Lillejord, S¹., Elstad, E.², & Kavli, H³. (2018). Teacher evaluation as a wicked policy problem. *Assessment in Education: Principles, Policy & Practice*, 1-19.

DOI: 10.1080/0969594X.2018.1429388

<http://www.tandfonline.com/eprint/X78kMCCeFaaz5qT3DJ2t/full>

Abstract

While it is generally assumed that the aim of teacher evaluation is to *formatively* support teachers' professional development, research finds that teacher evaluation practices are predominantly *summative*. This paper describes a Norwegian governmental policy experiment aiming to overcome this fallacy through a bargaining process, where experience-based knowledge was combined with research evidence. When preparing to introduce teacher evaluation, the Ministry of Education and Research commissioned a group of researchers and a group representing practitioners to identify teacher evaluation practices that are conducive for educational quality. Drawing on experiences from the policy experiment, the article discusses three approaches to teacher evaluation: the political, the administrative, and the professional. The analysis indicates that successful implementation of interventions needs a new educational infrastructure and professional school leadership. One conclusion is that teacher evaluation cannot be successfully implemented through traditional linear approaches. A more productive approach is to treat it as a wicked problem.

Key words: teacher evaluation; school leaders; Scandinavian bargaining model; formative assessment; teaching profession

Introduction

Teacher evaluation has been defined as 'the process of arriving at judgements about an individual's past or present performance, against the background of his work environment, and about his future potential for an organization' (Castetter, 1976 p. 232). Murphy, Hallinger, & Heck (2013) find that this definition, by foregrounding the personnel function, positions teacher evaluation in the hierarchical architecture of schooling, with principals as supervisors and teachers as subordinates. Over the last decades, more collaborative and participatory approaches to evaluation are developed, intended to support professional practice in general (Shulha, Whitmore, Cousins, Gilbert & al Hudib, 2016) and in schools (Darling-Hammond, 2013). When describing less top-down teacher evaluation practices,

¹ The Norwegian Knowledge Centre for Education, Oslo, Norway; Department of Education, University of Oxford, Oxford, England

² Department of Teacher Education and School Research, University of Oslo, Oslo, Norway

³ Norwegian Ministry of Education and Research, Oslo, Norway

some researchers use teacher *assessment* (e.g McMahon & Jones, 2015), combine evaluation and assessment (Smagorinsky, 2014) or use teacher *appraisal* (e.g Flores, 2012).

In one perspective, teacher evaluation may be perceived as an abstract policy idea that circulates globally (Popkewitz, 2000), entangled in a neoliberal discourse ignited by the report *A Nation at Risk* (National Commission on Excellence in Education, 1983), blaming schools for the recession in the US economy (Hursh, 2007), and holding teachers accountable for their students' learning outcome. In another, it is a practice yielding information leaders can use to enhance teaching quality (Goe, Biggers & Croft, 2012). In theory, therefore, the purpose of teacher evaluation is twofold: to serve as a professional development process and as a quality assurance mechanism (Kraft & Gilmour, 2016). Difficulties in reconciling these two approaches (Popham, 1988; Firestone, 2014) may explain why teacher evaluation is a contested practice (Cohen & Goldhaber, 2016; Elstad et al., 2015b) and a hotly debated topic (Collins & Amrein-Beardsly, 2014).

Different countries have tried different approaches to evaluate teachers' work, and when Norwegian politicians during election year 2013 contemplated national initiatives for more systematic teacher evaluation, policymakers in the Ministry of Education and Research initiated a process that will be described in detail later, of gathering knowledge from research and practice. The assumption was that ownership to a joint knowledge base might facilitate the process of implementing teacher evaluation nationwide.

While each country responds in its own way to international policy discourses (Hudson, 2011), researchers often find that across countries, systems for teacher evaluation are reduced to bureaucratic problem solving and fail to strengthen teachers' professional knowledge base (Darling-Hammond, 2013; Taut & Sun, 2014) or improve schools (Hallinger, Heck & Murphy, 2014). This article presents a policy experiment that sheds light on why this happens.

The three authors had different roles in the policy experiment. The first author was the lead author of the systematic review, the second author headed the work in the group representing practitioners and the third author contributed to designing the policy experiment. We first explain the policy context and background for the experiment that drew on the Scandinavian bargaining model, then frame teacher evaluation as a wicked problem and discuss three approaches to teacher evaluation: the political; the administrative; and the professional. The article analyses teacher evaluation as a *wicked problem* by asking: why is the formative purpose of teacher evaluation frequently reduced to administrative routines and technical practices? We finally indicate measures needed to strengthen the teaching profession.

Background and policy context

Educational policy in Norway (for instance curriculum guidelines) is developed at the national level and adapted locally by autonomous local education authorities. The public-school system is pervasive; only about three percent of primary and lower secondary school pupils and seven percent of upper secondary pupils attend private schools. The Organization

for Economic Co-operation and Development (OECD) has commented on the high level of trust and lack of national quality control in Norwegian education (OECD, 1987). Schools are expected to conduct self-evaluations, and some of them do, but Norway has no nation-wide systematic evaluation of teachers' work and no national school inspectorate. The responsibility for quality in education rests on local municipalities (primary education) or counties (upper secondary) and the County Governor carries out inspections in accordance with the *Education Act* (1998) and *Kindergarten Act* (2005). Having reviewed Norwegian assessment and evaluation practices, the OECD more recently recommended that Norway connect teacher appraisal to professional learning and school development and integrate teacher evaluation in the unfinished National Quality Assessment System (Nusche, Earl, Maxwell & Schrewbridge, 2011).

Norway is not the only country to get such advice from the OECD. Researchers (e. g. Anagnostopoulos, Lingard & Sellar, 2016; Ozga, Dahler-Larsen, Segerholm & Simola, 2011) argue that the OECD defines education quality and steers education systems through assessment. Teacher evaluation is an interesting case, as it has been identified as a means to increase student's learning outcome in an era of outcomes-based education policy (OECD, 2009). A recurring problem is, however, that while the intention is to implement evaluation practices that support teachers' professional development and promote quality in schools, many systems are reduced to instruments of accountability, burdening and demotivating schools and teachers instead of invigorating and empowering them (Santiago, Benavides, Danielson, Goe & Nusche, 2013). This might happen if school leaders use bureaucratic levers of requirements and regulations when trying to solve problems that require professional skill and expertise (Mehta, 2013). The described policy experiment aimed to avoid these pitfalls by

(a) combining research evidence and knowledge from practice and

(b) anchoring the implementation process in broad, participative bargaining processes.

Before outlining the design of the policy experiment, we briefly present the policy context, research on implementation failure and the theory of wicked problems.

