Teaching evaluation as a contested practice: Teacher resistance to teaching evaluation schemes in Norway

Eyvind Elstad*, Eli Lejonberg* & Knut-Andreas Christophersen**

Abstract
Norwegian education authorities pursue the objective that students in upper secondary schools should evaluate teaching. This kind of evaluation scheme has, however, elicited resistance from teachers. The aims of this article are to explain the advent of teaching evaluation policies in Norway and to explore which factors explain teacher resistance to teaching evaluation schemes. Structural equation modelling of teacher survey data shows which factors are statistically associated with the concept of teacher resistance to teaching evaluation. Teacher stress and resistance are positively associated with the perceived controlling purpose of teaching evaluation rather than the evaluation itself. Teacher resistance is clearly negatively associated with their acknowledgment of students’ feedback and their perception of the communication with leadership. Further, the model shows moderately strong negative pathways between the perceived control purposes of teaching evaluation and acknowledgement of students’ feedback on the one hand and the perceived control purposes of teaching evaluation and perception of clear communication of goals on the other hand. Teachers’ acknowledgement of the students’ feedback is weakly negatively associated with teacher stress. The implications for practice and further research are discussed.

Keywords: teacher evaluation, teacher appraisal, teaching evaluation, teacher stress, leadership, education policy, Norway

Introduction
Teacher evaluation is at the centre of current education policies in many countries and is part of an international trend in which different variants of the evaluation of teachers’ work have been implemented via initiatives from education authorities in a number of countries since the start of this millennium (Isoré 2009; Flores 2010). This is part of a global edu-business (Ball 2012). Teacher evaluation has very different features. On the one end of the scale there is a mandatory, high-stakes national teacher evaluation system (Tornero and Taut 2010; Taut et al. 2011). On the other end of the scale there is development-oriented teacher evaluation. Teacher evaluation schemes unfold differently across several countries. Portugal has established a

*Department of Teacher Education and School Research, University of Oslo, Norway.
**Department of Political Science, University of Oslo, Norway.

Correspondence to: Eyvind Elstad, Department of Teacher Education and School Research, University of Oslo, Norway. Email: eyvind.elstad@ils.uio.no

Citation: Education Inquiry (EDUI) 2015, 6, 27850, http://dx.doi.org/10.3402/edui.v6.27850
national teacher performance appraisal system (Flores 2010). A similar scheme was introduced in the Flemish Community of Belgium in 2007 (Delvaux et al. 2013). Teaching evaluation is also an emerging phenomenon in Asia. Schemes are implemented in Singapore (Ong Kelly, Yun, Ling Chong and Sheng Hu 2008), some parts of China (Liu and Zhao 2013), for instance Hong Kong, Shanghai and Beijing (Sun et al. 2012), the U.S.A. (Ellett and Teddlie 2003), New Zealand (Fitzgerald, Youngs and Grootenboer 2003) and Japan (Katsuno 2010). Several states in the U.S.A. have implemented teacher evaluation systems. The Organisation for Economic Co-operation and Development (OECD) recommends teacher appraisal as part of a standard policy package (OECD 2009). This recommendation was also submitted to Norwegian education authorities (OECD 2011). Teaching evaluation is one of many possible instruments that can be used to provide feedback to teachers with a view to improving teaching practices.1

Teacher evaluation is a highly controversial practice among teachers (e.g. Beerens 2000; Peterson and Peterson 2006; Isoré 2009; Flores 2010). Resistance among teachers is observed (Monyati, Steyn and Kamper 2006) and “the literature indicates that the success of evaluation and accountability systems depends ... on the perceptions held by those individuals affected by the evaluation, and on their acceptance of the system” (Tornero and Taut 2010, 133). In the Norwegian context, teachers’ critical voices related to the ongoing implementation of evaluation schemes have been raised in the media (e.g. Jambak 2015).

A new phenomenon is privately-led online evaluations of school teachers where everyone can submit ratings and reviews (Hinz 2011). Submissions are anonymous, and the public online ratings of teaching quality can be read by everyone. Both private online evaluations and evaluations led by education authorities have met resistance. This resistance to teaching appraisal schemes may slow down the fulfilment of the potential of teacher appraisal schemes, and it is important to explore the antecedents of teacher resistance to better understand how a teaching evaluation scheme should be improved. However, there is general doubt as to whether teacher evaluation can contribute to teachers’ professional development (Tuytens and Devos 2009, 2014).

The purpose of this article is twofold: firstly, to explain the advent of teaching evaluation policies in Norway and, secondly, to explore statistical associations with the teachers’ perception of control via teaching evaluation, their perception of the school management’s goals, and their acknowledgment of the feedback provided by students with regard to their own teaching as antecedents of resistance. Stress related to teaching evaluation schemes is the mediating variable, and resistance to the teaching evaluation scheme is the dependent variable. We will use the term “teaching evaluation” to refer to upper secondary students’ anonymous evaluation of their teachers’ educational practice via the local education authority’s evaluation scheme.
The advent of teaching evaluation policies in Norway

Norway is a latecomer to national teacher evaluation policy. In this part of the article, we aim to explain why. A curious phenomenon is that the students’ association was the instigator of teacher evaluation practice in Norway. The evaluation of teachers by students with the aid of surveys is a phenomenon that can be justified with reference to four different visions: effective administration of the school sector, professionalisation of the teaching vocation, co-determination by students in issues that pertain to them, and effective and quality-enhancing quasi-market competition (Olsen 2007; Elstad 2013). All of these ideal types will have an effect on how teaching evaluation can be used and justified in various contexts. However, the argumentation in favour of teaching evaluation in the Norwegian context has passed through a number of stages: first came the argument related to student co-determination, followed by administrative monitoring of the teachers’ work, then by information portals that provide public access to quality indicators of teaching at the school level and, finally, the view of teaching evaluation as a tool for professional development among teachers (Elstad 2009; Lien 2009).

These four ideal types are combined in arguments related to the different practices in which teaching evaluation occurs, although the relative weight of these four concerns may shift over time, and their balance may vary according to the particular context. Their interrelations are highly complex, and the actual significance that teaching evaluation may have will depend on contextual factors. For example, there is a difference between teaching quality which depends on factors such as positive student contributions to the teaching process, the external framework for the teacher’s work and the teacher’s contribution to the teaching process and the quality of the teacher as such (Darling-Hammond 2013). Teaching evaluation as a practice may also influence the perception of roles in the teaching process. For example, teaching evaluation may modify the students’ perception of their role from believing that their learning outcomes largely depend on their own efforts to seeing these outcomes as depending on the teachers’ delivery of quality (Garmannslund, Elstad and Langfeldt 2008).

