

This item is the archived peer-reviewed author-version of:

Effects and side effects of Flemish school inspection

Reference:

Penninckx Maarten, Vanhoof Jan, de Maeyer Sven, Van Petegem Peter.- Effects and side effects of Flemish school inspection

Educational management administration and leadership - ISSN 1741-1432 - (2015), p. 1-17

Full text (Publishers DOI): <http://dx.doi.org/doi:10.1177/1741143215570305>

Effects and side effects of Flemish school inspection

Penninckx, M.; Vanhoof, J. ; De Maeyer, S.; Van Petegem, P.

Abstract

Despite the increased importance of school inspection in recent years, the current knowledge base does not provide a clear view on the effects and side effects of being inspected. More evidence is needed in more diverse educational contexts. This article responds to this need with a quantitative study on the effects and side effects of school inspections on Flemish schools. In total 2,624 respondents from 130 primary and secondary schools participated in the study. The article discusses the conceptual, instrumental, symbolic and strategic effects of inspection and its impact on self-efficacy and collective efficacy. Regarding side effects, the emotional impact is discussed alongside disturbing effects, misleading activities by schools and individual teachers and the impact on staff members' personal lives. Furthermore the article distinguishes between the responses of staff members from schools with a positive inspection judgement and those with a less favourable judgement, and between the responses of teachers and of members of the school management teams. This study is embedded in the Flemish educational context. The results of our study are put in perspective next to results of studies in other educational contexts. Several particularities of the Flemish educational context (and how they may have affected our results) are discussed.

Keywords: inspection; effects; side effects; primary schools; secondary schools; quantitative study

1. Introduction

In an international perspective, the importance of school inspection has intensified during the last decennium due to an increased demand for school accountability. In most countries, inspections have been established to control the quality of education provided by schools, but also to contribute to school improvement (OECD, 2013). The question as to what the effects of inspection are, and whether school inspections effectively contribute to school improvement has been subject of several studies (Chapman, 2001; Chapman and Earley, 2010; Ehren and Visscher, 2006). Initially, effectiveness research was concentrated in the UK

in the years after the establishment of the inspection agency OFSTED in 1992 (Learmonth, 2000). However, until about 2005 the effects of inspections on school improvement had hardly been investigated outside the UK (Ehren and Visscher, 2008).

More recently and in line with the increased role of school inspection in recent years, there has been renewed academic interest in the effects of inspections in different educational systems, e.g. Germany (Dedering and Müller, 2011), Ireland (McNamara and O'Hara, 2006) and the Netherlands (Ehren et al., 2013b). These studies investigated the effects of inspection on schools and whether inspection does respond to its school improvement function, but often also examined the occurrence of unintended effects of inspection. Indeed, the effects of inspections are unarguably accompanied by undesirable side effects, such as “intended strategic behaviour” (e.g. window dressing activities), “unintended strategic behaviour” (e.g. disturbing effects on normal school life) and “other side effects” (e.g. increased stress) (de Wolf and Janssens, 2007, p. 383). Notwithstanding the increased research efforts, it remains unclear what the effects and side effects of inspection are for the inspected schools, and whether or not the inspection drives schools to further improvement. Different studies found incoherent or even contradictory results: whereas some studies concluded that inspections are helpful to most schools (e.g. McCrone et al., 2007; Cuckle, Hodgson & Broadhead, 1998), other studies have concluded that the benefit for schools is generally rather small (e.g. Chapman, 2002; Wilcox and Gray, 1996). Moreover, in order to clarify the driving forces for school improvement, more research on the effects and side effects of inspection in diverse educational contexts is warranted (de Wolf and Janssens, 2007; OECD, 2013).

Also in Flanders (Belgium), the research evidence regarding this topic is currently rather limited. In our view, the Flemish educational system provides an interesting context for this kind of research, as inspection is the only accountability measure for schools towards the national authorities. This study aims to fill the current gap in the knowledge base on the (side)

effects of the Flemish inspection system and thereby to contribute to the international evidence base on this topic. Moreover it distinguishes between schools with different inspection judgements, as it has been argued that the inspection judgement determines the extent to which effects on schools occur (McCrone et al., 2007; Ouston and Davies, 1998). We also compare the responses of teachers and of members of the school management team. Earlier studies pointed out that the reported effects and side effects may depend on staff members' position in the school (Brunsden et al., 2006; Chapman, 2002). It has yet to be investigated whether or not the above-mentioned findings can be confirmed in the Flemish educational context.