Outcomes based policies and teacher evaluation

In 1990, a white paper introduced management by objectives (MbO) in the Norwegian education sector (White Paper no. 37). The idea was simple. Politicians should formulate goals and steer from a distance; administrators should interpret the goals, divide them into smaller, manageable goals, explain these to the schools, and oversee how teachers followed up on the politicians' intentions. An essential but initially under-communicated part of MbO was the reporting of results and the measurement and comparison of each employee's performance with set standards. In the wake of MbO, developed by Drucker (1954) for private business companies, new public management (NPM) emerged, modelled on multidivisional private sector corporations, with corporate headquarters overseeing business units, controlled through monitoring performance outcomes. Central to managerialism has been a rational-technical approach to decision making, with an aim to simplify and favouring general managerial skills over professional or technical content knowledge (Head & Alford, 2015). Both MbO and NPM assume the existence of an infrastructure that is lacking in

complex education systems. Therefore, neither of them is appropriate for wicked environments, characterised by uncertainty, complexity and disagreement.

When the Programme for International Student Assessment (PISA) shock hit Norway and other countries in 2001 (Lundgren, 2011), a focus on measurable results rapidly overshadowed other political visions for education. Drawing on research arguing that the teacher is the most important factor explaining improved student learning outcomes (Hanushek, 1971, 1999), politicians wanted to know how they could increase the 'value added' of teachers. Teacher evaluation emerged as one solution (Berliner, 2013) with three interest groups: politicians; administrative staff; and teachers. Typical political arguments were that teacher evaluation secures pupils' democratic participation (Elstad et al., 2015b) and that schools more readily will strive for improvement in a competitive market system (Hanushek & Raymond, 2005). Administrative arguments were that accountable and efficient school leaders should evaluate teachers' work and inform them about how to improve their teaching, while teachers expect teacher evaluation to support their broader mandate and professional learning (Smylie, 2014; Smylie, Murphy & Louis, 2016). As these interests are not easily reconciled, teacher evaluation turns into a policy implementation problem.

Implementation challenges

There is a growing realisation that major social issues of modern life are grounded in value perspectives. Gathering more information for scientific analysis or initiating further research is insufficient when the goal is to understand and resolve social problems (Head & Alford, 2015). Both public and private institutions now realise that a strong knowledge base is necessary, but does not in itself guarantee successful implementation.

Having reviewed the research on implementation failure, Decker et al. (2012) find that failure rates for organizational change have been hovering around 73 percent since the 1960s and range somewhere between 28 percent and 93 percent. One reason, they suggest, is that traditional linear problem-solving models, such as MbO, assume that the leaders' job is finished when the goals are formulated. This assumption no longer holds, they argue, because organisations are handling increasingly complex problems in increasingly complex contexts. Leaders can no longer formulate goals and expect others to finish the job; they must keep their attention on the entire implementation chain. Because the politics of implementation differs from the politics of enactment, McDonnell and Weatherford (2016) suggest that implementation should be understood as a *continuous* political process.

This insight yields the question: is teacher evaluation reduced to technical procedures and bureaucratic documentation, not fulfilling the formative intentions, because it is treated as a benign problem while it – more appropriately – should be treated as a 'wicked' problem? Characteristically, wicked problems are difficult to solve because contradictory intentions, for instance formative *and* summative ambitions, are embedded in the problem. Wicked problems are ambivalent, resist resolution and cannot be 'solved' in the sense that they disappear. Merton (1976) has shown that ambivalence is central to modern societies, and warns of unintended consequences when solving problems in ambiguous contexts. Evaluation and

assessment are practices with high probability for unintended consequences (Lavigne, 2014), causing potential problems for actors involved.

Teacher evaluation as a wicked problem

The term wicked problem (Rittel & Webber, 1973) is used to describe social problems in modern, pluralistic societies. According to Buchanan (1992), Horst Rittel borrowed the phrase from Karl Popper during the 1960s and modified it for his own use. Wicked problems are characterized by ambiguity and uncertainty (Head, 2008), generating conceptual difficulties and practical challenges. Efforts to solve one aspect of a wicked problem often breed new problems. The greater the disagreement between stakeholders, the more wicked the problem (Camillus, 2008). Table 1 highlights general characteristics of wicked problems.

[Insert Table 1 here]

In modern organisations, stakeholders with different values and priorities are struggling with problems that are difficult to come to grips with (Camillus 2008), with no 'right' answers. Head & Alford (2015, p. 717) distinguish between Type 1 situations, where both the definition of the problem and the likely solution are clear, Type 2 situations where the definition of the problem is clear but not the solution, and Type 3 situations where both problem definition and solution are unclear. Teacher evaluation appears to be a Type 3 situation, as goals are ambiguous or contradictory, the expected outcome is broad and vaguely defined (evaluation should contribute to professional development, quality in school, improved learning outcomes for students etc.) and various actors have their personal interpretations of how this outcome may be achieved. This kind of 'inescapable wickedness' is, according to Jordan, Kleingasser, & Roe, (2014, p. 415), frequent in education.

Head & Alford (2015) argue that efforts to deal with wicked problems are impeded by the public sector's characteristic ways of decision making, organising routines, financing and staffing. Hierarchical organisation and control, focused on input monitoring and process compliance, limit the opportunities to think expansively about complex policy issues. Instead, they suggest the following strategies: (a) going beyond technical/rational thinking; (b) collaborative working; (c) new models of leadership; and (d) reforming the managerial infrastructure of government. For the Norwegian Ministry of Education and Research, a way forward when designing a scheme for teacher evaluation that could be implemented nationally, was to draw on the Scandinavian bargaining model, based on influence, joint decision-making and mutual respect. The term Scandinavian bargaining model refers to inter-party collaboration between employer and employee organisations (and, in some instances, the state). The policy experiment follows Head & Alford (2015) points a, b and c, as it goes beyond technical/rational thinking, by collecting knowledge both from research and experience, is innovative in terms of collaborative policy leadership and therefore meets three of the four above suggested requirements for successful implementation strategy. Having presented the policy experiment, we will return to the question of infrastructure (point d).

Policy design for teacher evaluation

The policy experiment is anchored in the current Norwegian government's political platform (2013–2017) which opens for students in upper secondary to evaluate their teachers' instructional practice, and aims at implementing teacher evaluation more systematically nationwide. In Norwegian schools, teacher evaluation is rarely used as an instrument for quality assurance (Anderson, Terras & Dagfinrud, 2013) and schemes vary considerably nationwide. In some secondary schools, deputy heads perform appraisal interviews with groups of teachers (Abrahamsen, 2017), and in one county, student ratings are found to provide teachers in upper secondary with constructive feedback they can use to improve the quality of their teaching (Elstad et al., 2017). When planning the policy experiment, the Ministry of Education and Research drew upon EU and OECD recommendations defining teacher evaluation as a method of developing and improving the quality of teaching, strengthening teachers' status and recognition as professionals, raising the attractiveness of teaching, and supporting professional development for individual teachers throughout their careers. Guided by a Centre for Educational Research and Innovation, CERI/OECD project, *Governing Complex Education Systems*, the introduction of teacher evaluation was perceived as a *systemic* organisational change effort (Senge, 1990), and treated as a bargained solution amongst stakeholders.

