Comparing effects of school inspections in Sweden and Austria

David Kemethofer¹ · Jan-Eric Gustafsson² · Herbert Altrichter¹

Received: 31 October 2016 / Accepted: 4 June 2017 / Published online: 23 June 2017
© The Author(s) 2017. This article is an open access publication

Abstract  In recent years, school inspections have been newly introduced or adapted to the evidence-based governance logic in many European countries. So far, empirical research on the impact of school inspections has produced inconclusive results. Methodologically, it has mainly focussed on analysis of a national inspection model and used cross-sectional data. In the following paper, we investigate the effects of school inspections in a comparative and longitudinal perspective in Sweden and Austria (province of Styria). In these countries, inspections follow different governance approaches. While Austria represents a low-stakes system which does not link consequences for schools to inspection results, sanctions and even financial penalties are potential consequences of Swedish inspections. To discuss the effects of inspections, we asked principals in three consecutive years about the impact of inspections. Our results indicate that school inspections in Austria and Sweden do have a small to medium impact on school improvement and school effectiveness.

Keywords  School inspection · Inspection systems · Comparative research · School improvement · School effectiveness

David Kemethofer
david.kemethofer@jku.at

Jan-Eric Gustafsson
jan-eric.gustafsson@ped.gu.se

Herbert Altrichter
herbert.altrichter@jku.at

1 Linz School of Education, Department of Educational Research, Johannes Kepler University Linz, Altenbergerstraße 69, 4040 Linz, Austria
2 Department of Education and Special Education, University of Gothenburg, P. O. Box 300, 405 30 Gothenburg, Sweden
1 Introduction

During the last decades, many school systems in Europe have undergone extensive changes in their governance structures. Improvements of effectiveness, equity, and quality in education (Faubert 2009) have been proclaimed as the goals of these changes. Accountability of educational services (Barber 2004), output orientation and evidence-based decision making in education policy, school and classroom development have been the main operational principles guiding these governance reforms (Altrichter and Maag Merki 2016, pp. 21). In this scenario of unfolding an evidence-based governance of schooling, ‘new school inspections’ have occupied an important place.

“Inspection in its most general sense may be defined as the process of assessing the quality and/or performance of institutions, services, programmes or projects by those (inspectors) who are not directly involved in them and who are usually specially appointed to fulfil these responsibilities. Inspection involves visits made by inspectors, individually or in teams, to observe the institutions, services etc. concerned while they are actually functioning (i.e. in real time). The most common outcome of an inspection is a written report of the inspectors’ findings (Wilcox 2000, p. 15f). While school inspections have been in operation in many countries since the nineteenth century (van Bruggen 2010), inspection systems have been modernised (Ehren and Honingh 2011; Jones and Tymms 2014) and new inspectorates have been built up (e.g. in the German states; see Döbert et al. 2008) since the 1990s (with some differences in date) in order to cope with the increased demands for accountability and to control and promote the quality of schools more resolutely. As the concept of quality is ambiguous and relative to processes and outcomes (Harvey and Green 1993), a common and pragmatic approach is to define a set of criteria that create, facilitate and stimulate conditions for effective instruction (Scheerens 2016). Such factors describe school characteristics that explain why some schools perform better than others and thus allow quantifying quality. School inspections follow this approach by using quality frameworks to set benchmarks. These descriptions of ‘good education’ however range from equity-related perspectives to categories based on educational effectiveness research (Ehren et al. 2013, p. 26).

There are common features of European school inspections, but there are also differences (see e.g. Ehren et al. 2013; van Bruggen 2010). New school inspection is obviously a ‘travelling policy’; embedding this policy into different national contexts will—according to Ozga and Jones (2006)—result in varying constellations which cannot be expected to function and produce in identical ways.

A feature which may distinguish between ways of functioning and variation of results is the ‘evaluative context’ of national education systems into which inspections are embedded (and which they themselves communicate to the school system). In high-stakes systems, school inspection is meant to control quality and to promote quality development by using mechanisms of hierarchical control and/or market mechanisms. They represent ‘hard governance’ structures which operate through target-setting, indicators, benchmarks and evaluations (Grek et al. 2013, p. 495). In such systems, ‘accountability pressure’ is meant to be an important factor to promote improvement activities (see Reezigt and Creemers 2005, p. 410). In inspection arrangements, pressure on schools can be regulated by using certain elements like differentiated inspections, thresholds for distinguishing failing schools, using comparative student
performance information, administering sanctions and incentives and publishing the inspection results of individual schools (see Altrichter and Kemethofer 2015). Low-stakes or ‘soft governance’ regimes on the other hand emphasise rational insight (Böttger-Beer and Koch 2008), self-regulation and supportive context as the spring boards of improvement. By providing meaningful information to schools, inspections are meant to give new insights to in-school actors which they will use for quality development of schools and classrooms.

New inspections aspire to be effective innovations to increase the quality of the education system. Thus, it is important and necessary to ask for their actual effects. As a consequence, the guiding purpose of our research is to study the effects of ‘new inspection models’. Since we assume that ‘inspection’ will not function in the same way under different contextual conditions, we are also interested in ascertaining these effects in different ‘evaluative contexts’.

2 Theoretical framework and previous research

2.1 Effects of school inspections

Given the innovatory claim of inspections, it is no wonder that there is quite a body of research which ventures to check its actual effects. Initially, most studies originated from the Anglo-Saxon context (representing hard governance structures), but lately, the number of studies from other parts of the world is increasing.

Rosenthal (2004) studied the direct effects of school inspections on student achievement in English secondary schools and found small but negative effects in the year of the school visit. Shaw et al. (2003) present evidence that OFSTED inspections lead to slight improvements of student achievement in schools performing above or below the average. In the majority of the schools, however, no effect was found (see also Cullingford and Daniels 1999). The results by Luginbuhl et al. (2009) indicate that inspection may improve student performance. They found evidence of improvement of 2–3% of a standard deviation of the CITO test score in the two years following a Dutch school inspection. These effects, however, disappeared in a smaller random sample and may be connected to sampling characteristics. Ehren and Shackleton (2015) conclude from Dutch longitudinal data that the actual inspection effects on school and teaching conditions are limited. Schools with (very) weak inspection judgements, however, show an increase in improving their self-evaluation and capacity building. To sum up, school inspections in high-stakes systems may lead to both positive and negative effects on student performance which, in any case, were rather small.