We set out on a profound investigation of the effects and side effects in Flemish primary and secondary schools guided by three research questions:

- (1) What are the effects and side effects of inspections on schools?
- (2) What is the impact of the inspection judgement on the effects and side effects of the inspection?
- (3) What are the differences between teachers and the school management teams regarding their perception of the effects and side effects of the inspection?

2. Research context

In this paragraph we provide a thumbnail sketch of the particularities of the Flemish educational accountability context. Flemish schools enjoy a relatively large degree of autonomy when it comes to setting up processes with regard to their traditions and educational methods on how to achieve the “attainment targets” imposed by the government (OECD, 2013; Van Bruggen, 2010). In the absence of central examinations, these inspections are the only accountability measure that exists (Standaert, 2001). The lack of reliable output

data in terms of students' learning performances constrains the inspectors to adopt a process-oriented approach (Van Bruggen, 2010).

Compared to other countries, Flemish schools are not often inspected, i.e. (at least) once every ten years. The inspection gives schools a 'positive', 'restricted positive' or 'negative' judgement. A 'positive' judgement means that a school is considered to have the competencies and preparedness to continue working in an optimal manner and that no follow-up needs to be scheduled. The judgement 'restricted positive' denotes a second inspection is required three years after the initial inspection to determine whether or not identified shortfalls have been adequately addressed. Schools that show structural deficiencies are given a 'negative' judgement, which comes down to the school being obliged to set up an improvement plan and to have its progress 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).

As it is highly unlikely that schools will be closed down, or that staff members lose their job as an immediate result of inspection, the Flemish inspection system is generally regarded as a relatively 'low-stake'-inspection compared to other educational contexts (Van Bruggen, 2010). Low-stake inspection is considered to foster school improvement and to reduce undesirable side effects (Gärtner et al., 2009; Martin, 2005; Yeung, 2012).

Inspectors in the Flemish educational context do not have the legal right to give advice to schools on how they can improve their current practices. Inspectors have to analyze and report on the school's strengths and weaknesses, but need to refrain from any kind of recommendation to the school on how they might address identified weaknesses. The legislation makes a strict distinction between inspection (for control) and school counselling (for advice). The Inspectorate's operating assumption to pursue its development-oriented

function is that the objective analysis of the schools' own strengths and weaknesses provided by the inspectors will serve as an impetus for the schools to secure the strengths and address the identified weaknesses (Vanotterdijk, 2008).

3. Conceptual framework

This section provides an overview of effects and side effects discussed in the present study. An overview and a definition of each effect and side effect is included in Table 1. Rossi et al. (1999) distinguished between the *conceptual*, *instrumental* and *symbolic effects* of interventions in schools. A fourth effect type, *strategic effects*, was added by Visscher (2002). Based on the assumptions made by several scholars (e.g. Matthews and Smith, 1995; McCrone et al., 2007), a fifth effect type can be added, namely the effect on the feelings of efficacy within the school, in which we distinguish between the *effect on staff members' self-efficacy* and the *effect on the school's collective efficacy*. The beliefs staff members hold about their capabilities to face the challenges of today's education reality strongly influence students' motivation and learning outcomes (Thoonen, Slegers, Peetsma, & Oort, 2011; Tschannen-Moran & Hoy, 2001). The collective efficacy is a key aspect of the school culture and may contribute to explain the effect of teachers on student achievement (Goddard & Goddard, 2001).

This study investigates four kinds of side effects: *emotional side effects* before and during the inspection – i.e. an increase in stress, anxiety, tiredness, conflict, and a decrease in enthusiasm; the *impact on staff members' personal lives*; engaging in *misleading activities*; and *disturbing effects* on normal school life. The distinction between 'effects' and 'side effects' is not strict. Several effects (e.g. symbolic effects or the effect on self-efficacy and collective efficacy) may turn out to be negative, and thus be an undesirable side effect, while several side effects could have a positive outcome (e.g. when misleading behaviour leads to actual and lasting improvement).

Table 1 provides an overview of each of these (side) effects alongside evidence from other educational contexts. The present study focuses on the question whether or not these effects are generated in schools in the Flemish educational context.