To facilitate the process, the Ministry designed a project with several phases, starting with involving relevant stakeholders and establishing a knowledge base. The aim was that a parallel and participatory process would result in agreement on characteristics of high quality teaching, guidelines for teacher evaluation and requirements for teacher professionalism that could later be used in national recommendations. The Ministry simultaneously commissioned a systematic review on teacher evaluation from a group of researchers (Lillejord et al., 2014) and established a stakeholder group with teachers' and headmasters' unions, an employer association, a teacher educator and student association etc., to summarise assessment experiences from schools and municipalities (Elstad et al., 2015a). This policy design, where stakeholders are engaged in bottom-up processes, provides ownership and empowerment, but may also produce confounding compromises, complicating the implementation processes.

The Scandinavian bargaining tradition and the GNIST partnership

The hallmark of successful policy making is to reconcile disagreements. While contradictory interests give something to bargain *about*, common interests give something to bargain *for* (Schelling, 1980). Scandinavian countries have a long tradition of significant employee influence, collective agreements and centralised bargaining (Byrkjeflot, 2001). For long periods, the open economies in Scandinavia have had high productivity and work effort, small wage differentials, and a generous welfare state (Barth, Moene & Willumsen, 2014). The system is based on mutual respect, influence, and joint decision-making (Brandal, Bratberg, & Thorsen, 2013), aiming to reduce the negative effects of inherent inter-party conflicts of interest, which in some instances can disrupt processes (Moene & Wallerstein, 2002). This model for reconciling conflicts has been regarded with both astonishment and interest by non-Scandinavians (Barth, Moene & Willumsen, 2014; Lindbeck, 1997).

The tradition of negotiating compromises through giving and taking is fruitful for many purposes. However, when teacher evaluation is on the bargaining table, a relevant question is: Which are the common and potentially contradictory interests?

In line with the bargaining tradition, the Norwegian Ministry of Education and Research in 2009 initiated the establishment of GNIST ('SPARK'), a partnership for teacher professional development and arena for cross-collaboration between stakeholders (leaders of teacher and teacher educator unions, student teachers, headmaster and student associations, an employer association etc.). The partnership's mandate was to "strengthen the schools' academic platform and the prestige of the teaching profession" (GNIST, 2009, p. 3). One key initiative identified by the partners was teacher evaluation. The GNIST partnership was subsequently mobilised in this policy experiment with three phases, described below.

Phase 1: The two groups: Knowledge from experience and research

In autumn 2013, the GNIST partnership appointed a working group (Elstad et al., 2014) with a mandate to "...identify and assess important prerequisites for teacher evaluation contributing to the development of schools as learning organisations, to teachers' professional development, and with a positive effect on pupils' learning and development" (p.5). The report concluded with these preconditions for successful teacher evaluation:

- mutual trust between the parties;
- appraisal schemes should be systematic, predictable, and practically feasible;
- specific, immediate feedback is deemed better than general and delayed feedback;
- teacher appraisal should focus on development and not serve to monitor the teachers' work;
- teacher appraisal schemes should be based on research evidence.

In parallel to appointing the working group, the Ministry commissioned a systematic review from the Norwegian Knowledge Centre for Education (Lillejord et al. 2014). The review question was: Which teacher evaluation schemes may promote quality in school? Initial searches did not identify previous systematic reviews on the topic. The search process was complicated as the Ministry asked the group of researchers to include quantitative *and* qualitative studies, studies about formative *and* summative assessment, and information about process *and* outcome quality. Systematic searches (search period 2009-2013) yielded around 12.000 entries. Five expert researchers supported the review team in the process of assessing studies, following predefined criteria for inclusion and exclusion. Finally, 79 studies were included in the review that concluded with four preconditions for successful teacher evaluation:

- Active participation, responsibility and trust;
- Involved parties need scientific methods and data use literacy;
- Clear and direct communication (not too many goals);
- Responsible and competent leaders with attention to dialogue and follow-up.

The systematic review revealed that the amount of time and resources needed for successful teacher evaluation is frequently underestimated. Main findings were that evaluation is unsuccessful when the scope is too broad, the object unclear or no plans are in place for how to deal with unintended consequences that emerge during the evaluation. Also, key actors in successful implementation of teacher evaluation are active school leaders and local school authorities and results are better when teachers' professional learning is prioritised over bureaucratic evaluation procedures aiming at control. The two groups did not collaborate, but met twice during 2013 and 2014 to exchange information about work processes. For some of the practitioners it was unclear what they could expect from research. During the meetings, misunderstandings were clarified. Because of the meetings and overlapping findings from the two reports, the GNIST partners developed a sense of ownership to the overall process.

In the GNIST group, disagreements occurred, for instance on the name of the activity (teacher evaluation or *teaching* evaluation) and on the anonymity of students' feedback to teachers. Further, in February 2014, one teacher association (Norwegian Union of School Employees) resigned from the partnership as it had lost its confidence in the employer body (KS, the Norwegian Association of Local and Regional Authorities), due to a breakdown in negotiations over regulations of working hours. In April 2014, the Minister of Education and Research invited all participants to discuss the knowledge base (from research and practice) and plan for the next phase. All remaining GNIST partners adhered to the knowledge gathered during Phase 1, underlined the importance of avoiding the pitfalls and unintended consequences revealed in the systematic review, and agreed to move on to Phase 2.

Phase 2: Designing teacher evaluation

Next, the working group outlined guidelines for teacher evaluation to be piloted and tested in Norway, based on these recommendations from the partnership organisations, (Elstad et al., 2015a):

1. Students' evaluations of teaching via anonymous surveys;
2. Observation of teachers' work by school leaders or external experts, with subsequent feedback;
3. Students' learning progress: information about their learning, self-formation ('Bildung'), and development;
4. Feedback on teachers' educational practice.

Based on these recommendations a trial scheme for teacher appraisal was developed, inspired by teacher evaluation practices in upper secondary schools in two Norwegian counties (Vestfold and Akershus), endorsed by the local teachers' unions and perceived as an interesting model for a pilot. It included anonymous student surveys and follow-up through individual performance reviews and group sessions for each discipline.

