

Comparing effects and side effects of different school inspection systems across Europe

M.C.M. Ehren^{a*}, J.E. Gustafsson^b, H. Altrichter^c, G. Skedsmo^d, D. Kemethofer^c and S.G. Huber^d

^a*Institute of Education, University College London, London, UK;* ^b*Department of Education and Special Education, University of Gothenburg, Gothenburg, Sweden;* ^c*Institut für Pädagogik und Psychologie, Johannes Kepler Universität, Linz, Austria;* ^d*IBB – Institut für Bildungsmanagement und Bildungsökonomie, Pädagogische Hochschule Zug, Zug, Switzerland*

In this article, different inspection models are compared in terms of their impact on school improvement and the mechanisms each of these models generates to have such an impact. Our theoretical framework was drawn from the programme theories of six countries' school inspection systems (i.e. the Netherlands, England, Sweden, Ireland, the province of Styria in Austria and the Czech Republic). We describe how inspection models differ in the scheduling and frequency of visits (using a differentiated or cyclical approach), the evaluation of process and/or output standards, and the consequences of visits, and how these models lead to school improvement through the setting of expectations, the use of performance feedback and actions of the school's stakeholders. These assumptions were tested by means of a survey of principals in primary and secondary schools in these countries ($n=2239$). The data analysis followed a three-step approach: (1) confirmatory factor analyses, (2) path modelling and (3) fitting of multiple-indicator multiple-cause models. The results indicate that Inspectorates of Education that use a differentiated model (in addition to regular visits), in which they evaluate both educational practices and outcomes of schools and publicly report inspection findings of individual schools, are the most effective. These changes seem to be mediated by improvements in the schools' self-evaluations and the schools' stakeholders' awareness of the findings in the public inspection reports. However, differentiated inspections also lead to unintended consequences as principals report on narrowing the curriculum and on discouraging teachers from experimenting with new teaching methods.

Introduction

With many European education systems decentralising decision-making to schools and establishing new evidence-based accountability regimes, school inspections are becoming increasingly important and 'modern' in their operation. According to Altrichter, Kemethofer, and Schmidinger (2013), 'new inspection systems' concentrate on the evaluative functions, and aim to professionalise these evaluative functions by formalising and proceduralising them and by enriching their operation by using research instruments from social science. Inspection systems include school stakeholders in their data collection and presentation of inspection findings, thus alerting them to issues regarding school quality.

*Corresponding author. Email: m.ehren@ioe.ac.uk

By virtue of these features ‘new inspections’ fit well with the image of rationalised control, propagated by the proponents of evidence-based governance. School inspections are seen as an important tool in bridging the distance between the central government and the locally produced and delivered public services (Clarke 2011). Inspections are one way of ensuring that strategies and requirements formulated by the top levels of government translate into appropriate processes and structures in schools. Inspection systems can also be used to probe and explore the current state of schools according to specific quality criteria (Ehren et al. 2013). By producing data in a more localised or contextualised way than other monitoring instruments (e.g. large-scale student assessment), school inspections are expected to effectively promote school improvement as they often incorporate context-rich means for development actions (Gärtner and Pant 2011; Ehren et al. 2013).

Internationally, school inspections are institutionalised and practised in a variety of ways (Clarke 2011). Different national political contexts, education systems and governing apparatuses embed the inspection system into existing structures which, consequently, result in variations of governance procedures, inspection methodologies and mechanisms of impact. Clarke and Ozga (2011), for example, distinguish between ‘soft’ and ‘hard’ governance forms where soft governance operates through attraction, that is, drawing people in to take part in processes of mediation, brokering and ‘translation’, creating networks and partnerships of actors that rely on self-evaluations, giving good examples and learning from expert knowledge. Contrasting approaches of hard governance include target-setting, performance management, benchmarks and indicators, and data use to foster competition. In both ways of ‘hard’ and ‘soft’ governance, school inspections mirror a countries approach (e.g. by using thresholds to categorise failing schools in ‘hard’ governance systems).

Previous studies on the impact of school inspections have generally not looked at the impact of different inspection models, but instead primarily focused on how one inspection model leads to improvement. Klerks’ review (2013), for example, describes how school inspections may have an effect on behavioural change of teachers, on various school improvement measures and on student achievement results. Her study presents evidence that school inspections promote the improvement of educational quality in schools leading to better student outcomes. Other studies (e.g. Luginbuhl, Webbink, and De Wolf 2007) also found that test scores of pupils in primary education improved by 2–3% of a standard deviation in the two years following a school inspection.

Recent reviews (De Wolf and Janssens 2007; Klerks 2013) also highlight the unintended consequences of school inspections, such as a narrowing of the curriculum, teaching to the test and raising performance for the inspection, but allowing it to decline afterwards. These actions may negatively affect student achievement overall. Rosenthal (2004) reports, for example, a decrease in examination results in English secondary schools in the year following a school inspection. These studies neither specify which models of school inspections have an impact nor do they look at the mechanisms through which these models work. This knowledge is important in understanding which types of inspection are most effective in improving school quality. The study presented in this paper aims to enhance our understanding of effective school inspections.

We selected six Inspectorates of Education for this study that use different approaches; these differences fit within the continuum of ‘hard’ versus ‘soft’ governance (Altrichter and Kemethofer 2015). England and the Netherlands are typical

examples of hard governance as they focus on output data to schedule targeted inspection visits to potentially failing schools, and they actively inform the schools' stakeholders about performance of schools through the publication of inspections reports and (in the Netherlands) lists of failing schools. These two Inspectorates of Education also have a clear set of standards around school quality and output to inform inspection judgements. The move away from relying on schools' self-evaluations to inform inspection assessments, but instead primarily relying on student achievement data has increasingly shifted these inspections towards a 'hard' governance approach.

Contrasting models of school inspections that fit a 'soft' governance approach can be found in Austria, and to some extent in Sweden, Ireland and the Czech Republic where Inspectorates of Education regularly visit schools to provide feedback on strengths and weaknesses, often without an overall classification of schools as failing or well-performing and with no or limited consequences for failing schools. In Austria, self-evaluations and the school's development programme are an explicit starting point for school inspections.

In this article, we compare different inspection models in terms of their impact on school improvement and the processes each of these models generates to have such an impact. The following questions will guide our analysis:

- (1) What impact (intended and unintended consequences) do different types of school inspection models have on schools?
- (2) Through what causal mechanisms can this impact be explained?
- (3) Are these causal mechanisms different for different inspection models?

Theoretical model: distinguishing inspection models and their (mechanisms of) impact

As this study focuses on the impact of school inspections, identifying key characteristics of school inspections that have an impact on school improvement is central to our model. According to Ehren et al. (2013), the frequency of visits, the standards and thresholds used to evaluate schools during inspection visits, and the sanctions and rewards used to improve schools appear to be the dominant aspects of school inspections affecting change in schools. We will therefore focus on how inspection models differ on these key characteristics, and how these motivate different mechanisms of impact and lead to varied desired effects and unintended consequences.

Table 1 summarises the Inspectorates of Education in our study according to the key characteristics in Figure 1 (frequency of visits, standards, consequences and public reporting). These six models are very broad but capture the most prominent distinctions in how school inspections are implemented in these six countries. The distinction is also relevant for other countries in Europe as can be seen from van Bruggen's (2010) description of inspection profiles for SICI (the European Association of Inspectorates of Education).

The following section describes these different models in more detail, and explores the expected impact and the mechanisms of impact of these models.

Frequency of inspection visits: cyclical versus differentiated inspections

In differentiated inspection models, the frequency of inspections depends on an analysis of documents and/or student achievement results (including self-evaluation documentation) submitted to the Inspectorate of Education. If poor school quality is suspected,

Table 1. Summary of inspection characteristics in participating countries (Ehren et al. 2013, 32).

	The Netherlands	England	Sweden	Ireland	Austria (Styria)	The Czech Republic
<i>Frequency of visits</i>						
Cyclical inspections of all schools	Every 4 years	Every 5 years	Every 4–5 years	Every 5 years	Every 2–4 years	Every 3 years
Differentiated inspections	X	X	X			
<i>Standards</i>						
Legal aspects	X	X	X	X	X	X
Process quality	X	X	X	X	X	X
Output/outcomes	X	X				X
Threshold for distinguishing failing schools	X	X				X
<i>Consequences</i>						
(Advising on) sanctions	X	X	X			X
Interventions	X	X	X	X	X	X
<i>Reporting on individual schools to the general public</i>	X	X		X		

Notes: X indicates the presence of characteristic in each country at the time of our study in 2010. The descriptions apply to school inspections during the time of our study in 2010.

an inspection will be scheduled and more frequent inspections will usually be called for. Differentiated school inspections are generally implemented to increase the efficiency of school inspections by targeting inspection resources to potentially failing schools. In some of the countries we studied, Inspectorates of Education use both cyclical and differentiated inspections. In these cases, cyclical inspections include short visits which look at a set of basic inspection standards, while the entire inspection framework is used in differentiated school inspection visits.