<<Insert Table 1 >>

Next to providing descriptive evidence, this study aims to find out to what extent differences exist between schools that received a different inspection judgement. As mentioned in the Introduction, stronger positive effects are expected in schools with a less favourable inspection judgement (McCrone et al., 2007; Ouston and Davies, 1998). Furthermore, earlier studies have shown that principals and members of the management team have a more positive stance towards the effects of inspection on the school compared to teachers (Chapman, 2002; Matthews and Sammons, 2004; Scanlon, 1999). Other studies suggested that the side effects also depend at least partially on the position of members of staff in the school, for instance that the emotional burden before and during the inspection is larger for teachers compared to non-teaching colleagues (Brimblecombe et al., 1996; Brunsdon et al., 2006; Wilcox and Gray, 1996).

4. Research method and sample

This article reports on an online survey study of schools' perceptions of the (side) effects of inspection. The study sample included every Flemish primary and secondary school that was inspected during a predefined period in the school year 2012-2013. The survey was intended to reach all the staff members in a teaching or managing position in these schools. Schools received the questionnaire eight weeks after the inspection or, in a few cases, where the inspection report had arrived late, three weeks after acknowledgement of the inspection report.

In total data from 130 schools (54.3% of the total number of schools inspected during the predefined period) were retained for this study, totaling 2,624 respondents. We surveyed

both primary and secondary schools. 77.7% of the participating schools and 60.3% of the respondents were from primary education (the difference was due to the larger school size in secondary education). 82.8% of the respondents were teachers, the remaining 18.2% were either principals or members of the management team. In total, 78.7% of those taking part in the survey were women and 21.3% 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 management or teaching position, 73.6% are women and 26.4% are men (Vlaamse Overheid, 2013). Overall figures about the positions of staff members in schools are unavailable.

56.4% of the participants reported their school had received a 'positive' judgement, while 39.6% received a 'restricted positive' judgement and 4.0% of the responses were missing on this item. No schools with a 'negative' judgement participated in the study.

The various concepts in the theoretical framework were operationalised using 5-point-Likert scales, except for the emotional side-effects (as we will discuss later in this section). Principal axis factoring (with Oblimin rotation) revealed that each of the concept scales were unique factors. Table 2 includes an example item for each scale, in addition to information about the psychometric characteristics of the scale in question. This table demonstrates that these scales can be trusted for use in the analyses. Cronbach's alpha are 'moderate' (above 0.78) or 'high' (above 0.90) (Murphy and Davidshofer, 1988).

<< Insert Table 2 >>

Measures for the emotional side effects (stress, anxiety, enthusiasm, tiredness and conflict) were obtained in an alternative way. In order to measure the increase or decrease in the emotional experience due to the inspection, we compared the extent to which the

experienced emotion was experienced right before and during the inspection, with the extent to which the same emotion is experienced at a regular time without inspection. The respondents were asked to what extent they had experienced anxiety two weeks prior to the inspection (t1), during the inspection (t2) and two weeks after the inspection (t3) using 5-point-Likert scales. These estimates were compared to an estimate of anxiety during a regular day in the school (without inspection) (t0) using a t-test. Cohen's d is used as a measure to determine the impact size of eventual differences, with $|d| < .20$ considered an effect with a 'negligible' impact size, $|d| < .50$ 'small', $|d| < .80$ 'medium' and $|d| > .80$ a major impact size (Cohen, 1988).

We also used t-tests to find out whether the reported (side) effects depend on the inspection judgement and on the respondents' position in the school (research question 2 and 3).

5. Results

In this section we first discuss the results of each of the effects and side effects elaborated in the conceptual framework (research question 1). Next we provide figures on the effects in relation to the inspection judgement (research question 2) and on differences between teachers' responses and the school management teams' responses (research question 3).

5.1 Effects and side effects of school inspection

Table 3 gives an overview of the effects of inspection in our sample (with exception of emotional side effects, see Table 4).

<< Insert Table 3 >>

5.1.1 Intended effects

On average, schools report moderate conceptual effects ($M= 3.36$ on a scale from 1 to 5) as well as moderate instrumental effects ($M= 3.43$). Conceptual effects rather concern an increase in ‘reflection about the qualities of the school’ than an increased ‘knowledge about the strengths and weaknesses of the school’. An analysis at the item level reveals that the lowest scoring items (M =respectively 2.96 and 3.28) are ‘*the inspection gave me a better idea about my professional weaknesses*’ and ‘*the inspection gave me a better idea about the school’s weaknesses*’, while the items that obtained the highest scores (both $M=3.51$) are ‘*the inspection made me reflect about my current practices*’ and ‘*the inspection fostered the willingness to take a critical stance*’.