Teaching evaluation is practised in a number of Norwegian schools and municipalities, but up until 2013 it was not subject to national government initiatives. At the time of writing (May 2015), national policies were being prepared. The political platform for a government to be formed by the Conservative Party and the Progress Party (2013–2017) states that “the Government will permit students at the upper secondary level to evaluate teaching” (The Norwegian Government 2013). Yet the implementation of this political ambition will not come in the form of a governmental decree but as a negotiated solution in which the parties involved will have some influence over a trial scheme for teaching evaluation. For teaching evaluation to have beneficial effects, it is essential to understand the factors that induce teacher resistance.
In recent years, the teaching vocation has been referred to as a true profession by several trade unions, employers and Norwegian education authorities (Lejonberg and Elstad 2014). From a perspective that emphasises the teachers’ professional development, teaching evaluation can be included as an information element for reflection that promotes personal growth and learning (Darling-Hammond et al. 1983; Day et al. 2007). However, the proposal to introduce teaching evaluation in schools has been met with resistance (Garmannslund, Elstad and Langfeldt 2008), and the way in which teaching evaluation may promote professional development remains controversial (Tuytens and Devos 2011, 2014).

In the late 1980s, new policies for management of the school sector were implemented in Norway. Management by objectives was introduced as an innovative idea in the design of government policies for the school sector (Christensen and Lægreid 2011). Other fundamental changes were made in the early 1990s, e.g. by imposing statutory responsibility for school quality on the municipalities and counties (Hovdenak 2000). These changes included the decentralisation of decision-making authority and responsibility and were recommended by the OECD in 1988. Nevertheless, Norwegian authorities chose not to introduce national tests, or a transparent collation of measurements of learning outcomes at the school level (Jarning 2010; OECD 1988). Such measures were first introduced after the start of the new millennium. At that time, the centre-conservative government (2001–2005) established legitimacy for significant amendments to educational policy in the wake of disappointing PISA results. Norwegian educational policy took a new, clearly results-oriented turn (Elstad 2010). In 2004, national tests became a powerful instrument for drawing more attention to the students’ documented learning outcomes among teachers and school leaders. Despite the initial resistance, this implementation has become an accepted blueprint among people in general, teachers and politicians. Notwithstanding this, the Norwegian school system remains different from those of several other European countries, e.g. in the fact that students may not be streamed, that students are individually entitled to adapted teaching and that students are not sorted into tracks until they reach the age of 15. The idea of the ‘knowledge school’ is prominent in political debate, although certain ideas that have their origin in progressive educational theory still remain valid as principles (Telhaug, Mediås and Aasen 2006; Elstad, Turmo and Guttersrud 2011).

Over time, ideas of strongly centralised governance have alternated with the decentralisation of decision-making authority. For a long period, Norwegian schools have been subject to government control of teaching content through national curricula and a nationally managed examination that, it is claimed, ensure a national educational standard (Engeland and Langfeldt 2009). However, the municipalities and counties have been charged with the responsibility for policy enactment and the quality of education (since 1992). The Norwegian parliament has also introduced statutory provisions that require municipalities and counties to have systems in
place that are sufficient to ensure quality in the teaching provided to students (The Education Act 1998). In 2004, the responsibility for the remuneration of school employees was transferred to the municipalities. Thus, considerable decision-making authority was transferred from the government to the municipalities and counties (Engeland 2000). The previous centre-conservative government (2001–2005) delegated responsibility for quality assurance of schooling to the municipalities (with the argument that local challenges are best addressed at the local level).

Monitoring of the quality of education in 2006 and 2007 showed, however, that the municipalities and counties were varying considerably with regard to how they were choosing to handle their responsibilities (The Directorate of Education and Training 2006). However, decentralising the governance of education systems often carries the seeds of their own contradictions (Weiler 1990). The municipalities have chosen widely different methods for quality assurance of the teaching provided to students. The monitoring has occasionally revealed censurable conditions, e.g. in the manner in which the municipalities have established systems for quality assurance and systemic innovations. The Red-Green government (2005–2013) returned to stricter governmental control of conditions that have an effect on school quality.

In his 2008 New Year’s Address to the Norwegian people, former Prime Minister Jens Stoltenberg spoke of this development as “...a serious warning. School should be a place of learning and of fundamental knowledge of reading, writing and arithmetic. Teachers should have a clear responsibility for what students learn in school” (Stoltenberg 2008; Christophersen, Elstad and Turmo 2010). The return to governmental control was based on a review of the quality systems that revealed considerable shortcomings in the ways the municipalities and counties had attended to their responsibilities (The Directorate 2006). At the same time, large-scale international studies claim to reveal a disappointing sideways shift in the Norwegian performance scores (Kjærnsli and Olsen 2013). Norwegian political authorities harbour greater ambitions than students to score around the international average: “There is no reason why Norway should not have one of the world’s best education systems” (GNIST 2009). The central governmental level in the management structure of the Norwegian education system has therefore returned with a stronger desire to be a key initiator of new measures and system-wide innovation (for instance, teacher appraisal). The Ministry of Education and Research (2014) argues that, since the teacher is on the frontline in schools’ efforts to improve students’ learning outcomes, the focus should be on the teachers’ work.

**Student empowerment as an instigator of teaching evaluation**

Initiatives for anonymous teaching evaluation in Norway were launched by the student association The School Student Union of Norway (2003).² This association had argued that teaching evaluation should not be undertaken by the teacher alone...
but should take the form of a scheme in which the school management and students have access to the results. The association’s argument emphasised this opportunity to express opinions on teaching practices as a democratic right in school society and as an instrument to counterbalance the asymmetrical relationship between teachers and students (The School Student Union 2003). The vision underlying this idea is that the students should have the right to co-determination in issues that pertain to them in their school life. This vision has manifested as a political goal embedded in the programmes of several political parties for the parliamentary period 2013–2017 as well as in the political platform of the incumbent Norwegian government (The Norwegian Government 2013). A government initiative was undertaken to ensure that teaching evaluation is implemented in Norwegian schools. However, this initiative elicited resistance from teachers. It is therefore essential to investigate which factors are associated with teacher stress related to evaluation as well as to their resistance to being evaluated.

Teaching evaluation serves not only the democratically justified purpose of promoting student co-determination but also has a wider objective. The results of teaching evaluation may also include administrative monitoring of the teachers’ work, and when the results measured at the school level are collated in Internet portals (and/or reported in the media), student satisfaction with teaching is turned into a quality indicator that can be used as an argument in support of the quality of schools and to monitor quality development. The results of teaching evaluation can also be used to hold teachers accountable or as a basis for development goals. Quality measurements can be used, and have been used, to market particular schools to present and future students and their parents. Measurements of satisfaction can thus help engender competition between schools, and some have claimed that this type of competition gives rise to quality-enhancing processes within schools (Chubb and Moe 1991). There are examples of Norwegian municipalities that collate quality measurements and publish them at school level. Thus, there is an assumed linkage between the desire for adequate and effective administration of the school sector on one hand and the idea of the beneficial effects of competition on quality development on the other.

From educational government to policy networks

The national authorities’ renewed initiative (2009) is designed as a policy network: a ‘partnership for holistic teacher improvement’, christened GNIST (‘Spark’). The key stakeholders of the programme include the Ministry of Education and Research, which is the driving force in this partnership, the teaching and school management trade unions and the interest organisation for local authorities. The parties are engaged in so-called ‘national collaboration’ to achieve the goal of “strengthening the schools’ academic platform and the prestige of the teaching profession, and thus contribute to adequate recruitment and professionalism” (GNIST 2009, 3).
This partnership scheme can be seen in the context of the relatively strong role of consensus-seeking negotiation between the social partners in the Scandinavian countries, called the Scandinavian negotiation model (Barth, Moene and Willumsen 2014). The GNIST partnership is organised at the central level but also has ramifications at the regional and local levels (GNIST 2009). The central government agencies have assumed a leading role in the initiatives for new programmes, while the interest organisations have also participated in deliberations and efforts to establish consensus. However, in February 2014, one of the teachers’ associations resigned from the partnership after declaring that it had lost confidence in the interest organisation of the local authorities. This demonstrates that the expressed intention of negotiation-based consensus in these matters can be difficult to achieve.

