Studies in Educational Evaluation 47 (2015) 47–57

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/stueduc

From inspection to quality: Ways in which school inspection influences change in schools

J.-E. Gustafsson [a,*], M.C.M. Ehren [b], G. Conyngham [c], G. McNamara [c], H. Altrichter [d], J. O'Hara [c]

[a] University of Gothenburg, Department of Education and Special Education, P.O. Box 300, S-40530 Gothenburg, Sweden
[b] Institute of Education, 20 Bedford Way, WC1H 0AL London, UK
[c] Dublin City University, School of Education Studies, Dublin 9, Ireland
[d] Johannes Kepler Universität Linz, 4040 Linz-Auhof Österreich, Austria

ARTICLE INFO

Article history:
Received 24 September 2014
Received in revised form 5 July 2015
Accepted 13 July 2015
Available online

Keywords:
School inspection
School self-evaluation
Accountability
External evaluation
School improvement
School effectiveness

ABSTRACT

Inspection is employed by most European education systems as an instrument for controlling and promoting the quality of schools. Yet there is little research knowledge about how inspection drives the improvement of schools. The study reports on surveys to principals in primary and secondary education in six European countries to attempt to clarify how school inspection impacts on the improvement of schools. Based on an analysis of principals' perceptions the evidence suggests that inspection primarily drives change indirectly, through encouraging certain developmental processes, rather than through more direct coercive methods. Inspectorates that set clear expectations and standards have an impact on the increased utilization of self-evaluation and on developing the capacity of schools to improve in a variety of ways.

1. Introduction

Evaluation and accountability are perceived as key elements in attempting to ensure quality educational provision for all (Eurydice, 2004). In most of Europe, an important instrument of educational evaluation and accountability is school inspection. Inspectorates assess the quality of education and hold schools accountable for a broad range of goals related to student achievement, teaching, organization and leadership. In most cases these judgements are made against criteria and standards and may involve sanctions for poor performance. Naturally the ultimate purpose of all this activity is to improve schools in terms of the experience and performance of learners. However it is not clear that inspection impacts directly on these outcomes or, if it does, whether different models of inspection may be more effective in achieving them.

A review by Nelson and Ehren (2014) found little empirical research which attempted to link inspection directly to pupil achievement, and it was also observed that few studies took selection effects into account. Most studies showed slight or no effect of inspection on achievement (e.g., Hanushek & Raymond, 2005; Luginbuhl, Webbink, & De Wolf, 2009; Matthews & Sammons, 2004; Rosenthal, 2004). However, two recent studies with longitudinal design (Allen & Burgess, 2012; Hussain, 2012) do provide evidence of a link between the inspection findings and student achievement results, indicating that studies with strong design are needed to establish such effects.

As a first step to understanding the effects of school inspection on student achievement it is of great importance to gain more knowledge about the in-school processes which mediate between school inspection and the improvement of student performance (Husfeldt, 2011). The study presented in this paper intends to expand this knowledge base by identifying and empirically analysing the mechanisms which link school inspections to school improvement activities.

The research reported in this paper is designed to test a conceptual model (Ehren, Altrichter, McNamara & O'Hara, 2013) developed by closely examining the policy assumptions or 'programme theory' which underpins school inspection systems in six European countries. The synthesis of these six programme theories suggests that there are identifiable common methods or mechanisms of inspection through which inspection is expected to

* Corresponding author.
E-mail addresses: Jan-Eric.Gustafsson@ped.gu.se (J.-E. Gustafsson), m.ehren@ioe.ac.uk (M.C.M. Ehren), Gerry.Conyngham@dcu.ie (G. Conyngham), gerry.mcnamara@dcu.ie (G. McNamara), herbert.altrichter@jku.at (H. Altrichter), Joe.OHara@dcu.ie (J. O'Hara).

http://dx.doi.org/10.1016/j.stueduc.2015.07.002
0191-491X/

drive school improvement. These mechanisms which mediate between inspection and eventual impact were used to create a conceptual model. In the light of the conceptual model developed, the empirical and theoretical literature on the hypothesized mechanisms is reviewed. The paper then proceeds to analyze survey data from principals in six European countries, to investigate empirically their perceptions as to which mechanisms are most important in actually linking school inspections to improvement efforts in schools.

1.1. Previous research and theoretical framework

The conceptual framework builds from the programme theories of the Inspectorates of Education in six European countries (Austria (Styria), the Czech Republic, England, Ireland, the Netherlands and Sweden, (see Ehren et al., 2013). These countries were chosen because of their significant differences in school inspection systems representing the varied models of inspection to be found across Europe. The variations range across a wide spectrum, from using a low stakes approach involving inspecting schools on a regular basis without sanctions or rewards (for example Ireland and Austria) to school inspectorates utilizing directive and focused, medium/high stakes early warning analysis and customized inspections (e.g., the Netherlands, England), and from a centralized to a decentralized level of operation. Including a broad range of different models in our study allows us to examine how a range of contrasting inspection approaches influences the way in which schools respond to external monitoring.

The programme theories were elaborated for each country from a study of the key policy documents governing inspection and from interviews with relevant officials. The individual programme theories were then brought together to create a common theoretical framework for all six countries, including a description of the mechanisms through which inspectorates aim to achieve school improvement. These mechanisms represent the ways in which policy makers assume inspections will lead to the improvement of schools by acting, as it were, as the bridge between inspection and the ultimate objective of improved teaching and learning, hopefully leading to better learning outcomes.

The first part of the theoretical framework (see Ehren et al., 2013) includes descriptions of different aspects of school inspections in the countries involved including: the frequency of inspection visits, the use of regular cycles of full inspections or a differentiated approach to target potentially failing schools, the setting of standards and thresholds, the consequences of failure, and the feedback given during and after inspection. The second part elaborates hypotheses concerning three common mechanisms identified across the various systems through which school inspection is generally expected to drive change (Ehren et al., 2013, p. 14).Fig. 1 summarizes these mechanisms and presents the conceptual framework of the study. Below the hypothesized mediating mechanisms are discussed at greater length, and the available empirical research investigating the mechanisms is reviewed. It should be noted, though, that little empirical research has been conducted on the impact of inspection, and that the largest number of studies is from England (Nelson & Ehren, 2014).

1.1.1. Mechanisms linking the impact of school inspections to outcomes

The first common mechanism to link school inspections and their intended outcome is inspection frameworks setting expectations, norms and standards. These frameworks define expectations of quality for schools and their stakeholders. Schools are expected to attend to the requirements included in inspection standards and procedures and adapt their goals and ways of working to come into line with the normative image of high quality schools demanded by the inspectorate. These inspection frameworks are designed to inform and drive school policy, planning and practices.

