Norwegian education system. *Journal of Educational Administration and History*, 45(4), 336–353. <https://doi.org/10.1080/00220620.2013.822353>
- NOU (2019:3). Nye sjanser – bedre læring. Kjønnsforskjeller i skoleprestasjoner og utdanningsløp. [New chances - better learning. Gender differences in school performance and education] Norwegian Official Report. Oslo: Ministry of Education and Research.
- OECD. (2005). *Teachers matter: attracting, developing and retaining effective teachers*. OECD Publishing.
- OECD. (2011). *OECD reviews of evaluation and assessment in education*. OECD Publishing.
- OECD. (2013). *Synergies for better learning. An International Perspective on Evaluation and Assessment*. OECD Publishing.
- OECD. (2019). Improving school quality in Norway: the new competence Development model, implementing education policies. *OECD Publishing, Paris*. <https://doi.org/10.1787/179d4ded-en>
- Ozga, J., & Jones, R. (2006). Travelling and embedded policy: the case of knowledge transfer. *Journal of Education Policy*, 21(1), 1–17. <https://doi.org/10.1080/02680930500391462>
- Paige, M. A., & Amrein-Beardsley, A. (2020). 'Houston, we have a lawsuit': A cautionary tale for the implementation of value-added models for high-stakes employment decisions. *Educational Researcher*, 49(5), 350–359. <https://doi.org/10.3102/0013189X20923046>

- Patton, E., & Appelbaum, S. H. (2003). The case for case studies in management research. *Management Research News*, 26(5), 60–71. <https://doi.org/10.1108/01409170310783484>
- Pauffer, N. A., & Sloat, E. F. (2020). Using standards to evaluate accountability policy in context: school administrator and teacher perceptions of a teacher evaluation system. *Studies in Educational Evaluation*, 64, 100806. <https://doi.org/10.1016/j.stueduc.2019.07.007>
- Reddy, L. A., Dudek, C. M., Peters, S., Alperin, A., Kettler, R. J., & Kurz, A. (2018). Teachers' and school administrators' attitudes and beliefs of teacher evaluation: a preliminary investigation of high poverty school districts. *Educational Assessment, Evaluation and Accountability*, 30, 47–70. <https://doi.org/10.1007/s11092-017-9263-3>
- Reynolds, D. (2007). School effectiveness and school improvement (SESI): links with the international standards/accountability agenda. In T. Townsend (Ed.), *International handbook of school effectiveness and improvement* (pp. 471–484). Springer Netherlands.
- Skedsmo, G. (2018). Comparison and benchmarking as key elements in governing processes in Norwegian schools. In *Education policies and the restructuring of the educational profession* (pp. 137–158). Springer.
- Wilcox, K. C., & Lawson, H. A. (2018). Teachers' agency, efficacy, engagement, and emotional resilience during policy innovation implementation. *Journal of Educational Change*, 19, 181–204. <https://doi.org/10.1007/s10833-017-9313-0>
- Yin, R. K. (2012). Case study methods. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2. Research designs: quantitative, qualitative, neuropsychological, and biological* (pp. 141–155). American Psychological Association. <https://doi.org/10.1037/13620-009>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.