Inspectorates of Education that use differentiated inspections in addition to cyclical visits are expected to have greater impact on schools. The targeted approach allows a more efficient use of inspection resources, and allows some inspection resources to be redirected to support schools that are most in need of improvement. The (potentially) failing schools are expected to use the inspection feedback to improve the quality of their schools, while well-functioning schools that are visited on a regular basis are expected to continuously use the inspection standards in their school self-evaluations, in their own planning and in their daily practices.

All of the Inspectorates of Education in our study schedule cyclical school inspections. The Irish, Czech and Styrian Inspectorates of Education carried out whole school evaluations of all schools at the time of our study. Recent policy changes in Ireland have seen the introduction of ‘incidental inspections’; however, at this moment it is not clear if re-inspection will be part of this process. In Styria and the Czech Republic,

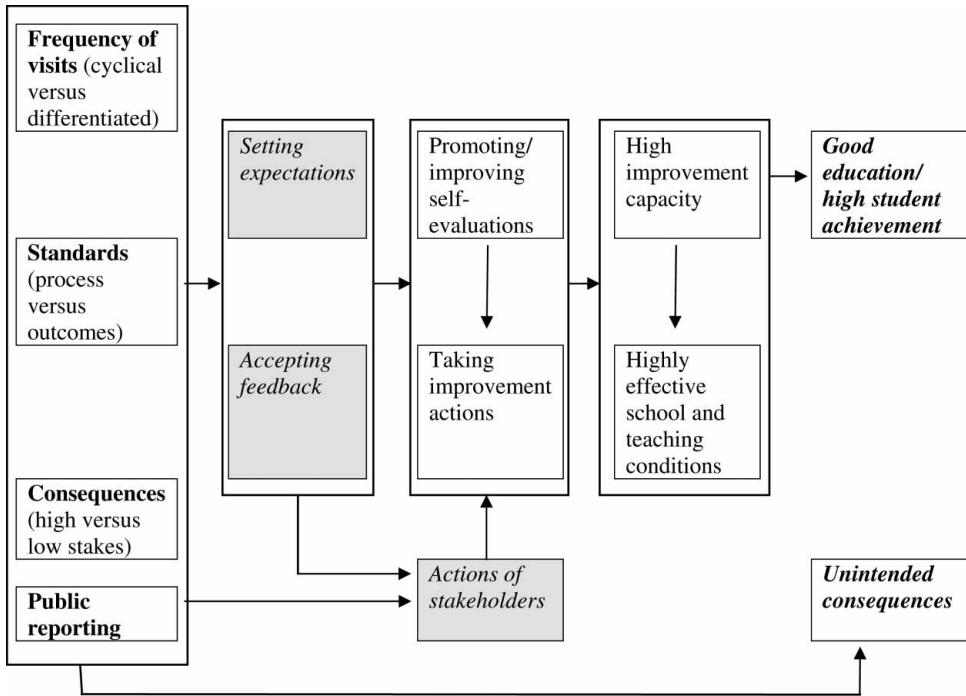


Figure 1. Framework of causal mechanisms of school inspections.

the original schedule of full cyclical inspections proved too ambitious and led to changes (e.g. a downscale of the schedule including additional criteria and strategies for selecting schools for inspection in Styria and a decrease of inspection days and the number of school inspectors in small schools in the Czech Republic). At the time of our study, however, these more differentiated approaches were not common practice.

The Inspectorates of Education in the Netherlands, England and Sweden have a longer tradition of implementing a differentiated schedule of visits in addition to cyclical visits. The Netherlands, for example, uses early warning analyses to schedule inspection visits to potentially failing schools, while each school also receives at least one inspection visit every four years in which specific areas of concern or national targets are evaluated. Similarly, the English Inspectorate of Education, the Office for Standards in Education (Ofsted), conducts regular inspection visits to all schools, while 40% of schools graded as satisfactory and all other schools graded as inadequate receive further monitoring inspections. In Sweden, regular supervision includes basic inspection visits to all schools once every four to five years, while schools that are evaluated as weak receive more extensive school inspections. The selection of schools for these ‘widened inspections’ is based on grades and results on national tests, observations made in previous inspections, complaints and questionnaire responses from students, parents and teachers.

Standards: process versus outcomes

Standards are used by Inspectorates of Education to assess the quality of schools. They lay out what is expected of schools and determine what aspects of schools are necessary

to look at when evaluating school quality. Inspection frameworks generally include standards to monitor the school's compliance to regulations, process standards which emphasise the principles and practices of good education, and standards on the school's outcomes of student achievement in core subjects. Thresholds are used by a number of Inspectorates of Education to grade schools as 'successful', 'satisfactory' or 'failing'.

Inspectorates of Education that evaluate outcomes in addition to processes are expected to have greater impact on schools, compared with Inspectorates that evaluate processes and compliance to legislation alone. This broader approach is expected to stimulate schools to focus on wider goals and to lessen incentives to engage in strategic behaviours that potentially lead to unintended consequences (Barber 2004; Ladd 2007).

All of the Inspectorates of Education in our study evaluate process standards, such as the quality of teaching and the learning climate of the school. They also monitor the schools' compliance to regulations, such as meeting safety regulations or having annual education plans. The Inspectorates of Education report on the strengths and weaknesses of each school according to the inspection framework and indicate where schools fail to comply with legislation. The Inspectorates of Education in the Netherlands, England and the Czech Republic additionally include outcomes of schools (using aggregated student test scores) in their evaluations and use a threshold to identify failing schools. These outcomes include cognitive scores on a number of subjects (generally mathematics, reading and writing), and, in England, also the cultural and social development of students. Thresholds include school ratings, such as in England where Ofsted grades schools as 'outstanding', 'good', 'satisfactory' or 'inadequate'. This type of overall summary assessment is not given to schools in Styria, Sweden and Ireland.

Consequences of school inspections: sanctions versus no sanctions

Schools that are evaluated as 'failing' may receive additional support, but also often face punitive consequences, such as sanctions or interventions (van Bruggen 2010; Ehren et al. 2013). Sanctions may include the public reporting of school status (online) or restructuring, merging of schools, reconstitution or even closure. School inspectors may increase their monitoring of these schools via specific improvement plans which the schools are required to implement. Consequences of school inspections can also include rewards for high-performing schools, financial bonuses or the opportunity to take over another school.

'High stakes' accountability systems work on the assumption that rational actors will be motivated by the threat of sanctions for failing to meet given standards. In this scenario, 'high stakes' pressure will increase the chance that schools seriously attend to inspection standards ('setting expectations') and to the inspection results and feedback ('accepting feedback'), and that stakeholders will become aware of inspection norms and results and feel obliged to act. According to Meyer and Rowan (2006), evidence is growing that the imposition of clear standards of performance, when coupled with the presence of sanctions, produces real conformity to technical rules and guidance. Schools seem to work harder to perform well when something valuable is to be gained or lost; information and feedback alone is seen as insufficient to motivate schools to perform to high standards (Malen 1999; Elmore and Fuhrman 2001; Nichols, Glass, and Berliner 2006). However, the more 'high stakes' a system is, the more a school will suffer if it does not deliver the expected results, which increases the chance that these schools will turn to strategic behaviour to improve

their results (Koretz 2003). Results of these studies suggest that consequences (sanctions but also rewards) have a positive effect on the improvement of schools, and ‘high stakes’ school inspections are more effective than ‘low stakes’ inspection models.

Aside from the Swedish Inspectorate, which may withdraw the licence and funding of independent schools and may temporarily close down public schools, the Inspectorates of Education in our study are not in a position to sanction schools directly. In the Netherlands, England and the Czech Republic, however, Inspectorates of Education may advise the Minister of Education to impose sanctions on failing schools (e.g. to remove the school from the Register of Educational Facilities, or to impose administrative or financial sanctions).

Yet Inspectorates of Education (with the exception of the Irish Inspectorate; however, policy here may be changing) do intervene in schools that are judged to be failing. Ofsted, for example, categorises schools as being in ‘special measures’ if the school is evaluated as inadequate and does not have the capacity to improve; or it gives a school ‘a notice to improve’ when it is performing below expectations. In the Netherlands, schools are monitored intensively when they are performing below expectation, and school boards are expected to develop an improvement plan based on inspection results. The Dutch Inspectorate monitors the implementation of this plan. In Sweden, struggling schools need to amend identified shortcomings within a specified time frame, and the implementation of improvements is inspected at a follow-up visit. In Styria and Ireland, all schools have to develop an improvement plan, even if they are not considered to be failing. This plan serves as a target agreement between the principal and the inspector, and school inspectors check on the implementation of these targets after one or two years. However, there are no consequences in place for schools that fail to implement these targets. In the Czech Republic, failing schools are monitored more frequently and are obliged to implement corrections that have been identified by the Inspectorate.