Three types of standards were identified by Ehren et al. (2013): legal standards, standards relating to the context and process

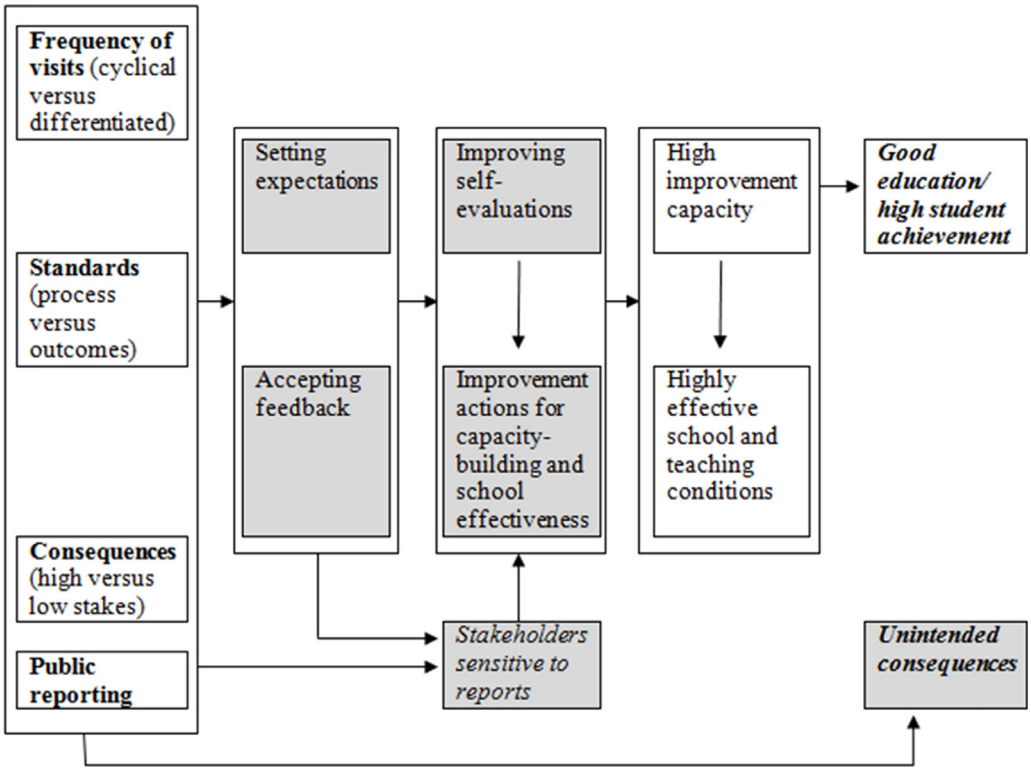

Fig. 1. Intended effects of school inspections-proposed conceptual model.

quality of the education provided and standards defining performance and results targets for schools. Standards related to the process quality of education are often inspired by research on school improvement and school effectiveness, and inspection frameworks usually include indicators such as opportunity to learn, achievement orientation, clear and structured teaching, and challenging teaching approaches.

Some studies suggest that sanctions and rewards have a positive effect on educational quality in schools, suggesting that policies, performance criteria and feedback alone may be insufficient to motivate schools to perform to high standards (Elmore & Fuhrman, 2001; Malen, 1999; Nichols, Glass, & Berliner, 2006). Schools may receive rewards, such as financial bonuses or awards, for good performance. Sanctions include, for example, naming and shaming of the school on the Internet or in the media (Elstad, 2009) or fines. Hanushek and Raymond (2002) point to rational choice theory to describe how standards and performance thresholds, and related sanctions and rewards, may influence actions in schools. They argue that school officials would select the action that they perceive to have the highest yield, given their planning horizon, budget and appetite for risk. Hanushek and Raymond (2002) for example found failing schools in the USA risking sanctions to make dramatic changes in the year after identification of these failures.

Responses to inspection tend, it is suggested in the literature, to be most focused and effective where funding is at stake or exposure is higher (Matthews & Sammons, 2004). Formal sanctions, like forced reconstitution of consistently low performing schools, were found to be more likely to promote responses than mere embarrassment from grading schools and reporting results publicly. In summary, the literature suggests that school responses to the pressure of inspection are likely to be influenced by, on the one hand, a clear awareness of the importance of the policies and standards and, on the other hand, by significant sanctions and rewards. In the cases of the countries in this research all could be said to have frameworks of required standards and performance criteria but the use of sanctions and rewards varied greatly.

However, high stakes accountability systems have also been shown to produce harmful consequences (Heubert & Hauser, 1999; Koretz, 2003; Stecher, 2002). Sanctions and rewards may discourage desirable behaviour or may stimulate unintended and undesirable behaviour. According to Elmore and Fuhrman (2001), schools operating under severe sanctions such as reconstitution and probation do not appear to be making fundamental changes in their core processes. Instead they seem to place considerable emphasis on test preparation when test scores are used to assess schools or to emphasize other elements in the organization of education that are assessed as part of school inspections. These quick fix solutions may lead to rapid improvement on the measures of the inspectorate rather than to genuine long term improvement.

The second hypothesized mechanism to promote intended effects is through the feedback that is provided to schools during inspection visits and/or in inspection reports, including the consequences of not following up on the feedback. Schools are expected to use the feedback to improve, and stakeholders are expected to take note of the feedback and hold schools accountable for use of the feedback for improvement.

Theories on organizational learning and school improvement point to the role of performance feedback in encouraging change within schools. During inspections visits, inspectors assess educational quality with respect to the standards in the framework and give feedback on the strong and weak points of the performance of schools on these standards. Most inspectorates also give schools advice on how to improve. In short, inspection visits are expected to lead to remedial actions as schools are made aware of the standards they have to comply with and are provided with feedback and support.

Ehren and Visscher (2008) found in a case study that all schools used the feedback received from the school inspectors to improve their functioning and, after six months, all schools were still carrying out improvement plans. According to Matthews and Sammons (2004) clear and explicit feedback to schools is successful in informing the improvement plan after school inspections, resulting in more effective action. Other studies too emphasize the importance of the way in which feedback is provided (e.g., Dedering & Muller, 2011; Dobbelaer, Prins, & van Dongen, 2013; McCrone et al., 2007). In contrast other studies indicate that many teachers are not willing to change their teaching after an inspection. Gärtner, Füsemann, and Pant (2009, p. 10) found in a German study that only a minority of the schools which had been inspected reacted actively to the inspection report. Verhaeghe, Vanhoof, Valcke, and Van Petegem (2010) also found that principals made little systematic use of feedback, which was interpreted to be due to lack of time, skills and support (cf. Van Petegem & Vanhoof, 2007). Studies also show that the effects of feedback depend on whether it is positive or negative. Ouston, Fidler, and Earley (1997) observed that school inspections led to school improvement only if schools had received a negative or only slightly positive assessment from the inspectorate. Visscher and Coe (2003) concluded that if feedback is experienced as threatening it will be associated with fewer effects and they also observed that feedback should be perceived as providing information and supporting self-determination, rather than as controlling. Thus, even though acceptance of feedback might be expected to have positive effects, it does seem that the impact is moderated by other factors.

The third hypothesized mechanism to promote the intended effects of school inspection is stakeholder pressure. The provision of information on the inspection process and outcome to a broad range of stakeholders will, it is hoped, force the school to act on recommendations. Each of the systems sees stakeholder involvement as being essential for improvement. Stakeholders, such as parents and school boards, are expected to facilitate school improvement through actions of 'voice', 'choice' and 'exit'. The informed parent or school board is expected to hold the school to account by using the inspection feedback to demand improvement in specific areas (voice). If improvement does not take place, parents can move their children to different schools (choice).

Here too the existing research indicates that stakeholder pressure is a complex mechanism. Research on school choice has shown that parents rarely use published information as the primary motive for their choice of school and that willingness to make school choice decisions is unevenly distributed among different social groups (Belfield & Levin, 2009; Buckley & Schneider, 2003; Denessen, Driessen, & Sleegers, 2005). Other studies (Ofsted, 2009) suggest that parents feel disempowered in relation to those who they feel are the 'professionals' in the field of school improvement, even when the parents are provided with full information about inspection outcomes.