With regard to the instrumental effects, a larger effect is reported on individual classroom practices compared to the school policy level: the respondents rather agree that the team has made changes in terms of its professional approach ($M= 3.39$), while there is less agreement that the school has made changes to the school policy ($M= 3.06$). Respondents report that certain things are handled in a different manner ($M= 3.58$), and that concrete actions for improvement are taken or planned as a response to the inspection ($M= 3.85$), but they are unsure whether or not in the end pupils will benefit from these actions ($M= 3.22$).

The mean scores for symbolic and strategic effects ($M=$ respectively 3.02 and 3.00) show that these are not common effects, but the relatively high standard deviation for strategic effects ($SD= 1.16$) indicates large differences between respondents.

There is generally a moderate positive impact on the individual self-efficacy of the respondents, and a slightly stronger positive impact on the collective efficacy ($M=$ respectively 3.36 and 3.49). Also for these effects, the standard deviations indicate considerable differences between respondents.

5.1.2 Side effects

The mean score of 2.79 in combination with the high standard deviation (SD=1.23) for the scale ‘impact on personal life’ means that for a large group of respondents, the inspection has no or only a very limited personal impact, but that for a considerable minority the inspection strongly affected their personal life: indeed, 34.5% of the respondents have a scale score between 1 and 2, while 22.0% have a scale score of 4 or higher. Furthermore, the low scores obtained for both scales regarding ‘misleading activities’ clearly indicate that, on average, neither schools in general, nor individual teachers do engage in activities intended to present a false picture to the inspectors. Finally, disturbing effects are rather uncommon in the Flemish educational context, although also here the standard deviation indicates rather strong differences between respondents (SD= 1.03).

Table 4 shows that the inspection leads to a very strong increase in stress before and during the inspection ($|d|$ = respectively 1.03 and 1.13). Nevertheless, two weeks after the inspection, the amount of stress has dropped to a negligible extent ($|d|$ = 0.09) (Cohen, 1988). There is also a strong increase in anxiety before and during the inspection ($|d|$ = respectively 0.84 and 1.01), while after the inspection some of the increased anxiety still remains ($|d|$ = 0.32). The inspection also leads to a strong decrease in professional enthusiasm two weeks prior to, and during, the inspection. Two weeks after the inspection this decrease is still considerable ($|d|$ = respectively 0.71, 0.87 and 0.46). The same applies to the increase in tiredness before, during and after the inspection, although each of the impact sizes is smaller than for enthusiasm ($|d|$ = respectively 0.44, 0.48 and 0.39). On average, no substantial increase or decrease in conflict between staff members was reported.

<< Insert Table 4 >>

5.2 Impact of the inspection judgement on the effects of inspection

Each of the aforementioned effects is statistically significantly influenced by the inspection judgement (see Table 5). The conceptual effect (the extent to which the inspection influences the thinking of staff members) and the instrumental effect (the extent to which the inspection results in decision-making and actions) are affected by the judgement in an opposite direction: a positive judgement leads to more conceptual effects, while a less favourable judgement spurs instrumental effects. However, the impact of the judgement on these effects can be considered 'negligible' (Cohen's $|d|$ = respectively 0.11 and 0.19). The judgement has a 'small' impact ($|d|=0.29$) on symbolic effects: in schools with a 'restricted positive' judgement, more symbolic effects are reported, albeit still to a small extent ($M= 3.18$).

In schools with a positive inspection judgement, more strategic effects are reported (impact size 'medium', $|d|= 0.64$). There is also a 'medium' impact of the inspection judgement on the feelings of self-efficacy ($|d|= 0.62$) and even a 'large' impact on the collective efficacy ($|d|= 1.20$). When the inspection judgement is 'positive', there is a substantial positive impact on the self-efficacy of individual staff members ($M= 3.61$) and even a strong positive impact on the collective efficacy ($M= 3.88$). In schools where the inspection judgement is 'restricted positive', these effects are neutral ($M=$ respectively 3.00 and 2.93).