Reporting: inspection systems with/without public reporting

The Inspectorates of Education in the Netherlands, England, Sweden and Ireland publish inspection reports on the Internet, in which they describe the functioning of individual schools according to inspection standards and identify areas for improvement. The Netherlands also publishes lists of failing schools and summaries of the inspection assessments of all schools. In Austria, head teachers have the duty to ‘demonstrably inform school partners’ (parents, students and teachers) and the school maintaining body (mostly communities) about the inspection results. Inspectors do not usually confirm that the inspection report was on the agenda of a parent–teacher meeting; however, they would do so if problems persisted or parents complained. In the Czech Republic, reports of thematic school inspections (e.g. summarising annual results of all the schools) are available, but reports of individual schools are not publicised.

Outcomes and intermediate processes

The intended outcomes of school inspections (see the right-hand column in Figure 1) vary across different inspection systems, but Inspectorates of Education generally refer to promoting good education in individual schools and/or the education system as a whole (see Ehren et al. 2013). Descriptions of ‘good education’ often reflect

high student achievement levels and the conditions that are expected to contribute to these levels, such as educational leadership, a productive climate and culture, achievement-oriented school policy, clear and structured teaching, a challenging teaching approach and high expectations of students (Ehren et al. 2013).

Causal mechanisms

The programme theory of each Inspectorate describes assumptions on the causal mechanisms underlying these intended effects of school inspections. It was reconstructed by six national teams of researchers who analysed inspection guidelines and documents, and interviewed representatives of their respective Inspectorates. Ehren et al. (2013) summarised the commonalities in these six programme theories to come up with a joint ‘programme theory’ of inspection for these six countries, that is, a framework which connects general normative assumptions about the relationships of different inspection characteristics, causal mechanisms, intermediate processes and outcomes which are identified across national contexts (see Figure 1).

The joint ‘programme theory’, which is depicted in Figure 1, shows that effective school and teaching conditions are expected to be generated through high-quality self-evaluations and enhanced improvement capacity in the school, where improvement capacity refers to the school’s capacity to implement change and to expand its learning capacity in order to optimise its effectiveness. Improvement actions implemented by the schools are considered to be the intermediate outcomes through which school inspections lead to high student achievement.

Inspection models may draw on different theories of change to explain how and through what processes school inspections may impact on schools. Ehren et al. (2013) have identified three main mechanisms in the programme theories of six European Inspectorates of Education which connect school inspections and school improvement. These mechanisms are: ‘setting expectations’, ‘performance feedback’ and ‘actions of stakeholders’. These mechanisms (indicated by grey boxes in Figure 1) are considered ‘causal’ in the sense that they are meant to stimulate those ‘intermediate outcomes’ that are to produce the desired ‘outcomes’.

Schools and their stakeholders are expected to align their views and expectations of what a good school encapsulates and the school internal processes with the standards in the inspection framework, particularly with respect to those standards the school failed to meet during the inspection visit. Schools are expected to use the inspection feedback for further improvement actions. Stakeholders should use the inspection standards and the inspection results, if reported publicly, to take actions that will stimulate school improvement. These causal mechanisms are described in more detail below.

Setting expectations

A central idea of proponents of school inspections is that schools will pay attention to the information included in inspection standards and procedures; they will reflect on it, process it and adapt their goals and their daily practice to meet the normative expectations of school quality communicated by the Inspectorate. In school inspections, values about what constitutes a good/bad school, good/bad teaching and teachers, important/unimportant knowledge and how to learn, etc. are present in both *what* is inspected and in *how* the inspections are carried out. These values are potentially influential in shaping education. Eventually, these expectations should drive the school’s

own planning, self-evaluations and daily practices, and, as a result, will be institutionalised into the school's work structures and culture.

Previous studies describe how inspection standards set expectations on good education for schools and their stakeholders, which are enforced through the broader accountability arena in which these standards are implemented (Ehren et al. 2013). Not only do teachers and principals change their behaviour to match inspection standards and procedures (Ehren et al. 2013; Ehren and Shackleton 2015), but different types of evaluative activities also impact how people think and feel about education (Segerholm 2011).

Neo-institutional theories can be used to understand the processes through which these inspection values shape education. These theories explain that organisations (and, organisational fields, as a higher level unit of analysis; see Powell 2007, 3) do not operate according to efficiency criteria alone, but that they also seek legitimacy from their environment for what they do, which in turn enhances their chances of acquiring resources and their survival prospects (Meyer and Rowan 1977, 340). In an inspection and evaluation context, legitimacy is typically derived from and enforced through external actors, such as an Inspectorate of Education. Conformity to inspection standards provides legitimacy to the school's operations. Standards therefore function as active agents in the development of routines, structures, positions and tools by creating normative pressures, coercion and by enhancing mimetic isomorphism (Meyer and Rowan 1992, 2006).

Normative pressures are created when inspection standards present and create socially acceptable definitions of quality. Conformity to inspection standards and expectations is also enhanced through 'mimetic isomorphism' when schools actively seek out examples of peers who successfully meet the expectations of the inspectorate and mimic their responses to proactively prevent the potential consequences of being categorised by the inspectorate as failing (DiMaggio and Powell 1991).

External expectations, and the adaptation of schools to them, can lead to both positive and negative consequences. Under inspection or the threat of inspection, schools run the risk of becoming institutions performing *for* inspection, creating similar proactive and reactive arrangements which are generated simply and solely to be assessed more favourably by the supervisor (Perryman 2006). The discourse around what constitutes a good school can therefore have a positive impact on school improvement, but can equally, when measures are flawed and standards are rigid, lead to unintended fabrication of documentation, staging and game-playing before and during inspection.

Performance feedback

The second mechanism by which inspections are expected to drive school improvement is based on theories of performance feedback and goal setting (see Visscher and Coe 2002). During inspection visits, inspectors assess the educational quality of schools with respect to inspection standards and give feedback on the performance of schools. Some Inspectorates also give advice or recommendations on how a school can improve or they directly instruct schools to take specific measures for improvement. Following Hattie and Timperley (2007, 103) actors adapt their actions and/or their perception of the situation according to their interpretation of the 'feedback information'. In any case, Inspectorates of Education assume that schools will reflect on the feedback, devise appropriate improvement strategies based on this feedback and put the strategies into action, thereby improving school quality.

In evidence-based governance systems, feedback takes place in a more complex multi-level system. Information about performance is available from various sources such as national student assessments and school inspections. In order for this feedback to be relevant on a systemic level, actors on the various levels must pay attention to this feedback and compare it to their personal, their institutional, and their national goals and standards. They must derive some idea for action and develop appropriate strategies to close the gap between performance and standards (Coe 2002). They must communicate their strategy and gain support for it. Finally, feedback on the organisational and system level is *just one* element in a more complex arena (Kuper 2005, 101); schools and teachers have many other factors to consider in their daily operations. Evaluative feedback is often seen as an ‘unspecific impulse’ for improvement and not a road map for innovation. Inspection feedback often does not offer specific cues which point the actors’ attention to the process of development and to options for improvement, but includes comparative information which may be interpreted with respect to the individual and institutional self-esteem. Feedback is often related to very complex tasks (Visscher and Coe 2003, 328) and may include threatening elements in a high stakes accountability system (Visscher and Coe 2002, 247).

Feedback as conceptualised in evidence-based governance models will therefore not automatically produce development; careful design of content, format and communication of feedback is essential as is implementation and support for teachers and schools who are to build up competences for making use of feedback. As with feedback on the interpersonal level, system-level feedback is most effective when it is experienced as non-threatening (see Visscher and Coe 2003, 328).

A number of qualitative studies have analysed if and how feedback on inspection standards actually leads to school improvement. Brimblecombe, Shaw, and Ormston (1996) and Chapman (2001), for example, describe that teachers seem to regard oral and written feedback from school inspectors as an important stimulus for school improvement, especially when given in a context of trust rather than punishment. Standaert (2001) confirms these findings when describing how feedback given in a private setting and fitting with a school’s culture seems to have a particularly positive impact. Ouston, Fidler, and Earley (1997) point out that school inspections promote school improvement if the inspection report details the areas in which the school has performed poorly. According to Matthews and Sammons (2004), clear and explicit reports and feedback to schools are effective in informing school improvement plans following school inspections.

There are, however, also studies which nurture scepticism about the impact of school inspections. Gärtner, Wurster, and Pant (2013; see also Gärtner 2011) found that feedback information is ‘only rarely used for autonomous school improvement’. Ehren and Visscher (2008) emphasise that feedback on its own does not often lead to improvement, but models of operation where feedback is combined with unsatisfactory scores, specific improvement suggestions and agreements on improvement do influence school improvement. A ‘directive approach’ seems to be most effective.