1.1.2. Intermediate mechanisms mediating the impact of school inspections on outcomes

According to the conceptual framework developed for this research (see Ehren et al., 2013) the three mechanisms proposed above are hypothesized to impact on outcomes via three further related tools identified by inspectorates which we describe as 'intermediate mechanisms'. These are promoting and improving school self-evaluation, actively building the capacity of schools to absorb, internalize and respond to inspection and demanding and monitoring specific school improvement actions. These mediating mechanisms are briefly described below.

Within the programme theories of most school inspection systems analyzed in this research, high-quality self-evaluation is

considered to be a critical element in the improvement of schools (Ehren et al., 2013, pp. 22-23). Self-evaluations, together with inspection, are seen as inseparable and integral parts of an improvement and accountability cycle. Self-evaluations conducted by schools are also an important part of school inspection systems.

Whitby (2010) noted in a review that the fact that school self-evaluation is used to inform school inspection may lead schools to see external inspection in a developmental perspective rather than a judgmental one. However, Whitby (2010) also noted a risk that self-evaluations may be written mainly to comply with expectations of the inspectorates (MacBeath, 2000; Meuret & Morlaix, 2003). The main conclusion of the review was that it is the amount of guidance and support that schools are provided for school self-evaluation and external inspection that affect the impact on school improvement.

It also has been observed that school self-evaluations may be of mixed quality and inconsistently used among schools (Blok, Sleegers, & Karsten, 2008; Hofman, de Boom, & Hofman, 2010; Karagiorgi, 2012; McNamara, O'Hara, Lisi, & Davidsdottir, 2011; Schildkamp & Visscher, 2010). The reception of school self-evaluation processes and results have been found to be influenced by whether the evaluation is externally imposed or internally developed (McNamara, O'Hara, Lisi, & Davidsdottir, 2011). Chances that schools will make use of self-evaluation results improve if the school team believes that self-evaluation can lead to quality improvement, and if teachers have a sense of ownership, feeling that they can influence measures taken based on the self-evaluation results (Bubb & Earley, 2008; Schildkamp, Vanhoof, Petegem, & Visscher, 2012). The principal's leadership also is an important determinant of whether the self-evaluation is followed by improvement actions or not (Emstad, 2011).

Capacity building is another intermediate mechanism which is widely emphasized. While research on school change has shown that altering teachers' practices is difficult (Fullan, 2002), research also has demonstrated that teachers' participation in professional learning activities are influenced by both personal characteristics and by leadership practices and organizational conditions. Geijsel, Sleegers, Stoel, and Kruger (2009) concluded that principals' transformational leadership is an important determinant to enhance the professional learning of teachers. Encouraging the participation of teachers in decision making and cooperation between teachers were found to be important mediators in capacity building.

The third intermediate mechanism is improving school effectiveness. Research has established several generalizations concerning factors contributing to student achievement, such as high teacher expectations, a challenging teaching approach, an orderly learning environment, feedback, and clear and structured teaching (Scheerens, 2009). Building upon the available capacity, schools leaders are expected to increase the effectiveness of their school by implementing and monitoring effective teaching and instruction, and by creating school level conditions conducive to improvement, such as distributed educational leadership, a productive climate and culture, and an achievement-oriented school policy.

According to the conceptual model these three intermediate mediating mechanisms are closely related in that promoting and improving self-evaluation impacts on capacity building which, in turn, informs policies and actions to enhance school effectiveness.

1.2. Research questions

The conceptual model with its hypothesized mechanisms forms the starting point of our empirical research. Of course the elaboration of the programme theories in each country demonstrates that the mechanisms, while present to some degree everywhere, vary in use and importance. This does not invalidate the conceptual model, but rather demonstrates the range of variables that need to be taken into account both when investigating inspection on a transnational level and still more the challenges involved in assessing the general applicability of research findings across such varied contexts. Nonetheless the model still has a coherence and operational logic that makes it useful for investigating mechanisms through which school inspections have an impact across a range of inspection systems.

The programme theories focus, by definition, on intended outcomes. However, the review of the literature has demonstrated many cases of unintended, typically negative, effects of school inspections. The theoretical framework is not designed to account for such effects, and to investigate these other theories are more relevant, such as, for example, neo-institutional theories (Powell, 2007; Scott, 2008). These theories assume that organizations not only operate according to efficiency criteria, but that they also seek legitimacy for their actions from other sources (Meyer & Rowan, 1977, p. 340). During an inspection, schools may therefore feel that their chances for legitimacy are enhanced if they show conformity to official rules of inspection and to normative pressures, even though this is not the actual practice of the school.

Utilizing data obtained from a questionnaire designed to elicit the perceptions of school principals in the participating countries concerning the ways in which inspection impacts on change in their schools the following research questions will be investigated in this paper: (1) through which mechanisms do school inspections in Europe promote school improvement? and (2) through which mechanisms do school inspections in Europe lead to unintended consequences?

2. Method

The researchers developed a survey for principals in primary and secondary schools in six European countries to identify their perceptions of the mechanisms linking school inspections to the actual improvement of schools. The survey included questions on the main variables in the conceptual framework which enables comparison between the responses of principals and analyses how these responses are related to one another. Principals are considered to be the best informants of changes in schools as a result of school inspection as they are generally the key actors in preparing the school for inspection and in implementing changes in response to inspection. It would have been useful to have information from the teachers as well, but questionnaires could only be distributed to teachers in a few of the countries participating in the study for practical and ethical reasons.

2.1. Selection of European inspectorates of education

The Inspectorates of Education selected for this study are located in Austria (Styria), the Czech Republic, England, Ireland, the Netherlands, and Sweden. These inspectorates share a similar practice – the embodied observation of classrooms and schools – but they differ in a number of areas, as was summarized by Ehren et al. (2013). However, the present study does not aim at comparisons between countries, and the purpose is not to develop a model which fits each and every country. The idea is rather to exploit the variation across countries in school inspection practices to test hypotheses about the mechanisms through which school inspections result in intended and unintended outcomes.

2.2. Selection of schools in each country

Table 1 provides a description of the target and actual sample of schools participating in the survey. Target samples in the Netherlands and England included schools in different inspection

Table 1
Summary of sample of schools in each country.

Country	Targeted sample	Actual sample
Netherlands	Three threshold groups (no risk, risk, high risk) which each consist of 100 schools (50 schools in primary education, 50 schools in secondary education), adding up to a selection of 300 schools in total. Targeted sample (taking into account drop out of schools): 408 primary schools, 359 secondary schools.	Actual sample primary schools: 73 (17.76%) Actual sample secondary schools: 15 (4.18%)
England	Targeted sample: 211 primary and 211 secondary schools that were closest to the threshold for monitoring inspections (the "treatment" in the England study). Logistic regression models were developed to give good estimates as to how close each school was to the threshold for monitoring inspections.	Actual sample primary schools: 62 (29.38%) Actual sample secondary schools: 42 (20.10%)
Sweden	The target sample included a random selection of 1167 primary schools, and 987 secondary schools from the population of 3468 primary schools and 1529 secondary schools.	Actual sample primary schools: 567 (48.59%) Actual sample secondary schools: 464 (47.01%)
Ireland	All schools were included in the target sample: 3200 primary schools and 729 secondary schools.	Actual sample primary schools: 220 (6.9%) Actual sample secondary schools: 51 (7%)
Austria (Styria)	All schools were included in the target sample: 503 primary schools and 194 secondary schools.	Actual sample primary schools: 345 (68.5%) Actual sample secondary schools: 149 (77.2%)
Czech Republic	TIMSS design used to sample of schools: 150 primary schools and 170 secondary schools	Actual sample primary schools: 50 (33%) Actual sample secondary schools: 66 (39%)

arrangements. These schools are expected to respond differently to inspection measures as they face different types of inspection visits and consequences. Schools in the most intensive inspection arrangement (high risk schools in the Netherlands and schools in England receiving monitoring visits) are, for example, inspected more frequently and have to submit improvement plans to the inspectorate. Including these schools in the sample provides a broad range of potential responses. The other four countries used a different sampling design as the inspectorate in these countries schedules regular visits to all schools instead of targeted visits. The designs include either a random sample of schools (Sweden) or use of the TIMSS sampling design (the Czech Republic). Austria (province of Styria) and Ireland have selected all schools due to the small numbers of schools in the population.