Turning to studies from soft governance contexts, we again come across inconsistent results. A questionnaire study in a one-year longitudinal control group design found no effect of school inspections on school quality as self-reported by principals and teachers in Berlin and Brandenburg (Gärtner et al. 2014; similar results by Böhm-Kasper and

---

1 There is also a body of research on unintended side effects of inspections which will not be discussed here in detail (e.g. de Wolf and Janssens 2007).
2 Office for Standards in Education, Children’s Services and Skills
3 Central Institute for Test Development
Selders 2013). Pietsch et al. (2014, p. 466) showed in two studies that inspection affects performance growth and performance trends of students in Hamburg. Their results indicate an improvement of nearly 20% of a standard deviation in reading; in mathematics a positive, but smaller effect appeared only in one study.

Pietsch et al. (2013, p. 15; see also Pietsch et al. 2015) attribute present inconclusive results to two shortcomings of previous effect studies: (1) The theoretical frameworks are not doing justice to the “complicatedness of the intervention”. They do not conceptualise the in-school processes in the course of inspections and “thereby exclude the concrete mechanisms which lead to goal achievement” (Husfeldt 2011, p. 10). (2) The methods used are “in most cases not appropriate to questions of causal analysis” (Pietsch et al. 2013, p. 15).

2.2 A theoretical framework for mediating processes in inspections

Ehren et al. (2013) developed a conceptual model describing the causal mechanisms of school inspections which are expected to lead to school development. For this purpose, they analysed official documents and interviewed inspection officials in six European countries (i.e. Austria/Styria, the Czech Republic, England, Ireland, the Netherlands and Sweden) to identify causal mechanisms of school inspections that are supposed to promote school improvement. Although these countries were chosen to represent different inspection regimes, a cross-case analysis of the six programme theories revealed that (a) all inspectorates share common goals (good education and high student achievement) and (b) refer to the same general mechanisms to meet these targets (see Fig. 1).

The left box in Fig. 1 includes inspection characteristics that may affect change in schools: Inspection methods, the standards used, the consequences applied and the specific features of reporting are considered critical issues which may alter the impact of inspections (see Ehren et al. 2013, pp. 15). Three mediating mechanisms were identified which inspectorial models assume to be causal in mediating developmental activities (see second column in Fig. 1; Ehren et al. 2013, pp. 21). Firstly, inspectorates are setting expectations about what a ‘good school’ should look like and how it is to operate. This is done by the quality frameworks which define inspection standards and criteria for good education. Accepting feedback, i.e. schools perceiving, accepting, appropriating and using inspection feedback for developing the quality of their work, is considered the second important mechanism (see Ehren et al. 2013, pp. 21). Actions of stakeholders, a third mediating factor, may influence development processes in schools by (a) creating external pressure on schools to fulfil the inspection criteria and/or (b) by providing support for schools. In order to stimulate such processes, many inspectorates ensure to communicate inspections standards, procedures and inspection reports to stakeholders.

The mediating mechanisms are preconditions for the school’s development activities which are represented in the Ehren et al. (2013, pp. 22) model by two factors, by promoting/improving self-evaluations and by taking development actions. These development activities, in turn, are supposed to result in improvement capacity and in

---

4 Translations of German sources are by the authors.
5 This variable is not included in the empirical model described later on due to case number limitations.
effective school and teaching conditions which may be seen as ‘interim results’ of development on occasion of inspections. Both, improvement capacity and effective school and teaching conditions represent quality indicators and therefore the preconditional for good education and high student achievement (Scheerens 2016).

What is the research base of this process model? The conceptual model is a ‘normative model’ in that it spells out the normative aspirations and goals which are associated with the inspection systems by the respective national authorities. Ehren et al. (2013) themselves critically discussed the assumptions of this ‘normative conceptual model’ by referring to state of the art research. They found evidence that supports some relationships described in the model; however, they also found some of the assumptions to be less substantiated by research. Using Dutch data, Ehren et al. (2015b) provided evidence that setting expectations is an essential factor for developmental processes in the wake of school inspections (see also Kemethofer et al. 2015 for Austria and Switzerland). According to them, inspection criteria influence school leaders’ behaviour and, as a consequence, there are indirect effects on school and teaching development. The relevance of ‘Setting expectations’ for the development of social institutions is also emphasised by theorists of Neo-Institutionalism (Meyer and Rowan 2006): Inspection standards, procedures and reports create normative pressure which stimulate schools—in their quest for legitimacy—to react in a way to school inspections that will enhance their legitimacy.

Accepting feedback is a core item of evidence-based models (Visscher and Coe 2002). Feedback research is well developed on the level of interpersonal interaction (Hattie and Timperley 2007); however, there are some doubts whether or not this body of knowledge can be transferred to the governance level of organisations, institutions and corporate actors (Altrichter et al. 2016). The extent to which schools accept and use performance feedback depends on the formal and content quality of the feedback (Visscher and Coe 2003), time and resources available at a school for processing feedback and on support and stimulation by the school management (Schildkamp and Visscher 2010). Penninckx et al. (2014) show that a constructive debate on inspection feedback in schools only takes place if inspection reports include concrete recommendations for improvement.
The inclusion of *Actions of stakeholders* in an inspection model assumes that a ‘third’ party will reinforce inspection expectations and may be justified by theories of social coordination (e.g. Schimank 2002). For Hamburg inspection data, Peters (2015) was able to show that school quality and good work with parents is correlated.

The entire process model was tested by Gustafsson et al. (2015). A path model based on data from six European countries provided some support for the hypothesised relations. In particular, it was found that *Setting expectations* and *Actions of stakeholders* had indirect effects on *Improvement actions for capacity building and school effectiveness*. However, in contrast to expectations, the model did not support the hypothesis that *Accepting feedback* was such a mechanism (similar results by Ehren et al. 2015b). While the path model approach produced some general support for the conceptual model, it cannot be claimed to provide evidence on causal effects of school inspections, since it was based on cross-sectional data which generally do not allow distinguishing correlation from causation. Another reason is that the model made no explicit connection between the presence of school inspection and the onset of the hypothesised mechanisms.

### 2.3 Research on the effects of different evaluative contexts

The hypothesis that inspections will produce different effects depending on the specific national context they are implemented in is well substantiated by empirical research. Ehren et al. (2015a) made use of the differences of the inspection systems in six European countries to compare the effects of four distinctive inspection elements. They found that inspections in *high-stakes* systems (i.e. accountability pressure, public reporting of inspections results, differentiated inspections, outcomes-orientation and sanctions) trigger more development activities in schools (reported by school leaders) than in *low-stakes* systems. At the same time, these hard governance approaches were also associated with more unintended side effects (e.g. narrowing the curriculum). Both findings were replicated by Altrichter and Kemethofer (2015).