Actions of stakeholders

Most inspection models deliberately include stakeholders (such as parents, local policy-makers or school boards) in their theories of action. The findings of school inspections are published with the intent of giving stakeholders a role in school improvement. The expectation is that parents, more than any other stakeholders, will use this information to evaluate the quality of their children’s schools. If a

school does not meet their expectations for quality, it is assumed that parents will address the school (or its administrators) about possible improvements (voice). Parents can also participate in school improvement by choosing schools that offer higher educational quality (choice) or transferring their child to a school with better inspection results (exit). This mechanism of choice, voice and exit is based on theories of social coordination (e.g. Schimank 2002; de Boer, Enders, and Schimank 2007), assuming that the inclusion of a ‘third’ party will reinforce inspection expectations and make it more likely that schools respond to the inspection.

What do we know about this hypothetical mechanism from empirical research? Studies from different countries show that parents do not use the public information about schools as much as many accountability models assume; they are often interested in matters other than inspection results (e.g. Dronkers and Veenstra 2001; Dutch Educational Council 2001; Karsten and Visscher 2001; Ladd and Walsh 2002; Cullen, Jacob, and Levitt 2006; Cullen and Jacob 2007; OECD 2008; Altrichter et al. 2011). Bell (2005) explains how parents have different ‘choice sets’ according to their socio-economic background which influence their choice of schools. These choice sets are determined by a complex interaction of their view of their child’s academic capacities, their own background and preferences for schools. Parents weigh, for example, the atmosphere, pedagogical climate, working methods, safety, clarity of regulations and reputation of the school in making a school choice decision, rather than relying on inspection data. Only 2% of the parents (mostly those who are more highly educated) use student learning results as a criterion for choosing a school (Dronkers and Veenstra 2001, 33).

In addition, parents rarely use inspection information in their communication with schools. When parents do suggest possible improvements, they usually voice purely organisational conditions such as changes to their child’s timetable. Parents do not generally interfere in matters relating to educational quality. Additionally, many schools do not seem inclined to use criticism and suggestions from parents to guide actual improvement (Dutch Educational Council 2001).

Therefore, acceptance and use of inspection feedback, and setting expectations seem to be the most promising mechanisms for change, particularly in (1) Inspectorates of Education that use *differentiated inspections* (in addition to regular cycles of visits), (2) that evaluate *output* (in addition to educational processes), (3) that have *consequences* in place for failing schools and (4) that *publish inspection reports* of individual schools. In the following section, we will compare inspection systems on these dimensions to investigate the impact of different systems on school improvement as it is perceived by school principals.

Methodology¹

School principals in primary and secondary education in the six countries, that is, Austria (region Styria, $N=540$), the Czech Republic ($N=165$), England ($N=290$), Ireland ($N=125$), the Netherlands ($N=88$) and Sweden ($N=1031$), were asked to participate in an online survey to collect comparative data on the mechanisms and processes of school inspections in the autumn of 2011.

Instrument

The questionnaire included 73 questions based on the framework in Figure 1. Questions about the school’s capacity for improvement and the school’s effectiveness and

teaching conditions were framed in terms of the time principals 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'). Questions on 'unintended consequences' asked whether school inspections lead to a narrowing of curricula and instructional processes, whether principals experience inspections as an administrative burden and the extent to which school leaders manipulate documents and data they send to the Inspectorate. Principals could respond to these questions on a 5-point scale (strongly agree–strongly disagree). Questions about 'setting of expectations', 'accepting feedback' and 'improving self-evaluations' are answered on a 5-point scale (strongly agree–strongly disagree), while questions about 'stakeholder action' are answered on a 4-point scale (not at all–to a great extent; see the [Appendix](#) for more details and www.schoolinspections.eu for the technical report describing data collection and instruments).

Approaches of analysis

The analysis followed a three-step approach: (1) confirmatory factor analyses (CFAs), (2) path modelling and (3) fitting of multiple-indicator multiple-cause (MIMIC) models. The CFAs were conducted to identify the relations between the manifest and the latent variables which, based on the theoretical model, are expected to be linked (Gustafsson et al. submitted). Thirty-seven items were used to construct 13 latent variables that represent the major concepts in the theoretical model. Among the 13 latent variables, 7 were narrow factors which were nested under broader factors, and these proved to be of little substantive interest. Finally, we include six main latent variables and additional items asking for unintended consequences in our analysis. Variables measuring unintended consequences were included as individual items in the analyses since a latent construct revealed poor consistency:

- *Setting expectations* is taken to be an independent variable influencing improvement actions.
- *Actions of stakeholders* refer to stakeholders' awareness of and sensitivity to school inspection reports and their willingness to pressure schools to take action for improvement. This variable is also an independent variable in the model.
- *Accepting feedback* is influenced by the two latent independent variables.
- *Improving self-evaluation* is an intermediate variable which influences improvement actions.
- *Improvement in capacity building* and *improvement in school effectiveness* are dependent variables in the model.
- *Improvement in school effectiveness* is measured by 10 items which focus on actions taken to improve effectiveness of teaching.
- *Unintended effects of school inspections* are represented by four individual items in the analysis since a latent construct revealed poor consistency.

Relations between these latent variables were analysed in a path model, based on the hypothetical relations in our theoretical model; the results indicated good model fit (Gustafsson et al. submitted). This model was the starting point to investigate how different inspection models influence the levels (i.e. mean/intercept) of the variables which measure mechanisms (setting expectations, actions of stakeholders and

accepting feedback) and (intermediate) outcomes (improving self-evaluation/capacity building/school effectiveness, unintended consequences).

The research questions presented in our theoretical framework above are framed in a way that implies comparisons between categories of countries which differ with respect to the four inspection models (differentiated inspections, outcomes-oriented inspections, amount of sanctions and interventions associated with inspections, and non-public versus public reporting of inspection results). The participating countries can be grouped in such a way that they are either included or not included in each category. The categories are not mutually exclusive as countries can fall into more than one category, and some of the categories may also be related (see Gustafsson et al. submitted). Since there are only six countries observed, it is necessary to analyse each of the four categories of inspection models separately.

A potential problem with this analytical approach is that the number of observations varies dramatically across countries, from a low of 88 in the Netherlands to a high of 1031 in Sweden. In order to correct for this imbalance, case weights have been used. The weights have been adjusted so that each country has the same size ($N=339$), and so that the weights of all six countries add up to the total actual sample size.

We analysed relations between the different inspection models and the variables in the theoretical framework by coding the categories representing different inspection models with dummy variables, and entering these one at a time into the model as an independent variable. This is a simple form of a so-called MIMIC model (e.g. Brown 2006). This modelling approach is parsimonious, because it assumes that the same measurement model holds true in all the countries, and that the relations among the latent variables are the same within each group. A MIMIC model therefore is an efficient approach to investigate the effects of the different inspection models on the means and intercepts of the latent variables. These models can be estimated with the Mplus programme (Muthén and Muthén 1998–2012), which takes case weights into account. Mplus also considers the effects on estimates of standard errors caused by the countries forming cluster samples, using the so-called Complex option.

One set of MIMIC models was fitted in which the latent variables were assumed to be freely correlated in an oblique measurement model. These models estimate the total effects of the school inspection characteristics on each latent variable, without controlling for the effects of any other variables.

Another set of MIMIC models was fitted in which the dummy variable was added to the variables in the path model. These models estimate the direct effects of the school inspection models on the latent variables, controlling for the effects of the other variables in the model.

We tested the total and direct effects by making the following comparisons:

- countries with differentiated inspection models (the Netherlands, England and Sweden) against countries with cyclical inspections only (Ireland, Austria and the Czech Republic);
- countries that measure both output and processes (the Netherlands, England and the Czech Republic) against countries that only measure processes (Ireland, Austria and Sweden);
- countries in which Inspectorates have the opportunity to sanction schools (the Netherlands, England, Sweden and the Czech Republic) against countries without punitive consequences (Ireland and Austria); and

- countries that publicly publish inspection reports of individual schools (England, the Netherlands, Ireland and Sweden) against countries where such reports are not made publicly available (Austria and the Czech Republic).

Models were estimated with the MLR estimator, which supports case weights, and takes non-normality of data into account when estimating standard errors. Model fit was evaluated with standard techniques, such as the Chi-square goodness-of-fit test, the root mean square error of approximation (RMSEA) and the Comparative Fit Index (CFI; see, e.g. Brown 2006). Four different models were fitted, and in each of these a dummy variable representing the presence or absence of each inspection model (i.e. *differentiated inspections*, *outcomes*, *sanctions* and *public reporting*) was entered as an independent variable. The path model without any independent variable had excellent fit (Chi-square = 778.04, $df = 583$, CFI = 0.986, RMSEA = 0.012, 90% CI 0.010–0.014) and the MIMIC models also had very good fit, with RMSEA values around .020 and CFI values at 0.95.

Results: direct and total effects of different inspection models on school improvement

Table 2 presents the direct and total effects² of the different school inspection models on variables representing school improvement and unintended consequences.