According to Camillus (2008), traits of wicked problems include confusion, discord, and lack of progress. In hindsight, such sentiments may have characterised Phase 2. Even though all three authors attended meetings and were involved in parts of the process, we find it difficult to identify one single event that caused the process to slow down. It could be a growing

realisation in the GNIST group that the work ahead was overly complex, partly related to research findings in the systematic review, about potential pitfalls and unintended consequences of teacher evaluation. Some of the partners may have wanted not to be interpreted as distrusting teachers; some teacher representatives may have felt that evaluation indirectly signalled a lack of trust in teachers' work and therefore did not embrace the process whole-heartedly. Or, the process may have revealed that there were more conflicting than common interests to bargain for.

The policy experiment rests on the assumption that successful implementation of political decisions presupposes ownership and commitment to bargained solutions. Munneke, Andriessen, Kanselaar and Kirschner (2007) have reviewed research indicating that serious argumentation only emerges when serious problems are encountered. Normally, when people reason in groups, they take the correctness of their supportive theories for granted and most evidence brought to the table is not evidence, but anecdotal pseudo-evidence. Because thinking about alternative theories, counterarguments, and rebuttals is very difficult, more time is spent on supporting theories and less attention given to the opponents' views. The main goal is to persuade the other, not explore together. Much policy debate takes the form of bargaining to achieve temporary compromises (small-scale symbolic wins), while the wicked, underlying differences persist. This might also have happened in the policy experiment described in this article.

During this phase, the divergent interests between the stakeholders were not *on* the bargaining table, but rather *under* the table and difficult to grasp. Political goals are broad and ambitious; supposed to 'trickle down' to administrators who are responsible for the successful implementation of political visions. Administrators realise that their job is to transform the idea into a feasible project, reduce the ambitions and narrow the scope while remaining faithful to the overarching ideals. Teachers know, from experience, that they are left with the challenge of realising ambitions that are often at odds with what may possibly be achieved in their everyday 'reality'. Consensus on the surface may have concealed the uncertainty underneath, causing uncertainty and hesitance. Before presenting the pilot scheme (Phase 3), we will outline two categories of problems that, according to our interpretation, seem to have influenced the deliberations in the GNIST working group during Phase 2.

First category of problems: Tensions between summative and formative purposes of teacher evaluation

The most consistent finding in the systematic review (Lillejord et al., 2014), reported from countries as diverse as Belgium (Delvaux et al., 2013), Chile (Taut & Sun, 2014), China (Zhang & Ng, 2011), and Portugal (Flores, 2012), was that teacher evaluation procedures seem to fail when they violate insights from evaluation research such as (a) sufficient resources (expertise, money, time) must be made available and efforts balanced against expectations (Cousins & Earl, 1992); (b) evaluation must be planned and conducted in ways that are experienced by the person(s) being evaluated as valid and reliable (Smagorinsky, 2014), and performed by *knowledgeable* evaluators (Darling-Hammond, 2014); (c) evaluation should concentrate on specific parts of the work while not neglecting that each part

is integral to a bigger whole (Darling-Hammond, 2013; 2014); (d) evaluation should not only measure outcomes but contribute to strengthening the learning environment (Cousins, Whitmore & Shulha, 2013); and (e) because evaluation has a huge impact on practice, it should be planned and conducted in ways that minimise unintended consequences (Lavigne, 2014).

[Insert Table 2 here]

Researchers found that while negotiations over teacher evaluation schemes were consensual at the national or central level, implementation problems occurred locally. Divergence between political visions, administrative procedures and expected teacher professional development reduced teacher evaluation to technical procedures, paper work and little attention to teachers' professional learning. The sudden lack of progress in the group may be ascribed to a suspicion that this might also happen in Norway.

Second category of problems: Stakeholder groups with divergent interests

The GNIST partners (representing teachers, school leaders, district and National level authorities) had different expectations to the bargaining process. In a system based on MbO, politicians are formulating vision goals. When implementing their interpretation of the political visions for teacher evaluation (accountability, improved learning outcomes, transparent democratic processes and better schools), administrators at district- and school level use available tools and familiar quality assurance procedures (Pupala, Kascak & Tesar, 2016). Teachers expect formative feedback from school leaders, but frequently find themselves victims to bureaucratic procedures, narrowing the scope of their job and constraining their professional discretion (Kraft & Gilmour, 2016). Figure 1 illustrates this:

[Insert Figure 1 here]

The political-bureaucratic discourse is holistic and normative, favouring general tools and procedures. Teachers, however, continuously relate to concrete problems in need of specific solutions and immediate attention. While politicians and bureaucrats give direction from a distant *system* perspective, teachers sort out complex *life-world* issues (Habermas, 1984). Such tensions, inherent to MbO as a steering system, driven by increased control, were not initially identified by the bargaining partners as contradictory interests, but appear to have surfaced when more concrete practices and procedures were specified in Phase 2. Discussions about whether future teacher evaluation activities should be labelled *teacher* or *teaching* evaluation may seem trivial, but might have had significant symbolic bearings as indicators of trust or distrust for some stakeholders. The theory of wicked problems addresses these subtle mechanisms of inter-party tensions by showing that disagreements may have multiple explanations, be symptoms of other problems than the one discussed and that solutions are experienced as good or bad, not right or wrong.

To gain further insight, Phase 3 of the experiment was therefore designed as a pilot.

Phase 3: Trial scheme

The trial scheme proposed by the GNIST working group was designed as a follow-up study at eight schools (The Norwegian Directorate of Education and Training, 2016), with supervision from a consultancy firm (IMTEC), actively guiding teachers and school leaders in improvement processes. IMTEC is expected to collect, organise, and analyse schools' experiences related to the testing of a system labelled 'evaluation of professional development' in schools. The questions guiding Phase 3 are developed by the Directorate for IMTEC to answer in a report after the trial scheme:

1. What impact did anchoring at school and school owner level have for participation in evaluation for professional development?
2. How was the trust between participants, based on agreement to use the information for development and not control, safeguarded during the period schools participated in the evaluation of professional development?
3. Are there signs that the four elements mentioned in Phase 2 were used formatively and systematically? Did they contribute to teacher professional learning and, more broadly, school development? Did they strengthen the general feedback culture in schools? Did they develop schools as learning organisations?

The report from IMTEC (Stranden, 2017) does not provide sound empirical answers to the questions, and serves to confirm a slowdown in the process that was successfully brought through Phase 1 by the Scandinavian bargaining model, where the participants agreed on a joint knowledge base from research and practice, to Phase 2, when unclear responsibilities and conflicting interests between the stakeholders, in line with the theory of wicked problems, may have contributed to a sense of confusion and lack of progress. As previously mentioned, no single event appears to have caused the process to slow down, more a growing suspicion that the task at hand was more complicated than anticipated, and the realisation that teacher evaluation may have unintended consequences. Apparently, at this stage, a complex implementation process with several stakeholders with divergent interests needs a different approach.