*High-stakes* and *low-stakes* systems seem to differ in their impact on school development, however, they may be similar in their way of functioning (if this is conceptualised in the way Ehren et al. 2013 did). This is indicated by the very good model fit achieved by Gustafsson et al. (2015, p. 53). Also, Altrichter and Kemethofer (2015, p. 49) found some different paths of effects, however, no profoundly different overall models for English, Swedish und Austrian data.

### 2.4 Research questions

In this study, we compare the inspection models of Austria (Styria) and Sweden which allow not only contrasting different evaluative contexts but also two different types of educational governance by itself (Windzio et al. 2005): Altrichter and Kemethofer (2015, pp. 45) found that Swedish and Austrian school leaders differed significantly in the amount of accountability pressure experienced; according to their analysis, Austria (Styria) represents a *low-stakes* approach and Sweden can be seen as a *medium-to high-stakes* system. Comparing these two countries admits to study the effects of new inspection models under different contextual conditions.
As a consequence, the following research questions with respect to effects have been phrased:

Q1a. What effect do school inspections have on quality indicators of schools?
Q1b. Are there different effects on quality indicators in Sweden and Austria?

For Q1a, we expect—according to the proclaimed reform goals—that school inspections will have an impact on quality indicators. With respect to Q1b, we expect on the basis of prior research that there will be stronger effects on quality indicators reported by Swedish principals. Taking up the critique by Pietsch et al. (2013), we note that more elaborate theoretical frameworks and robust methods for causal analysis must be applied for discussing these questions.

Previous studies regularly used cross-sectional data. Our data allows studying some additional questions with respect to the theoretical framework:

Q2a. Which mediating processes are affected by school inspections?
Q2b. Are there differences with respect to mediating processes between Sweden and Austria?

Q2a is a replication of Gustafsson et al. (2015) with longitudinal data. In our analysis, we assume that the main effects are to be found after the inspection; thereby, we follow the traditional idea that schools will react to the inspection procedure and results. Beyond that, we expect effects that occur before inspections are sustainable and may last for a couple of years. Following Ehren et al. (2013), we also suppose that school inspections stimulate all three mediating mechanisms. With respect to Q2b, we expect that similar effects appear in both countries.

3 School inspection in Sweden and Austria

In this section, brief descriptions of the inspection models in the Austrian province of Styria and Sweden are given. Information refers to the state of the inspection systems in 2011 (i.e. the date of the first round of data collection).

3.1 Sweden

School inspection in Sweden has a long tradition going back more than 150 years (Gustafsson et al. 2014, p. 462). After having shut down inspections in 1990, concepts of New Public Management accompanied by the results of international large-scale assessments at the beginning of the twentieth century led to new policy adjustments. At this time decentralisation, recentralisation and marketisation became significant means for solving problems in schools (Hanberger et al. 2016; Lindgren et al. 2012). A key concept of the new governance structure was a re-introduction of school inspections in 2003. Since then, control was one major element for school improvement resulting in severe assessments and inspections (Lindgren et al. 2012). According to Hall and Sivesind (2015), the move towards hard governance can be dated with 2007. In this year, the Swedish Schools Inspectorate (SSI) became responsible for the execution of
inspections; however, first inspections did not take place before October 2008. Since 2011, the Education Act and Ordinance empowered the inspectorate to impose penalties. Based on interpretations of the Education Act and further regulations, the SSI performed school inspections including regular supervisions of all schools and thematic quality audits for a sample of schools (Segerholm and Hult 2016).

Lindgren et al. (2012, p. 538) summarised that “(...) the shift in focus from the inspection of soft areas, like norms and values, to hard areas, like knowledge and attainment, is interrelated with both the inspection techniques and a growing increasing reliance on evidence, objective data and knowledge collected using standardised processes independent of the local context.” School inspections strongly changed towards accountability pressure and a control-oriented evaluation. The SSI’s purpose is to hold schools accountable for achieving national objectives and ensure systematic quality work (Hanberger et al. 2016).

The inspectorate can visit municipal and independent schools whenever it decides; in practice, regular supervisions take place within a five-year cycle. Regular supervisions are based on previous inspection results and background information of the schools: Before the school visit, prior results are studied, and information of the school is gathered. Schools therefore have to fill in a template and provide background information. In addition, questionnaires are administrated to staff and stakeholders to collect supplementary data. On the basis of the materials sent in by the schools, the inspectorate decides whether there is a ‘basic supervision’ for well-functioning schools, a ‘widened supervision’, or a ‘deepened supervision’ for schools with clear problems. Schools are visited by two inspectors for one to two days, or if required even longer. The inspection includes lesson observations and interviews with school representatives (principal, teachers, pupils, school nurses, board members in charge). After the school visit, the principal receives oral feedback and a preliminary report with the possibility for reactions. The final report is made available on the internet and sent to the media. Within three months, schools have to react by producing an action plan that has to be accepted by the SSI (Segerholm and Hult 2016).

Based on the results, the inspectorate has the right to apply sanctions and even financial penalties. Teachers may receive a warning if their behaviour is not in line with the certification regulation. Besides regular visits, thematic quality audits take place every year. These audits focus on a specific issue including a small and randomly sampled number of schools. Two inspectors visit the chosen schools and provide individual feedback for the schools as well as a summary report for the general public (Gustafsson et al. 2014).

3.2 Austria (Styria)

In Austria, there is a long tradition of school inspection going back to the second half of the nineteenth century (Scheipl and Seel 1985). Traditionally, the role of inspectors included both administrative and supervisory tasks for the schools in their region. Although inspection criteria varied over time, the focus had always been a mixture of school quality attributes on the one hand and legal and administrative aspects on the other. In the wake of the PISA shock at the beginning of the twenty-first century, education policy reacted by ‘modernising’ the governance in education. Traditional ways of gathering information on the quality of schools seemed to lag behind new
demands. As a consequence, some regional inspectorates in Austria started to develop new approaches of school inspections. In Austria, school inspection is based on a central legal framework which however leaves room for different experimental approaches in the Austrian provinces (Altrichter et al. 2013). In the following description, we focus on the specific inspection system in the Austrian province of Styria (which is the province from which the Austrian data originate).

In reaction to a critique by the Austrian Court of Audit in 2003/2004 on the lack of comparable criteria in school inspections, the inspectorate in the province of Styria established a working group to devise a structured and more systematic procedure to inspect schools. In the school year 2007/2008, a ‘team inspection model’ which was inspired by inspection approaches in the Netherlands and the German state of Lower Saxony was introduced. From this time on, all schools in Styria received school inspections in regular intervals until in the school year, 2013/2014, a new type of quality management was introduced and replaced school inspections (see Altrichter 2017).