The partnership model is based on a continuous negotiation process that relies on mutual trust, in which the parties are to seek compromises in areas where their interests may be contradictory as well as partly overlap.

One of the initiatives included in the partnership’s efforts for quality development in schools is the evaluation of teaching. In Norwegian schools, this evaluation has had a weak and delayed entry as an instrument for quality assurance. Some teacher-initiated and school-initiated trials of letting students evaluate teaching have been undertaken since 2006, and some municipal initiatives for the introduction of systematic schemes for the evaluation of teachers have been implemented since 2008. However, the schemes have been shown to vary considerably among municipalities and counties and, in some places, they are non-existent. This area has been devoid of any national governance. The OECD’s country-specific evaluations have advised Norway to implement a national framework for “teacher appraisal” schemes with a developmental orientation (OECD 2011). Against this background, the GNIST partnership identified the evaluation of teachers as one of its key initiatives, and it proposed trial programmes under the auspices of the partnership (GNIST 2014).

In the autumn of 2013, the GNIST partnership appointed a working group charged with the task of preparing a report on the preconditions for a teacher appraisal scheme (GNIST 2013). The parties agreed on the following preconditions: the cooperation between the employees and the employers must be based on mutual trust; the appraisal scheme must be systematic, predictable and practically feasible; specific, immediate feedback is deemed better than general and delayed feedback, and teacher appraisal should focus on development and not serve to monitor the teachers’ work (GNIST 2009). Moreover, the teacher appraisal scheme is supposed to be based on research evidence. This working group (of which one of the authors was a member) proposed a teacher appraisal scheme based on: (1) students’ evaluations of teaching; (2) observation of teaching by school leaders or colleagues with subsequent feedback; and (3) learning outcomes measured via national tests.
and exams. The working group rejected measurements of ‘value added’ as an instrument for teacher appraisal (GNIST 2015).

Following an initiative by the teachers’ trade unions, the working group chose to concentrate on an appraisal scheme that had been implemented in the Akershus and Vestfold counties, and which the teachers’ trade union had deemed acceptable. Since 2010, these counties have practised a scheme that involves an anonymous survey among students (see Appendix 2). A follow-up conversation with a supervisor is part of the teacher appraisal system in Akershus County. The teachers’ trade unions have been involved in negotiations regarding the practical application of this scheme. The scheme operated by Akershus County is denoted as a consensual model in the GNIST partnership for how an anonymous survey of satisfaction with teaching etc. should operate. It is therefore interesting to investigate which factors are empirically associated with teacher resistance to teaching evaluation schemes in which the students declare their degree of satisfaction with the teaching, and to investigate which factors are empirically associated with teachers’ stress related to such an evaluation.

**Theoretical framework and hypotheses**

Teacher resistance to the teaching evaluation scheme induced by the local education authorities is the dependent variable in our theoretical model. A teaching evaluation exposes a teacher’s potential vulnerability. Some teachers perceive critical feedback from students as a problem. A possible psychological mechanism could be cognitive dissonance: when teaching evaluation cannot produce any good results, the scheme itself is disparaged. Teaching evaluation practice may trigger teacher stress and this stress is often used as an argument against the implementation of teaching evaluation systems (Kelly, Ang, Chong and Hu 2008; Flores 2012). There is a recognised link between teaching evaluation schemes and teacher stress (Austin, Shah and Muncer 2005; Kyriacou and Sutcliffe 1978; Litt and Turk 1985); therefore, we hypothesise a relationship between stress and teacher resistance.

Hypothesis 1: Stress related to teaching evaluation is associated with teacher resistance to teaching evaluation.

Even if the evaluation is not used as a method of control, teachers might nevertheless perceive it as a control measure, which might have a negative effect on their effort and ability to derive useful information from the evaluation. This means that the teachers’ perceptions of the ways in which management exercises its role may have an impact on whether the teachers regard the scheme as useful for their own development or merely as a method of controlling them. Indeed, these perceptions can be combined (Isoré 2009; Katsuno 2010). We hypothesise that teachers’ perceptions of the controlling purpose of evaluation schemes induce stress among teachers.

Hypothesis 2a: The control purpose is associated with stress.

Hypothesis 2b: The stronger the teachers experience the control aspect the more they will resist being evaluated.
Students in upper secondary schools may have mature as well as immature assessments of their teachers’ educational practices. Teachers who have a good relationship with their students may have more confidence in their students’ judgements of teaching quality than teachers who have a less positive relationship. Good relationships between teachers and students are regarded as a significant feature of the quality of schools. In our context, we expect that the teachers’ acknowledgment of the students’ feedback is negatively associated with teacher resistance.

Hypothesis 3a: Students’ feedback is associated with stress.

Hypothesis 3b: The more the teachers trust the feedback, the less reluctant they are to be evaluated.

Teaching evaluation is implemented by school management and followed up with performance reviews. Teaching evaluation is thus the responsibility of the school management. The typical procedure is that the results from teaching evaluations are analysed in a performance review where a middle manager or the principal discusses the average results with the teacher in question. School leaders may influence teachers’ perceptions of teacher evaluation policy (Tuytens and Devos 2009, 2010; Derrington 2014; Timperley and Robinson 1997). This means that the teachers’ perception of the way in which the management exercises the communication of school goals may have an influence on teacher resistance. The management’s communication of what they expect from the teacher in the context of the teaching evaluation scheme will generate teacher resistance. Our fourth hypothesis is that the communication with the leadership is associated with teacher stress and resistance.

Hypothesis 4a: The teachers’ perception of the way in which the management exercises the communication of school goals is associated with teachers’ stress.

Hypothesis 4b: The more useful teachers perceive communication with the school management to be, the less reluctant they are to be evaluated.

Based on these hypotheses, we draw the following theoretical model (Fig. 1):

![Figure 1. The hypothesised research model.](image)
Sample

The GNIST working group that is to propose a trial scheme for teacher appraisal focused on the way teaching evaluation has been implemented in upper secondary schools in Akershus County. This approach, which includes anonymous surveys among students and follow-ups through individual performance reviews and group sessions for each discipline, is regarded as an appropriate practice and thus as an interesting model for further trials and pilot schemes under the auspices of the GNIST partnership. The relevant teachers’ unions have endorsed the implementation of the scheme in Akershus County. We selected five upper secondary schools with general studies programmes spread out over the entire county. Students in these schools perform at approximately the national average in key disciplines and to some extent better than the national average. The average age among the teachers is somewhat higher than the average of all teachers in general studies programmes, while the gender distribution is approximately equal to that of the reference population: all teachers in general studies programmes in Akershus County. All five of the schools included in the sample are regarded as well-established and well-managed, with a workforce that has remained relatively stable over several years. The empirical patterns that emerge from our study may thus provide some insight into statistical associations between the dependent variable and the independent variables in these types of schools. We also believe that these patterns are valid beyond our actual sample.

