School Effectiveness and School Improvement
An International Journal of Research, Policy and Practice

ISSN: 0924-3453 (Print) 1744-5124 (Online) Journal homepage: https://www.tandfonline.com/loi/nses20

Explaining effects and side effects of school inspections: a path analysis

Maarten Penninckx, Jan Vanhoof, Sven De Maeyer & Peter Van Petegem

To cite this article: Maarten Penninckx, Jan Vanhoof, Sven De Maeyer & Peter Van Petegem (2016) Explaining effects and side effects of school inspections: a path analysis, School Effectiveness and School Improvement, 27:3, 333-347, DOI: 10.1080/09243453.2015.1085421

To link to this article: https://doi.org/10.1080/09243453.2015.1085421

SCHOOL EFFECTIVENESS AND SCHOOL IMPROVEMENT, 2016
VOL. 27, NO. 3, 333–347
http://dx.doi.org/10.1080/09243453.2015.1085421

Routledge
Taylor & Francis Group

Explaining effects and side effects of school inspections: a path analysis

Maarten Penninckx, Jan Vanhoof, Sven De Maeyer and Peter Van Petegem

Edubron Research Unit, University of Antwerp, Antwerp, Belgium

ABSTRACT

There are large differences between schools with regard to how they are affected by a school inspection. This study provides quantitative evidence about the extent to which perceived effects and side effects of an inspection are related to the inspection's judgement on the school, to features of the inspection, and to school features. This study draws on quantitative data collected from 2,202 teachers in Flemish primary and secondary schools that had recently been inspected. Using path analysis, it investigates the existence and strength of relationships between the perceived effects and side effects of inspection, on the one hand, and the perceived inspection quality, the perceived school's policy-making capacities, and the inspection judgement, on the other hand. The analysis reveals that most effects and several side effects are predominantly explained by the perceived inspection quality. Emotional side effects during inspection are best predicted by the school's policy-making capacities.

ARTICLE HISTORY
Received 9 November 2014
Final version received
22 July 2015

KEYWORDS
School inspection; effects; side effects; path analysis; quantitative study

Problem statement

Mechanisms for school evaluation have been established in many education systems, as external pressure on schools is deemed necessary to ensure the educational quality provided by schools (Sun, Creemers, & De Jong, 2007). The predominant arrangement for external school evaluation is school inspection (Organisation for Economic Co-operation and Development [OECD], 2013). We define a school inspection as:

> *an evaluation of the quality of schools including minimally a site visit leading to a judgement on whether the quality of schools is meeting the expected standards, by persons with specific expertise who are not directly or indirectly involved in the school.*

Apart from its role in holding schools accountable for their actions, most Inspectorates pursue development goals (Faubert, 2009; OECD, 2013).

There has been considerable research interest in the effects of inspection on schools. Both the intended effects and the (undesirable) side effects have been documented (and summarized by, e.g., Chapman & Earley, 2010; Ehren & Visscher, 2006; Learmonth, 2000; Muijs, Harris, Chapman, Stoll, & Russ, 2004). However, a great deal of the literature on the effects and side effects of inspections is of a descriptive nature. In contrast, little attention has been paid to profound analysis aimed at explaining features that may strengthen the intended effects on schools and maximally reduce the side effects (Chapman & Earley, 2010; De Wolf & Janssens, 2007). The few examples of explanatory studies indicate by and large that differences between schools regarding the effects

CONTACT: Maarten Penninckx Maarten.penninckx@uantwerp.be

and side effects of inspection are explained by a mixture of school characteristics, inspection characteristics, and the inspection judgement itself (Ehren & Visscher, 2006; Matthews & Sammons, 2004). School characteristics deemed important are the school's innovative capacity, shared leadership, as well as professional relationships between staff members (Lowe, 1998; McCrone et al., 2007; Scanlon, 1999), while relevant inspection features are the inspection's transparency, its psychometric quality, and the inspectors' behaviour (Chapman, 2001b; Cuckle & Broadhead, 1999; McCrone et al., 2007).

Not only is the evidence regarding these explaining features currently still rather limited, the available evidence is also largely embedded in English education settings. Validation of the results from these studies in other countries is required. Research in several educational contexts is an imperative, as it is yet unclear whether different inspection systems lead to different effects on schools, and whether or not different features matter in terms of the occurrence of these effects (De Wolf & Janssens, 2007; OECD, 2013).

The present study contributes to the knowledge base by presenting insights into the explaining features of the effects of inspection in Flemish (Belgium) schools. The Flemish educational context provides an interesting background for this kind of research. The reported effects of inspection are not entangled with the effects of other accountability measures as the Flemish education system is characterized by the absence of central examinations. We refer to the next section for a concise description of the Flemish educational accountability system.

The study is designed to serve an explanatory purpose, and starts from the following research question: To what extent can differences between teachers' perceptions regarding the effects and side effects of school inspection be explained (a) by the inspection judgement, (b) by characteristics of the inspection, and (c) by characteristics of the school? Given our aim to explain the effects and side effects of inspection, it is important to obtain a view of the school staff members' perception of the relevant inspection processes, as well as the teachers' assessment of the effects and side effects with regard to their school. These perceptions are important given that teachers are the key figures in establishing effects in the school after the inspection, and because they are also best placed to report on their personal experience of any side effects (MacBeath, 1999; Sandbrook, 1996; Vandenberghe, 2010). Furthermore, this study also draws on perceptions of staff members on their school, as it has recently been pointed out that these perceptions are not affected by the fact that schools have been inspected (Gärtner, Wurster, & Pant, 2014). As a consequence of our interest into perceptions of staff members, the analyses on the above research question are limited to the individual level.

Research context

Given the context-inclusive nature of this study, we first provide a thumbnail sketch of the accountability context in Flemish education. Like many Inspectorates in other educational contexts, the Flemish Inspectorate has explicitly set out on both accountability- and improvement-oriented aims (Vanhoof & Van Petegem, 2007). Its ambition is to control and to contribute to the schools' educational quality at the same time (Michielsens, 2007; Onderwijsinspectie, 2013). Every school is inspected at least once every 10 years. An inspection leads to a judgement on the school (either "positive", "restricted positive", or "negative") which applies to the whole school or to a specific department in the school. The Inspectorate does not conduct full inspections. Instead, two or three education areas (e.g., mathematics, arts education) and a number of process indicators (e.g., staff policy, pupil evaluation) are selected for an in-depth inspection. There is a "sub-judgement" for each of the selected education areas, which implies that it may be the case that one department in the school is assessed as "positive" while the school in general receives a "restricted positive" judgement. A "positive" judgement means a school (department) is considered to have the competencies and preparedness to continue working in an optimal manner, and that no follow-up inspection is required. A "restricted positive" judgement denotes that a second

inspection is required in 3 years' time, to determine whether or not the identified shortfalls have been adequately addressed. Schools that show structural deficiencies are given a "negative" judgement, which most often implies that the school needs to create an improvement plan by which it is obliged to have itself monitored by an external agency. During the school year 2012–2013, the judgements "positive", "restricted positive", and "negative" were given to respectively 60.6%, 36.1%, and 3.3% of the inspected primary and secondary schools (Onderwijsinspectie, 2013).