The results in this table indicate that *differentiated inspections* had a significant positive direct effect on ‘setting expectations’ and on ‘improving self-evaluations’, and there was also a significant negative direct effect on ‘accepting feedback’. Even though there was no direct effect of differentiated inspections on ‘change in capacity building’ and ‘change in school effectiveness’, there was a highly significant total effect on these two variables, due to indirect effects. The *use of outcome measures* had a direct effect on ‘setting expectations’ and also on ‘change in school effectiveness’, and it also exerted a significant negative effect on ‘accepting feedback’ and a non-significant negative effect on ‘change in capacity building’. Because positive and negative indirect effects balanced each other out, there was no significant total effect on ‘change in capacity building’ or on ‘change in school effectiveness’. *Sanctions* had a similar pattern of direct effects as had differentiated inspections, with positive effects on ‘setting expectations’ and ‘improving self-evaluations’, and a negative effect on ‘accepting feedback’. *Public reporting* had a strong positive effect on ‘actions of stakeholders’, and on ‘improving self-evaluations’, and via indirect effects it also exerted influence on ‘change in capacity building’ and on ‘change in school effectiveness’.

These results show that the different school inspection models are associated with a differentiated pattern of influence on the mechanisms that generate impact of school inspections. Thus, whether there is public reporting or not influences stakeholders’ actions directly, and it also directly influences the schools’ self-evaluations. These factors in turn have effects on the principals’ improvement actions. Whether the school inspections are differentiated or not directly influences ‘setting expectations’, as well as self-evaluations, which in turn influences improvement actions in the form of increased capacity building and efforts to increase school effectiveness. Outcome-oriented school inspections had a positive direct effect on ‘setting expectations’, a negative direct effect on ‘accepting feedback’, a tendency towards a negative effect on ‘changes in capacity building’ and a positive direct effect on efforts to increase school effectiveness. The total effect of outcome-oriented school inspections on

Table 2. Total and direct effects of school inspection models on school improvement.

	Differentiated inspections		Outcomes		Sanctions		Public reporting	
	Beta	<i>t</i> -Value	Beta	<i>t</i> -Value	Beta	<i>t</i> -Value	Beta	<i>t</i> -Value
<i>Direct effects</i>								
Setting expectations	0.32	3.23	0.27	2.28	0.25	2.36	0.23	1.72
Actions of stakeholders	0.31	1.35	0.10	0.41	0.04	0.17	0.58	6.85
Accepting feedback	-0.25	-2.13	-0.27	-3.22	-0.20	-2.04	-0.23	-1.91
Improving self-evaluations	0.25	3.51	0.18	1.60	0.18	2.37	0.18	3.79
Change in capacity building	0.23	1.57	-0.20	-1.56	0.05	0.37	0.16	1.06
Change in school effectiveness	0.03	0.39	0.11	2.86	0.05	0.57	-0.04	-1.48
<i>Total effects</i>								
Setting expectations	0.32	3.10	0.27	2.20	0.25	2.26	0.23	1.74
Actions of stakeholders	0.30	1.29	0.10	0.40	0.03	0.13	0.57	6.81
Accepting feedback	-0.06	-0.62	-0.17	-3.16	-0.12	-1.96	0.06	0.67
Improving self-evaluations	0.35	3.55	0.25	1.60	0.25	1.93	0.31	2.85
Change in capacity building	0.40	5.03	-0.02	-0.07	0.19	1.47	0.33	2.50
Change in school effectiveness	0.31	4.91	0.18	1.35	0.19	2.00	0.23	2.51

Note: Bold print indicates significance at least at $p \leq .05$.

improvement actions therefore was non-significant. Inspection systems with sanctions had a similar pattern of influences, the total effect on changes in capacity building being non-significant and marginally significant on changes in school effectiveness.

Surprisingly, 'accepting feedback' does not fit with the overall picture. In the programme theories of the participating Inspectorates of Education, inspection is meant to stimulate school development as schools receive, interpret and use feedback to devise action improvement strategies. However, we found a negative relation between three of the four inspection models and 'accepting feedback', and only one positive relation between inspection models publishing individual school reports and 'accepting feedback'. It seems that public reporting generates more acceptance of inspection feedback, whereas differentiated and outcome-oriented inspections and inspections with sanctions are associated with less acceptance of feedback.

Direct and total effects of different inspection models on unintended consequences

The four items asking about unintended consequences of school inspections were included as manifest dependent variables in the same models reported above. Again, direct and total effects of the four inspection models (represented by dummy variables) on the four items were estimated, and the results are shown in Table 3.

Table 3. Total and direct effects of inspection characteristics on unintended consequences.

	Differentiated inspections		Outcomes		Sanctions		Public reporting		
	Beta	<i>t</i> -Value	Beta	<i>t</i> -Value	Beta	<i>t</i> -Value	Beta	<i>t</i> -Value	
<i>Direct effects</i>									
Discourage teachers to experiment with new teaching methods	0.28	3.53	0.15	1.26	0.27	3.64	0.07	0.70	
Narrowing curriculum and instructional strategies	0.36	5.20	0.23	1.74	0.23	2.25	0.28	3.94	
Documents present an overly positive picture of the quality of our school	-0.37	-2.39	-0.38	-2.66	-0.46	-3.20	0.27	1.84	
Preparation is about putting protocols and procedures in writing	0.08	0.75	0.05	0.43	0.16	1.96	-0.05	-0.37	
<i>Total effects</i>									
Discourage teachers to experiment with new teaching methods	0.26	2.89	0.17	1.30	0.28	3.47	0.05	0.45	
Narrowing curriculum and instructional strategies	0.42	5.83	0.32	2.37	0.30	2.48	0.32	2.61	
Documents present an overly positive picture of the quality of our school	-0.31	-1.51	-0.37	-2.08	-0.45	-2.54	0.22	1.29	
Preparation is about putting protocols and procedures in writing	0.03	0.21	0.06	0.47	0.17	1.60	-0.14	-1.06	

Note: Bold print indicates significance at least at $p \leq .05$.

‘Discouragement of teachers from experimenting with new teaching methods’ was affected by *differentiated inspections* and by the *amount of sanctions* in an inspection system. ‘Narrowing curriculum and instructional strategies’ was directly affected by the same two models of school inspections, and also by *public reporting*. The statement ‘documents present an overly positive picture of the quality of our school’ was to a smaller extent endorsed by principals when school inspections were differentiated, outcomes-oriented and sanctions-oriented. No significant effects were observed for the statement ‘preparation is about putting protocols and procedures in writing’. The estimated total effects were similar to the direct effects for the items measuring unintended consequences of school inspections. The results thus indicate that school inspections that are differentiated and use a higher level of sanctions seem to be associated with unintended consequences of inspection.

Two cases

The previous section showed how Inspectorates of Education that fit a ‘hard’ governance approach and use a differentiated, high stakes, output-oriented inspection methodology have the highest impact on school improvement but also generate the most

unintended consequences. This section aims to elaborate on those results by building on the work of Jones and Tymms (2014) to single out the Inspectorates of Education in England and comparing it to school inspections in Austria which essentially use a 'soft' approach to governance.

England: Ofsted school inspections

Ofsted is charged with the requirement to inspect all public primary and secondary schools in England (Smith 2000). Ofsted provides information publicly about the relative performance of schools measured against published inspection criteria regarding student achievement, teaching quality, student behaviour and safety, leadership quality and school management. Schools are to be inspected once every four years against these inspection criteria, and evaluations lead to an overall assessment of the effectiveness of the school.

There is a strong emphasis in the Ofsted framework on student achievement as inspectors analyse performance data extensively prior to the visit and use it as a starting point for inspection; many schools feel that their output data are the main source of information used to inform the Inspectorate's judgement of the entire framework. Student achievement data are also published and made accessible by Ofsted through Data View, a digital tool to help schools' stakeholders compare regional and local performance of schools.

Schools that are found lacking are, at the time of our study, categorised as being in either 'special measures' if the school is failing and does not have the capacity to improve, or given a 'notice to improve' if they are performing below expectations. The timings and frequency of inspections vary according to whether the school is satisfactory, has a notice to improve or is in special measures. Risk assessments are also undertaken to schedule inspection visits, making Ofsted a model of differentiated inspections.

School inspections in England are high stakes, as schools that receive a notice to improve and that are in special measures face threat of closure by the Secretary of State. The high stakes context is additionally enhanced by the existence of a number of informal pressures, such as key prizes for high-performing schools and principals of these schools, performance management systems of principals and teachers which are tied to the inspection assessment, and the publication of inspection reports which promote competition between schools.