As long as this model of school inspection was in operation, the inspection teams consisted of two to three inspectors (in very small schools: one person). Depending on the size of a school, the visit lasted up to three days. The selection of schools was up to the ‘district inspector’, i.e. the administrative superior of the respective schools which was also a member of the inspection team. Inspection visits were to take place in regular intervals every two to four years. All schools were informed of an upcoming inspection visit well in advance. The district inspector arranged a meeting with the schools’ principal and staff to explain the focus and procedure of the inspection and to ensure that relevant information was available during the inspection visit. The visit included classroom observations, group interviews with teachers, students and parents’ representatives, a site inspection, an individual interview with the mayor and a meeting with the school leader. All activities were structured by forms.

Following the inspection visit, the inspection team prepared a written report to sum up the strengths and weaknesses of the school and point to potential fields for development. A preliminary draft report was presented by the district inspector and discussed in a staff meeting some days after the school visit. Afterwards, a final version of the report was sent to the school. The principal had to communicate the results of the school inspection to relevant school partners and stakeholders. On the basis of the inspection report, the school leader had to prepare a school development plan which served as a target agreement between the school and the inspectorate (Altrichter et al. 2013).

Since the Styrian model neither used any thresholds for labelling schools according to their performance nor other sanctions, it may be taken as a soft governance approach. School inspections aimed to promote quality development not through accountability pressure but by providing meaningful information and support to schools in order to promote rational insight.

4 Methodology

The data we use for discussing our research questions originate from the European project ‘Impact of school inspections on teaching and learning’ (ISI-TL; Ehren et al. 2013). School principals in primary and secondary education were asked to participate in an online survey in three consecutive years (2011, 2012 and 2013) to collect
information on the mechanisms and processes of school inspection. The inspection of different subsets of schools within this period allows analysis of the effects of school inspections in a longitudinal design. In Austria and Sweden, different sampling strategies were used. In Austria, all schools in the province of Styria (about 700) were sampled from which 148 schools participated in all three years, 193 schools filled in the questionnaire in two out of three years and another 190 responded at least once. In Sweden, a random sample of the population of primary and secondary schools comprised 2154 schools in total. Three hundred three principals completed the questionnaire once, 495 twice, and 419 responded at all three occasions. Details of the sampling strategy and the detailed response rates in both countries are provided in the technical report of the ISI-TL project (see http://www.schoolinspections.eu/impact). Furthermore, the data analysed and the models used are available from the authors upon request.

The questionnaire included 73 questions which operationalize the conceptual model by Ehren et al. (2013). Improvement capacity is based on information about ‘teacher participation in decision making’, ‘teacher co-operation’ and ‘transformational leadership’ (5 items, e.g. teachers are involved in making decisions about educational matters such as teaching methods, curriculum and objectives). Another set of questions investigates the extent the school adheres to Effective school and teaching conditions (5 items, e.g. the school has a safe and orderly social environment that is conducive to learning). In contrast to Gustafsson et al. (2015), we use information on the actual status of a school’s quality instead of development activities.

The questionnaire also included the mediating mechanisms of school inspections: Setting expectations was operationalized by four items (e.g. the inspection standards affect the evaluation and supervision of teachers), Accepting feedback by four items (e.g. the feedback received from the school inspectors was useful) and Actions of stakeholder by three items (e.g. the parents’ representatives of the school are sensitive to the contents of the school inspection report). Principals scored all these items on a 5-point scale ranging from strongly disagree to strongly agree. The development of the latent constructs and their measurement quality are described by Gustafsson et al. (2015).

The main outcome variable is Effective school and teaching conditions which is hypothesised to be influenced by Improvement Capacity (Geijsel et al. 2009). These two concepts must serve as dependent variables and indicators of impact of inspections since it was not possible to get access to other indicators of success (such as student performance data or inspection judgements) in our cross-national study. Since research has established several generalisations concerning school effectiveness factors (such as high teacher expectations, a challenging teaching approach, an orderly learning environment, feedback and clear and structured teaching) contributing to student achievement (Scheerens et al. 2007), this may be taken as best approximation available.

The mediating mechanisms as described above should on the one hand have an impact on these two dependent variables and on the other be influenced by whether the school was inspected or not. Following the theoretical framework from Ehren et al. (2013), Promoting/improving self-evaluation is assumed to affect our outcome variables and to be influenced by the mediating mechanisms.

6 The latent variables used in the model are indicated by capital initials and italics.
As explained above, the process model was tested by Gustafsson et al. (2015) in a cross-sectional design. In order to create a better basis for causal inference, we can take advantage of the longitudinal design of the study (e.g. Rindfleisch et al. 2008). A simple way to do that is to compare the responses of each school before and after inspection. If there is a change in the predicted direction for the different aspects measured, we have a reasonably strong basis for arguing that the change was an effect of the school inspection. With this design, the schools are their own controls, and if we can assume that the schools are stable from one year to another in aspects not related to school inspections, this implies a strong control of biasing influence from omitted variables which are time-invariant school characteristics.

Technically, estimation of models based on such an approach can be conducted in several different ways, such as for example with fixed-effects regression techniques (e.g. Angrist and Pischke 2008). In this study, estimation procedures based on latent variable growth modelling procedures implemented in the Mplus 7 program (Muthén and Muthén 1998-2015) were used because of their flexibility and versatility, the availability of tests of model fit and access to powerful procedures for missing data modelling.

The first round of measurement in both countries was conducted in the autumn (in most cases in October) of the academic year 2011/2012, the second in the autumn of 2012/2013 and the third in the autumn of 2013/2014. According to a traditional idea of school, inspection effects of the inspection activities are expected to show the year after the inspection was conducted, possibly also lasting the following couple of years if the effects are ‘sustainable’. This implies that the 2011/2012 wave of measurement would reflect effects of inspections conducted in 2010/2011, the 2012/2013 wave effects of inspections conducted in 2011/2012 and the 2013/2014 wave effects of inspections conducted in 2012/2013.

For Sweden 228, 202 and 228 schools were inspected in the years 2010/2011, 2011/2012 and 2012/2013, respectively. The schools (N = 658) included in these three waves form the Swedish sample used in the analysis. For Austria, the situation was more complicated, because in the school year 2012/2013, school inspections were terminated due to new political strategies for school development approaches (Kemethofer and Altrichter 2015). However, the data collection was conducted as usual in 2013/2014. Hence, for the sampled Austrian schools, there were three points of measurement as well. The difference is that no new set of inspected schools was added to the data in the last year of the study. This reduces the power of the study but does not change the logic of the analyses. For Austria, 53 schools were inspected in 2010/2011 and 40 schools in 2011/2012, so the sample analysed included 93 schools.