Comparable to other education contexts, education standards are set by the government. These education standards are conceived as minimal attainment targets that pupils should achieve at the end of a specified stage during their education. Nevertheless, it is left to schools to assess whether or not pupils have reached these targets: There is no standardized exam, nor do inspectors test pupils during their visit.

Conceptual framework

In this paragraph, we provide an overview of the concepts that are included in the present study, together with the relevant evidence that is available from studies in other educational contexts. The effects, side effects, and explanatory variables and their expected relationships are summarized in Figure 1.

Perceived effects and side effects

We distinguish two kinds of intended effects. Based on Rossi, Lipsey, and Freeman (1999), we define the "*conceptual effect*" as "the influence of the inspection on the understanding or reflection of principals and teachers in schools". Inspection is expected to provide schools with a valuable analysis of their strengths and weaknesses, and to spur reflection in the school on their quality (Matthews & Sammons, 2004). Although it has indeed been argued that teachers may show an increased reflection on their school's qualities (Brimblecombe, Shaw, & Ormston, 1996; Chapman, 2002; Dimmer & Metiuk, 1998), studies have shown that the inspection itself most often does not

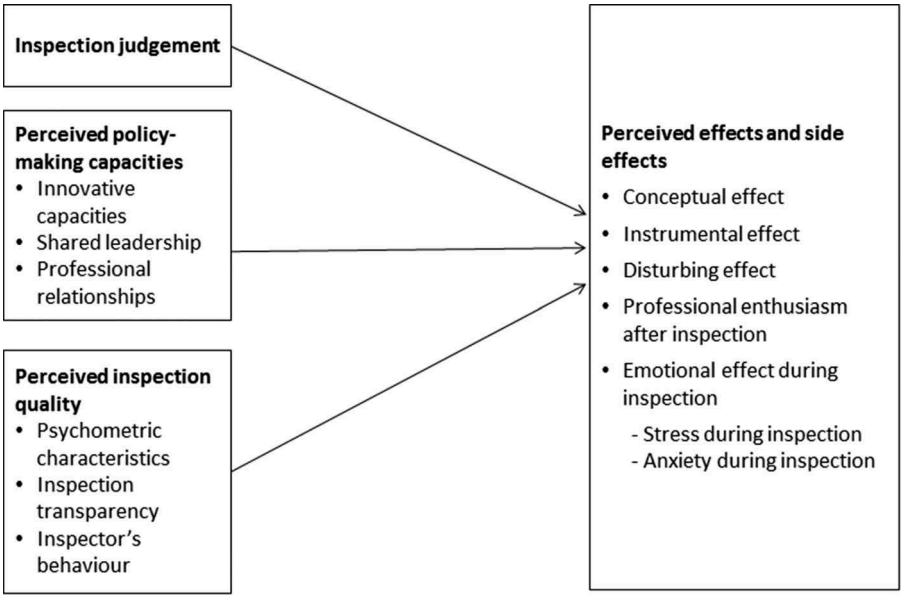

Figure 1. Conceptual framework.

lead to new insights into the school's own functioning (e.g., Chapman, 2002; McCrone et al., 2007; Wilcox & Gray, 1996). We define the "*instrumental effect*" as related to "the decisions taken as a result of the inspection and the actions that are based upon these decisions" (Rossi et al., 1999). Schools are expected to act on the views and expectations provided by the inspection and to use the inspector's feedback for setting up improvement actions (Ehren, Perryman, & Shackleton, 2015). Descriptive evidence shows that inspection may lead schools to change their policies, but that there is a large variety between schools, ranging from no instrumental effects to strong ones (Lowe, 1998; McCrone et al., 2007; Ouston, Fidler, & Earley, 1997).

Inspection comes at a cost, as different side effects may occur (De Wolf & Janssens, 2007; Ehren & Swanborn, 2012). Due to the inspection, teachers may have less time for their normal tasks, for example, lesson preparation or engagement in projects (e.g., Kogan & Maden, 1999; Lonsdale & Parsons, 1998; Stoll & Fink, 1998). This side effect – prior to, and during the inspection – is referred to as the *disturbing effect on normal school life*. Furthermore, descriptive research attention regarding side effects has mainly focused on the *emotional impact during the inspection* (e.g., Brunsden, Davies, & Shevlin, 2006; Chapman, 2002; Nicolaidou & Ainscow, 2005; Perryman, 2007; Scanlon, 1999). There are, however, numerous ways in which this emotional impact is conceptualized in different studies (De Wolf & Janssens, 2007). As "emotional impact" is hard to grasp directly, we opted to measure the two feelings that are mentioned most in this kind of research, namely, an increase in stress and in levels of anxiety during the inspection. Finally, there is also a side effect in that teachers may be left with a *loss of professional enthusiasm*, even when the inspection is over (Ferguson, Earley, Ouston, & Fidler, 1999; Webb, Vulliamy, Häkkinen, & Hämälainen, 1998).

Explanatory variables

In this section, we discuss the currently available evidence regarding the explanatory features of the inspections' effects and side effects. We focus on the impact of the inspection judgement in terms of these effects, the schools' policy-making capacities, and the inspection quality. Because of the difficulty to clearly align the latter two explanatory variables, we will make use of underlying variables to conceptualize the schools' policy-making capacities and the inspection quality.

Perceived inspection judgement

On the basis of earlier research in other educational contexts (Emmelot, Karsten, Ledoux, & Vermeulen, 2004; McCrone et al., 2007; Ouston & Davies, 1998), strong effects are expected in schools which receive an unfavourable inspection judgement. Ehren (2010) found the strongest reaction to exist in schools in which many weak points were noted, although only when the inspectors had discussed with the school how those weak points could be addressed. However, the research evidence is inconsistent: Other studies found no significant correlation between the inspection judgement and the number of measures that were taken as a response to the inspection (Gärtner, Hüsemann, & Pant, 2009; Wurster & Gärtner, 2011).

The inspection judgement is given to the schools at the end of the inspection week. Consequently, the expected impact of the inspection judgement is limited to the conceptual effect, the instrumental effect, and the professional enthusiasm of teachers 2 weeks after the inspection, while no relationship is expected with regard to the emotional impact of inspection during the inspection, or with regard to the disturbing effect (before and during the inspection).