Discussion and way forward

The purpose of this study was to shed light on the question why teacher evaluation often is reduced to bureaucratic routines and technical practices and fails to support teachers' professional learning. One reason, we have indicated, may be that teacher evaluation is perceived as a benign and technical problem, while it should be treated as a wicked problem. The policy experiment has shown that the bargaining approach is well suited to establish agreement at the initial stages, as consensus was established in the group during Phase 1. However, the fact that problems occurring at later stages require alternative approaches should have been anticipated and planned for. Partners with diverse interests may endorse the knowledge base and agree on the general intention to establish a system for teacher evaluation. At the same time, however, they envision different outcomes and end results of the process. Tensions emerge when the more specific questions are raised about *what* should be done, *how* it should be done and by *whom*.

We have identified a slowdown in the implementation process during Phase 2 of the policy experiment, and have tried to indicate why this may have happened. Here, we will elaborate on what was learned from the policy experiment and what must be taken into consideration if future politicians want to implement teacher evaluation more systematically at the national level in Norway.

When analysing educational organisations, Weick (1976) noticed that in schools, two *loosely coupled* systems appear to work separately from each other. In this analytical perspective, one system consists of teachers, parents, students and curriculum, another of the principal, vice-principal and superintendent. The two systems are somehow attached, but each part preserves its own identity. While teachers have the responsibility for the schools' 'technical core', the principal holds the authority of office. Loosely coupled organisations are stable, allow for self-determination and provide a sense of autonomy and efficacy. However, Meyer & Rowan (1977) observe that decoupled organisations tend to employ a logic of confidence and good faith – in place of coordination, inspection and evaluation. Quite often, teachers' autonomy and reluctance to change is cited as the prime reason for reform implementation failure in schools. Instead, Mehta (2013, 463) argues that a root problem in the educational sector is that it is organised as a bureaucracy rather than as a profession, and suffers under the absence of an 'educational infrastructure' to support practice. While he relates this to the US late 1800 and early 1900, Ramirez & Boli (1987) show that in most Western countries education was bureaucratised when mass schooling was institutionalised. According to Weber (1947), the hallmark of a successful bureaucracy is impartiality. The bureaucrat is an expert on general and domain-specific rules and regulations, and as bureaucrats are socialised into an ethos of rule-following, one bureaucrat may – in principle – be replaced by another bureaucrat who would reach the same conclusion.

If school leaders historically have belonged to a different 'system' than teachers, they may identify more with bureaucrats than with the profession and even perceive themselves as detached, top-down general managers. Many problems in education can therefore be ascribed to the current educational infrastructure where politicians, administrators and teachers, work in a 'layered' and loosely coupled system largely based on linear top-down communication and principal-agent relationships. The three groups have different knowledge bases, divergent interests and problem-solving procedures that will be discussed below.

Tyack (1974) and Tyack & Hansot (1982) show how Taylorism, between 1890 and 1920, inspired the development of the modern school system. Following business models, (mostly male) superintendents were empowered as CEOs to lead schools, while low status (mostly female) teachers worked in relative isolation from each other, following rules and programmes prescribed by the superintendents. One by-product of this institutional form has been relatively weak professionalisation among teachers (Elmore, 2000) and the idea that the complex practice in need of attention in education was not teaching, but administration. The training of administrators provided an opportunity for education departments to produce a new class of mostly male professionals who brought findings from research to the teachers. Mehta (2013, p. 471) comments that already John Dewey did not believe in the model where

administrative 'experts' dictated educational methods to passive teachers and warned against partitioning researchers and teachers, as the two groups have a shared interest in understanding and improving practice.

Tschannen-Moran (2009) argues that school leaders adopt a bureaucratic orientation at the expense of cultivating professionalism in schools. The OECD observes that, in Chile, school leaders have traditionally played more of an administrative and managerial role than a pedagogical leadership role, and asks if school leaders have the necessary competencies to lead the effective implementation of teacher evaluation (Santiago et al., 2013). Kraft & Gilmour (2016) find that not only do 19 of 24 interviewed principals express concern about their ability to provide meaningful feedback to teachers, they also regard this as outside of their area of expertise. If school leaders see themselves as administrators in a top-down system and lack the necessary competence to give formative feedback to the teachers, we are narrowing in on the question why teacher evaluation is predominantly summative and why the policy process lost its energy during Phase 2, when the more concrete questions about who should do what were raised.

Rittel and Webber (1973) claim that because governance, since the 1960s, is built on goal-setting and measuring outcomes, modern professions, initially expected to solve definable problems, have become victims of their own success. Formulating actionable goals is a complicated task, and it has turned out to be 'terribly difficult, if not impossible', to make various systems operational (Ritter & Webber, 1973, p. 157). One solution is to move beyond dichotomies (Blömeke, Gustafsson & Shavelson, 2015) and perceive teacher professionalism as a developing continuum. This aligns with insights from the theory of wicked problems and the argument that leaders cannot communicate goals and step aside, because successful implementation requires leaders' continuous attention.

Education is a highly politicised field with a growing, but still weak, knowledge base and no strong tradition for collaborative professional work. Having reviewed the literature on assessment of professionalism, Goldie (2013) finds that peer assessment is the most commonly used instrument among medical doctors. In the systematic review that represents one part of the knowledge base for the policy experiment (Lillejord et al., 2014) peer assessment was frequently mentioned as an underutilised resource in teacher evaluation. Studies also identified assessment literacy as a prerequisite for teacher evaluation to be successful. Sahlberg (2010) describes the ambivalence in ongoing reforms where teachers are trapped in a dilemma between schooling for social capital and moral purpose with student-centred pedagogy and learning on the one side, and, on the other, efficiency-driven education with teacher-centred instruction and achievement. He calls for intelligent accountability (O'Neill, 2002), which combines internal accountability, self-evaluations, critical reflection and school-community interaction with external accountability. When discussing how teachers develop assessment capacities, Livingston and Hutchinson (2016) suggest ongoing enquiry-based training and mentoring focused on pupils' *and* teachers' learning. This may distribute the responsibility to the professional collective of teachers and school leaders and move the attention away from the teacher being individually accountable for the students'

learning outcome. One implication is that steering from a distance must be supplemented with organising schools for continuous development – two approaches that require different structures and leadership skills.

Ideally, teacher evaluation should support teacher professionalism. However, argue Cohen & Mehta (2017), teaching is not a full-fledged profession because professions have features that regulate and organise the quality of work; standards for training and licensing, prohibiting those without such credentials from joining the profession; agreement on definitions of problems the profession can solve; appropriate approaches to tackling those problems; and mechanisms to ensure that standards are applied in practice. Professional work therefore requires adherence to a joint and continuously renewed knowledge base, ethical standards, collaboration and a self-critical approach to practice (Gardner & Valentine, 2014; Simons & Ruijters, 2014). In professions, leaders and employees typically share knowledge base, standards, principles and guidelines.