As the inspection judgement is given at the end of the inspection week, it does not make sense to investigate the impact of the inspection judgement on side effects that occur before or during the inspection (misleading behaviour, impact on personal life, disturbing effect and emotional effects before and during the inspection). Two weeks after the inspection, each of the emotional side effects is stronger in schools with a less favourable inspection judgement, albeit only to a medium (enthusiasm, $|d|= 0.57$), small (stress, anxiety and conflicts, $|d|=$ respectively 0.47, 0.29 and 0.48) or negligible extent (tiredness, $|d|= 0.17$).

<<Insert Table 5>>

5.3 Differences between teachers and management teams regarding their perception of (side) effects

Table 6 provides an overview of differences between the responses of teachers and members of the management team (including principals). A number of statistically significant differences arose from the data on the effects of inspection, i.e. differences regarding conceptual effects, instrumental effects and the impact on self-efficacy (estimated higher by members of the management team) and regarding strategic effects (estimated higher by teachers). However, Cohen's d measure for impact size revealed that these differences were either negligible ($|d| < 0.20$) or small ($|d| < 0.50$) (Cohen, 1988). The same applies to the side effects of the inspection: only the impact of the inspection on the professional enthusiasm prior to and during the inspection reveals appreciable differences between teachers and the school management teams: teachers report slightly lower professional enthusiasm in both periods ($|d| =$ respectively 0.22 and 0.20). No figures about misleading activities at teacher level are included in Table 6 as this side effect was only surveyed with regard to teachers.

<<Insert Table 6 >>

6. Discussion

This study contributes to the existing evidence base on the (side) effects of inspection by identifying insights into the occurrence of these effects in the Flemish educational context. The above evidence expands the current knowledge base and adds several nuances to it and, in some cases, it contradicts earlier findings.

Regarding the first research question, we found that the inspection has, on average, only moderate conceptual and instrumental effects on schools, and small symbolic and strategic effects. There are moderate positive effects on the feelings of self-efficacy and slightly stronger positive effects on collective efficacy within the inspected schools. Moderate conceptual effects (e.g. Chapman, 2002; Ouston et al., 1997; Wilcox & Gray, 1995, 1996) and moderate instrumental effects (e.g. Lowe, 1998; Ouston et al., 1997; Wilcox & Gray, 1996) were also reported by earlier studies, but other studies found a larger instrumental effect of inspections on schools (e.g. Cuckle et al., 1998; McCrone et al., 2007). However, the ambiguity of the results in earlier studies (Chapman and Earley, 2010; Ehren and Visscher, 2006) and the lack of clear conceptualization observed in most studies in the field of the effects of inspections (Ehren and Visscher, 2006) restrict the extent to which our results can be compared to earlier studies.

Even though the Flemish inspection operates at school level and does not make judgements about individual classroom practices, we found that classroom-related instrumental effects at teacher level are stronger than instrumental effects at school policy level. This finding contradicts earlier findings from studies in other contexts (Case et al., 2000; Chapman, 2002). Possibly, it is the strong process-oriented approach adopted by the Flemish inspectors that implies that classroom practices come under the inspection spotlight and are consequently affected by the inspection. It is self-evident that this hypothesis should be the subject of further research.

Related to the conceptual effects, the inspection increases reflection about the school's qualities, but it does not directly result in new insights into the schools' strengths and weaknesses (in line with the results of some qualitative studies which also reported increased reflection and consultation amongst staff (Chapman, 2002; Hosker & Robb, 1998; Kelchtermans, 2007)). Our data do not support the operating assumption of policy-makers and

the Inspectorate in the Flemish educational context, i.e. that the inspection will lead to (new) information about weaknesses in the school, which in turn will be a major lever for schools to engage in change (Vanotterdijk, 2008). The results endorse the idea of Macbeath (2006), Matthews and Sammons (2004) and Woods and Jeffrey (1998) that inspection needs to go hand-in-hand with constructive advice to ensure profound conceptual and instrumental effects. Nevertheless a close eye needs to be kept on the accountability purpose of inspection when inspectors cross the line between the ‘watchdog’ and ‘guidedog’ roles (Macnab, 2004; Martin, 2005; Ouston and Davies, 1998).