Perceived schools' policy-making capacities

Several school characteristics are suggested to explain part of the (side) effects of inspection (Chapman, 2001b; Diamant, 2007; Ehren & Visscher, 2006; Gray, 2000; Gray & Wilcox, 1995; Grubb, 1999). However, it often concerns assumptions which as yet have not been substantiated by much empirical data (Chapman & Earley, 2010; De Wolf & Janssens, 2007). The results of studies that provided indications about the relationship between school features and the (side) effects of

inspection mostly point to the schools' policy-making capacities. Based on Vanhoof, Van Petegem, Verhoeven, and Buvens (2009), we define the "school's policy-making capacities" as "the extent to which schools use the available autonomy for policy-making to come to a continuous process of retaining or changing their work in order to improve their educational quality and attain both the external and self-imposed objectives". Earlier research has shown that the concept of "school's policy-making capacities" is the result of a complex interaction between different factors, and is not easy to measure. The current study focuses on three aspects of the policy-making capacities which have been proven to have a large impact on this concept: innovative capacity, shared leadership, and the quality of the professional relationships within the school (Vanhoof, Deneire, & Van Petegem, 2011). The study will test whether each of these aspects indeed makes a unique contribution to the (perception) of the schools' policy-making capacities. The *innovative capacity* relates to the extent to which a school is open to innovation, copes with changes, and is able to implement innovations successfully (Geijsel, 2001). The effects of the inspection are assumed to depend on the extent to which the school incorporates a professional learning culture and an open attitude towards innovation (Chapman, 2001a; Ehren & Visscher, 2006; Lowe, 1998; Stoll, Bolam, McMahon, Wallace, & Thomas, 2006). Second, the (side) effects of the inspection are expected to depend on the participation of staff members in the development of the school policy (Dimmer & Metiuk, 1998; Standaert, 2001), the extent to which the management style of the principal matches the ideas of staff members (Lowe, 1998), and on the extent to which clear expectations are set and communicated with staff members (Ehren, Altrichter, McNamara, & O'Hara, 2013). Each of these features fits into the concept of "*shared leadership*" defined as the extent to which the school staff actively participates in decision-making processes (Harris, 2000, 2004). Finally, the *professional relationships* are also assumed to explain the effects and side effects of inspection (Kelchtermans & Vandenberghe, 1998; Scanlon, 1999). This concept addresses the extent to which professional relations between school staff are experienced as supportive.

Perceived inspection quality

Next to the inspection judgement and schools' policy-making capacities, "inspection quality" is one of the major factors that explain effects and side effects (Ehren & Visscher, 2006; Matthews & Sammons, 2004). In many studies, it is not clear what is exactly covered by the concept "inspection quality". The perception of inspection quality depends on many factors, and each person involved may have a different view on what is meant by "inspection quality". When discussing the inspection quality, scholars generally refer to either one or more of the three factors, namely, the quality of the inspectors' behaviour, the inspection's psychometric quality, and the transparency of the inspection. In this study, we will test the assumption that the existence of a concept (called "perceived inspection quality") relies on these three variables. The *(perceived) inspector's behaviour* (as friendly, trustworthy, supportive, and respectful towards the school) is suggested to be paramount for the effects on school development (Chapman, 2001b; Learmonth, 2000; MacBeath, 2006; Sandbrook, 1996; Weiner, 2002; Winch, 2001), yet only a few studies have provided evidence of this relationship (Lowe, 1998; Office for Standards in Education, Children's Services and Skills [Ofsted], 2007). Kelchtermans and Vandenberghe (1998) reported an effect in terms of the inspectors' attitude on the willingness to use the inspection findings for development, as well as on the experienced side effects. Furthermore, there is a great deal of literature about the importance of the *perceived psychometric characteristics* (reliability and validity) of the inspection process and about its impact on the effects and side effects of the inspection. This perception depends, amongst others, on the assessment of the inspection methods used, the extent to which the inspectors have grasped the school reality and taken the school context into account, and the expertise and credibility of the inspectors (Braun, Ball, Maguire, & Hoskins, 2011; Chapman, 2001a; Cuckle & Broadhead, 1999; Kelchtermans & Vandenberghe, 1998; McCrone et al., 2007; Sandbrook, 1996; Webb et al., 1998). Both policy-makers and scholars have advocated the need for a strengthened reliability and validity in terms of the inspection for accountability purposes, but also for its impact on school development (e.g., Gärtner & Pant, 2011;

MacBeath, 2006; Van Bruggen, 2010). Although some studies have provided indications of a relationship between the perception of psychometric strength and the (side) effects of inspections (Brimblecombe et al., 1996; Cuckle & Broadhead, 1999; Lowe, 1998; McCrone et al., 2007; Ouston & Davies, 1998; Scanlon, 1999; Wilcox & Gray, 1996), this link has never been the main focus of empirical research. Finally, the *transparency of the inspection* reflects issues raised about the importance of the transparency of the inspection process (Leeuw, 2002; Van Bruggen, 2010), as well as of the inspection judgement on the school (Brauckmann & Pashiardis, 2010; Gärtner & Pant, 2011; Matthews & Smith, 1995). There are indications that efforts to increase transparency would increase schools' satisfaction with the inspection process as well as their willingness to accept the inspectors' findings (Kelchtermans & Vandenberghe, 1998; McCrone, Coghlan, Wade, & Rudd, 2009; Sandbrook, 1996), while the lack of transparency would lead to increased stress and anxiety (Brimblecombe, Ormston, & Shaw 1995; Wilcox & Gray, 1996).

Research methods

This article reports on an online survey of the perception of teachers regarding the above-mentioned aspects. Using path analysis, we tested the existence and strength of the relationships presented in the conceptual framework. In order to rule out sample mistakes, we sent out the survey to every teacher in each primary and secondary school that was inspected during a predefined period in the school year 2012–2013 (Creswell, 2005). Respondents received the questionnaire 8 weeks after the inspection. Only the responses from teachers in schools with a response rate higher than 25% (or in which the number of participating teachers exceeds 20) were retained for further analysis. In total, this study draws on the responses of 2,202 teachers from 137 schools (57.3% of the total number of schools inspected in the predefined period), of which 101 schools (77.4%) and 1,313 respondents (59.6%) were from primary education and the remaining schools and respondents were from secondary schools. Of those taking part in the survey, 80.3% were women and 19.7% were men. These figures indicate a good representation with regard to the target population: 71.3% of the Flemish schools are primary schools, and 28.7% of the schools in Flanders provide secondary education. Of all staff members in a teaching or management position, 73.6% are women and 26.4% are men (Vlaamse Overheid, 2014) – separate figures for only teachers are not available, but overall male staff members are overrepresented in the management functions.