Table 1 summarizes the target sample and actual response rates. The table shows that Sweden and Austria have particularly high response rates, while England has low response rates. In the Netherlands, a limited number of secondary school principals responded to the survey. The varying number of participating schools in the different countries is not necessarily a problem, because the aim is not to make inferences about the different countries. However, since varying response rates might result in the inspection arrangements of certain countries dominating the pattern of results case-weights which give equal weight to each of the countries have been used in analysing the data. This method has the disadvantage that it reduces the statistical power of the analysis, and it also requires estimation techniques which provide correct estimates of standard errors from weighted data. Given the large sample and availability of appropriate estimation methods in the Mplus programme (Muthén & Muthén, 1998–2010) we found this to be the best solution to the problem of allowing all countries equal influence on the statistical analysis.

2.3. Data collection

In each country an online questionnaire was administered to principals in primary and secondary education from September to December 2010. The questionnaire included questions on the hypothesized mechanisms of effects of inspection (setting expectations, stakeholder sensitivity to reports, and accepting feedback) and the intermediate mediating mechanisms (improving self-evaluation, improvement actions for capacity-building and improvement actions for school and teaching conditions) in our theoretical framework. Items to measure the improvement of capacity-building of schools were inspired by the Dutch School Improvement Questionnaire (see Geijsel et al., 2009). Items measuring improvement actions for effective school and teaching conditions were inspired by Scheerens (2009) and were adapted from the ICALT questionnaire which was developed by the inspectorates of education in several European countries to measure the quality of teaching and learning, using a shared framework of indicators. Questions on intermediate mechanisms of inspection were inspired by the United Kingdom's National Foundation for Educational Research (NFER) survey 'Evaluation of the impact of Section 5 inspections' (McCrone et al., 2007).

Questions about improvement actions refer to actions the school has taken to develop its capacity to improve and specifically to enhance effective school and teaching conditions; questions are framed in terms of changes in the amount of time principals have spent during the previous academic year to improve the school's functioning in these areas (using a 5-point scale ranging from 'much less time' to 'much more time').

The questionnaire was translated from English into Czech, Dutch, German and Swedish and adapted to include context-specific terminology (e.g., when referring to national inspectorates of education) to allow principals to answer questions in their own language. The questionnaire was also pilot tested in each country. More information on the data collection and the questionnaire can be found on the project website: (link to website).

2.4. Methods of analysis

Structural equation modelling (SEM) (see, e.g., Brown, 2006) was used to analyze the data. One reason for preferring this analytical approach was that it allows estimation of a smaller set of latent variables from a larger set of observed variables, where the latent variables are hypothesized to correspond to constructs in a theoretical model. Another reason for using SEM was that this technique allows estimation of path models among latent variables, where hypothesized relations among chains of variables may be estimated, and where direct and indirect effects may be computed. It is, furthermore, possible to test the fit of the model against the observed data. A good fit provides support for the hypothesized model, even though the correctness of the model is not proven, because there may be alternative models which have equally good, or better, fit.

In the analysis, data from all countries was pooled ($N = 2226$). Given that the current study focuses on the general mechanisms identified in the theoretical model, the aim was not to make a comparative analysis. The hypothesized model was therefore fitted

to the pooled within-groups covariance matrix, which was accomplished through regressing all the manifest variables in the model on dummy variables for the countries. In this way the country means on the manifest variables did not influence the results. The Mplus 6.1 programme (Muthén & Muthén, 1998–2010) was used to estimate the model with the MLR estimator, taking missing data and case weights into account. The missingness was due to a small amount of internal missing data, and also to the fact that only those principals who had recent experiences of school inspections were asked a subset of items pertaining to impressions and evaluations of the inspection.

The fit of the model was estimated with standard measures: the chi-square goodness-of-fit test, the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the Root Mean Square of Approximation (RMSEA) (see, e.g., Brown, 2006). RMSEA values less than 0.05 represent a 'close fit', and models with values above 0.08 should be rejected. For CFI a value of around 0.95 is required to accept a model, and the TLI should be higher than 0.90.

2.5. The measurement model

In the first step of the modelling procedure the measurement model was fitted to the data, during which step relations among latent and observed variables were specified and tested. The questionnaire was constructed in such a way that each construct typically included four to six items. However, because of limitations in computing power, and because of a need to reduce the length of the questionnaire for future use, only two to three items that best reflected each construct were included in the construction of the latent variables.

Another challenge in the construction of the measurement model was the fact that most questionnaire items were classified to measure both a broad construct (e.g., improvement of capacity building) and a narrow (sub) construct (e.g., transformational leadership). Given that both broad and narrow constructs were potentially useful in the construction of the model, there was a need to include both types of latent variables in the path model. This was done through fitting what has been called 'nested-factor' (NF) measurement models (Gustafsson & Balke, 1993; another label is 'bifactor models', see, e.g., Reise, 2012). These are multidimensional, orthogonal models, in which one dimension typically is related to all items, while other items are related to subsets of items representing a narrow latent variable. Such models are conceptually powerful, but they sometimes offer methodological challenges in the form of identification problems. Table 2 presents an overview of the hypothesized latent variables.

The first latent variable (setting expectations) is hypothesized to capture the extent to which the principal sees the inspection norms and standards as influencing professional development, evaluation and supervision of teachers, implementation of long-term improvements and self-evaluation of the school, among other things. The second latent variable (stakeholder sensitivity to reports) represents the extent to which the principal sees school-boards, parents and other stakeholders as being aware of and sensitive to the content of the inspection reports. The third latent variable (accepting feedback) represents the degree to which the principal perceives the feedback received from inspectors as insightful and useful, and the extent to which they are prepared to act on it. The items taken to be indicators of these three latent variables all offered a 5-category Likert scale for responses, with alternatives ranging from 'strongly agree' to 'strongly disagree'. The responses were coded in such a way that 'strongly agree' was scored highest.

The fourth latent variable (improving self-evaluation) is intended to capture the extent to which principals, compared to the previous year, put more or less emphasis on developing and using self-evaluation in the school. The fifth latent variable (improvement actions for capacity building) captures if principals compared to the previous year have increased or decreased the amount of time spent on building capacity in the school, for example through involving teachers in decision making, and improving collaboration and transformational leadership. The next three latent variables are narrower, capturing different aspects of the broader latent variable 'improvement in capacity building'. The ninth latent variable (improvement actions for school effectiveness) represents the degree to which the principals have increased or decreased the amount of time spent on trying to improve learning outcomes, for example through making effective use of teaching time, focusing on content, monitoring the progress of individual students and the school, and improving the extent to which teachers give feedback to students. The next four latent variables represent narrow dimensions nested under the broad latent variable 'improvement in school effectiveness' (see Table 2). The items taken to be indicators of these variables all presented a 5-category Likert scale, with response alternatives ranging from 'much less' to 'much more'. The responses were coded in such a way that 'much more' was scored highest.