The study was implemented by one of the authors of this article attending the teachers’ joint meetings. The main features of the study were explained, and the teachers were informed that participation was voluntary. All teachers (n = 268) who had been through the teaching evaluation process were invited to participate. None of the teachers exercised their right to refuse to participate, and the response rate was therefore 89.33% (n = 255) in four of the five schools. In one school, an extraordinary situation occurred that obligated some of the personnel to attend an alternative arrangement on the day the survey was scheduled to take place. As a result, only 30.23% of the permanent teaching staff (n = 13) of that school participated. However, we have no reason to believe that this situation caused a sampling bias. The teachers completed a paper-based questionnaire by ticking off pre-determined response alternatives. The teachers entered no information that could reveal their identity, and the survey is thus fully anonymised. We may therefore assume that the teachers have entered truthful and well-considered judgments in completing the response categories and that the sample is representative of teachers in Akershus County.

Measurement instruments

A questionnaire (Appendix 1) was constructed based on measurement instruments previously reported in the literature, and on new developments. In the survey, the students responded to items on a seven-point Likert scale, where the alternative ‘four’ represented a neutral midpoint. The concepts were measured with two to three single items. The analysis reported below is based on seven measurement instruments.
The internal consistency (Cronbach’s alpha) for each concept is satisfactory; Cronbach’s alpha is between .78 and .90.

**Analysis**

Structural equation modelling (SEM) was used to analyse the relationships between the variables. Assessments of fit between the model and data are based on the following indices: root mean square error of approximation (RMSEA), normed fit index (NFI), goodness-of-fit index (GFI) and comparative fit index (CFI). RMSEA < .05 and NFI, GFI and CFI > .95 indicate a good fit and RMSEA < .08 and NFI, GFI and CFI > .90 indicate an acceptable fit (Kline 2005).

After excluding respondents with missing values, all parameters were estimated on 206 respondents. The values of RMSEA, NFI, GFI and CFI indicate that the structural model in Fig. 2 has an acceptable fit. Ellipses represent the latent variables, circles

![Diagram of the structural model](image)

**Figure 2. Estimated model (N = 206).** TR = teacher resistance against teaching evaluation schemes; CC = Communication with the leadership; PCP = Perceived control purposes of teaching evaluation (abbreviated PCP), STR = stress induced by teaching evaluation and TS = Acknowledgement of the feedback from students.
represent measurement errors and rectangles represent the observed measured variables. The structural model consists of terms with paths (arrows) between them. The path arrows indicate theoretical common causes and the figures (standardised regression coefficients) reflect the measured strength of the connections. The strength increases with the numerical value. The measurement and structural models were analysed with IBM SPSS Amos 22.

Results
The structural equation model shows the direct effect (the arrows) between the variables. The analysis shows:

- Hypothesis 1: Stress is not significantly (p > .05) associated with teacher resistance to teaching evaluation (b_{STR \rightarrow TR} = -0.07).
- Hypothesis 2a: The direct effect of the control purposes on teacher stress is significant (p < .05), strong and positive (b_{PCP \rightarrow STR} = 0.39). The higher the level of perceived control purposes the teachers report, the more stressed they report they are.
- Hypothesis 2b: The direct effect of the control purposes on teacher resistance is significant (p < .05), moderate and positive (b_{PCP \rightarrow TR} = 0.22). The higher the level of perceived control purposes the teachers report, the more resistant they seem to be.
- Hypothesis 3a: The direct effect of the teachers’ acknowledgement of the students’ feedback on teacher stress is significant (p < .05), moderate and negative (b_{TS \rightarrow STR} = -0.19). The more the teachers trust the feedback, the less stressed they report they are.
- Hypothesis 3b: The direct effect of the teachers’ acknowledgement of the students’ feedback on teacher resistance is significant (p < .05), strong and negative (b_{TS \rightarrow TR} = -0.49). The more the teachers trust the feedback, the less reluctant they are to be evaluated.
- Hypothesis 4a: Perceived clear communication with the leadership is not significantly (p > .05) associated with teacher stress (b_{CC \rightarrow STR} = -0.03).
- Hypothesis 4b: The direct effect of the perceived clear communication with the leadership on teacher resistance is significant (p < .05), moderate and negative (b_{CC \rightarrow TR} = -0.24). The clearer teachers perceive the communication with the school management to be, the less reluctant they are to be evaluated.

Discussion and implications for practice
Norwegian education authorities pursue the objective that students in upper secondary schools should evaluate the teaching (The Norwegian Government 2013). The acceptance of the evaluation scheme is key to any beneficial effects of
this system-wide innovation (Tornero and Taut 2010). The teaching evaluation schemes have, however, elicited resistance from teachers in Norway as well as other countries. Critical issues concerning teacher evaluation have arisen worldwide (Isoré 2009; Flores 2010, 2012) and there is widespread recognition of the fact that contextual factors have an effect on teachers’ attitudes to the implementation of a teaching evaluation scheme. Teachers’ perceptions of teacher evaluation, the role of school culture and leadership and the context of teaching as a profession are examples of controversial issues which touch on power relations. We consider the context as a mediating factor in the policy enactment work done in schools (Ball, Maguire and Braun 2012, 40). Policy enactment involves the “translation of texts into action and the abstraction of policy ideas into contextualised practices” (Ball et al. 2012, 3). This means that the management of change is crucial to the success of systemic innovation and shows that individual school leadership and culture might be important factors in teacher trust, perceptions of the usefulness of student feedback, and relative perceptions of stress (Derrington 2014; Sun et al. 2012; Timperley and Robinson 1997). Further, the systemic innovation of teacher appraisal was decided at the meso level (Akershus County Council), and this meso-level institution decided to publish the school mean scores. Publications of school mean scores induce the public’s perception of pupils’ consumer satisfaction as an important quality indicator (Ball 2012). The vision for quality-enhancing competition based on the transparency of quality indicators partially contradicts the vision of teacher appraisal being a tool for professional development. This and other contradictions are critical issues for successful leadership, and strategic decisions within the superintendency and school management might easily become an exercise in strategic navigation through difficult waters.