Conclusion

Based on experiences from a Norwegian policy experiment, we have argued that teacher evaluation has traits of a wicked problem and should not be treated as a benign problem. Ashforth, Rogers & Pratt (2014) have identified organisational responses to ambivalence as avoidance, domination, compromise, and holism. In the described policy experiment, several of these responses were observed, related to the fact that teacher evaluation is expected to reach two goals not easily reconciled: professional development *and* quality assurance – in Popham's (1988) words – a *dysfunctional* marriage of formative and summative approaches. The expected outcome of the process is vaguely defined (professional development for teachers, quality schools, improved learning outcome for students etc.), and various interest groups and stakeholders have their personal interpretation of goals and outcome.

According to Ingvarson (2005), policy makers can have quite naïve expectations about how easy it is to bring about educational change. This may result in a technical approach to professional development, based on the assumption that the goal is reached if skilled administrators guide teachers step by step in pre-prescribed processes towards the realisation of the political vision for education. While administration today is modernised, certain bureaucratic ground rules persist. March (1984) ascribes the success and longevity of the bureaucracy to adaptivity and loose coupling between what is said and done.

One ambition of the study was to shed light on the question: Why is teacher evaluation often reduced to bureaucratic routines and technical procedures? The analysis indicates that when the policy issue at hand is wicked and ambivalent, traditional bargaining approaches only help part of the way as compromises fail to give direction and impetus. It appears that the political problem-solving approach (holistic compromises) and the administrative problem-solving approach (reducing complexity) produce more problems for the teaching profession than they solve. The policy experiment has also revealed that a core problem in the case of teacher evaluation is the historic loose-coupling between school leaders and teachers. The top-down

linear MbO-approach, combined with NPM, appears to have broadened the distance between the two 'systems'.

Governments everywhere face the challenge of translating national policy into implementation strategies that work on the ground and fit the varying conditions in different parts of a country (Kettl, 2015). When political decisions are holistic compromises, and there are no policy feedback loops in place, several local interpretations of centrally initiated intentions may be expected. We have argued that MbO and NPM both assume the existence of an infrastructure that is lacking in educational institutions, where loose-coupling appears to be a major impediment to implementation efforts. Because wicked problems always are only temporarily 'solved'; there is a need for an organisational infrastructure with feed-back loops where professionals continuously relate to the problems, knowing that they are not expected to solve them once and for all. Realising that all solutions are temporary lies at the heart of continuously learning, self-renewing systems and requires that school leaders exercise their authority with a professional orientation; perceive teachers as competent and knowledgeable contributors (Spillane & Diamond, 2007) and extend adaptive discretion to teachers in the conduct of their work (Tschannen-Moran, 2009).

Transformational leadership, where the leader formulates a vision, sets the direction, and inspires others to follow, does not work when circumstances are wicked (Head & Alford, 2015). Top-down bureaucratic structures tend to adopt an implicit distrust towards subordinates; an orientation that is counterproductive when the goal is productive learning and development (Lillejord & Dysthe, 2008). A lesson learned from the GNIST partnership is that relational trust among the parties involved in this kind of managerial structure lubricates the endeavour of common interests. Another lesson learned is that narrowing the loose-coupling by adopting practices that engage school leaders more directly in teacher's work is fundamental to the further development of the teaching profession.

Following this argument, a possible next step when moving teacher evaluation forward, is to perceive teaching an intellectual activity (Stroupe, 2016) and engage teachers more directly in the evaluation process. Professionals' knowledge consists of research (theory) and acknowledged good practice, and Darling- Hammond (2015) suggests that educators should design systems for teacher evaluation based on their knowledge of how students are learning in relation to how teachers are teaching. This has support in implementation research arguing that leaders must keep their attention on the entire implementation chain (Decker et al., 2012). One approach is to establish an educational infrastructure where knowledge from research and experience supports professional work and give the responsibility of evaluating, improving and professionalising teaching jointly to teachers and school leaders.

There is, currently, increased interest in how the teaching profession may become a more mature profession (Sachs, 2016). This article has shown that in parallel to this debate, the question where school leaders are positioned in relation to the profession should be raised. Formative teacher evaluation presupposes knowledgeable evaluators who understand the complex and contextual nature of teaching and can engage in informed dialogue about it.

Professionalisation of teachers presupposes a collective effort within the profession to establish a joint research- and experience-based knowledge base that will serve as a centrepiece of a modernised educational infrastructure. This work needs leadership.

References

- Abrahamsen, H. (2017). Redesigning the role of deputy heads in Norwegian schools – tensions between control and autonomy? *International Journal of Leadership in Education*, DOI: 10.1080/13603124.2017.1294265
- Anagnostopoulos, D., Lingard, B., & Sellar, S. (2016). Argumentation in Educational Policy Disputes: Competing Visions of Quality and Equity. *Theory Into Practice*, 55(4), 342-351.
- Anderson, S. K., Terras, K. L., & Dagfinrud, M. (2013). Appraising Teacher Quality in Norwegian Schools. *International Journal of Business, Humanities and Technology* 3(8), 16-42.
- Ashforth, B. E., Rogers, K. M., Pratt, M. G., & Pradies, C. (2014). Ambivalence in organizations: A multilevel approach. *Organization Science*, 25(5), 1453-1478.
- Barth, E., Moene, K. O., & Willumsen, F. (2014). The Scandinavian model—an interpretation. *Journal of Public Economics*, 117, 60-72.
- Berliner, D. C. (2013). Problems with value-added evaluations of teachers? Let me count the ways! *The Teacher Educator*, 48(4), 235-243.
- Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (2015). Beyond dichotomies. *Zeitschrift für Psychologie*, 223(1), 3-13. DOI:10.1027/2151-2604/a000194
- Brandal, N., Bratberg, Ø., & Thorsen, D. E. (2013). *The Nordic model of social democracy*. London: Palgrave Macmillan.
- Byrkjeflot, H. (2001). *The Nordic model of democracy and management*. Bergen: Fagbokforlaget.
- Buchanan, R. (1992). Wicked problems in design thinking. *Design Thinking*, 8(2), 5–21.
- Camillus, J. C. (2008). Strategy as a wicked problem. *Harvard Business Review*, 86(5), 98–108.
- Castetter, W. B. (1976). *The personnel function in educational administration*. New York, NY: Macmillan.
- Cohen, J., & Goldhaber, D. (2016). Building a More Complete Understanding of Teacher Evaluation Using Classroom Observations. *Educational Researcher*, 45(6), 378-387.
- Cohen, D. K., & Mehta, J. D. (2017). Why Reform Sometimes Succeeds: Understanding the Conditions That Produce Reforms That Last. *American Educational Research Journal* 54(4), 644-690.
- Collins, C., & Amrein-Beardsley, A. (2014). Putting growth and value-added models on the map: A national overview. *Teachers College Record*, 116(1), 1-32.
- Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational evaluation and policy analysis*, 14(4), 397-418.