The hypothesized latent variables correspond to the constructs of the conceptual model, and they were tested empirically against the collected survey data in a series of confirmatory factor analysis (CFA) models. They were all empirically supported with good model fit (see the Results section below) and acceptable relations between the latent and manifest variables, which are presented in Table 2.

As is clear from our review of the literature, school inspections may also have unintended consequences, and the questionnaire included a set of questions about such effects. Attempts were made to fit a latent variable to the following four items: 'I discourage teachers from experimenting with new teaching methods that do not fit the scoring rubric of the Inspectorate' (Q46), 'School inspections have resulted in narrowing curriculum and instructional strategies in my school' (Q47), 'The latest documents, facts and figures we sent to the inspectorate present a more positive picture of the quality of our school then how we are really doing' (Q49), and 'Preparation for school inspection is mainly about putting protocols and procedures in writing that are in place in the school and gathering documents and data' (Q50). However, the measurement characteristics of these latent variables were unacceptably poor. The items were therefore added to the model as observed variables which were regressed on the latent variables of the path model.

3. Results

Our conceptual model specifies the major hypothesized processes and mechanisms through which school inspections may influence teaching and learning. This model was the starting point for specifying hypothesized paths of influence among the latent variables. These were estimated and tested with a model based on the 37 items and the 13 latent variables defined in the measurement model presented above.

The path model includes two independent variables. One of these is 'setting expectations', reflecting the hypothesis that principals' adoption of the norms and standards used to define quality education by the inspection authority is essential for taking action. The other independent variable is 'stakeholder sensitivity to reports', which is taken to be correlated with 'setting expectations'. It is expected that principals who are aware that stakeholders have information about the extent to which the school meets the norms and standards will feel both pressure and support to take improvement actions to align the schools' practices with the norms and standards. These two independent variables

Table 2
Questionnaire items, latent variables, and factor loadings.

Latent variables and item formulations	Loading on broad factor	Loading on narrow factor
Setting expectations		
The inspection standards affect:		
a. The evaluation and supervision of teachers	0.57	
b. The implementation of long term improvements	0.74	
d. The development of the school plan in which goals for the next academic year are outlined	0.69	
e. The areas of professional development of teachers	0.69	
f. Self evaluation of the school	0.74	
We use the inspection rubric to set new priorities for the future	0.39	
Stakeholder sensitivity to reports		
The school's Boards of Management/Boards of Governors is very aware of the contents of the school inspection report	0.51	
The Parents' Representatives of the school are sensitive to the contents of the school inspection report	0.74	
The Student Representatives of the school are sensitive to the contents of the school inspection report	0.53	
Accepting feedback		
The feedback provided to the school during the last inspection visit was insightful	0.75	
We use the inspection rubric to set new priorities for the future	0.30	
Overall the school was happy with the feedback it received	0.66	
The feedback received from the school inspectors was useful	0.89	
The school in the main will act on the feedback received from the inspectors	0.65	
Improving self-evaluations		
Compared to last academic year, I spent [much less … much more] time on the following aspects of self-evaluation and quality assurance in your school:		
a. The self-evaluation process as a whole	0.78	
b. Developing the quality of our self-evaluation process	0.85	
c. Involving other people in the self-evaluation process	0.68	
Improvement actions for capacity building		
Teacher participation in decision making		
Compared to last academic year, I spent [much less … much more] time involving teachers in making decisions about:		
a. Using new teaching methods	0.54	0.32
b. The curriculum over the different school years	0.63	0.31
c. New educational objectives for the school	0.69	
Teacher cooperation		
Compared to last academic year, I spent[much less … much more] time on improving teachers' collaboration in the following ways:		
a. Discussing new teaching methods with each other	0.56	0.37
b. Discussing assessment results of students with each other	0.51	0.37
Transformational leadership		
Compared to last academic year, I spent [much less … much more] time in my leadership role doing the following:		
a. Communicating the school's vision to the staff, pupils, parents and others	0.37	0.55
b. Referring explicitly to the school's objectives during decision-making processes	0.42	0.54
Improvement actions for school effectiveness		
Opportunity to learn		
Compared to last academic year, I spent [much less … much more] time on improving the extent to which teachers:		
a. Make effective use of teaching time within lessons	0.58	0.37
b. Use specific learning objectives, detailed for specific teaching units and subgroups of individual students, to inform teaching	0.59	0.33
d. Teach content in greater depth during regular teaching hours	0.57	0.10
Assessment of students		
Compared to last academic year, I spent [much less … much more] time on making sure teachers use assessments for the following purposes:		
a. Testing to monitor students' progress	0.45	0.54
b. Using assessment results to set learning goals for individual/groups of students	0.49	0.52
Assessment of school		
Compared to last academic year, I spent [much less … much more] time on using assessment results for the following purposes:		
a. Comparing the school with other schools	0.15	0.56
b. Monitoring the school's progress from year to year	0.37	0.55
Structured teaching		
Compared to last academic year, I spent [much less … much more] time on improving the extent to which teachers		
a. Make use of active teaching methods	0.63	0.44
b. Check whether students understand the lesson content	0.64	0.63
c. Give clear instructions and explanations	0.61	0.36

are therefore hypothesized to influence the principals' willingness to accept and act upon feedback from school inspection. Schools are assumed to use the feedback to improve, with support and pressure from stakeholders. The improvement processes are hypothesized to be informed by self-evaluation, and to be focussed upon capacity building among the teachers, which in turn leads to improvements in school effectiveness. Paths were therefore hypothesized from 'improving self-evaluation' to 'improvement actions for capacity building' and from there to 'improvement actions for school effectiveness'. The latter latent variable therefore also is the dependent variable of the model, even though it conceptually has status as an intermediate variable.

It thus is possible to make a straightforward translation of the conceptual model into a hypothesized set of relations among the main latent variables. However, given that it was not possible to derive explicit hypotheses for relations with the narrow latent variables, a parsimonious model was first fitted, and then a few more relations were added to the model. With these amendments, the model fitted the data very well (Chi-square = 778.04, df = 583, CFI = 0.986, TLI = 0.979, RMSEA = 0.012, 90% CI 0.010 – 0.014). Fig. 2 presents the relations which were statistically significant at the .05 level in the form of standardized regression coefficients. The four items measuring unintended consequences were included in the analysis but are not shown in the diagram to increase readability.