A separate growth model was defined for each construct, which was measured at the three waves of data collection. The model was specified in terms of a structural equation model (SEM). This implies that an intercept factor was defined as a latent variable with a fixed relation of unity to all three measures. Furthermore, a change factor was defined as another latent variable which had a fixed relation of unity to the variable observed the year after the inspection, and also to all the following observations. The change factor thus had the value of 1 for the relation to the 2011/2012, 2012/2013 and 2013/2014 observations for those inspected in 2010, to the 2012/2013 and 2013/2014 observations for those inspected in 2011/2012 and to the 2013/2014 observation for those inspected in 2012/2013. The model specification makes the
assumption that the effect of the inspection appears the year after the inspection, and then remains at the same level. An alternative model could be that the effect disappears after one year. However, comparisons between these models on the Swedish data clearly favoured the model making the assumption of lasting effects, so this model was used in all analyses.

The models were specified and estimated as multiple-group models using the Mplus software (Muthén and Muthén 1998-2015). Given that not all principals participated in all three waves of data collection, there was missing data. We used the maximum likelihood (ML) modelling methods implemented in Mplus to deal with this. These procedures are based on the assumption that the data is missing at random (MAR) which implies that missingness is assumed to be random, given the information in the data. This is a much less restrictive assumption than the missing completely at random (MCAR) assumption, and with the longitudinal design, the MAR-based ML modelling should be able to control for differences between the samples participating in the three waves of measurement (Lüdtke et al. 2007).

However, there was also another source of missingness. The three variables directly related to school inspections (i.e. Setting expectations, Actions of stakeholders sensitive to reports, and Accepting feedback) were at the first wave of data collection only administered to those participants who were inspected in 2010/2011, while these items were administered to all participants at the two other waves of data collection. Thus, this missingness partly is by design, albeit unplanned, and treated in the same way as the missingness due the principals’ decisions to participate or not.

The models were set up in identical ways for the different scales. The models for Sweden and Austria also were logically identical, except that a three-group model was estimated for Sweden and a two-group model was estimated for Austria. This is due to the abovementioned missing of inspected schools in Austria in 2012/2013. All estimated parameters were constrained to be equal across groups within each country.

To evaluate goodness-of-fit of the models to data, we use the chi-square goodness-of-fit test. We use the $p$ values from these tests to evaluate the acceptability of the models. However, as there are several shortcomings associated with the chi-square statistic (Schermelleh-Engel et al. 2003), we additionally refer to the ratio of chi-square and the number of degrees of freedom (df). In particular, this statistic eases the comparison between results for Austria and Sweden, given that the models for the two countries had different dfs. There is no absolute standard to identify a “good” or “acceptable” model, but the ratio should be as small as possible. Schermelleh-Engel et al. (2003) recommend a ratio of $\leq 3$, though some authors would accept a ratio of $\leq 5$ (Hayduk 1987). For Sweden, where most analyses involved around 600 schools, we also used the root mean square error of approximation (RMSEA) to take into account the fact that the chi-square test tends to detect even trivially small deviations between the hypothesised model and the data when the sample is large. This goodness-of-fit index measures the amount of deviation between the hypothesised model and the actual model on an absolute scale, and it offers fairly simple rules for accepting and rejecting models. Following Little (2013), we accept models as fitting the data acceptably when RMSEA $<0.08$. For Austria, the analyses involved only around 90 schools, and for these data, the RMSEA is not applicable.

To make the estimates comparable across scales and the two samples, we divided them by the standard deviation of the intercept parameter, thereby creating an effect
size measure similar to Cohen’s $d$. Thompson (2007) argues to interpret effect sizes in direct and explicit reference to related studies. Hattie (2009, p. 240) for example used a value of $d = .15$ as a benchmark. In consideration of the fact that small effects may be meaningful as well, we interpret effects of $d = 0.15–0.34$ as small, $d = 0.35–0.50$ as medium and $d > 0.50$ as large.

5 Results

Goodness-of-fit statistics for the models are presented in Table 1. For the main outcome variable Effective school and teaching conditions, the chi-square test showed model fit to be good in both Austria and Sweden, and for Sweden, this was also the case for Improvement Capacity. For the other scales, the fit of the model was less good. However, for Promoting/improving self-evaluations, the RMSEA estimate was 0.078, which indicates acceptable fit. The RMSEA index showed that fit was poor for Setting expectations, Actions of stakeholders and Accepting feedback with values of 0.105, 0.098 and 0.093, respectively. As mentioned above, these three mediating mechanisms directly related to school inspections were at the first wave of data collection only administered to those participants who were inspected in 2010/2011. This may have caused a lack of comparability of the responses to these particular items for schools inspected at the different waves of data collection that could not be corrected for by the missing-data modelling technique. We will look further into this issue below, by investigating alternative model specifications for these three variables.

The estimates of the effect of school inspection are presented in Table 2. For Effective school and teaching conditions, there was a significant positive effect for Sweden, with a small effect size. The estimated effect was not significant for Austria. The lack of statistical significance is likely to be due to the small sample. The estimated effect of inspections on Effective school and teaching conditions was small too but somewhat higher than for Sweden.

For Improvement capacity, a significant treatment effect was estimated for Sweden, too, with a moderate effect. The Austrian data did not indicate any effect, and here too, this may be due to the small sample size. For Promoting self-evaluations, there were positive, albeit non-significant, estimates in both countries. However, the lack of significance in both countries suggests that these positive estimates are chance effects.

The results for Accepting feedback were positive and significant in Austria and Sweden, with a moderate effect in Sweden and a large effect in Austria. However, given the poor fit of this model, these estimates cannot be accepted without further investigation. Inspection of modification indices suggested that for the group of schools that was inspected first (i.e. in 2010), both the variance of the change parameter and the residual variance of the first measurement was smaller than for the two other groups of schools. Allowing these modifications, the re-estimated model had acceptable fit (Chi-square = 27.50, df = 17, $p < .05$, RMSEA = .056). However, the estimate of the effect of school-inspection was the same in this model ($0.11, t = 2.29$) as in the unmodified model, showing that the estimate was robust against this source of model misfit.