Cousins, J. B., Whitmore, E., & Shulha, L. (2013). Arguments for a common set of principles for collaborative inquiry in evaluation. *American Journal of Evaluation*, 34(1), 7-22.

Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement*. New York: Teachers College Press.

Darling-Hammond, L. (2014). One piece of the whole: Teacher evaluation as part of a comprehensive system for teaching and learning. *American Educator*, 38(1), 4.

Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher*, 44(2), 132-137.

Decker, P., Durand, R., Mayfield, C. O., McCormack, C., Skinner, D., & Perdue, G. (2012). Predicting implementation failure in organization change. *Journal of Organizational Culture, Communication and Conflict*, 16(2), 29-49.

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.

Drucker, P. (1954). *The Practice of Management*. New York: Harper.

Education Act (1998). Retrieved from <https://lovdata.no/dokument/NL/lov/1998-07-17-61>

Elmore, R. F. (2000). *Building a new structure for school leadership*. Albert Shanker Institute.

Elstad, E., Lysa, I., Baklien, R., Standal, E. O., Johansen, A. G., Filmberg, ... Rimeslåttén, A. (2014). Rapport til drøfting i GNIST-partnerskapet. (Report to the GNIST partnership) GNIST, Oslo. (ISBN 978-82-999655-0-7). <http://udirbeta.udir.no/wp-content/uploads/2015/11/rapport-fra-arbeidsgruppa.pdf>

Elstad, E., Lysa, I., Standal, E. O., Buan, A., Hvidsten Dahl, S., Trudeng, M., & Furulund, M. (2015a). Forslag til hvordan et system for lærervurdering kan utformes og prøves ut i Norge (How can a system for teacher evaluation be designed and tried out in Norway?). GNIST, Oslo. (ISBN 978-82-999655-0-7).

Elstad, E., Lejonberg, E., & Christophersen, K. A. (2015b). Teaching evaluation as a contested practice: Teacher resistance to teaching evaluation schemes in Norway. *Education Inquiry*, 6(4), 375-399.

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 of Educational Research Online*, 9 (1), 99-117. <http://www.j-e-r-o.com/index.php/jero/article/view/735>

Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher* 43(2), 100-107.

- Flores, M. A. (2012). The implementation of a new policy on teacher appraisal in Portugal: how do teachers experience it at school? *Educational Assessment, Evaluation and Accountability*, 24(4), 351-368.
- Gardner, H. K., & Valentine, M. A. (2014). Instrumental collaboration: Why autonomous professionals collaborate and how they benefit. *Harvard Law School Working Paper*.p 1-54
- Goe, L., Biggers, K., & Croft, A. (2012). *Linking Teacher Evaluation to Professional Development: Focusing on Improving Teaching and Learning. Research & Policy Brief*. Washington DC: National Comprehensive Center for Teacher Quality.
- Goldie, J. (2013). Assessment of professionalism: A consolidation of current thinking. *Medical Teacher*, 35(2), 952956, doi:10.3109/0142159X.2012.714888.
- GNIST (2009) GNIST – partnership for a coherent and comprehensive effort for teachers. Oslo: Norwegian Ministry of Education and Research.
- Habermas, J. (1984). *The Theory of Communicative Action. Vol 2*. Cambridge MA. MIT Press.
- 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.
- Hanushek, E. (1971). Teacher characteristics and gains in student achievement; estimation using micro data. *American Economic Review*, 61, 280–288.
- Hanushek, E. (1999). The Evidence on Class Size In S. E. Mayer & P. E. Peterson (Eds.), *Earning and Learning: How Schools Matter*. Washington. D.C.: Brookings Institution, 131–168.
- Hanushek, E. A., & Raymond, M. E. (2005). Does School Accountability Lead to Improved Student Performance? *Journal of Policy Analysis and Management* 24, 297–327.
- Head, B. W. (2008). Wicked Problems in Public Policy. *Public Policy*, 3(2), 101–118.
- Head, B. W., & Alford, J. (2015). Wicked Problems: implications for public policy and management. *Administration & Society*, 47(6), 711–739.
- Hudson, C. (2011). Evaluation – the (not so) softly-softly approach to governance and its consequences for compulsory education in the Nordic countries. *Education Inquiry*, 2(4). 671-687.
- Hursh, D. (2007). Assessing No Child Left Behind and the rise of neoliberal education policies. *American educational research journal*, 44(3), 493-518.
- Ingvarson, L. (2005). *Getting professional development right*. Australian Council for Educational Research. Retrieved from http://research.acer.edu.au/professional_dev/4

- Jordan, M. E., Kleinsasser, R. C., & Roe, M. F. (2014). Wicked problems: Inescapable wickedness. *Journal of Education for Teaching*, 40(4), 415-430.
- Kettl, D. F. (2015). Syncing the instruments and missions of government: Re-thinking the roots of decay. *Asia Pacific Journal of Public Administration*, 37(4), 247–252.
- Kindergarten Act (2005). Retrieved from <https://lovdata.no/dokument/NL/lov/2005-06-17-64>
- Kraft, M. A., & Gilmour, A. F. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly*, 52(5), 711- 753.
- 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.
- Lillejord, S., & Dysthe, O. (2008). Productive learning practice—a theoretical discussion based on two cases. *Journal of Education and Work*, 21(1), 75-89.
- Lillejord, S., Børte, K., Ruud, E., Hauge, T. E., Hopfenbeck, T. N., Tolo, A., Fischer-Griffiths, P. & Smeby, J.-C. (2014). *Former for lærervurdering som kan ha positiv innvirkning på skolens kvalitet: En systematisk kunnskapsoversikt*. Oslo: Kunnskapssenter for utdanning. www.kunnskapssenter.no
- Lindbeck, A. (1997). The Swedish experiment. *Journal of Economic Literature*, 3, 1273–1319.
- Livingston, K., & Hutchinson, C. (2016). Developing teachers' capacities in assessment through career-long professional learning. *Assessment in Education: Principles, Policy & Practice*, 24(2), 290-307.
- Lundgren, U. P. (2011). PISA as a political instrument. In M. A. Pereyra, H. G. Kotthoff, & R. Cowen (Eds). *Pisa Under Examination. Changing Knowledge, Changing Tests, and Changing Schools* Dordrecht: Sense Publishers, 17–30.
- March, J. G. (1984). How we talk and how we act: Administrative theory and administrative life, in T. J. Sergiovanni & J. E. Corbally (eds.), *Leadership and Organizational Cultures*. Urbana: University of Illinois Press, 18-35.
- McDonnell, L. M., & Weatherford, M. S. (2016). Recognizing the political in implementation research. *Educational Researcher*, 45(4), 233-242.
- McMahon, S., & Jones, I. (2015). A comparative judgement approach to teacher assessment. *Assessment in Education: Principles, Policy & Practice*, 22(3), 368-389.
- Mehta, J. (2013). From bureaucracy to profession: Remaking the educational sector for the twenty-first century. *Harvard Educational Review*, 83(3), 463–488.
- Merton, R. K. (1976). *Sociological ambivalence and other essays*. N.Y: The Free Press. Macmillan Publishing.

- Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American journal of sociology*, 83(2), 340-363.
- Moene, K. O., & Wallerstein, M. (2002). *Social democracy as a development strategy*. Princeton, NJ: Princeton University Press.
- Munneke, L., Andriessen, J., Kanselaar, G., & Kirschner, P. (2007). Supporting interactive argumentation: Influence of representational tools on discussing a wicked problem. *Computers in Human Behavior*, 23(3), 1072–1088.
- Murphy, J., Hallinger, P., & Heck, R. H. (2013). Leading via teacher evaluation: The case of the missing clothes?. *Educational Researcher*, 42(6), 349–354.
- National Commission on Excellence in Education. (1983). *A nation at risk: A report to the nation and the Secretary of Education*. Washington, DC: U.S. Department of Education.
- Nusche, D., Earl, L., Maxwell, W., & Schrewbridge, C. (2011). *OECD Reviews of Evaluation and Assessment in Education. Norway*. Paris: OECD. Retrieved from <https://www.oecd.org/norway/48632032.pdf>
- OECD (1987). *Reviews of national policies for education. Norway. Draft of February 1988* (Copy from the National Library of Norway, Depot).
- OECD (2009). *Teacher Evaluation. A Conceptual Framework and examples of Country Practices*. OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes <http://www.oecd.org/edu/school/44568106.pdf>
- O'Neill, O. (2002). *A question of trust: The BBC Reith Lectures 2002*. Cambridge University Press.
- Ozga, J., Dahler-Larsen, P., Segerholm, C., & Simola, H. (Eds.). (2011). *Fabricating quality in education: Data and governance in Europe*. Routledge.
- Popham, W. (1988). The dysfunctional marriage of formative and summative teacher evaluation. *Journal of Personnel Evaluation in Education*, 1(3), 269–273.
- Popkewitz, T. S. (2000) "Globalization/Regionalization, Knowledge, and the Educational Practices: Some Notes on Comparative Strategies for Educational Research," in T. S. Popkewitz (Ed.) *Educational Knowledge: Changing Relationships between the State, Civil Society, and the Educational Community*, p. 3-27. Albany, NY: State University of New York Press.
- Pupala, B., Kascak, O., & Tesar, M. (2016). Learning how to do up buttons: Professionalism, teacher identity and bureaucratic subjectivities in early years settings. *Policy Futures in Education*, 14(6), 655-665.
- Ramirez, F. O., & Boli, J. (1987). The political construction of mass schooling: European origins and worldwide institutionalization. *Sociology of education*, 2-17.

- Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4(2), 155–169.
- Sachs, J. (2016). Teacher professionalism: why are we still talking about it? *Teachers and Teaching*, 22(4), 413-425.
- Sahlberg, P. (2010). Rethinking accountability in a knowledge society. *Journal of Educational Change*, 11(1), 45-61.
- Santiago, P., Benavides, F., Danielson, C., Goe, L. and Nusche, D. (2013): *Teacher Evaluation in Chile. Main conclusions*. OECD Reviews of Evaluation and Assessment in Education.
- Schelling, T. C. (1980). *The Strategy of Conflict*. Boston, MA: Harvard University Press.
- Senge, P. (1990). *The fifth discipline: The art and science of the learning organization*. New York: Currency Doubleday.
- Shulha, L. M., Whitmore, E., Cousins, J. B., Gilbert, N., & al Hudib, H. (2016). Introducing evidence-based principles to guide collaborative approaches to evaluation: Results of an empirical process. *American Journal of Evaluation*, 37(2), 193-215.
- Simons, P. R. J., & Ruijters, M. C. (2014). The real professional is a learning professional. In *International Handbook of Research in Professional and Practice-based Learning* (pp. 955-985). Springer Netherlands.
- Smagorinsky, P. (2014). Authentic teacher evaluation: A two-tiered proposal for formative and summative assessment. *English Education*, 46(2), 165-185.
- Smylie, M. A. (2014). Teacher Evaluation and the Problem of Professional Development. *Mid-Western Educational Researcher*, 26(2), 97-111
- Smylie, M. A., Murphy, J., & Louis, K. S. (2016). Caring School Leadership: A MultiDisciplinary, Cross-Occupational Model. *American Journal of Education*, 123(1), 1-35.
- Spillane, J. P., & Diamond, J. B. (Eds) (2007). *Distributed Leadership in Practice* N. Y.: Teacher College Press.
- Stranden, K. (2017). *Profesjonsutvikling i skolen. Rapport fra et utviklingsprosjekt initiert av Utdanningsdirektoratet* (Professional development in school. Report from a developmental project initiated by the Directorate for Education and Training) <https://www.udir.no/tall-og-forskning/finn-forskning/rapporter/profesjonsutvikling-i-skolen/>
- Stroupe, D. (2016). Beginning Teachers' Use of Resources to Enact and Learn from Ambitious Instruction. *Cognition and Instruction*, 34(1), 51-77.
- Taut, S., & Sun, Y. (2014). The development and implementation of a national, standards-based, multi-method teacher performance assessment system in Chile. *Education policy analysis archives*, 22 (71), 1-33.

- Tschannen-Moran, M. (2009). Fostering teacher professionalism in schools: The role of leadership orientation and trust. *Educational Administration Quarterly*, 45(2), 217-247.
- Tyack, D. B. (1974). *The One Best System*. Cambridge, MA: Harvard University Press.
- Tyack, D. & Hansot, E. (1982). *Managers of virtue: Public school leadership in America 1820-1980*. New York. Basic Books.
- Weber, M. (1947). *The Theory of Social and Economic Organization*. Translated by A. M. Henderson and Talcott Parsons. N. Y.: Macmillan Publishing.
- Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21(1) 1-19.
- White Paper (Stortingsmelding) no. 37 (1990-91). *Om organisering og styring i utdanningssektoren* (Organization and steering in the education sector).
- Zhang, X. F., & Ng, H. M. (2011). A case study of teacher appraisal in Shanghai, China: In relation to teacher professional development. *Asia Pacific Education Review*, 12(4), 569-580.