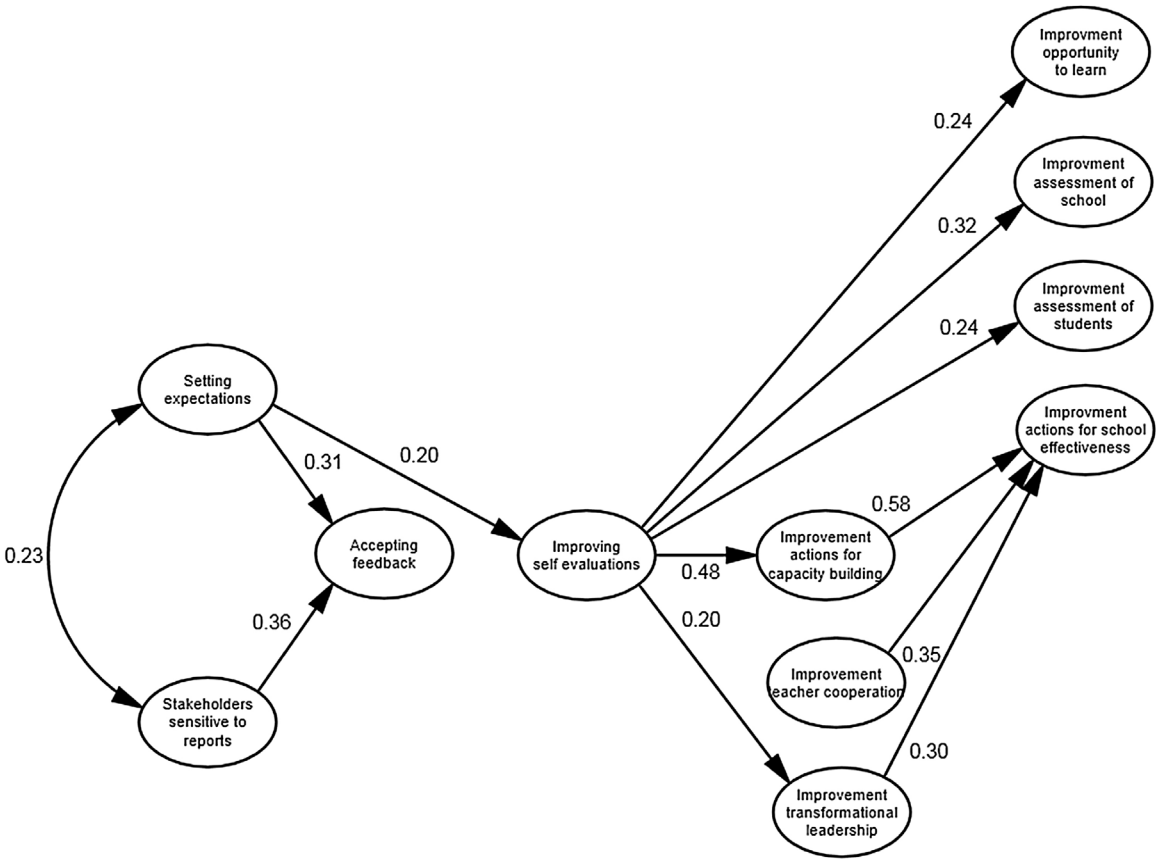

Fig. 2. Path model based on the conceptual framework.

Fig. 2 shows that 'setting expectations' is directly related to 'accepting feedback' and to 'improving self-evaluation'. This supports the hypothesis that school inspection drives improvement through a process of setting expectations for schools. Setting expectations also promotes the utilization of self-evaluation in schools, which in turn impacts on the schools' actions to improve capacity building in key areas, particularly transformational leadership. These improvement efforts are in turn related to actions to improve the effectiveness of the school's teaching and learning. Enhancing self-evaluation also impacts indirectly on actions to improve school effectiveness by way of the development of transformational leadership. Improving self-evaluation also has direct effects on improvement efforts with respect to opportunity to learn and assessment of the students.

Fig. 2 also shows that 'stakeholder sensitivity to reports' is related to 'accepting feedback'. Thus, the degree to which the schools' stakeholders are sensitive to reported outcomes of inspection influences the readiness with which principals see feedback from the inspection as useful for improvement efforts. However, contrary to what was hypothesized in the conceptual framework, the school's acceptance of inspection feedback does not seem to be an important condition for improvement after school inspections. Schools that feel that the inspectorate sets clear expectations regarding quality education and have stakeholders who are sensitive to inspection reports do indicate a higher level of accepting feedback, but accepting feedback does not of itself lead to any improvement actions.

The path diagram only presents estimates of direct effects of one latent variable upon another latent variable. While the pattern of direct effects indicates the importance of the independent variable 'setting expectations' in determining actions taken to 'improving self-evaluation', these estimates do not provide a summary estimate of their impact. However, this can be obtained through computation of the indirect effects that a particular independent variable exerts via effects on one or more other variables. 'Setting expectations' had a total effect .11 ($t = 3.85$) on 'improvement actions to build capacity' which was mediated via 'improving self-evaluation'; a total effect of .04 ($t = 2.45$) on 'improvement in transformational leadership', which also was mediated via 'improving self-evaluation'; a total effect of .09 ($t = 3.68$) on 'improvement actions for school effectiveness', which was mediated via 'improving self-evaluations' and 'improvement actions to build capacity'. 'Setting expectations' also had significant total effects around .05 to .07 on 'improvement in opportunity to learn' and 'improvement in assessment' of both students and the school, which were mediated via 'improving self-evaluations'.

These results therefore support the hypothesis that the principal's adoption of, and adaptation to the requirements stipulated in inspection standards and procedures is a determinant of actions taken to improve the school in terms of capacity and school effectiveness.

We now turn to the issue of unintended consequences of school inspections. Q46 asks about the effects of school inspection on experimentation with new teaching methods. There was a positive total effect (.09) of 'setting expectations' which was due both to a small (.02) but significant indirect effect via 'improving self-evaluations' and to a direct effect of 'setting expectations' (.07). These relations imply that principals who have high values on the 'setting expectations' latent variable tend to agree that they discourage teachers from experimenting with new teaching methods which are not in line with the inspectorate's scoring rubric.

Q47 asks if school inspections have resulted in narrowing of curriculum and instructional strategies. There was no significant total effect of 'setting expectations', which was because there was a

highly significant negative indirect effect (−.08) via 'accepting feedback', and a positive direct effect (.06). These results indicate that principals who have adopted the standards and practices of the inspectorate tend to see the inspections as causing a narrowing of the curriculum and instructional strategies. However, the principals who are willing to accept feedback see less negative consequences than do those who are less willing to accept feedback. 'Stakeholder sensitivity to reports' also has a highly significant negative total effect (−.10) on the narrowing of the curriculum, due to a negative indirect effect via 'accepting feedback'.

Q49 asks principals if the material sent to the inspectorate presents an overly positive picture of the school. For this item there were no effects, either from 'setting expectations' or from 'stakeholder sensitive to reports'.

Q50 states that preparation for school inspections is mainly about gathering documents and data and putting in writing the schools' policies and procedures. For this item there was a negative direct effect of 'accepting feedback', and negative indirect effects of both 'setting expectations' (−.09) and 'stakeholder sensitivity to reports' (−.11). These results demonstrate that principals who accept feedback tend to reject the view that preparation for inspections is mainly about gathering documents and data.

In summary these analyses of unintended consequences suggest that the mechanism of setting expectations not only exerts influence on improvement processes but also results in narrowing of curriculum and instructional strategies, and discourages experimentation with new instructional methods. The results also show that principals who are ready to accept feedback do not see any such consequences, and that they reject statements implying a degree of 'window dressing'.