For Setting expectations, a positive and significant estimate was observed for Austria. In Sweden, in contrast, there was a medium, non-significant negative effect of school inspections. However, given that the fit of the model for Sweden was poor,
this result requires closer scrutiny. Here too, the modification indices were the starting point for model revision, and they suggested that the group of schools inspected in the first round had a lower mean on the first measure of Setting Expectations, while the two groups inspected thereafter had higher means on the second measure of Setting expectations. The re-estimated model had acceptable fit (chi-square = 33.76, df = 17, \( p < .01 \), RMSEA = 0.071). In this case, the estimated change parameter turned from being negative and insignificant (see Table 2) to being positive and insignificant (0.14, \( t = 1.53 \)). This brings findings from Austria and Sweden more in line with one another, but given the ad hoc nature of the analysis, no strong conclusions can be drawn concerning effects of school inspections on this variable.

There was no significant effect of school inspections on Actions of stakeholders in either of the countries. Fit was not good in either Austria or Sweden. The model for Sweden was modified in a similar manner as the model for Setting expectations which resulted in good fit (chi-square = 16.94, df = 17, \( p < .46 \), RMSEA = 0.000). In the modified model, the estimate of the effect of school inspection was close to zero (−0.03, \( t = −0.07 \)).

The models discussed so far have investigated effects of school inspections for one variable at a time. To investigate whether there were any direct or indirect effects of mediating processes on our outcome variables, we ran a growth model in which we could relate the different variables to one another. The basic idea was to join together the six separately estimated growth curve models into one model with six correlated growth curves. For Sweden, the correlations among the effects on the mediating variables were very high (0.75–0.90). Correlations between the estimates for intermediary factors and outcomes indicated only one significant effect between Accepting feedback and Improvement capacity (\( r = .37 \)). However, the model fit was poor. For Austria, the model failed to converge due to the small number of observations.

### 6 Conclusion and discussion

In many education systems, school inspections have been introduced as a major instrument for sustaining and improving school quality. School inspections mostly follow the same basic ideas; however, there is a wide variation between

---

**Table 1** Goodness-of-fit of the growth models

|                     | Austria Chi-square | df | \( p \) | \( \chi^2/df \) | Sweden Chi-square | df | \( p \) | \( \chi^2/df \) |
|---------------------|-------------------|----|--------|----------------|------------------|----|--------|----------------|
| Setting expectations | 17.00             | 11 | .108   | 1.55           | 63.69            | 20 | .000   | 3.18           |
| Actions of stakeholders | 35.67         | 11 | .000   | 3.24           | 57.87            | 20 | .000   | 2.89           |
| Accepting feedback  | 37.75             | 11 | .000   | 3.43           | 53.50            | 20 | .000   | 2.68           |
| Promoting/improving self-evaluations | 44.41 | 11 | .000   | 4.04           | 46.63            | 20 | .001   | 2.33           |
| Improvement capacity | 22.29             | 11 | .022   | 2.03           | 16.81            | 20 | .666   | .84            |
| Effective school and teaching conditions | 12.84           | 11 | .304   | 1.17           | 24.78            | 20 | .210   | 1.24           |
countries when it comes to the implementation and special elements of inspections. Surprisingly, there is only little comparative research on the impact of school inspections. In the wake of the PISA results, Sweden and Austria have undergone several changes in the governance of education, among them the development of ‘new school inspections’. While Austria followed a soft governance approach, the Swedish accountability system includes sanctions, publications of inspection reports and differentiated inspections and can be considered as medium- to high-stake accountability regime. A process model of inspections by Ehren et al. (2013) was used to conceptualise how impact of school inspections may be generated in schools. In this sense, our study focused on inspection effects in Sweden and Austria as well as on checking the empirical plausibility of a normative model of inspection effects. For discussing our research questions, longitudinal data is used which was collected by means of a survey to principals in three consecutive school years. Due to the longitudinal design of the study, we were also able to discuss some questions with respect to specific processes meant to contribute to inspection impact. Hence, our study expands existing cross-sectional and comparative research on effects of school inspections in different contexts.

There are some limitations that need to be addressed before discussing our results. First of all, our data only represents the perspective of principals; their retrospective self-reports may be influenced by memory errors or social desirability. Nevertheless, we consider principals as an adequate source of information for school quality, if access to performance data and of inspection judgements is lacking. A second limitation is missing data. We dealt with this problem by using missing-data modelling approaches. Furthermore, due to the small number of inspected schools, the power of our model is restricted and particularly so for Austrian data. To estimate direct or indirect effects on our outcome variables, more observations would have been necessary. In addition, it would have been beneficial to follow schools over a more extended period of time to study long-term and cumulative effects of school inspections. Besides that, the termination of school inspections in Austria after the second year of the study had to be accounted for in modelling the data. Finally, our results

|                          | Austria Estimate | t-value | d-value | Sweden Estimate | t-value | d-value |
|--------------------------|------------------|---------|---------|-----------------|---------|---------|
| Setting expectations     | 0.25             | 2.02    | 0.37    | −0.08          | −1.75   | −0.24   |
| Actions of stakeholders  | 0.03             | 0.10    | 0.06    | 0.06           | 1.30    | 0.17    |
| Accepting feedback       | 0.43             | 2.58    | 1.11    | 0.11           | 2.32    | 0.29    |
| Promoting/improving self-evaluations | 0.26 | 1.68 | 0.45 | 0.04 | 1.39 | 0.16 |
| Improvement capacity     | −0.05            | −0.74   | −0.10   | 0.08           | 3.69    | 0.25    |
| Effective school and teaching conditions | 0.08 | 1.15 | 0.26 | 0.05 | 2.36 | 0.17 |

Note. Significant estimates ($p \leq .05$) in bold
have to be interpreted with some caution as the model fit statistics for most models are inadequate according to conservative guidelines.

Our results indicate that—as to research question 1a—school inspections in Austria and Sweden do have a small to medium impact on the major impact indicator, *Effective school and teaching condition*. While the effect estimate reached statistical significance in Sweden only, the higher effect size in Austria suggest that the lack of significance in this country was due to low power. In line with Pietsch et al. (2014), we were thus able to show that using more resolute causal analysis techniques unveils a positive effect of inspection systems in both countries on implementation of improved conditions for learning. The conceptual model by Ehren et al. (2013) hypothesises that such improvement would be mediated via *Improvement capacity*; for Sweden, there was, indeed, a significant effect of school inspection on this indicator. For Austria, no such effect could be observed. The impact of school inspections on the mediating process *Promoting and improving self-evaluation* was positive in both Austria and Sweden, but the effect did not reach statistical significance in neither of the educational systems.

Research question 1b focuses on differences in effects of school inspections between Austria and Sweden, and we hypothesised stronger effects in Sweden. There was one more significant effect in Sweden than in Austria and t-values generally were higher in Sweden than in Austria. However, these differences seem to be expressions of the considerably larger Swedish sample rather than of any generally stronger effects in Sweden. Rather, it seems that school inspections in the two countries have different quality foci. Our results indicate that inspections in Sweden promote *Improvement capacity* while in Austria, there was a tendency for *Self-evaluation* to be a quality tool affected by inspections. Both indicators, however, may lead to good school and teaching conditions.