We found generally a strong increase in stress and anxiety, and – to a smaller extent – in tiredness, as well as a decreased level of professional enthusiasm before and during the inspection. Therefore, our results are in line with earlier results with regard to the emotional impact of the inspection prior to, during, and after the inspection (Brimblecombe & Ormston, 1995; Brunsten et al., 2006; Ferguson et al., 1999). However, the present study adds several nuances, such as the extent to which levels of anxiety and tiredness, and particularly professional enthusiasm remain affected after the inspection. On average hardly any effect was reported on conflicts in the school teams. Other side effects such as misleading activities were on average rather limited, in contrast with the evidence collected in other educational contexts (Brimblecombe et al., 1996; Case et al., 2000; Perryman, 2009). In addition, earlier evidence regarding disturbing effects (e.g. Kogan and Maden, 1999; McCrone et al., 2007) can not be confirmed in our study, as we did not find evidence of a substantial impact. Jeffrey and Woods’ (1998) and Perryman’s (2006) finding that inspection impacts on the personal lives of staff members in schools, are confirmed for some staff members, but were not reported by others. The lack of coherence of some of our results with earlier research regarding the side effects of inspection may be explained by the rather low-stakes context of Flemish inspections, compared to the high-stakes UK context in which most studies have

been conducted (Van Bruggen, 2010). However, our results are not supportive towards several scholars' assumption that lower stakes inspection approaches lead to a smaller emotional impact on school staff (Gärtner et al., 2009; Martin, 2005; Yeung, 2012). Two particularities of the Flemish educational context are worth discussing in this regard. First the long period between inspections adds to the lack of experience of staff members with external evaluation. This lack of experience does not only bring uncertainty about what is to be expected – fear of the unknown is a major source of stress and anxiety (Brimblecombe and Ormston, 1995; Wilcox and Gray, 1996) – but also artificially increases (the perception of) the stakes for schools. In a similar vein, the fact that inspection is the sole accountability measure for Flemish schools may increase the emotional side effects because schools lack reliable benchmarking information in order to judge their own quality. This may raise uncertainty about the school's output performance in relation to the expected standards, which is unarguably associated with increased stress and anxiety (Brimblecombe and Ormston, 1995; Sandbrook, 1996). A final possible explanation derives from the finding that only limited disturbing effects were reported, which means that the preparation for the inspection is added to the regular workload. The increased workload may also partially explain the increase in stress and the decrease in professional enthusiasm, as earlier studies have shown (Case et al., 2000; Chapman, 2002; Perryman, 2009).

Based on earlier studies, we expected that an unfavourable inspection judgement would lead to stronger effects (McCrone et al., 2007; Ouston and Davies, 1998). However, this hypothesis can not be generalized to every effect. We therefore stress the importance of conceptually distinguishing in future research at least between the conceptual and the instrumental effects of inspections. Although differences were small, we found larger conceptual effects in schools with a positive inspection report, while there is more action (instrumental effect) on the inspection results in schools given a 'restricted positive'

judgement. The stronger instrumental reaction in those schools is possibly explained by the imposed need to act. The finding that there is more positive impact on self-efficacy and collective efficacy in schools with a positive judgement, supports the findings of Perryman (2009) and McCrone et al. (2007), but contradicts the conclusion of Case et al. (2000). The latter study found a decrease in the feeling of self-efficacy, regardless of the inspection judgement. Furthermore, respondents from schools with an unfavourable inspection judgement reported more severe emotional effects after the inspection, but differences with respondents from other schools were, by and large, rather limited.

We found that conceptual effects are smaller, while both the symbolic and instrumental effects are larger in schools that received a 'restricted positive' judgement. These findings lead to the assumption that the weaknesses identified by the inspectors strongly reinforce what either the management team or some of the teachers already knew, but in order to address these deficiencies, they needed the inspection's judgement to convince others in the school to take action. We presume that improvement in schools with an unfavourable inspection judgement is generally not directly influenced by the inspection, but rather indirectly by the legitimacy it adds to positions that had been taken earlier by the principal and/or other staff members in the school.

We found only small differences in the perception of the effects of inspection between teachers and members of the management team. The finding in other educational contexts that teachers report less effects of inspection (Chapman, 2002; Matthews and Sammons, 2004; Scanlon, 1999) could therefore not be confirmed in our study. We found that some of the side effects are slightly more severe for teachers than for members of the management team, while earlier findings in other educational contexts indicate larger differences (Brimblecombe et al., 1996; Brunsten et al., 2006; Wilcox and Gray, 1996).

We have refrained in this study from drawing conclusions regarding causal effects. It is methodologically challenging to measure the precise effect of inspection, as inspection becomes quickly entangled with other external and internal impulses on the schools (de Wolf and Janssens, 2007; Matthews and Sammons, 2004). Only seldom is an effect of inspection the result of one single school or inspection feature (Ehren and Visscher, 2008). While our quantitative approach served the goal of describing the effects and side effects, a more qualitative approach may be required to identify the processes in the school that lead to certain outcomes.