The various concepts in the conceptual framework were operationalized using different items with Likert scales (from 1 = *entirely disagree* to 5 = *entirely agree*, with 3 as a neutral score). An exception was made for the measurement of emotional effects during the inspection, and professional enthusiasm after the inspection (see below). Another exception was made for the inspection judgement, which was a dichotomous variable (1 = *positive*, 0 = *restricted positive*. No schools with a "negative" judgement agreed to participate in the study). In 33 schools (24.1%) in our sample, relatively separate entities within the schools received a different inspection judgement (e.g., in primary schools a different judgement for the nursery versus the primary department; or for secondary schools a different judgement for different education areas). Therefore, we relied on the participants' feedback about the inspection judgement. Hence the inspection judgement is to be considered a personal rather than a school characteristic. Of the participants, 56.4% reported the judgement as being "positive", while 39.6% perceived the judgement as being "restricted positive", and 1.5% of the responses were missing on this item. The instruments used to measure the three perceived school features were derived from existing questionnaires on the policy-making capacities of schools (Vanhoof et al., 2011). On the basis of pilot research, six scales were developed to survey the effects of inspection and the inspection features. Table 1 provides an overview of the psychometric characteristics of each scale. Principal axis factoring (with Oblimin rotation) revealed that each of those scales were unique factors. Cronbach's alpha's indicated that the internal consistency is "moderate" (above 0.80) or "high" (above 0.90), which implies that the scales can be trusted for use in the analyses we intended to carry out (Murphy & Davidshofer, 1988).

Table 1. Psychometric characteristics of the scales.

Scales	Number of items	Cronbach's alpha	Example item
School's policy-making capacities			
Shared leadership	6	.92	In our school the expertise of individual staff members is used when decisions need to be made.
Innovative capacity	6	.85	Our school has a positive attitude towards innovations.
Professional relationships	5	.90	Staff members in our school trust each other.
Inspection quality			
Psychometric characteristics	5	.91	The inspectors disposed of sufficient expertise to make a fair judgement about the quality of our school.
Transparency	6	.86	It was clear for me which criteria were used to make a judgement about the school.
Inspector behaviour	5	.85	The inspectors were polite towards me.
Effects			
Conceptual effect	7	.89	The inspection made me reflect about our school policy.
Instrumental effect	5	.80	The inspection has led to concrete actions for improvement.
Disturbing effect	5	.90	Because of the inspection we paid less attention to the pupils.

As we had to rule out a general feeling of stress, anxiety, and professional enthusiasm, an alternative calculation was developed for emotional effects during and after the inspection. On different Likert scales, both an estimate of the stress/anxiety during the inspection (t1) and separately an estimate during a regular period without inspection (t0) were examined. The score for the effect of the inspection on stress/anxiety during the inspection was obtained by the subtraction *t1–t0*. Similarly, the effect on enthusiasm 2 weeks after the inspection (t2) was compared with the estimate at t0 (*t2–t0*).

Using structural equation modeling (SEM), we looked at whether or not the hypothesized relationships fit the empirical data (Muthén & Muthén, 2003). We opted for a path analysis because it allows the reduction of measurement error by having multiple indicators for latent variables, and due to its ability to test models overall rather than individual coefficients. The analyses were conducted using the software package Mplus. We accounted for the eventual impact of the hierarchic structure in our data (model = complex, with "school" as the cluster variable; maximum likelihood estimation with robust standard errors). The level of analysis was the individual teacher. As we have measured "individual perceptions of (side) effects and of explanatory characteristics", we refrained from aggregating different scores at school level. When testing the model, we used the comparative fit index (CFI) and the root mean square error of approximation (RMSEA) as fit indices. CFI compares the assumed model with a model without assumed relationships. It has an upper bound of 1. Any value greater than .95 is considered to be an adequate level of model fit (Hu & Bentler, 1999). RMSEA provides an indication of the model fit with the real situation in the population, should that be known. Values lower than .05 are considered a close fit, .05 to .08 a "fair fit" (Byrne, 1998). The model was refined based on modification indices.

Results

In this paragraph, we first provide descriptive results regarding each of the effects and of the hypothesized explaining variables. The second section focuses on the explanatory results.

Descriptive results

Descriptive results are summarized in Table 2 and Table 3. Teachers in our sample report, on average, moderate conceptual and instrumental effects in terms of being inspected (M = 3.34 and 3.41, respectively). Most teachers do not agree that the inspection has disturbing effects (M = 2.30), but the high standard deviation indicates large differences between respondents. The

Table 2. Descriptive results (general).

Scale	Valid *N*	*M*	*SD*
Effects			
Conceptual effect	2,113	3.34	0.79
Instrumental effect	1,970	3.41	0.71
Disturbing effect	1,936	2.30	1.02
Inspection quality			
Psychometric characteristics	2,042	3.81	0.85
Transparency	1,888	3.79	0.78
Inspector behaviour	1,724	4.28	0.74
School Policy-making capacities			
Shared leadership	2,159	3.71	0.66
Innovative capacity	2,177	3.81	0.84
Professional relationships	2,084	3.57	0.82

Note: Answer categories: 1 = entirely disagree; 2 = disagree; 3 = neither agree nor disagree; 4 = agree; 5 = entirely agree.

Table 3. Descriptive results regarding emotional effects during and after inspection.

	t1/2–t0[1]			t0		t1/t2				
Scale	Valid *N*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*t*	*p*	Cohen's *d*
Emotional effects										
Increased stress during inspection	2,119	1.22	1.16	2.20	0.93	3.42[2]	1.16	−48.62	***	−1.17
Increased anxiety during inspection	2,127	0.92	1.10	1.20	0.54	2.12[2]	1.21	−38.25	***	−1.05
Increased enthusiasm after inspection	2,128	−0.40	0.98	4.13	0.67	3.73[3]	1.06	19.05	***	0.47

[1]Scores could vary from –4 (maximum decrease of stress/anxiety/enthusiasm during/after the inspection compared to regular levels) to +4 (maximum increase), with 0 = neutral effect; [2]t1, during the inspection; [3]t2, 2 weeks after the inspection.

inspection leads to strong emotional side effects. The stress and anxiety scores during inspection differ significantly and with a high impact from the scores during regular periods without inspection size ($|d|$ = 1.17 and 1.07, respectively). Two weeks after the inspection, teachers' professional enthusiasm is still significantly lower than during a regular period without inspection ($|d|$ = 0.47).

The inspection is generally considered as valid and reliable (mean score for psychometric characteristics M = 3.81) and as transparent (M = 3.79). Teachers are generally very satisfied with the inspectors' behaviour regarding the school and its staff members (M = 4.28). Our data further show that teachers have, on average, a positive attitude towards the policy-making capacities of their own school. This applies to the shared leadership (M = 3.71), the innovative capacity (M = 3.81), and, to a slightly lesser extent, professional relationships (M = 3.57) within the school.