With respect to the *mediating mechanisms* which are hypothesised to stimulate school development in research question 2a, a surprising result emerges from our analysis: Contrary to prior cross-sectional studies, school inspections have a positive and significant effect on *Accepting feedback* in both countries. For *Actions of stakeholders*, no evidence for effect was found in both countries. For the third mechanism, *Setting expectations*, a positive and significant effect, showed up in Austria and a negative but non-significant effect in Sweden. However, after modification of the model, the estimate was close to zero.

Our explanation is that the difference in findings reflects different analytic approaches: Previous studies on intermediary mechanism used cross-sectional data and highlighted the effect of inspections on *Setting expectations*. Theoretically and empirically, it makes sense to assume that ‘expectations’ are affected by inspections primarily before and during inspections (see also Gärtner et al. 2009; Ehren et al. 2015a, pp. 20): Before the inspection visit, schools check their operations against the quality criteria (communicated by the mechanism *Setting expectations*) and make appropriate adjustments.

In contrast, *Accepting feedback* is, by definition, taking place after the inspection when schools process and react to the inspection results. This effect can only be unveiled by repeated observations. A possible consequence is that inspections may have effects both before and after the inspection visits, however, through different mechanisms.

To turn to research question 2b: Our results indicate—in conformity with other comparative studies (e.g. Altrichter and Kemethofer 2015; Ehren et al. 2015a)—
different effects of school inspections. This result is in line with our expectation that variation in evaluative contexts makes a difference. In comparing hard and soft governance approaches, the effect of inspections for Accepting feedback was much higher in Austria. This effect is in line with the findings of Ehren et al. (2015a) and Altrichter and Kemethofer (2015) and may be explained by specific governance approaches. Feedback is more likely to be accepted in low-stakes systems, while high-stakes environments produce much accountability pressure which is not conducive for processing and using the informational messages of inspection feedback. A possible explanation for the different effects found for Setting expectations may be due to that more accountability pressure principals in Sweden (and in high-stakes systems in general) may place more attention to inspection criteria before school inspections in order to prevent sanctions, while in low-stakes systems, inspection criteria may be viewed as long-term targets which are considered important over a longer time span.

Our study firmly illustrates the importance of using repeated observations and comparative approaches when measuring the effects of treatments such as school inspections. Conflicting results from cross-sectional and longitudinal studies are most likely due to different methodological approaches. While cross-sectional studies are not able to distinguish between correlation and causality, there should be much less risk of biased inference due to reversed causality and omitted variables in our analysis. In future research, it is a key issue to include teacher information on teaching conditions, student performance data and other independent quality information in analysing the intended effects of school inspections. In consideration of the costs of school inspections, it seems necessary to include also an appreciation of unintended effects (see de Wolf and Janssens 2007; Altrichter and Kemethofer 2015).

Acknowledgements Open access funding provided by Johannes Kepler University Linz. This research was funded by the EU Lifelong Learning Programme, grant number 511490-LLP-1-2010-1-NL-KA1-KA1SCR

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Altrichter, H. (2017). The short flourishing of an inspection system. In J. Baxter (Ed.), School inspectors. Policy implementers and policy shapers in national policy contexts (pp. 207–230). London: Springer.
Altrichter, H., & Kemethofer, D. (2015). Does accountability pressure through school inspections promote school improvement? School Effectiveness and School Improvement, 26(1), 32–56.
Altrichter, H., & Maag Merki, K. (2016). Steuerung der Entwicklung des Schulwesens. In H. Altrichter & K. Maag Merki (Eds.), Handbuch Neue Steuerung im Schulsystem (pp. 1-27). Wiesbaden: Springer VS.
Altrichter, H., Kemethofer, D., & Schmidinger, E. (2013). Neue Schulinspektion und Qualitätsmanagement im Schulsystem. Erziehung und Unterricht, 169(9–10), 961–978.
Altrichter, H., Moosbrugger, R., & Zuber, J. (2016). Schulentwicklung und Unterrichtsentwicklung durch Datenrückmeldung. In H. Altrichter & K. Maag Merki (Eds.), Handbuch Neue Steuerung im Schulsystem (pp. 235–277). Wiesbaden: Springer VS.
Angrist, J. D., & Pischke, J. S. (2008). Mostly harmless econometrics: an empiricist’s companion. Princeton: Princeton University Press.
Barber, M. (2004). The virtue of accountability: system redesign, inspection, and incentives in the era of informed professionalism. Journal of Education, 185(1), 7–38.
Böhm-Kasper, O., & Selders, O. (2013). Schulinspektionen sollten regelmäßig durchgeführt werden? *Die Deutsche Schule, 12.* Beiblatt, 121–153.

Böttger-Beer, M., & Koch, E. (2008). Externe Schulevaluation in Sachsen. In W. Böttcher, W. Bos, H. Döbert, & H. G. Holtappels (Eds.), *Bildungsmonitoring und Bildungscontrolling in nationaler und internationaler Perspektive* (pp. 253–264). Münster: Waxmann.

Cullingford, C., & Daniels, S. (1999). Effects of Ofsted inspections on school performance. In C. Cullingford (Ed.), *An inspector calls* (pp. 59–69). London: Kogan.

De Wolf, I. F., & Janssens, F. J. G. (2007). Effects and side effects of inspections and accountability in education: an overview of empirical studies. *Oxford Review of Education, 33*(3), 379–396.

Döbert, H., Rürup, M., & Dedering, K. (2008). Externe evaluation von Schulen in Deutschland. In H. Döbert & K. Dederig (Eds.), *Externe evaluation von Schulen* (pp. 63–151). Münster: Waxmann.

Ehren, M. C. M., & Honingh, M. E. (2011). Risk-based school inspections in the Netherlands. *Studies in Educational Evaluation, 37*, 239–248.

Ehren, M. C. M., & Shackleton, N. (2015). Mechanisms of change in Dutch inspected schools: comparing schools in different inspection treatments. *British Journal of Educational Studies, 64*(2), 185–213.

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

Ehren, M. C. M., Gustafsson, J.-E., Altrichter, H., Skedsmo, G., Kemethofer, D., & Huber, S. G. (2015a). Comparing effects and side effects of different school inspection systems across Europe. *Comparative Education, 51*(3), 375–400.

Ehren, M. C. M., Perryman, J., & Shackleton, N. (2015b). Setting expectations for good education: how Dutch school inspections drive improvement. *School Effectiveness and School Improvement, 26*(2), 296–327.