4. Discussion and conclusions

The purpose of this research is to identify and analyze the ways in which school inspection impacts on the work of schools, driving change and bringing about improvement. In the theoretical framework, which was based on analyses of the programme theories of the six inspectorates, it was hypothesized that certain mechanisms stressed by most inspection rubrics might be important in influencing schools to improve. These were the setting of clear expectations, standards and norms by the inspectorate, the provision and utilization of post inspection feedback and guidance, and the pressure for improvement brought to bear by key stakeholders. It was further hypothesized that these mechanisms for change might be operationalized in the schools by what was defined as intermediate mechanisms. These include wider and more systematic self-evaluation in schools, greater capacity to improve through enhanced transformational leadership by principals, greater collaborative work among teachers and improvement of effective school and teaching conditions such as the use of better instructional and assessment methods to monitor student progress and enhance outcomes. Finally, the theoretical framework included the hypothesis that inspection may also lead to unintended and potentially negative consequences with schools becoming averse to taking risks and thus limiting new approaches to instruction or emphasizing more easily measurable learning outcomes at the expense of creativity and experimentation.

The analysis of data from across the six countries provided partial support for the hypotheses through the path model analysing the relationships between them and their influence on improvement actions. The results of the path model indicate that clear inspection expectations are determinants of improvement actions. Such expectations are also significantly related to wider use of self-evaluation, which indicates that schools see systematic self-evaluation as a vital developmental strategy when responding to school inspection. In fact the significant role of self-evaluation as integral part of the inspection process across the systems studied is a very noteworthy feature of this research.

Stakeholders' sensitivity to inspection reports is also related to the setting of expectations in schools. The hypothesis was that schools feeling pressured by parents, school boards, and students to act on inspection findings, are more ready to accept feedback and inspection standards. However contrary to the expectation in the conceptual framework, stakeholders seem to have most impact on the response of schools early in the improvement cycle through motivating schools to accept feedback and inspection standards, rather than influencing actions after the publication of the inspection report. This is difficult to separate out clearly, however, since setting criteria, conducting inspection, giving feedback and publishing reports are essentially cyclical processes.

Accepting feedback did not influence any of the mediating variables, even though it is interesting to note that principals who are ready to accept feedback reject statements about 'window dressing' in preparations for inspection. However, previous research indicates that acceptance and use of feedback is moderated by whether the feedback is positive or negative, which may explain why there was no overall effect of this variable on improvement actions. Regrettably these data do not include information about the outcome of the inspection for the different schools, so they do not allow us to test this hypothesis. This is, however, an important task for further research.

Also in agreement with the conceptual framework, there is a strong relationship between implementing self-evaluation, taking improvement actions for capacity building and implementing transformational leadership. These capacity building actions have, in turn, a strong relationship with improvement actions for school effectiveness.

Neo-institutional theories would predict that accountability pressure in uncertain situations may induce schools to turn to standard solutions and shy away from experimentation, flexibility and adaptation to specific contexts (DiMaggio & Powell, 1983). The findings of this study seem to support this line of reasoning, since the mechanism of 'setting expectations' not only exerts influence on improvement processes but also results in narrowing of curriculum and instructional strategies, and discourages experimentation on instructional methods.

In summary the results of the study largely support the hypothesized mechanisms suggesting that the impact of school inspections on the quality of education is driven by the setting of expectations, standards and norms, with self-evaluation and encouragement of capacity building and better teaching and learning as mediating mechanisms. The results still leave some questions open, but the study offers ideas as to how to answer these in further research.

Although the study has allowed testing of the conceptual model against large-scale survey data, there are a number of limitations which need to be discussed. One limitation is the use of a cross-sectional design, which is not optimal for purposes of making causal inferences as such designs do not take into account the time frame in which causes precede effects. This suggests that the relations in our path model may not represent causal links between the variables, but that these should rather be interpreted as associations. A longitudinal design would be better suited for attempts to make inferences about causal relations. However, in the questionnaire we asked the principals to make assessments of the change in amount of time spent on different kinds of activities and efforts. Even though such retrospective assessments cannot provide information of the same level of quality as a longitudinal study does, they do provide information about change that should approximate that obtained from a longitudinal design. Capturing such changes to a certain extent strengthens the claims that we are empirically evaluating mechanisms, rather than just estimating associations among variables

A second limitation of the study is that it entirely relies on the perceptions and reports of the principals. Self-reports, and particularly so when done in retrospect, are known to be less than perfectly reliable and valid, being subject to influences from, for example, memory loss and social desirability bias. It is difficult to assess the impact that this may have had on the results, but it cannot the ruled out that bias in reporting has affected the pattern and size of relations in the model. In future research it therefore is essential to include information also from other sources, such as teachers and external observers. In the current study we also have not had access to information about student achievement, but expect that such information will be available for some of the participating countries at a later point in time.

A third limitation is the low response rate in some of the participating countries, which may threaten the validity of inferences from the study. However, as has already been emphasized the study does not have comparative purposes, but the aim is rather to take advantage of the range of school inspection arrangements in different countries. Nevertheless, the non-response reduces the statistical power of the study, as does the weighting used to assign equal weight to the participating countries. Absence of significant relations should therefore be interpreted with caution.

A fourth limitation concerns the generalizability of the findings, which is threatened both by the non-response of participants and by the fact that only six school-inspections systems were included in the study. While these six systems represent considerable variation, they certainly cannot be considered representative of all school-inspection systems. However, it should be emphasized that even though this is a large-scale quantitative study based on representative sampling, the main focus is on understanding the mechanisms through which school inspections relate to school improvement.

To avoid the limitations discussed above, it would seem essential in further research to adopt designs which allow stricter tests of the hypothesized mechanisms, such as longitudinal and quasi-experimental designs. It will be important to extend the range of outcomes investigated to also include student achievement. It also is essential to involve multiple categories of informants and to use a varied set of observational techniques. Qualitative methods, for example in the form of case studies, would also be useful for getting insight into the mechanisms through which school inspections exert influence on processes and outcomes of schooling.

Let us, finally, emphasize some implications of our findings. In terms of influence on inspection policy and practice the finding that inspection results in significant impact on schools where clear expectations are set and where stakeholders are knowledgeable about and engaged with the process suggests that the creation and maintenance of this context is a key task for the inspectorates. The research also suggests that processes such as self-evaluation, transformational leadership and collaborative staff activities are important and effective. Again these processes seldom flourish without encouragement and resources to facilitate schools in developing capacity in these areas. It follows that engagement in moving schools' teachers to embrace such processes needs to be central to inspection policy.

Acknowledgements

The research reported herein has been conducted within the project "Impact on school inspections on teaching and learning" (project leader Melanie Ehren), co-financed by the EC Eurydice Lifelong Learning Programme.

References

Allen, R., & Burgess, S. (2012). *How should we treat under-performing schools? A regression discontinuity analysis of school inspections in England* Bristol: University of Bristol, Centre for Market and Public Organisation, Bristol Institute of Public Affairs.

Blok, H., Sleegers, P., & Karsten, S. (2008). Looking for a balance between internal and external evaluation of school quality: Evaluation of the SVI model. *Journal of Education Policy, 23*(4), 379–395.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.

Belfield, C. R., & Levin, H. N. (2009). Market reforms in education. In G. Sykes, B. Schneider, D. N. Plank, & T. G. Ford (Eds.), *Handbook of education policy research* (pp. 513–527). New York: Routledge.

Bubb, S., & Earley, P. (2008). *From self-evaluation to school improvement: The importance of effective staff development*. Reading, Reading: CfBT Education Trust.

Buckley, J., & Schneider, M. (2003). Making the grade: Comparing DC charter schools to other DC public schools. *Educational Evaluation and Policy Analysis, 25*, 203–216.

Dedering, K., & Muller, S. (2011). School Improvement through Inspections? First Empirical Insights from Germany. *Journal of Educational Change, 12*(3), 301–322.