Explanatory results

The fit indices for the measurement model and the structure model indicate a good fit (RMSEA: 0.044; CFI: 0.961). All the assumed relationships in the structure model are statistically significant. As described in the conceptual framework, three latent variables were identified. The variances in the schools' shared leadership, innovative capacity, and professional relationships are explained by the latent variable "school policy-making capacities" (parameters .82, .76, and .77, respectively). The "perceived inspection quality" explains the variance in the perception of psychometric characteristics (.89), inspection transparency (.68), and inspector's behaviour (.60). Variance in the experience of stress and anxiety during the inspection is explained by the general "emotional effect during inspection" (.69 and .73, respectively). The resulting model – presented in Figure 2 – corresponds

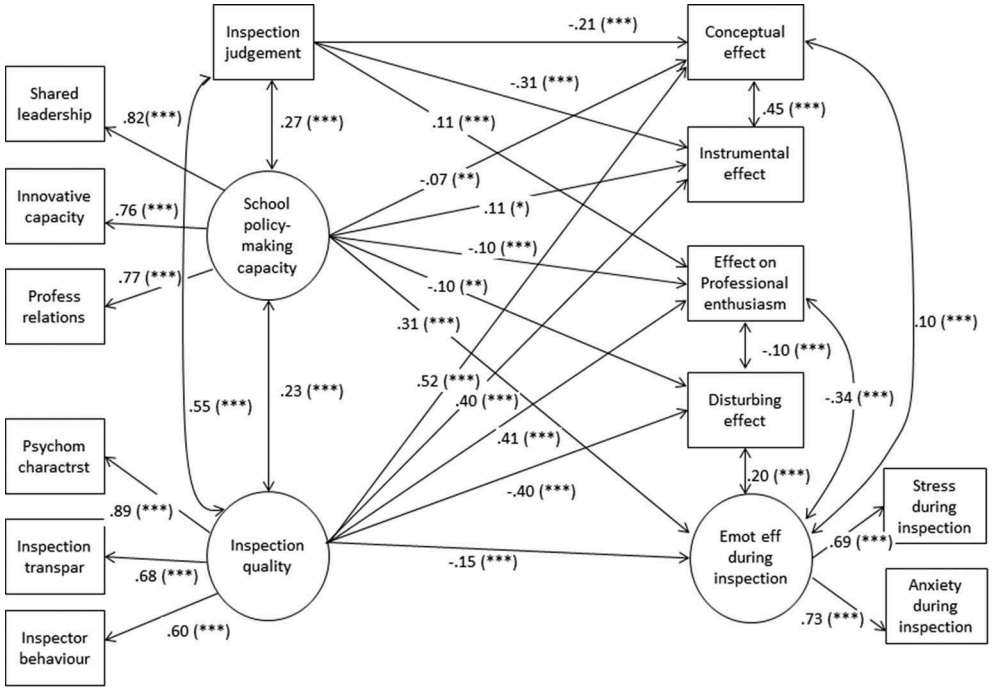

Figure 2. Path model with standardized parameter estimates and levels of significance.
Note: *significant at $p < .05$ level. **significant at $p < .01$ level. ***significant at $p < .001$ level.

with the conceptual framework. On the basis of modification indices, five correlations between (side) effects and three correlations between the explaining variables were added to the measurement model. An overview of estimates, standard errors, and levels of significance are included in Table 4.

The perceived inspection quality is the strongest predictor for most of the effects. This perception strongly affects the variance within the conceptual effect (parameter .52) and has a moderate covariance with the instrumental effect (.40) and with the effect on professional enthusiasm (.41), as well as (negatively) with the disturbing effect (–.40) after control for other variables. The inspections' conceptual effect also shows a small covariance with the inspection judgement (–.21). The inspection judgement further has a medium covariance with the instrumental effect (–.32) and a small correlation with teachers' professional enthusiasm after the inspection (.11). The conceptual effect, the instrumental effect, the effect on professional enthusiasm, and the disturbing effect, each show a small covariance with the schools' policy-making capacities (parameters –.07, .11, –.10, and –.10, respectively).

The variance in the emotional effect during the inspection shows a different picture as, in contrast with the aforementioned (side) effects, it is best predicted by the schools' policy-making capacities. The medium positive covariance between the schools' policy-making capacities and the emotional impact in terms of the experience of stress and anxiety during the inspection (.31) indicates that there is a stronger increase in stress and anxiety in schools with strong policy-making capacities. Furthermore, we found that emotional effects are allayed when the inspection quality is perceived as being high (–.15).

Figure 2 also presents the statistically significant correlations between different independent variables. A strong correlation (.55) is observed between the inspection judgement and the perceived inspection quality, alongside a small correlation (.23) between inspection quality and school's policy-making capacities (.23), and between the inspection judgement and school's policy-making capacities (.27), after controlling for other variables. Medium correlations are found

Table 4. Estimates, standard errors, and levels of significance.

	Estimate	*SE*	*p*
School policy-making capacity BY			
Shared Leadership	.76	.015	.000
Innovative Capacity	.82	.017	.000
Professional relationships	.77	.020	.000
Inspection Quality BY			
Psychometric characteristics	.89	.017	.000
Inspection transparency	.68	.026	.000
Inspector behaviour	.60	.028	.000
Emotional effect during inspection BY			
Stress during inspection	.69	.032	.000
Anxiety during inspection	.73	.033	.000
Emotional effect during inspection ON			
School policy-making capacity	.31	.039	.000
Inspection quality	–.15	.049	.001
Disturbing effect ON			
School policy-making capacity	–.10	.035	.004
Inspection quality	–.40	.038	.000
Effect on professional enthusiasm ON			
Inspection judgement	.11	.029	.000
School policy-making capacity	–.10	.026	.000
Inspection quality	.41	.034	.000
Instrumental effect ON			
Inspection judgement	–.31	.033	.000
School policy-making capacity	.11	.042	.000
Inspection quality	.40	.040	.000
Conceptual effect ON			
Inspection judgement	–.21	.030	.000
School policy-making capacity	–.07	.026	.010
Inspection quality	.52	.032	.000
School policy-making capacity WITH			
Inspection judgement	.27	.051	.000
Inspection quality	.23	.055	.000
Inspection quality WITH			
Inspection judgement	.55	.032	.000
Emotional effect during inspection WITH			
Conceptual effect	.10	.027	.000
Effect on professional enthusiasm	–.34	.034	.000
Disturbing effect	.20	.035	.000
Instrumental effect WITH			
Conceptual effect	.45	.024	.000
Disturbing effect WITH			
Effect on professional enthusiasm	–.10	.030	.001

between conceptual and instrumental effects (.45), and between the emotional effect and the effect on professional enthusiasm after the inspection (–.34). The latter negative correlation indicates that a stronger increase in stress and anxiety during the inspection correlates with a lower sense of professional enthusiasm 2 weeks after the inspection. The emotional effect during the inspection also shows a small correlation with the disturbing effect (.20) and with the conceptual effect (.10), whereas a stronger disturbing effect goes together with a lower professional enthusiasm after the inspection (–.10) (distinctions between "large", "medium", and "small" correlations as defined by Cohen, 1988).

Conclusion and discussion

This study set out to provide explanations for the differential (side) effects of inspections in schools. It is commonly accepted that these effects are determined by the inspection judgement, by

inspection features, and by school features (Ehren & Visscher, 2006; Matthews & Sammons, 2004), but there is currently no evidence available on the extent to which these explaining features matter (Chapman & Earley, 2010; De Wolf & Janssens, 2007).