It will be an advantage if the model chosen for policy enactment of a teaching evaluation scheme enjoys acceptance among teachers. If accepted, there will be potential for such a scheme to have a positive effect on teachers’ traditional cultures of privatisation and the way in which the teachers individually and collectively relate to the feedback from their students. In the Norwegian context, such acceptance depends on a negotiated solution involving the teachers’ trade unions, employers and central education authorities (Barth, Moene and Willumsen 2014). The first purpose of this article was to explain the advent of teaching evaluation policies in Norway. In one way, it may seem odd that trade unions should be able to exert such a large influence at the cost of the authority that is normally bestowed on leadership positions in organisations. Some researchers have looked in wonderment at what is referred to as the Scandinavian negotiation model and to some extent assess it as a hindrance to effectiveness and management efficiency (Acemoglu, Robinson and Verdier 2012). Other researchers, however, have claimed that the Scandinavian negotiation model may help complex organisations function by promoting co-determination and involvement at all levels of the organisation (Brandal, Bratberg
and Thorsen 2013). The inclusion of teachers’ trade unions and the student association in these negotiations features in an age when current rhetoric in public documents on how schools can improve calls for ‘clear’ and ‘forceful’ leadership (Ministry of Education and Research 2003, 2014). This is a paradox. In a theoretical perspective that emphasises teacher empowerment (Carl 2009) and student empowerment (Seigel and Rockwood 1993), this negotiation model of governance may demonstrate the strength of the Norwegian system of involving employees in deciding how this kind of scheme should be implemented. The model of teaching evaluation explained in this article stems from a county in which teaching evaluation is practised in accordance with methods that are recognised by the teacher unions in Norway (GNIST 2015). This fact could bolster the likelihood of acceptance by the relevant trade unions. On the other hand, such negotiations involving teachers’ trade unions may be regarded as a waste of time and a hindrance to management authority (Acemoglu, Robinson and Verdier 2012). However, there is also reason to question how genuine is the influence of the teacher unions. As mentioned, one of the unions chose to back out of the joint collaborative forum due to disagreements with other actors. If teacher unions are included but not heard when disagreements occur, in such forums their presence can be said to function mainly as legitimisation of policy. If one part chooses to leave the working group due to such reasons, it seems relevant to question whether the decisions are in fact consensus-based.

The second purpose of this study was to explore the statistical associations between the teachers’ perception of control via teaching evaluation, their perception of the school management’s goals, their acknowledgment of the feedback provided by students with regard to their own teaching, the stress induced by teaching evaluation schemes, and the resistance to this kind of evaluation scheme. Structural equation modelling of the teacher survey data showed which factors were statistically associated with the concept of teacher resistance to teaching evaluation. We found that teacher stress was not associated with teacher resistance to teaching evaluation (Hypothesis 1), while perceived control purposes were clearly associated with stress (Hypothesis 2a). This is quite surprising, and further research is needed to understand the mechanisms of this phenomena. Interviews with teachers are useful for studying this topic in more depth. However, possible explanations for this finding can be that teachers are negative and resistant to evaluation schemes for reasons other than they are stressed. We can imagine teachers perceiving evaluation processes as time-consuming or as not useful and thereby not worth the effort. This would be a more rational cost-benefit analytical approach to resistance than the more intuitive affective-based understanding of resistance we initially suggested. Abrahamsen and Aas (2014) found that middle managers following up on evaluation results in Norwegian schools also find it challenging to carry out teaching evaluation processes. If the school management expresses such hesitations, it might contribute to lowering the evaluated teachers’ stress; however, it could also strengthen their
condemnation of the evaluation scheme as a waste of time. That the intention with teacher evaluation in Norway is to contribute to professional development among teachers could further support our interpretation. If the teachers are familiar with the expressed intentions, they have no reason to feel stressed by teaching evaluations. Yet, if they do not perceive that the evaluation scheme contributes to their professional development, they have good reason to respond negatively to the implementation of such schemes.

Although the declared purpose is development-oriented in our context, we also find the perceived control purposes of teaching evaluation to be positively associated with teacher stress and resistance in our material (Hypotheses 2a and 2b). Teachers may misinterpret the official policy (Datnow and Castellano 2000; Tuytens and Devos 2009). However, control purposes are embedded in the reporting practice (for instance, the average school results of teaching evaluation are published on the Internet and reported to the County Executive Board). The school management has access to information on how the students perceive the quality of their teaching, and unofficial teacher scores may exist backstage. A teacher may have grounds to fear negative evaluation results. Such perceptions of hidden control purposes of teaching evaluation might lie at the core of teacher resistance. The person being evaluated may fear that a negative teaching evaluation may have consequences over time (Garmannslund, Elstad and Langfeldt 2008). Although the principal has certain management prerogatives, a Norwegian teacher can hardly be dismissed on the basis of nothing more than low scores in a teaching evaluation. However, a teacher may fear other consequences in time, e.g. lack of a local wage promotion, negative reputation as a teacher etc. If teachers without permanent employment are included, we can imagine that their results could influence their further career. Further, a negative teaching evaluation may influence the teacher’s self-esteem (Kilburg 1980). An implication for practice should be that school principals and superintendents should communicate the non-controlling purpose of the evaluation scheme very clearly and build relational trust among the teachers when it comes to the practice evaluation.

We found that teachers’ acknowledgement of the students’ feedback is negatively associated with teachers’ stress and resistance (Hypotheses 3a and 3b). The model shows moderately strong and negative pathways between the acknowledgement of students’ feedback on one hand and the perception of clear communication of goals on the other. Further, the model shows moderately strong and negative pathways between the acknowledgement of students’ feedback on the one hand and the perceived control purposes of teaching evaluation. One interpretation is that teachers whose educational practices are perceived as very poor by the students may fear their evaluation and perceive it as control and therefore show resistance. However, this interpretation is too hasty. The confidence in the students depends on whether they are able to provide adequate assessments of their teacher’s educational practice. Students’ level of maturity with regard to their ability to
provide well-considered feedback may vary. There are studies showing that some groups of students may disparage everything associated with school (Bø and Hovdenak 2011). What we do not know is how trustworthy students are as evaluators. More research is needed on this topic.

Teachers’ perception of clear communication is not associated with stress (Hypothesis 4a). Teachers’ perception of clear communication is negatively associated with teachers’ resistance (Hypothesis 4b). Further, the model shows moderately strong and negative pathways between perceived control purposes of teaching evaluation and acknowledgement of students’ feedback on one hand. Some teachers appreciate the teaching evaluation scheme, and those teachers may also have positive attitudes regarding the school managers’ communication of goals. Other teachers are resistant of the teaching evaluation scheme, but those teachers also perceive the communication as clear.

Limitations

As with related studies, this study has clear limitations from a methodological (e.g. cross-sectional) standpoint and from a conceptual perspective (e.g. parsimonious modelling). We acknowledge these limitations and argue that they can serve as the foundation for future studies. There are, as in other studies, insufficient controls in the research. However, controlled experiments are normally beyond the realm of possibility for educational researchers. Still, studying development processes at some schools, which in many respects have substantial similarities apart from the one aspect being studied, can set us on the trail of causal processes.

Multiple factors may influence our behaviour. Longitudinal and quasi-experimental studies are needed in order to determine causality. Cross-sectional studies only represent a momentary glimpse of an organisation, and do not allow for the testing of causal relationships among antecedents. Reverse causation may play a role. Omitted variables may have influenced the overall model. More longitudinal research is needed in order to address the complexity of the interactional dynamics between leaders and teachers and the associated impact on teacher motivation. Another limitation of this study is the use of self-reported questionnaire data. The subjective component of these data is undeniable. Independent judgements can provide interesting data about a teacher’s performance, but it is difficult to carry out this process while honouring promises of anonymity. The heavy reliance on teachers’ self-reporting is questionable.