As our results indicate, Ofsted has a significant impact on school improvement. Jones and Tymms (2014) explain how the criteria and descriptors set out in the inspection framework illustrate the standards of performance expected of schools. In recommending priorities for future action during inspection visits, offering a challenge to the school's own views on priorities for improvement, and monitoring progress, inspection standards are clearly communicated and continuously endorsed in conversations between school staff and inspectors. Principals will, according to Jones and Tymms, often accompany the inspectors as they observe lessons. A dialogue will occur between the head and the inspector in order to gauge whether the two of them make the same judgements, and principals will be coached on how to interpret the standards in assessing their teachers. Ofsted then actively recruits and trains principals from high-performing schools to take on the role of school inspectors and participate in Ofsted inspection teams.

The high stakes context of school inspections has, however, created a situation where 'What does Ofsted want?' dominates the thinking and behaviour of schools,

whatever category they are in.³ For many schools, Ofsted and the particular national focus of the moment are their agenda, despite the fact that current government policies aim to create more school autonomy. Also, the high stakes context will motivate schools to avoid falling below inspection thresholds and to ensure their schools meet Ofsted's expectations.

Austria: team inspections in Styria

The Austrian school Inspectorate originated in the second half of the nineteenth century (Scheipl and Seel 1985). Several supervisory functions for schools were traditionally amalgamated in this role: inspectors were both administrators and responsible for quality control of schools in their regions. Criteria for assessing 'quality' have varied over time but always included some mixture of educational effectiveness and quality on one hand, and legal and administrative appropriateness on the other.

In the wake of the Programme for International Student Assessment (PISA) shock Austrian inspectorates began to discuss alternative ways of fulfilling their functions and to search for a new balance between administration, quality control and developmental support. In 2005, the Inspectorate of the Austrian state of Styria introduced the new 'team inspection' approach which was modelled after inspections procedures from the Netherlands and Lower Saxony.

The inspection process began with an analysis of school documents (e.g. school programme, results of self-evaluation). Then, with advance notice, inspection teams of two to three inspectors visited the school for one to three days to inspect the site, observing classroom teaching, interviewing students, parents, teachers, school leaders and the regional mayor. Data collection was done according to a set of fixed procedures and forms which represented the relevant 'quality criteria'. The inspection results and additional data (e.g. student feedback data; however, no state-wide student performance assessment) were condensed into an inspection report which explained the strengths and weaknesses of the school and included 'very concrete recommendations' for development. A preliminary version of the report was discussed with school staff. The school management's duty was to produce a 'school development plan' detailing measures to improve identified issues. After one to three years, the Inspectorate was to check whether the development plan was put into practice (see Altrichter, Kemethofer, and Schmidinger 2013).

School inspections in the Austrian state of Styria are low stakes, as the inspection framework does not include threshold levels for identifying 'failing schools' or use rewards or sanctions. Inspectors focus their attention on schools and their development, rather than on communicating performance to the public.

Research indicates that this inspection approach was accepted by the majority of school leaders. In their view, inspection processes and reports pointed to important issues for development and helped to communicate to teachers important measures to improve classroom and school quality (Altrichter, Kemethofer, and Schmidinger 2013). On one hand, this low stakes approach seemed to trigger fewer activities of self-evaluation and school development than the English model. On the other hand, Austrian school leaders were more attentive to the messages of inspection feedback and saw fewer unintended consequences than English ones (Altrichter and Kemethofer 2015). In 2012, the regional 'team inspection' approach was terminated because a new central legislation for quality management in schools was introduced (Altrichter 2012).

Conclusion and discussion

School inspections are seen in many education systems as a method of sustaining and improving school quality and they are part of a central quality management system. Differences in inspection models we studied include those in the use of differentiated inspections, in inspections of school outcomes (in addition to evaluating only educational practices in schools or compliance to legislation), in inspections with or without punitive sanctions, and in the presence or lack of public reporting of inspection results.

These different inspection models may ‘work’ differently in generating impact and may vary according to the degree of influence on school improvement as well as on various unintended consequences. This paper presents the results of a survey of principals in primary and secondary education in six European countries (the Netherlands, England, Ireland, Sweden, the Czech Republic and Styria/Austria) on the impact and mechanisms of impact of these different school inspection models. We used MIMIC modelling to compare the impact of various inspection approaches as described above.

Even though the study has allowed a comparison of inspection models across six countries, there are a number of limitations that need to be addressed. The first limitation is that the study entirely relies on the perceptions and reports of the principals. Self-reports, particularly retrospective ones, 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 bias in reporting cannot be ruled out. We expect, however, that potential bias is consistent across the six countries and therefore did not affect the pattern of relations in our model. Another 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 inspection systems were included in the study. While these six systems represent considerable variation across two dimensions of interest, they certainly cannot be considered representative of all school inspection systems. However, it should be emphasised that the six Inspectorates of Education vary across two dimensions of interest: soft versus hard governance, and the comparison is therefore still relevant in enhancing our understanding of the mechanisms through which two distinct inspection approaches impact school improvement.

The results of our study indicate that Inspectorates of Education that use a differentiated model (in addition to regular visits), in which they evaluate both educational practices and outcomes of schools and publicly report the inspection findings of individual schools are the most effective. Principals in these systems report the greatest changes in capacity building and in improved school and teaching conditions. These changes seem to be mediated by improvements in the schools’ self-evaluations and the awareness of the schools’ stakeholders of the findings in the public inspection reports. However, differentiated inspections also lead to unintended consequences since principals report a narrowing of the curriculum in the school and the discouragement of teachers from experimenting with new teaching methods.

An interesting issue emerges when it comes to the ‘causal mechanisms’ by which inspection systems intend to produce their effects. Three of the four inspection models we studied (differentiated inspections, outcomes-orientation and sanctions, which may well be taken to indicate ‘high stakes’ inspection approaches) influence ‘setting expectations’. These models, at the same time, *reduce* the likelihood that principals pay attention to the inspection feedback and derive improvement strategies based on this feedback.

We will explain these effects of ‘setting expectations’ and the lack of effect of inspection feedback using institutional and social coordination theories. As outlined by neo-institutional theories, a school’s quest for legitimacy and the normative pressure created by inspection frameworks seem to be important drivers of schools’ reactions to inspection. The more clearly the inspection communicates its standards and the more normative pressure accompanies them, the more school leaders undertake and report self-evaluative and developmental activities. Such activities may similarly make inspection feedback unnecessary as schools are already aware of their strengths and weaknesses or find it difficult to use feedback when it implies changes in set (teaching) processes and (school organisational) structures.

The initial inspection framework is expected to influence the type of improvement and effects generated. According to Powell (1991), initial choices preclude future options, including those that would have been more effective in the long run. Organisational memory and learning processes shape the future course of organisations as they become committed to routines that are shaped by early and often arbitrary successes. Positive feedback initially magnifies and reinforces those routines, but once these practices are institutionalised, feedback is less likely to stimulate change to a technological path that is neither guaranteed to be efficient, nor easily altered (Powell 1991). Established ideas of the way things are done can be very beneficial as they guide action and predict the behaviour of others, and efforts to change these established patterns (even if these efforts are called for in inspection feedback) are often resisted because they threaten individuals’ sense of security, increase the cost of information processing and disrupt routines.

The dominant effect of ‘setting expectations’ may also be explained by the fact that the mechanism of ‘accepting feedback’ is less responsive to accountability pressure than ‘setting expectations’ and ‘action of stakeholders’. It takes more elaborate reflective and constructive processes to develop sound actions strategies from critical feedback. If there is high pressure to show progress in a short time span, then schools might be more likely to turn to strategic behaviour instead of thoroughly analysing and using feedback. This finding resonates with much of the research reported hitherto (see Nelson and Ehren 2014 for an overview): data feedback sounds like a plausible mechanism for stimulating rational school improvement, but it may be more difficult than expected to transfer the inspection feedback to complex multi-level contexts like schools.

Thus, we may be ill-advised if we expect that schools will use the inspection data and feedback for school development. It is more likely that inspections have an effect *before* an inspection visit, when schools prepare for the assessment by implementing self-evaluations and taking improvement actions to align their school with the inspection standards. The effects of school inspections may have a prescriptive rather than evaluative value, especially if the inspection model includes sanctions for failing schools. In such models, schools are affected by school inspections when they are motivated and learn how to self-evaluate their processes and products and to orchestrate data-based development processes.

We argue that, to make inspection work in a way that is beneficial to the overall system, improvement of educational quality is better thought of as a culture change rather than the implementation of an inspection instrument. Brennan and Shah (2000) explain that a culture and system of evaluation needs to be built on a fundamental understanding about the way, in a given institution, system or nation, a particular scheme of quality management will work.

Our Austrian case description underpins this conclusion as it showed us how national cultural characteristics have an impact on the possibility of using certain evaluation approaches. The inspection model embedded in a bureaucratic low stakes culture shows different results (fewer development activities) and different intermediary mechanisms (more attention to inspection feedback). The fact that school inspections have been abolished despite the current strong European narrative of increased activity and significance for Inspectorates of Education (see Grek et al. 2013) also points to cultural influence.