This study clearly shows the crucial role of the perceived inspection quality in explaining the conceptual and instrumental effects of the inspection. Hence, if school development is one of the aims of inspection, it is vitally important that sufficient care is taken of the teachers' perception of its quality or, more specifically, of the psychometric quality (reliability and validity), of the transparency of the inspection process, and of the criteria used for determining the inspection judgement, and of a friendly, trustworthy, supportive, and respectful approach towards staff members in the inspected schools. Also the inspection judgement and the school characteristics explain part of the variance in the conceptual and instrumental effect, albeit to a lesser extent.

Our data reveal a mixed picture regarding the explaining features of the inspections' side effects. The perceived inspection quality explains a substantial part of the variance with regard to the disturbing effects and the variance within the effect on professional enthusiasm after the inspection. Contrastingly, the schools' policy-making capacities are the most important predictor with regard to the emotional side effect during the inspection: the stronger these policy-making capacities, the more severe the emotional impact in terms of increased stress and anxiety (see further in this paragraph for discussion). Earlier research provided some indications of a link between features of the schools' policy-making capacities and the emotional response of school staff (Brimblecombe et al., 1995; Sandbrook, 1996; Scanlon, 1999), but this study is a valuable contribution to the evidence base regarding the direction and the strength of this relationship.

The model resulting from our data includes some notable relationships. For example, after controlling for the other variables in the model, teachers in schools with stronger policy-making capacities tend to have a greater increase in stress and anxiety during the inspection. They also report a more negative impact on their professional enthusiasm 2 weeks after the inspection. We hypothesize that teachers commit more strongly to the inspection process in schools that dispose of a strong self-reflective attitude in which teachers are commonly aware of their weaknesses, strengths, and opportunities for further development. This stronger commitment leads to a more elaborate preparation and the experience of higher stakes (while lower stakes are seen as a cause of limited emotional responses; Gärtner et al., 2009; Yeung, 2012). Another possible explanation for this finding is that teachers in those schools identify particularly strongly with their school, endorsing the idea that inspection (even though it only concerns school-based assessment) carries some kind of personal evaluation with it. It goes without saying that these assumptions need further validation, and that each of these unexpected relationships deserve further clarification through future studies.

Furthermore, being inspected leads to less conceptual but more instrumental effects, as perceived by teachers who regard their school having strong policy-making capacities. It may be that these schools have a stronger reflective culture (in times without inspection), which could explain why it is more difficult for the inspection to contribute to school-based reflection. Contrastingly, these schools are more successful in making use of the inspection in terms of actions for school development. In that respect, despite the smaller conceptual effect, the claim from several scholars that schools with strong policy-making capacities benefit more from inspection than do weaker schools (because weaker schools need more than just an evaluation to engage in processes of school development) (Ehren, 2010; Harris, Chapman, Muijs, Russ, & Stoll, 2006) is substantiated by our empirical data. For future research on the effects of inspection, we advocate to distinguish between conceptual and instrumental effects.

Some relationships within the measurement model had to be added. For instance, the path model shows that the perception of the inspection quality has a large correlation with the inspection judgement. From an international perspective, it is not the first time a correlation between the perceived inspection quality and the inspection judgement has been observed (Gärtner et al., 2009; McCrone et al., 2007). This strong relationship might be caused by a different

approach, or by a different stance taken by inspectors during the inspection week, when they feel the inspection judgement at the end of the inspection week will be favourable for the school. This anticipatory judgement may effectively lead to a more friendly approach towards the school's staff, and to more openness about the processes and criteria for inspection. Alternatively, it may just be easier for staff to be satisfied with the inspection quality when the result for their school was positive. An in-depth study into the relationship between the inspection judgement and (the perception of) the inspection quality is warranted in terms of probing its underlying features. Another added relationship relates to perceived inspection quality and perceived policy-making capacity. This relationship is harder to explain. It might be that staff members in school with strong policy-making capacities already initially hold a more open and welcoming stance towards the inspectors. Finally, the relationship between the inspection judgement and the policy-making capacities confirms that schools with stronger policy-making capacities generally receive more favourable inspection judgements.

This study has looked into the conceptual and instrumental effects as the intended effects of inspection. However, it cannot be assumed right away that these conceptual and instrumental effects equal "school improvement". Changes in schools' ideas about themselves and even changes to current practices and policies do not necessarily equal "improvement" (Biesta, 2009; Brimblecombe et al., 1996). It has been argued that everyone may understand something different with regard to the term "school improvement" (Hoy & Miskel, 2008) and that improvement from the inspection perspective does not always match with schools' ideas about improvement (Gilroy & Wilcox, 1997; Perryman, 2007). Furthermore, it remains unclear as to what extent initiatives taken by the school as a response to an inspection judgement are actually supported by the school team. When actions are the result of something imposed by a "threatening" power (in this case the inspectors), the sustainability of these actions may be rather limited (Van den Berg & Vandenberghe, 1999). Moreover, we have been able to enquire the extent to which schools engage in conceptual and instrumental effects, but in agreement with McCrone et al. (2007) and Van Bruggen (2010), we think it remains methodologically challenging to investigate whether or not these effects will lead to "better education" in terms of improved outcomes (e.g., learning results or pupil well-being). Responses may not always be effective, or can even be negative when they harm pupil learning (Booher-Jennings, 2005; Rosenthal, 2004; Shaw, Newton, Aitkin, & Darnell, 2003).

Yet another issue for further research is the extent to which these findings may be generalized. It was the purpose of this study to contribute to the knowledge base regarding the effects and side effects of inspection by adding insight into the explaining features of these (side) effects in one specific research context. One may therefore rightfully argue that our findings are particular to the Flemish context, and that they need validation in other educational contexts. Nevertheless, to our knowledge, this kind of research has currently not yet been conducted in other educational contexts. Therefore, the results of this study may definitely contribute to hypotheses for future research in other contexts, and serve as a comparative basis for the results of such research. Another limitation is that no schools that received a "negative" inspection judgement were willing to participate in our study. Specific research in this small segment of schools (3.3% of the schools received this judgement) may provide a useful addition to the results of this study.

Revisiting the original research question of this study, we suggest that the intended effects of inspection are predominantly explained by the perception of the inspection quality. The same holds for the impact on the professional enthusiasm of teachers 2 weeks after the inspection, and the inspection's disturbing effect on normal school life. The school's policy-making capacities are the strongest predictor of emotional side effects during the inspection, with teachers in schools with stronger policy-making capacities reporting a greater increase in stress and anxiety.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Maarten Penninckx is a PhD student at the University of Antwerp and a researcher at the Edubron Research Unit (www.edubron.be). His current work focuses on the effects and side effects of mechanisms for school improvement and school accountability.

Jan Vanhoof is as associated professor on the staff of the Institute of Education and Information Science of the University of Antwerp. He is a member of the Edubron Research Unit. His current research activities focus on school policy and quality care in general and on school self-evaluation and data-driven school policy in particular.

Sven De Maeyer is assistant professor at the Institute for Education and Information Science. He is a member of the Edubron Research Unit. His research focuses on educational measurement and methodological issues in educational sciences.