Faubert, V. (2009). School evaluation: current practices in OECD countries and a literature review. [http://www.oecd-ilibrary.org/education/school-evaluation_218816547156. Retrieved: 6 March 2013.](http://www.oecd-ilibrary.org/education/school-evaluation_218816547156. Retrieved: 6 March 2013.)

Gärtner, H., Hüsemann, D., & Pant, H. A. (2009). Wirkungen von Schulinpskierung aus Sicht betroffener Schulleitungen. *Empirische Pädagogik, 23*(2), 1–18.

Geijsel, P., Sleeppers, 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.

Grek, S., Lawn, M., Ozga, J., & Segerholm, C. (2013). Governing by inspection? *Comparative Education, 49*(4), 486–502.

Gustafsson, J.-E., Ehren, M. C. M., Conyngham, G., McNamara, G., Altrichter, H., & O’Hara, J. (2015). From inspection to quality: Ways in which school inspection influences change in schools. *Studies in Educational Evaluation, 47*, 47–57.

Gustafsson, J.-E., Landt, R., & Myrberg, E. (2014). Inspection of Swedish schools. *Education Inquiry, 5*(4), 461–479.

Hall, J. B., & Sivesind, K. (2015). State school inspection policy in Norway and Sweden. *Journal of Education Policy, 30*(3), 429–458.

Hanberger, A., Carlbaum, S., Hult, A., Lindgren, L., & Lundström, U. (2016). School evaluation in Sweden in a local perspective: A synthesis. *Sweden. Education Inquiry, 7*(3), 349–371.

Harvey, L., & Green, D. (1993). Defining quality. *Assessment & Evaluation in Higher Education, 18*(1), 9–34.

Hattie, J. (2009). Visible learning: a synthesis of over 800 meta-analyses relating to achievement. London: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL. Essentials and advances.* Baltimore: The John Hopkins University Press.

Husfeldt, V. (2011). Wirkungen und Wirksamkeit der externen Schulevaluation. *Zeitschrift für Erziehungswissenschaft, 14*(2), 259–282.

Jones, K., & Tymms, P. (2014). Ofsted’s role in promoting school improvement. *Oxford Review of Education, 40*(3), 315–330.

Kemethofer, D., & Altrichter, H. (2015). Schulqualität Allgemeinbildung (SQA) in der Einschätzung von Schulleitungen allgemeinbildender Pflichtschulen. *Erziehung und Unterricht, 165*(7–8), 675–690.

Kemethofer, D., Skedsmo, G., Altrichter, H., & Huber, S. G. (2015). Steuerung durch Erwartungen? *Schulverwaltung spezial, 17*(1), 26–28.

Lindgren, J., Hult, A., Segerholm, C., & Römberg, L. (2012). Mediating school inspection. *Education Inquiry, 3*(4), 569–590.

Little, T. D. (2013). *Longitudinal structural equation modeling.* New York: The Guilford Press.
Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Psychologische Rundschau, 58(2), 103–117.

Luginbuhl, R., Webbink, D., & de Wolf, I. F. (2009). Do inspections improve primary school performance? Educational Evaluation and Policy Analysis, 32(3), 221–237.

Meyer, H.-D., & Rowan, B. (2006). The new institutionalism in education. Albany: State University of New York Press.

Muthén, L. K., & Muthén, B. O. (1998-2015). Mplus user’s guide (7th ed.). Los Angeles: Muthén & Muthén.

Ozga, J., & Jones, R. (2006). Travelling and embedded policy: the case of knowledge transfer. Journal of Education Policy, 21(1), 1–17.

Penninx, M., Vanhoof, J., De Maeyer, M., & Van Petegem, P. (2014). Exploring and explaining the effects of being inspected. Educational Studies, 40(4), 446–470.

Peters, S. (2015). Eltern als Stakeholder von Schule. In M. Pietsch, B. Scholand, & K. Schulte (Eds.), Schulinspektion in Hamburg. Der erste Zyklus 2007–2013 (pp. 341–365). München: Waxmann.

Pietsch, M., Janke, N., & Mohr, I. (2014). Führt Schulinspektion zu besseren Schülerleistungen? Zeitschrift für Pädagogik, 60(3), 446–470.

Pietsch, M., van den Ham, A.-K., & Köller, O. (2015). Wirkungen von Schulinspektion. In M. Pietsch, B. Scholand, & K. Schulte (Eds.), Schulinspektion in Hamburg. Der erste Zyklus 2007–2013 (pp. 117–136). München: Waxmann.

Rindfleisch, A., Malter, A. J., Ganesan, S., & Moorman, C. (2008). Cross-sectional versus longitudinal survey research: concepts, findings, and guidelines. Journal of Marketing Research, 45(3), 312–325.

Rosenthal, L. (2004). Do school inspections improve school quality? Economics of Education Review, 23, 143–151.

Scheerens, J. (2016). Educational effectiveness and ineffectiveness. A critical review of the knowledge base. Dordrecht: Springer.

Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). Review and meta-analyses of school and teaching effectiveness. https://www.utwente.nl/bms/omd/Medewerkers/artikelen/scheerens/rev-meta-analysis 2007.PDF. Retrieved: 12 Oct 2015.

Schildkamp, K., & Visscher, A. J. (2010). The use of performance feedback in school improvement in Louisiana. Teaching and Teacher Education, 26, 1389–1403.

Schimank, U. (2002). Handeln und Strukturen. Weinheim: Juventa.

Segerholm, C., & Hult, A. (2016). Learning from and reacting to school inspection—two Swedish case narratives. Scandinavian Journal of Educational Research, 1–15.

Shaw, I., Newton, D. P., Atikin, M., & Damell, R. (2003). Do OFSTED inspections of secondary schools make a difference to GCSE results? British Educational Research Journal, 29(1), 63–75.

Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. Psychology in the Schools, 44(5), 423–432.

Van Bruggen, J. C. (2010). Inspectorates of education in Europe. http://www.sici-inspectorates.eu/Members/Comparative-Analyses/Inspectorates-of-Education-in-Europe. Retrieved: 21 Aug 2013.

Visscher, A. J., & Coe, R. (2002). School improvement through performance feedback. London: Routledge. Visscher, A. J., & Coe, R. (2003). School performance feedback systems. School Effectiveness and School Improvement, 14(4), 321–349.

Wilcox, B. (2000). Making school inspection visits more effective: the English experience. Paris: UNESCO-IIEP http://unesdoc.unesco.org/images/0012/001202/120289E.pdf. Retrieved: 23 Dec 2015.