Further, factors outside the school system could also influence teachers’ perception of the utility of teaching evaluation and should have been included. Only a limited number of antecedents were examined. A final limitation is that the moderate (comparable to those of corresponding studies) number of participants leaves room for uncertainty about whether the samples are representative. In sum, these limitations provide directions for future research.
Implications for further studies

Despite its limitations, this study contributes to our understanding of the factors influencing teacher resistance to evaluation schemes. The results of this study provide a valid description of the conditions that determine how teachers in the general studies programme in Akershus County will perceive the teaching evaluation scheme. Teachers in vocational and other programmes at the upper secondary level may perceive such schemes differently from their colleagues in general studies programmes; more research is required in this area. The recruitment of students also varies between these two types of programmes. For example, the teachers’ assessments of the students’ level of maturity with regard to their ability to provide well-considered feedback may vary. In addition, the professional culture among teachers in vocational programmes is also quite different from the one among teachers in general studies programmes. We therefore need more research on teaching evaluation in different contexts.

The overall aim of the evaluation scheme is better teaching quality and better student attainment. The question of whether teaching evaluation schemes actually serve to weaken or reinforce the teachers’ efforts and engagement at work is context-specific and remains controversial in available research. More research is needed for better understanding the schools’ internal mechanisms. Today, there is insufficient evidence to conclude that an anonymous teaching evaluation of the kind presented in this article will actually produce better learning outcomes, and this remains an issue for future empirical investigation.

In its election programme, the governing Conservative Party has committed itself to permitting students in both lower and upper secondary schools to evaluate teaching (The Conservative Party 2013). As a consequence, national initiatives for implementing teaching evaluation in lower secondary schools may become relevant as future policy. Students at the lower secondary level are younger than those in upper secondary schools, and some issues remain to be clarified with regard to how age affects the assessment of various aspects of teaching quality as well as how teachers relate to evaluation schemes involving primary and lower secondary students. In which situations will the teaching evaluation scheme be deemed meaningless by teachers? Future research may help us better understand this aspect of teacher resistance.

As far as we have been able to establish, all Norwegian counties that provide upper secondary education have used teaching evaluation schemes. If the context (including whether the county engages in negotiations with the teachers’ trade unions to establish acceptance of the scheme) should prove to have a significant effect, there may be county-specific and country-specific variations in how teachers relate to teaching evaluation.

Future research needs to take into account the critical issues surrounding teacher evaluation. Issues such as teachers’ perceptions of teacher evaluation, the role of
school culture and leadership, and the context of teaching as a profession are some examples that research should explore further. We should be cautious about using results from empirical studies undertaken in other countries (e.g. Taut et al. 2011) as a basis for drawing conclusions on teacher resistance. The specific context is an active force (Ball et al. 2012, 40). We know less about how school cultures vary with teacher resistance. Tuytens and Devos (2014) emphasise the importance of school culture and found that communication about the new teacher evaluation practice differed among schools; some perceived the evaluation positively and others negatively. The management of change is crucial to the success of systemic innovation and this shows that individual school leadership and culture might be important factors in teacher trust, perceptions of usefulness of student feedback, and relative perceptions of stress. Further, enactment at the meso level is also an avenue for further research.

Since we suspect that the culture for collaboration among teachers may vary between subject groups and educational programmes, we recommend that a study comprise several types of subjects and vocational studies to provide us with a better understanding of the characteristics of the collective dimension in teacher resistance. Further research should aim to investigate how teacher resistance may vary with the climate of education, the relationship between school leaders and teachers, and how teacher unions work.

**Acknowledgement**
This research was supported by a grant (no. 237863) from the Research Council of Norway.

_Eyvind Elstad, Dr. Polit., Professor, Department of Teacher Education and School Research, University of Oslo. His research interests are teacher professionalism, teacher education and school leadership._

_Eli Lejonberg, PhD Candidate, Department of Teacher Education and School Research, University of Oslo. Her research interests are mentoring, teacher professionalism and teacher education._

_Knut-Andreas Christophersen, Associate Professor, Department of Political Science, University of Oslo. His research interests are research methodology and statistical analysis._
Notes

1. The observation of teachers’ work by school managers or external experts is another tool for improving educational practices. A third instrument is to measure the added value in students’ learning progress. This would require measurement of student performance by testing.

2. In 2007, the School Student Union of Norway (an interest organisation for students at the upper secondary level) launched an initiative for mandatory teaching evaluation. This initiative was communicated to the political level. In the same year, the Youth City Council Meeting in Oslo adopted the proposal that students should “have the opportunity to evaluate their teachers and educational programmes with a view to providing constructive feedback”. A decision by Oslo City Council later confirmed this proposal. Thus, teaching evaluation was first implemented in Oslo during the 2008/2009 academic year under the auspices of Oslo Municipality.

3. Only a small minority of all Norwegian upper secondary schools are privately owned, and private schools can undertake teaching evaluation in ways other than through negotiations with trade unions. We are aware of some private schools that also use results from teaching evaluation for external marketing purposes.
References

Abrahamsen, H. and Aas, M. (2014). Avdelingsledere—en ny profesjon innenfor skoleledelse. In, Profesjonsutvikling i skolen, Eyvind Elstad and Kristin Helstad (eds.), 300–314. Oslo: Universitetsforlaget.

Acemoglu, D., Robinson, J. A. and Verdier, T. (2012). Can’t we all be more like Scandinavians? Asymmetric growth and institutions in an interdependent world. NBER working papers 18441.

Austin, V., Shah, S. and Muncer, S. (2005). Teacher stress and coping strategies used to reduce stress. Occupational Therapy International, 12(2), 63–80.

Ball, S. J. (2012). Global Education Inc. New policy networks and neo-liberal imaginary. London: Routledge.

Ball, S. J., Maguire, M. and Braun, A. (2012). How schools do policy. Policy enactments in secondary schools. London: Routledge.

Barth, E., Moene, K. O. and Willumsen, F. (2014). The Scandinavian model—An interpretation. Journal of Public Economics, 117, 60–72.

Beerens, D. R. (2000). Evaluating teachers for professional growth. Thousand Oaks, CA: Corwin Press.

Brandal, N., Bratberg, Ø. and Thorsen, D. E. (2013). The Nordic model of social democracy. London: Palgrave Macmillan.

Bo, A. K. and Hovdenak, S. S. (2011). Faglig og personlig støtte. Tidsskrift for ungdomsforskning, 11(1), 69–85.

Carl, A. E. (2009). Teacher empowerment through curriculum development: theory into practice. New York: Juta and Company Ltd.

Christophersen, K. A., Elstad, E. and Turmo, A. (2010). Is teacher accountability possible? The case of Norwegian high school science. Scandinavian Journal of Educational Research, 54(5), 413–429.

Chubb, J. E. and Moe, T. M. (1991). Politics, markets and America’s schools. British Journal of Sociology of Education, 12(3), 381–396.

Christensen, T. and Lægreid, P. (2011). The Ashgate research companion to new public management. Farnham: Ashgate.

Darling-Hammond, L., Wise, A. E. and Pease, S. R. (1983). Teacher evaluation in the organizational context: a review of the literature. Review of Educational Research, 53(3), 285–328.