Peter Van Petegem is full professor at the Institute of Education and Information Science of the University of Antwerp. He leads the research unit Edubron and the Center of Expertise in Higher Education.

References

Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*, *21*, 33–46.

Booher-Jennings, J. (2005). Below the bubble: "Educational Triage" and the Texas accountability system. *American Educational Research Journal*, *42*, 231–268.

Brauckmann, S., & Pashiardis, P. (2010). The clash of evaluations: In search of the missing link between school accountability and school improvement – Experiences from Cyprus. *International Journal of Educational Management*, *24*, 330–350.

Braun, A., Ball, S. J., Maguire, M., & Hoskins, K. (2011). Taking context seriously: Towards explaining policy enactments in the secondary school. *Discourse: Studies in the Cultural Politics of Education*, *32*, 585–596.

Brimblecombe, N., Ormston, M., & Shaw, M. (1995). Teachers' perceptions of school inspection: A stressful experience. *Cambridge Journal of Education*, *25*, 53–61.

Brimblecombe, N., Shaw, M., & Ormston, M. (1996). Teachers' intention to change practice as a result of OFSTED school inspections. *Educational Management, Administration and Leadership*, *24*, 339–354.

Brunsden, V., Davies, M., & Shevlin, M. (2006). Anxiety and stress in educational professionals in relation to Ofsted. *Education Today*, *56*, 24–31.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence.

Chapman, C. (2001a). Changing classrooms through inspection. *School Leadership & Management*, *21*, 59–73.

Chapman, C. (2001b). Unlocking the potential: Inspection as a mechanism for school improvement. *Improving Schools*, *4*(3), 41–50.

Chapman, C. (2002). Ofsted and school improvement: Teachers' perceptions of the inspection process in schools facing challenging circumstances. *School Leadership & Management*, *22*, 257–272.

Chapman, C., & Earley, P. (2010). School inspection/external school evaluation. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed., pp. 719–725). Oxford: Elsevier.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Creswell, J. W. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Cuckle, P., & Broadhead, P. (1999). Effects of Ofsted inspection on school development and staff morale. In C. Cullingford (Ed.), *An inspector calls: Ofsted and its effects on school standards* (pp. 176–187). London: Kogan Page.

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*, 379–396.

Diamant, I. (2007). *An investigation into the extent to which schools use external evaluation information and its implications for their capacity to improve* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. U232959.)

Dimmer, T., & Metiuk, J. (1998). The use and impact of OFSTED in a primary school. In P. Earley (Ed.), *School Improvement after Inspection? School and LEA Responses* (pp. 51–61). London: Paul Chapman.

Ehren, M. C. M. (2010). Effecten van toezicht op het basisonderwijs [Effects of inspection on primary education]. *Pedagogische studieën*, *87*, 165–182.

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

Ehren, M. C. M., Perryman, J., & Shackleton, N. (2015). Setting expectations for good education: How Dutch school inspections drive improvement. *School Effectiveness and School Improvement*, *26*, 296–327. doi:10.1080/09243453.2014.936472

Ehren, M. C. M., & Swanborn, M. S. L. (2012). Strategic data use of schools in accountability systems. *School Effectiveness and School Improvement*, *23*, 257–280.
Ehren, M. C. M., & Visscher, A. J. (2006). Towards a theory on the impact of school inspections. *British Journal of Educational Studies*, *54*, 51–72.
Emmelot, Y., Karsten, S., Ledoux, G., & Vermeulen, A. (2004). *Ervaringen met het vernieuwde onderwijstoezicht* [Experiences with the new school inspection system]. Amsterdam: SCO/Kohnstamminstituut.
Faubert, V. (2009). *School evaluation: Current practices in OECD countries and a literature review* (OECD Education Working Paper No. 42). Paris: OECD.
Ferguson, N., Earley, P., Ouston, J., & Fidler, P. (1999). New heads, OFSTED inspections and the prospects for school improvement. *Educational Research*, *41*, 241–249.
Gärtner, H., Hüsemann, D., & Pant, H. A. (2009). Wirkungen von Schulinspektion aus Sicht betroffener Schulleitungen. Die Brandenburger Schulleiterbefragung [The effects of school inspection from the viewpoint of school principals affected]. *Empirische Pädagogik*, *23*(1), 1–18.
Gärtner, H., & Pant, H. A. (2011). How valid are school inspections? Problems and strategies for validating processes and results. *Studies in Educational Evaluation*, *37*, 85–93.
Gärtner, H., Wurster, S., & Pant, H. A. (2014). The effect of school inspections on school improvement. *School Effectiveness and School Improvement*, *25*, 489–508. doi:10.1080/09243453.2013.811089
Geijsel, F. (2001). *Schools and innovations: Conditions fostering the implementation of educational innovation*. Nijmegen: University Press.
Gilroy, P., & Wilcox, B. (1997). OFSTED, criteria and the nature of social understanding: A Wittgensteinian critique of the practice of educational judgement. *British Journal of Educational Studies*, *45*, 22–38.
Gray, J. (2000). *Causing concern but improving: A review of schools' experiences on special measures* (Research Report No. 188). London: Department for Education and Employment.
Gray, J., & Wilcox, B. (1995). Inspection and school improvement: Rhetoric and experience from the bridge. In J. Gray & B. Wilcox (Eds.), *Good school, bad school: Evaluating performance and encouraging improvement* (pp. 133–148). Buckingham: Open University Press.
Grubb, W. N. (1999). Improvement or control? A US view of English inspection. In C. Cullingford (Ed.), *An inspector calls: Ofsted and its effect on school standards* (pp. 70–96). London: Kogan Page.
Harris, A. (2000). *School improvement. What's in it for schools?* London: Routledge Falmer.
Harris, A. (2004). Distributed leadership and school improvement: Leading or misleading? *Educational Management Administration & Leadership*, *32*, 11–24.
Harris, A., Chapman, C., Muijs, D., Russ, J., & Stoll, L. (2006). Improving schools in challenging contexts: Exploring the possible. *School Effectiveness and School Improvement*, *17*, 409–424.
Hoy, W. K., & Miskel, C. G. (2008). *Educational administration: Theory, research, and practice* (8th ed.). New York, NY: McGraw-Hill.
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.
Kelchtermans, G., & Vandenberghe, R. (1998, April). *Internal use of external control and support for quality improvement: The response to a national policy by primary schools*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.
Kogan, M., & Maden, M. (1999). An evaluation of evaluators: The Ofsted system of school inspection. In C. Cullingford (Ed.), *An inspector calls: Ofsted and its effect on school standards* (pp. 9–32). London: Kogan Page.
Learmonth, J. (2000). *Inspection: What's in it for schools?* London: Routledge Falmer.
Leeuw, F. L. (2002). Reciprocity and educational evaluations by European inspectorates: Assumptions and reality checks. *Quality in Higher Education*, *8*, 137–150.
Lonsdale, P., & Parsons, C. (1998). Inspection and the school improvement hoax. In P. Earley (Ed.), *School improvement after inspection? School and LEA responses* (pp. 110–125). London: Paul Chapman.
Lowe, G. (1998). Inspection and change in the classroom: Rhetoric and reality? In P. Earley (Ed.), *School improvement after inspection? School and LEA responses* (pp. 97–109). London: Paul Chapman.
MacBeath, J. (1999). *Schools must speak for themselves*. London: Routledge.
MacBeath, J. (2006). *School inspection and self-evaluation: Working with the new Relationship*. New York, NY: Routledge.
Matthews, P., & Sammons, P. (2004). *Improvement through inspection: An evaluation of the impact of Ofsted's work*. London: Ofsted.
Matthews, P., & Smith, G. (1995). OFSTED: Inspecting schools and improvement through inspection. *Cambridge Journal of Education*, *25*, 23–34.
McCrone, T., Coghlan, M., Wade, P., & Rudd, P. (2009). *Evaluation of the impact of Section 5 inspections – Strand 3: Final report for Ofsted*. Slough: National Foundation for Educational Research.
McCrone, T., Rudd, P., Blenkinsop, S., Wade, P., Rudd, S., & Yeshanew, T. (2007). *Evaluation of the impact of Section 5 inspections*. Slough: National Foundation for Educational Research.
Michielsens, P. (2007). Onderwijstoezicht in evolutie [Schools inspections in evolution]. *Tijdschrift voor Onderwijsrecht en Onderwijsbeleid*, *2007*(4), 283–290.