Thinking about the effectiveness of school inspections therefore needs to go beyond an evaluation of the type of models that are effective and their different mechanisms of impact. Inspection systems challenge the intrinsic value system of the teaching profession and give weight to extrinsic values and standards; they change the balance of power in and between education systems, schools and teachers. These changes in values and power will highly depend on the existing status quo in schools, particularly in the notions of school quality, the roles and responsibilities in shaping and implementing such notions, as well as in the attitudes and knowledge of school staff towards the external evaluation of their school's quality.

An increasingly accepted theory is that inspection systems tend to be adapted when education systems mature. Maturity in this sense refers, according to Barber (2004), to the self-evaluation literacy and innovation capacity of schools and their stakeholders to improve on their own, and the availability of high-quality data (e.g. from national student achievement tests) to inform school-level evaluation and improvement. According to this hypothesis, mature systems have a diminished need for top-down, standards-based inspections, and instead rely increasingly on the profession to review and develop their own quality. Such developments are often paralleled by increased decentralisation and autonomy of schools and a scaling down of school inspections to minimal monitoring of school quality.

Interestingly such developments seem to result in complex combinations of both 'hard' and 'soft' governance within the same education system, as national governments and the Inspectorates of Education increasingly move towards 'hard' governance, building on notions of new public management with a strong focus on output of and competition between schools, and a simultaneous 'outsourcing' of soft governance approaches for learning, feedback and capacity building to schools and the education profession.

The context of these systems is, however, often one in which Inspectorates of Education have traditionally measured and incentivised centralised standards of school quality, which explains why, in our study, inspections that fit such 'hard' governance approaches still have a strong impact on self-evaluations and capacity building, as schools hold on to self-evaluation frameworks aligned to (former) inspection standards and frame feedback and learning around notions of inspection-defined good practices.

Moving towards more effective models of school inspections therefore needs explicit consideration of the current status quo and how it can be improved through incremental changes. Our two typical models of school inspections (differentiated, high stakes, output-oriented versus cyclical, low stakes, process-oriented) and our description of how these models can have an impact via three mechanisms of change (feedback, setting expectations and stakeholder involvement) provide a framework to structure such considerations.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes

1. Details of sampling design, participation rates and technical details of the CFA and path modelling are presented by Gustafsson, submitted and can also be found on www.schoolinspections.eu.
2. Direct effects include the estimates of the standardized regression coefficients for the dummy variable (representing each of the four inspection models) with respect to each of the latent variables in the model. Total effects are effects of the dummy on each latent variable, which is the sum of the direct effect, and all the indirect effects via other variables in the model.
3. <http://www.theguardian.com/education/2014/oct/17/ofsted-tells-teachers-what-not-to-do-in-effort-to-dispel-inspection-myths>

Notes on contributors

Melanie Ehren is a senior lecturer at the Institute of Education, working on research into effects and unintended consequences of high stakes testing, accountability systems, and school inspections on school improvement and teaching and learning in schools. She has studied the effectiveness of accountability systems and external evaluations on school improvement in many different countries and contexts, for example, as part of the Organisation for Economic Co-operation and Development and UNICEF country review teams and as a project coordinator of a number of EU-funded comparative studies.

Jan-Eric Gustafsson has since 1986 been Professor of Education at the University of Gothenburg. His substantively oriented research focuses on individual differences in cognitive abilities, and on determinants of educational outcomes at individual and system levels. His methodologically oriented research focuses on conceptual and technical issues of measurement within classical and modern test theory, and on the application of latent variable models to both non-nested and nested data.

Herbert Altrichter is Full Professor of Education and Educational Psychology at Johannes Kepler University, Linz, Austria, and member of the executive committee of the European Educational Research Association (EERA). His research interests include educational governance studies, school development and system reform, evaluation, teacher education and qualitative research methodology.

Guri Skedsmo is working part time as Associate Professor at the Department of Teacher Education and School Research, University of Oslo in Norway and is a Senior Researcher at the Institute of the Management and Economics of Education at the University of Teacher Education Zug in Switzerland. Her research interests include school governance, evaluation, and school leadership. The last 10 years she has been actively involved in developing professional development programmes for school leaders and she teaches at the Master Programme in Educational Leadership and the National Programme for Principals at the University of Oslo. She is also co-editor of the book series *Policy Implications of Research in Education* (Springer).

Stephan Huber is the Head of the Institute for Management and Economics of Education (IBB) at the University of Teacher Education of Switzerland Zug (PHZG). His research is on organisational education, system counselling, education management, school effectiveness, school improvement, school management, professionalisation of teachers and educational leadership personnel (school leaders), and international comparative research on education.

David Kemethofer studied Sociology at Linz University, Austria, and currently works at the Department of Education and Educational Psychology at Linz University, Austria. His main research interests are educational improvement and quality assurance and evaluation with a focus on school inspections in an international perspective.

References

- Altrichter, H. 2012. "Veränderungen von Schulaufsicht und Schulinspektion in Österreich." Working Paper AT4. Project Impact of School Inspection on Teaching and Learning. Linz: JKU.
- Altrichter, H., J. Bacher, M. Beham, G. Nagy, and D. Wetzelhütter. 2011. "The Effects of a Free School Choice Policy on Parents' School Choice Behaviour." *Studies in Educational Evaluation* 37 (4): 230–238.
- Altrichter, H., and D. Kemethofer. 2015. "Does Accountability Pressure through School Inspections Promote School Improvement?" *School Effectiveness and School Improvement* 26 (1): 32–56. doi:10.1080/09243453.2014.927369.
- Altrichter, H., D. Kemethofer, and E. Schmidinger. 2013. "Neue Schulinspektion und Qualitätsmanagement im Schulsystem." *Erziehung und Unterricht* 163 (9–10): 961–978.
- Barber, M. 2004. "The Virtue of Accountability: System Redesign, Inspection, and Incentives in the Era of Informed Professionalism." *Journal of Education* 85 (1): 7–38.
- Bell, C. 2005. *All Choices Created Equal? How Good Parents Select "Failing" Schools*. New York: National Center for the Study of Privatization in Education, Columbia University.
- de Boer, H., J. Enders, and U. Schimank. 2007. "On the Way towards New Public Management? The Governance of University Systems in England, the Netherlands, Austria, and Germany." In *New Forms of Governance in Research Organisations*, edited by D. Jansen, 137–152. Dordrecht: Springer.
- Brennan, J., and T. Shah. 2000. "Quality Assessment and Institutional Change: Experiences from 14 Countries." *Higher Education* 40 (3): 331–349.
- Brimblecombe, N., M. Shaw, and M. Ormston. 1996. "Teachers' Intention to Change Practice as a Result of OFSTED School Inspections." *Educational Management & Administration* 24 (4): 339–354.
- Brown, T. A. 2006. *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.
- van Bruggen, J. C. 2010. "Inspectorates of Education in Europe; Some Comparative Remarks about Their Tasks and Work." SICI Report. Accessed from www.sici-inspectorates.org.
- Chapman, C. 2001. "Changing Classrooms through Inspection." *School leadership and management* 21 (1): 59–73.
- Clarke, J. 2011. "OFSTED Keywords D1." SIEG Meeting 29–29 March, 2011, Umeå. <http://governingbyinspection.com/tag/working-paper/>.
- Clarke, J., and J. Ozga. 2011. Working Paper 4: Inspection as Governing. Accessed December 2014. <http://www.education.ox.ac.uk/governing-by-inspection/working-papers/>.
- Coe, R. 2002. "Evidence on the Role and Impact of Performance Feedback in Schools." In *School Improvement through Performance Feedback*, edited by A. J. Visscher and R. Coe, 27–39. Lisse: Swets & Zeitlinger.
- Cullen, J. B., and B. A. Jacob. 2007. "Is Gaining Access to Selective Elementary Schools Gaining Ground? Evidence from Randomized Lotteries." Working Paper 13443. Cambridge: National Bureau of Economic Research. <http://www.nber.org/papers/w13443.pdf>.
- Cullen, J. B., B. A. Jacob, and S. Levitt. 2006. "The Effect of School Choice on Participants: Evidence from Randomized Lotteries." *Econometrica, Econometric Society* 74 (5): 1191–1230.
- De Wolf, I. F., and F. J. G. Janssens. 2007. "Effects and Side Effects of Inspections and Accountability in Education: An Overview of Empirical Studies." *Oxford Review of Education* 33 (3): 379–396.
- DiMaggio, P. J., and W. W. Powell. 1991. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." In *The New Institutionalism in Organizational Analysis*, edited by W. W. Powell and P. J. DiMaggio, 63–82. Chicago, IL: The University of Chicago Press.
- Dronkers, J., and R. Veenstra. 2001. "Schoolprestatie-indicatoren in het voortgezet onderwijs: start, reacties en vervolg." In *Het oog der natie: scholen op rapport; standaarden voor de publicatie van schoolprestaties*, edited by A. B. Dijkstra, S. Karsten, R. Veenstra, and A. J. Visscher, 21–36. Assen: Koninklijke Van Gorcum BV.
- Dutch Educational Council. 2001. *De markt meester? Een verkenning naar marktwerking in het onderwijs*. Den Haag: Onderwijsraad.