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

Datnow, A. and Castellano, M. (2000). Teachers’ responses to Success for All: how beliefs, experiences, and adaptations shape implementation. American Educational Research Journal, 37(3), 775–799.

Day, C., Flores, M. A. and Viana, I. (2007). Effects of national policies on teachers’ sense of professionalism: findings from an empirical study in Portugal and in England. European Journal of Teacher Education, 30(3), 249–265.

Delvaux, E., Vanhoof, J., Tuytens, M., Vekeman, E., Devos, G. and Van Petegem, P. (2013). How may teacher evaluation have an impact on professional development? A multilevel analysis. Teaching and Teacher Education, 36, 1–11.

Derrington, M. L. (2014). Teacher evaluation initial policy implementation: principal and superintendent perceptions. Planning and Changing, 45(1/2), 120–137.

Ellett, C. D. and Teddlie, C. (2003). Teacher evaluation, teacher effectiveness and school effectiveness: perspectives from the USA. Journal of Personnel Evaluation in Education, 17(1), 101–128.
Elstad, E. (2009). Styring, ansvarliggjøring og ansvarsfraskrivelse. In Læreplan i et forskningsperspektiv, Erling Lars Dale (ed.), 116–153. Oslo: Universitetsforlaget.

Elstad, E. (2010). PISA i norsk offentlighet: politisk teknologi for styring og bebreidelsesmanøvrering. In Eyvind Elstad, E. and Kirsten Sivesind (eds.), PISA: Sannheten om skolen? 100–121, Oslo: Universitetsforlaget.

Elstad, E., Turmo, A. and Guttersrud, Ø. (2011). Problems induced by amalgamation of pedagogical progressivism and educational accountability: oral exams with prior preparation time in Norwegian secondary schools. Problems of Education in the 21st Century, 30, 22–34.

Elstad, E. (2013). New discourses of teacher professionalism: a Norwegian case. Sovremennye issledovaniya sotsialnykh problem, 4, 1–20.

Engeland, Ø. (2000). Skolen i kommunalt eie – politisk styrt eller profesjonell ledet skoleutvikling. Oslo: Universitetet i Oslo, Unipub.

Engeland, Ø. and Langfeldt, G. (2009). Forholdet mellom stat og kommune i styring av norsk utdanningspolitikk 1970–2008. Acta Didactica Norge, 3(1), 1–16.

Fitzgerald, T., Youngs, H. and Grootenboer, P. (2003). Bureaucratic control or professional autonomy? Performance management in New Zealand schools. School Leadership and Management, 23(1), 91–105.

Flores, M. A. (2010). Teacher performance appraisal in Portugal: the (im)possibilities of a contested model. Mediterranean Journal of Educational Studies, 15(1), 41–60.

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.

Garmannslund, P. E., Elstad, E. and Langfeldt, G. (2008). Lærernes opplevelse av måling og rangering av kvalitetsaspekter ved undervisning og læringsprosesser. In Ansvarlighet i skolen, Gjert Langfeldt, Eyvind Elstad and Stefan T. Hopmann (eds.), 250–270. Oslo: Cappelen Forlag.

GNIST. (2009). GNIST – partnerskap for helhetlig lærersatsning. https://www.regjeringen.no/globalassets/upload/kd/vedlegg/udir_gnist_avtale_msign.pdf

GNIST. (2013). Oppdragsbrev. Unpublished manuscript.

GNIST. (2014). Rapport. Oslo: GNIST. http://www.gnistweb.no/aktuelt/271/nye-rapporter-med-lærervurdering-som-tema.html

GNIST. (2015). Rapport. Oslo: GNIST. Unpublished manuscript.

Heneman, III. H. G. and Milanowski, A. T. (2003). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. Journal of Personnel Evaluation in Education, 17(2), 173–195.

Hinz, A. (2011). Attitudes of German teachers and students towards public online ratings of teaching quality. Electronic Journal of Research in Educational Psychology, 9(2), 745–764.

Hovdenak, S. S. (2000). 90-tallsreformene: et instrumentalistisk mistak?. Oslo: Gyldendal Akademisk.

Hulpia, H., Devos, G. and Rosseel, Y. (2009). Development and validation of scores on the distributed leadership inventory. Educational and Psychological Measurement, 69(6), 1013–1034.

Isoré, M. (2009). Teacher evaluation: current practices in OECD countries and a literature review. OECD Education Working Papers No. 23. OECD Publishing.

Jambak, T. (2015). En dårlig lærers bekjennelser, Aftenposten 27. March http://www.aftenposten.no/meninger/kronikker/En-darlig-larers-bekjennelser-7516999.html

Jarning, H. (2010). Resultater som teller - Kunnskapskontroll og resultater i skolens århundre. In Eyvind Elstad and Kirsten Sivesind (eds.), PISA: Sannheten om skolen? 199–221. Oslo: Universitetsforlaget.

Katsuno, M. (2010). Teacher evaluation in Japanese schools: an examination from a micropolitical or relational viewpoint. Journal of Education Policy, 25(3), 293–307.
Kelly, K. O., Ang, S. Y. A., Chong, W. L. and Hu, W. S. (2008). Teacher appraisal and its outcomes in Singapore primary schools. *Journal of Educational Administration*, 46(1), 39–54.

Kilburg, R. (1980). Metaphysics and professional resistance to program evaluation. *Evaluation and Program Planning*, 3, 185–190.

Kjærnsli, M. and Olsen, R. V. (2013). *Fortsatt en vei å gå. Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012*. Oslo: Universitetsforlaget.

Kyriacou, C. and Sutcliffe, J. (1978). Teacher stress: prevalence, sources, and symptoms. *British Journal of Educational Psychology*, 48(2), 159–167.

Lejonberg, E. and Elstad, E. (2014). Bruk av profesjonsbegrepet i kampen om utdanningspolitikkens innhold. In *Profesjonsutvikling i skolen*, Eyvind Elstad and Kristin Helstad, 39–56. Oslo: Universitetsforlaget.

Lien, T. (2009). *Elevevaluering av lærere. Utviklingsverktøy eller kontrollinstrument?* Oslo: University of Oslo.

Litt, M. D. and Turk, D. C. (1985). Sources of stress and dissatisfaction in experienced high school teachers. *The Journal of Educational Research*, 78(3), 178–185.

Liu, S. and Zhao, D. (2013). Teacher evaluation in China: latest trends and future directions, *Evaluation and Accountability*, 25(3), 231–250.

Ministry of Education and Research. (2003). *White Paper 20 (2003–2004). Kultur for læring*. Oslo: The Ministry of Education and Research.

Ministry of Education and Research. (2014). *Strategy: the teacher promotion. Pulling towards knowledge in school*. Oslo: Ministry of Education and Research.

Monyatsi, P., Steyn, T. and Kamper, G. (2006). Teacher appraisal in Botswana secondary schools: a critical analysis. *South African Journal of Education*, 26(2), 215–228.

OECD. (1988). *Review of educational policy in Norway: examiner’s report and questions [ED(88)2]*. Paris: OECD.

OECD. (2009). *Teacher evaluation. A conceptual framework and examples of country practices*. Paris: OECD.