Muijs, D., Harris, A., Chapman, C., Stoll, L., & Russ, J. (2004). Improving schools in socioeconomically disadvantaged areas – A review of research evidence. *School Effectiveness and School Improvement*, *15*, 149–175.
Murphy, K. R., & Davidshofer, C. O. (1988). *Psychological testing: Principles and applications*. Englewood Cliffs, NJ: Prentice Hall.
Muthén, L. K., & Muthén, B. O. (2003). *Mplus: Statistical analysis with latent variables: User's guide*. Los Angeles, CA: Authors.
Nicolaidou, M., & Ainscow, M. (2005). Understanding failing schools: Perspectives from the inside. *School Effectiveness and School Improvement*, *16*, 229–248.
Office for Standards in Education, Children's Services and Skills. (2007). *Review of the impact of inspection*. London: Author.
Onderwijsinspectie. (2013). *Onderwijsspiegel 2013* [Education Mirror]. Brussel: Onderwijsinspectie/Vlaams Ministerie van Onderwijs en Vorming.
Organisation for Economic Co-peration and Development. (2013). *Synergies for better learning: An international perspective on evaluation and assessment* (OECD Reviews of Evaluation and Assessment in Education). Paris: Author.
Ouston, J., & Davies, J. (1998). OFSTED and afterwards? Schools' responses to inspection. In P. Earley (Ed.), *School improvement after inspection? School and LEA responses* (pp. 13–24). London: Paul Chapman.
Ouston, J., Fidler, B., & Earley, P. (1997). What do schools do after OFSTED school inspections – Or before? *School Leadership and Management*, *17*, 95–104.
Perryman, J. (2007). Inspection and emotion. *Cambridge Journal of Education*, *37*, 173–190.
Rosenthal, L. (2004). Do school inspections improve school quality? Ofsted inspections and school examination results in the UK. *Economics of Education Review*, *23*, 143–151.
Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (1999). *Evaluation: A systematic approach*. Thousand Oaks, CA: Sage.
Sandbrook, I. (1996). *Making sense of primary inspection*. Buckingham: Open University Press.
Scanlon, M. (1999). *The impact of Ofsted inspections*. Slough: National Foundation for Educational Research.
Shaw, I., Newton, D. P., Aitkin, M., & Darnell, R. (2003). Do OFSTED inspections of secondary schools make a difference to GCSE results? *British Educational Research Journal*, *29*, 63–75.
Standaert, R. (2001). *Inspectorates of education in Europe: A critical analysis*. Leuven: Acco.
Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change*, *7*, 221–258.
Stoll, L., & Fink, D. (1998). The cruising school: The unidentified ineffective school. In L. Stoll & K. Myers (Eds.), *No quick fixes: Perspectives on schools in difficulty* (pp. 189–206). Bristol: Falmer Press.
Sun, H., Creemers, B., & De Jong, R. (2007). Contextual factors and effective school improvement. *School Effectiveness and School Improvement*, *18*, 93–122.
Van Bruggen, J. C. (2010). *Inspectorates of Education in Europe: Some comparative remarks about their tasks and work*. Brussels: SICI.
Van den Berg, R., & Vandenberghe, R. (1999). *Succesvol leiding geven aan onderwijsinnovaties* [Guiding educational innovations succesfully]. Alphen aan den Rijn: Samson.
Vandenberghe, R. (2010). Vernieuwingsbeleid in het Vlaamse onderwijs: Een kritische analyse [Innovation policy in Flemish education: A critical analysis]. *Tijdschrift voor Onderwijsrecht en Onderwijsbeleid*, *2010*, 458–476.
Vanhoof, J., Deneire, A., & Van Petegem, P. (2011). *Waar zit het beleidsvoerend vermogen in (ver)scholen?* [Where are the policy-making capacities concealed in schools?]. Antwerpen: Wolters Plantyn.
Vanhoof, J., & Van Petegem, P. (2007). Matching internal and external evaluation in an era of accountability and school development: Lessons from a Flemish perspective. *Studies in Educational Evaluation*, *33*, 101–119.
Vanhoof, J., Van Petegem, P., Verhoeven, J., & Buvens, I. (2009). Linking the policymaking capacities of schools and the quality of school self-evaluations: The view of school leaders. *Educational Management Administration & Leadership*, *37*, 667–686.
Vlaamse Overheid. (2014). *Vlaams onderwijs in cijfers 2013–2014* [Flemish education in figures 2012–2014]. Retrieved from http://www.ond.vlaanderen.be/onderwijsstatistieken
Webb, R., Vulliamy, G., Häkkinen, K., & Hämälainen, S. (1998). External inspection or school self-evaluation? A comparative analysis of policy and practice in primary schools in England and Finland. *British Educational Research Journal*, *24*, 539–556.
Weiner, G. (2002). Auditing failure: Moral competence and school effectiveness. *British Educational Research Journal*, *28*, 789–804.
Wilcox, B., & Gray, J. (1996). *Inspecting schools: Holding schools to account and helping schools to improve*. Buckingham: Open University Press.
Winch, C. (2001). Towards a non-punitive school inspection regime. *Journal of Philosophy of Education*, *35*, 683–694.
Wurster, S., & Gärtner, H. (2011, September). *Handling school inspection and their results: Differentiation of school types*. Paper presented at the European Conference on Educational Research (ECER), Berlin.
Yeung, S. Y. S. (2012). A school evaluation policy with a dual character: Evaluating the school evaluation policy in Hong Kong from the perspective of curriculum leaders. *Educational Management Administration & Leadership*, *40*, 37–68.