Denessen, E., Driessen, G. W. J. M., & Sleegers, P. (2005). Segregation by choice? A study of group-specific reasons for school choice. *Journal of Education Policy, 20*(3), 347–368.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*(2), 147–160.

Dobbelaer, M. J., Prins, F. J., & van Dongen, D. (2013). The impact of feedback training for inspectors. *European Journal of Training and Development, 37*(1), 86–104.

Ehren, M. C. M., Altrichter, H., McNamara, G., & O'Hara (2013). Impact of school inspections on improvement of schools – Describing assumptions on causal mechanisms in six European countries. *Educational Assessment, Evaluation and Accountability, 25*(1), 3–43.

Ehren, M. C. M., & Visscher, A. J. (2008). The relationship between school inspections, school characteristics and school improvement. *British Journal of Educational Studies, 56*(2), 205–227.

Elmore, R. F., & Fuhrman, S. H. (2001). Holding schools accountable: Is it working? *Phi Delta Kappan, 83*(1), 67–72.

Elstad, E. (2009). Schools which are named, shamed and blamed by the media: School accountability in Norway. *Educational Assessment, Evaluation and Accountability, 21*, 173–189.

Emstad, A. B. (2011). The principal's role in the post-evaluation process – How does the principal engage in the work carried out after the schools self-evaluation? *Educational Assessment, Evaluation and Accountability, 23*(4), 271–288.

Eurydice (2004). *Evaluation of schools providing compulsory education in Europe*. http://www.eurydice.org/portal/page/portal/Eurydice

Fullan, M. (2002). The change leader. *Educational Leadership, 59*(8), 16–20.

Gärtner, H., Füsemann, D., & Pant, H. A. (2009). Wirkungen von Schulinspektion aus Sicht betroffener Schulleitungen. *Empirische Pädagogik, 23*, 1–18.

Geijsel, P., Sleegers, P. J. C., Stoel, R. D., & Kruger, M. L. (2009). The effect of teacher psychological and school organizational and leadership factors on teachers' professional learning in Dutch schools. *The Elementary School Journal, 109*(4), 1–22.

Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28*(4), 407–434.

Hanushek, E. A., & Raymond, M. E. (2002). Lessons about the design of state accountability systems. *Paper prepared for 'Taking Account of Accountability: Assessing Policy and Politics', Harvard University*.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved performance? *Journal of Policy Analysis and Management, 24*(2), 297–327.

Heubert, J. P., & Hauser, R. M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington: National Academy Press.

Hofman, R. H., de Boom, J., & Hofman, H. W. (2010). Quality control in primary schools: Progress from 2001–2006. *School Leadership & Management, 30*(4), 335–350.

Husfeldt, V. (2011). Wirkungen und Wirksamkeit der externen Schulevaluation. *Überblick zum Stand der Forschung. Zeitschrift für Erziehungswissenschaft*. http://dx.doi.org/10.1007/s11618-011-0204-5

Hussain, I. (2012). *Subjective performance in the public sector: Evidence from school inspections: London School of Economics and Political Science*. Centre for Economic Performance.

Karagiorgi, Y. (2012). School self-evaluation at an embryonic stage: Depicting teachers' experiences with a participative project. *International Journal of Leadership in Education: Theory and Practice, 15*(2), 199–220.

Koretz, D. M. (2003). Using multiple measures to address perverse incentives and score inflation. *Educational Measurement, 22*(2), 18–26.

Luginbuhl, R., Webbink, D., & De Wolf, I. (2009). Do inspections improve primary school performance? *Educational Evaluation and Policy Analysis, 31*(3), 221–237.

Malen, B. (1999). On rewards, punishments, and possibilities: Teacher compensation as an instrument for education reform. *Journal of Personnel Evaluation in Education, 12*(4), 387–394.

MacBeath, J. (2000). *Self-evaluation in European schools: A story of change*. London: Routledge Falmer.

Matthews, P., & Sammons, P. (2004). *Improvement through Inspection. An evaluation of the impact of Ofsted's work*. London: Ofsted.

McCrone, T., Rudd, P., Blenkinsop, S., Wade, P., Rutt, S., & Yeshanew, T. (2007). *Evaluation of the impact of section 5 inspections*. Slough: NFER.

McNamara, G., O'Hara, J., Lisi, P. L., & Davidsdottir, S. (2011). Operationalising self-evaluation in schools: Experiences from Ireland and Iceland. *Irish Educational Studies, 30*(1), 63–82.

Meuret, D., & Morlaix, S. (2003). Conditions of success of a school's self-evaluation: Some lessons of an European experience. *School Effectiveness and School Improvement, 14*(1), 53–71.
Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology, 83*(2), 340–363.
Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
Nelson, R., & Ehren, M. C. M. (2014). *Review and synthesis of evidence on the (mechanisms of) impact of school inspections*. http://schoolinspections.eu/wp-content/uploads/downloads/2014/02/Review-and-synthesis-of-evidence-on-the-mechanisms-of-impact-of-school-inspections.pdf
Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives, 14*(1). Available from http://epaa.asu.edu/epaa/v14n1
Ofsted (2009). *Parents' perceptions of Ofsted's work—a report of the NFER for Ofsted*. London: Ofsted.
Ouston, J., Fidler, B., & Earley, P. (1997). What do schools do after OFSTED school inspections-or before? *School Leadership & Management, 17*(1), 95–104.
Powell, W. W. (2007). The new institutionalism. In S. Clegg & J. R. Bailey (Eds.), *The International Encyclopedia of Organization Studies*. Sage http://www.stanford.edu/group/song/papers/NewInstitutionalism.pdf (acessed 1.8.12).
Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696.
Rosenthal, L. (2004). Do school inspections improve school quality? *Economics of Education Review, 23*(2), 143–151.
Scheerens, J. (2009). *Review and meta-analyses of school and teaching effectiveness*. The Netherlands: Department of Educational Organization and Management, www.iqb.hu-berlin.de/lehre/dateien/rapportScherens.pdf
Schildkamp, K., & Visscher, A. (2010). The utilisation of a school self-evaluation instrument. *Educational Studies, 36*(4), 371–389.
Schildkamp, K., Vanhoof, J., Petegem, P., & Visscher, v. A. J. (2012). The use of school self-evaluation results in the Netherlands and Flanders. *British Educational Research Journal, 38*(1), 125–152.
Scott, W. R. (2008). *Institutions and organizations. Ideas and interests*. Los Angeles: SAGE.
Stecher, B. M. (2002). Consequences of large-scale, high-stakes testing on school and classroom practices. Tests and their use in test-based accountability systems. In L. S. Hamilton, B. M. Stecher, & S. P. Klein (Eds.), *Making sense of test-based accountability in education*. Santa Monica: Rand Cooperation, http://www.rand.org/pubs/monograph_reports/MR1554/
Van Petegem, P., & Vanhoof, J. (2007). Towards a model of effective school feedback: School heads' points of view. *Educational Research and Evaluation, 13*(4), 311–325.
Verhaeghe, G., Vanhoof, J., Valcke, M., & Van Petegem, P. (2010). Using school performance feedback: Perceptions of primary school principals. *School Effectiveness and School Improvement, 21*(2), 167–188.
Visscher, A. J., & Coe, R. (2003). School performance feedback systems: Conceptualisation, analysis, and reflection. *School Effectiveness and School Improvement, 14*(3), 321–349.
Whitby, K. (2010). *School inspections: Recent experiences in high performing education systems*. Reading, England: CfBT Education Trust.