OECD. (2011). *OECD reviews of evaluation and assessment in education Norway 2011*. Paris: OECD.

Olsen, J. P. (2007). The institutional dynamics of the European university. In *University Dynamics and European Integration*, Peter Maassen and Johan P. Olsen (eds.), 25–54. Dordrecht: Springer.

Ong Kelly, K., Yun Angela Ang, S., Ling Chong, W. and Sheng Hu, W. (2008). Teacher appraisal and its outcomes in Singapore primary schools. *Journal of Educational Administration*, 46(1), 39–54.

Peterson, K. D. and Peterson, C. A. (2006). *Effective teacher evaluation. A guide for principals*. Thousand Oaks, CA: Corwin Press.

Seigel, S. and Rockwood, V. (1993). Democratic education, student empowerment, and community service: theory and practice. *Equity and Excellence in Education*, 26(2), 65–70.

Stoltenberg, J. (2008). 2008 New Year’s Address. *The Prime Minister’s Office*. http://www.regjeringen.no/nb/dep/smk/Statsministerens-kontor/Statsminister_Jens_Stoltenberg/Taler-og-artikler/2008/Nyttarstale-2008.html?id=495221

Sun, M., Youngs, P., Yang, Chu, H. and Zhao, Q. (2012). Association of district principal evaluation with learning-centered leadership practice: evidence from Michigan and Beijing, *Evaluation and Accountability*, 24(3), 189–213.

Taut, S., Santelices, M. V., Araya, C. and Manzi, J. (2011). Perceived effects and uses of the national teacher evaluation system in Chilean elementary schools. *Studies in Educational Evaluation*, 37(4), 218–229.
Telhaug, A. O., Mediås, A. and Aasen, P. (2006). The Nordic model in education: education as part of the political system in the last 50 years. *Scandinavian Journal of Educational Research, 50*(3), 245–283.

The Conservative Party. (2013). *New ways, better solutions. The conservative party’s program 2013–2017*. Oslo: The Conservative Party.

The Directorate of Education and Training. (2006). *Rapport fra felles nasjonalt tilsyn på opplæringsområdet 2006*. Oslo: The Directorate of Education and Training.

The Education Act. (1998). *Lov om grunnskolen og den vidaregåande opplæringa. https://lovdata.no/dokument/NL/lov/1998-07-17-61?q=oppl%C3%A6ringslova*

The Norwegian Government. (2013). *Political platform for the government*. Oslo: Regjeringen. https://www.regjeringen.no/nb/dokumenter/politisk-plattform/id743014/

The School Student Union of Norway. (2003). *Forberedende komite for politisk måldokument*. Unpublished note.

Timperley, H. S. and Robinson, V. M. J. (1997). The problem of policy implementation: the case of performance appraisal. *School Leadership and Management, 17*(3), 333–345.

Tornero, B. and Taut, S. (2010). A mandatory, high-stakes national teacher evaluation system: perceptions and attributions of teachers who actively refuse to participate. *Studies in Educational Evaluation, 36*(4), 132–142.

Tuytens, M. and Devos, G. (2009). Teachers’ perception of the new teacher evaluation policy: a validity study of the Policy Characteristics Scale. *Teaching and Teacher Education, 25*(6), 924–930.

Tuytens, M. and Devos, G. (2010). The influence of school leadership on teachers’ perception of teacher evaluation policy. *Educational Studies, 36*(5), 521–536.

Tuytens, M. and Devos, G. (2011). Stimulating professional learning through teacher evaluation: an impossible task for the school leader? *Teaching and Teacher Education, 27*(5), 891–899.

Tuytens, M. and Devos, G. (2014). The problematic implementation of teacher evaluation policy: school failure or governmental pitfall? *Educational Management Administration and Leadership, July 2014*(42), 155–174.

Weiler, H. N. (1990). Comparative perspectives on educational decentralization: an exercise in contradiction? *Educational Evaluation and Policy Analysis, 12*(4), 433–448.
Appendix 1. Overview of the constructs and indicators

### Descriptive Statistics

| Teacher resistance, tr \[\text{adapted from Hulpia, Devos, and Rosseel 2009}, \alpha_c = .93\] | Item | Mean | Standard deviation |
|---|---|---|---|
| In general, I think that the evaluation system for teaching is an excellent scheme. (Reversed) | 3.65 | 1.663 |
| Akershus County should continue using the existing system of teaching evaluation. (Reversed) | 3.9 | 1.636 |
| The system of teaching evaluation works as it should. (Reversed) | 3.85 | 1.524 |

| Stress induced by teaching evaluation, str \[\text{adapted from Heneman III and Milanowski 2003}, \alpha_c = .90\] | Item | Mean | Standard deviation |
|---|---|---|---|
| It is extremely stressful to me that teaching evaluation demonstrates my weaknesses as a teacher. | 2.22 | 1.433 |
| I am concerned that the management may possibly confront me with poor results revealed by the evaluation of my teaching. | 2.64 | 1.719 |
| The thought that my teaching will be continuously evaluated is highly stressful to me. | 2.58 | 1.721 |

| Perceived control purposes of teaching evaluation, pcp \[\text{adapted from Kelly et al. 2008}, \alpha_c = .88\] | Item | Mean | Standard deviation |
|---|---|---|---|
| The purpose of teaching evaluation is to obtain an overview of which teachers are good and which are poor. | 3.47 | 1.801 |
| The purpose of teaching evaluation is to establish competition among the teachers. | 2.34 | 1.618 |
| The purpose of teaching evaluation is to monitor the teachers’ classroom work. | 3.96 | 1.924 |

| Acknowledgement of the feedback from students, ts \[\text{adapted from Kelly et al. 2008}, \alpha_c = .88\] | Item | Mean | Standard deviation |
|---|---|---|---|
| I trust the judgment of those students who contribute to the evaluation of my teaching | 4.28 | 1.564 |
| Those students who contribute to the evaluation of my teaching are competent to evaluate my teaching | 4.08 | 1.599 |
| Those students who contribute to the evaluation of my teaching have sufficient insight into the teaching profession to evaluate the work of a teacher | 3.06 | 1.544 |
Appendix 2. Student Survey in Akershus County

All items were on a 5-point Likert scale: 1 = strongly disagree, 2 = slightly disagree, 3 = neither agree nor disagree, 4 = somewhat agree, and 5 = strongly agree

| Category/concept           | Example item                                                                 |
|---------------------------|-----------------------------------------------------------------------------|
| Student motivation        | I do my best in the lessons of this subject                                 |
| Adjusted teaching         | The teacher is good at helping me when I feel stuck                         |
| Working atmosphere        | The teacher is the leader when we have lessons                              |
| Structure                 | The teacher comes to class prepared                                         |
| Teacher–student relationship | The teacher takes pupils’ views seriously                                   |
| Teacher commitment        | The teacher shows high professional commitment                              |
| Use of digital tools      | The teacher uses digital tools in a way that provides good learning         |
| Assessment for learning   | The teacher explains how we must perform to achieve different grades       |
| Self-evaluation           | The teacher helps me actively participate in assessing my own schoolwork   |

1 How strongly do you agree with the statements below? Rank your responses from 1 (fully disagree) to 7 (fully agree)