- Ehren, M. C. M., H. Altrichter, G. McNamara, and J. O'Hara. 2013. "Impact of School Inspections on School Improvement; Describing Assumptions on Causal Mechanisms in Six European Countries." *Educational Assessment, Evaluation and Accountability* 25 (1): 3–43. doi:10.1007/s11092-012-9156-4.
- Ehren, M. C. M., and N. Shackleton. 2015. "Mechanisms of Change in Dutch Inspected Schools: Comparing Schools in Different Inspection Treatments." *British Journal of Educational Studies*. doi:10.1080/00071005.2015.1019413.
- Ehren, M. C. M., H. Altrichter, G. McNamara, and J. O'Hara. 2013. Impact of School Inspections on School Improvement: Describing Assumptions on Causal Mechanisms in Six European Countries. *Educational Assessment, Evaluation and Accountability* 25 (1): 3–43. <http://dx.doi.org/10.1007/s11092-012-9156-4>.
- Ehren, M. C. M., and A. J. Visscher. 2008. "The Relationship between School Inspections, School Characteristics and School Improvement." *British Journal of Educational Studies* 56 (2): 205–227.
- Elmore, R. F., and S. H. Fuhrman. 2001. "Research Finds the False Assumption of Accountability." *Phi Delta Kappan* 67 (4): 9–14.
- Gärtner, H. 2011. "Die Wirkung von Schulinspektion auf Schulentwicklung. Eine quasi-experimentelle Feldstudie." In *Wirkungen und Wirksamkeit der externen Schulevaluation*, edited by C. Quesel, V. Husfeldt, N. Landwehr, and P. Steiner, 145–161. Bern: h.e.p.
- Gärtner, H., and H. A. Pant. 2011. "Validity of Processes and Results of School Inspection." *Studies in Educational Evaluation* 37 (2–3): 85–93.
- Gärtner, H., S. Wurster, and H. A. Pant. 2013. "The Effect of School Inspections on School Improvement." *School Effectiveness and School Improvement*. doi:10.1080/09243453.2013.811089.
- Grek, S., M. Lawn, J. Ozga, and C. Segerholm. 2013. "Governing by Inspection? European Inspectorates and the Creation of a European Education Policy Space." *Comparative Education*. doi:10.1080/03050068.2013.787697.
- Gustafsson, J.-E., M. C. M. Ehren, G. Conyngham, G. McNamara, H. Altrichter, and J. O'Hara. submitted. "School Inspections and School Improvement: Testing Assumptions on Causal Mechanisms." *Studies in Educational Evaluation*.
- Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112.
- Jones, K., and P. Tymms. 2014. "Ofsted's Role in Promoting School Improvement: The Mechanisms of the School Inspection System in England." *Oxford Review of Education* 40 (3): 315–330. doi:10.1080/03054985.2014.911726.
- Karsten, S., and A. Visscher. 2001. "Ervaringen met het openbaar maken van schoolprestaties in Engeland en Frankrijk." In *Het oog der natie: scholen op rapport; standaarden voor de publicatie van schoolprestaties*, edited by A. B. Dijkstra, S. Karsten, R. Veenstra, and A. J. Visscher, 36–54. Assen: Koninklijke Van Gorcum BV.
- Klerks, M. 2013. "The Effect of School Inspections: A Systematic Review." *School Effectiveness and School Improvement*. Accessed January 2014. www.schoolinspections.eu
- Koretz, D. M. 2003. "Using Multiple Measures to Address Perverse Incentives and Score Inflation." *Educational Measurement* 22 (2): 18–26.
- Kuper. 2005. *Evaluation im Bildungssystem*. Stuttgart: Kohlhammer.
- Ladd, H. F. 2007. "Holding Schools Accountable Revisited." Spencer Foundation Lecture in Education Policy and Management, presented at the 2007 APPAM Fall Research Conference, Washington DC, November (Vol. 8).
- Ladd, H. F., and R. P. Walsh. 2002. "Implementing Value-added Measures of School Effectiveness: Getting the Incentives Right." *Economics of Education Review* 21 (1): 1–17.
- Luginbuhl, R., D. Webbink, and I. De Wolf. 2007. "Do School Inspections Improve Primary School Performance?" CPB-paper, nr. 83. <http://www.cpb.nl/nl/pub/cpbreeksen/discussie/83/disc83.pdf>.
- 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.
- Matthews, P., and P. Sammons. 2004. *Improvement through Inspection*. London: Ofsted.
- Meyer, J. W., and B. Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83 (2): 340–363.

- Meyer, J. W., and B. Rowan. 1992. "The Structure of Educational Organizations." In *Organizational Environments. Ritual and Rationality*, edited by J. W. Meyer and W. R. Scott, 71–97. Newbury Park, CA: Sage.
- Meyer, H.-D., and B. Rowan. 2006. *The New Institutionalism in Education*. Albany, NY: State University of New York Press.
- Muthén, L. K., and B. O. Muthén. 1998–2012. *Mplus user's guide*. 7th ed. Los Angeles, CA: Muthén & Muthén.
- Nelson, R., and M. C. M. Ehren. 2014. "Review and Synthesis of Evidence on the (Mechanisms of) Impact of School Inspections." Accessed from <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., G. V. Glass, and D. C. Berliner. 2006. "High-stakes Testing and Student Achievement: Does Accountability Pressure Increase Student Learning?" *Education Policy Analysis Archives* 14 (1). <http://epaa.asu.edu/epaa/v14n1>.
- OECD. 2008. *Measuring Improvements; Best practices to Assess Value Added of Schools*. Paris: OECD.
- Ouston, J., B. Fidler, and P. Earley. 1997. "What do Schools do after OFSTED School Inspections-or before?" *School Leadership & Management* 17 (1): 95–104.
- Perryman, J. 2006. "Panoptic Performativity and School Inspection Regimes: Disciplinary Mechanisms and Life under Special Measures." *Journal of Education Policy* 21 (2): 147–161.
- Powell, W. W. 1991. "Expanding the Scope of Institutional Analysis." In *The New Institutionalism in Organizational Analysis*, edited by W. W. Powell and P. J. DiMaggio, 183–203. Chicago: The University of Chicago Press.
- Powell, W. W. 2007. "The New Institutionalism." In *The International Encyclopedia of Organization Studies*, edited by S. R. Clegg and J. R. Bailey, 1–11. Preprint Accessed December, 2015. <http://www.stanford.edu/group/song/papers/NewInstitutionalism.pdf>.
- Rosenthal, L. 2004. "Do School Inspections Improve School Quality? Ofsted Inspections and School Examination Results in the UK." *Economics of Education Review* 23 (2): 143–151.
- Scheypl, J., and H. Seel. 1985. *Die Entwicklung des österreichischen Schulwesens von 1750 bis 1938*. Graz: Leykam.
- Schimank, U. 2002. *Handeln und Strukturen. Einführung in die akteurtheoretische Soziologie*. Weinheim: Juventa.
- Segerholm, C. 2011. "Values in Evaluation: The what and how values in Swedish school inspection." Paper presented at the American Evaluation Association Conference, Anaheim, November 2–5.
- Smith, G. 2000. "Research and Inspection: HMI and OFSTED, 1981–1996 – A Commentary." *Oxford Review of Education* 26 (3/4): 333–352.
- Standaert, R. 2001. *Inspectorates of Education in Europe: A Critical Analysis*. Leuven: Acco.
- Visscher, A. J., and R. Coe. 2002. *School Improvement through Performance Feedback*. London: Routledge.
- Visscher, A. J., and R. Coe. 2003. "School Performance Feedback Systems: Conceptualisation, Analysis, and Reflection." *School Effectiveness and School Improvement* 14 (3): 321–349.

Appendix: Overview of latent variables

Latent construct	Example item	Number of items	Scale
Setting expectations	The inspection standards affect the evaluation and supervision of teachers	7	Strongly agree (1)–strongly disagree (5)
Stakeholders sensitive to reports	The school's Board of Management/Boards of Governors is very aware of the contents of the school inspection report	3	
Accepting feedback	The feedback received from the inspectors was useful	5	
Promoting/improving self-evaluation	Compared to last academic year, I spent less/more time on the self-evaluation process as a whole	3	Much less (1)–much more (5)
Improvement in capacity building	Compared to last academic year, I spent less/more time involving teachers in making decisions about using new teaching methods	7	
Improvement in school effectiveness	Compared to last academic year, I spent less/more time on improving the extent to which teachers make effective use of teaching time within lessons	10	Much less (1)–much more (5)