The results provided by this study contribute to the discussion on how inspection can be organized in a way that spurs school improvement with a minimal amount of side effects for schools. We found that the operating policy rhetoric in Flemish education that ‘providing schools with an analysis of strengths and weaknesses will lead to school improvement’ is without foundation, and that the Flemish inspection – although called low-stakes – still has strong emotional side effects.

Reference list

- Bandura A. (1997) *Self-efficacy: The exercise of control*, New York: Freeman.
- Brimblecombe N and Ormston M. (1995) Teachers' perceptions of school inspection: A stressful experience. *Cambridge Journal of Education* 25: 53-61.
- Brimblecombe N, Shaw M and Ormston M. (1996) Teachers' intention to change practice as a result of OFSTED school inspections. *Educational Management, Administration and Leadership* 24: 339-354.
- Brunsdon V, Davies M and Shevlin M. (2006) Anxiety and stress in educational professionals in relation to Ofsted. *Education Today* 56: 24-31.
- Case P, Case S and Catling S. (2000) Please show you're working; a critical assessment of the impact of Ofsted inspection on primary teachers. *British Journal of Sociology of Education* 21: 605-621.
- Chapman C. (2001) Changing classrooms through inspections. *School Leadership and Management* 21: 59-73.
- 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 and Earley P. (2010) School inspection/external school evaluation. In: Peterson P, Baker E and McGaw B (eds) *International Encyclopedia of Education (Third Edition)*. Oxford: Elsevier, 719-725.
- Cohen J. (1988) *Statistical power analysis for the behavioral sciences*, Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cuckle P, Hodgson J and Broadhead, P (1998). Investigating the relationship between OfSTED inspections and school development planning. *School Leadership & Management* 18: 271-283.
- de Wolf IF and Janssens FJG. (2007) Effects and side effects of inspections and accountability in education: An overview of empirical studies. *Oxford Review of Education* 33: 379-396.
- Dederling K and Müller S. (2011) School improvement through inspections? First empirical insights from Germany. *Journal of Educational Change* 12: 301-322.
- Dimmer T and Metiuk J. (1998) The use and impact of OFSTED in a primary school. In: Earley P (ed) *School Improvement after Inspection? School and LEA Responses*. London: Paul Chapman Publishing, 51-61.
- Ehren MCM, Altrichter H, McNamara G, et al. (2013a) 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 MCM, Perryman J and Shackleton N. (2013b) Setting expectations for good education; how Dutch school inspections drive improvement. *ORD Conference*. Brussels.
- Ehren MCM and Visscher AJ. (2006) Towards a theory on the impact of school inspections. *British Journal of Educational Studies* 54: 51-72.
- Ehren MCM and Visscher AJ. (2008) The relationship between school inspections, school characteristics and school improvement. *British Journal of Educational Studies* 56: 205-227.
- Ferguson N, Earley P, Ouston J, et al. (1999) New heads, OFSTED inspections and the prospects for school improvement. *Educational Research* 41: 241-249.
- Fitz-Gibbon CT and Stephenson-Forster NJ. (1999) Is Ofsted helpful? An evaluation using social science criteria. In: Cullingford C (ed) *An inspector calls. Ofsted and its effect on school standards* London: Kogan Page, 97-118.
- Gärtner H, Hüsemann D and Pant HA. (2009) Wirkungen von Schulinspektion aus Sicht betroffener Schulleitungen. Die Brandenburger Schulleiterbefragung. *Empirische Pädagogik* 23: 1-18.
- Goddard, R. D., & Goddard, Y. L. (2001). A Multilevel Analysis of the Relationship between Teacher and Collective Efficacy in Urban Schools. *Teaching and Teacher Education*, 17(7), 807-818.
- Gray J and Gardner J. (1999) The impact of school inspections. *Oxford Review of Education* 25: 455-468.
- Hosker H and Robb S. (1998) Raising standards and raising morale: a case study of change. In: Earley P (ed) *School Improvement after Inspection? School and LEA Responses*. London: Paul Chapman Publishing, 86-96.

- Jeffrey B and Woods P. (1998) *Testing teachers: The effect of school inspection on primary teachers*, London: Falmer Press.
- Kelchtermans G. (2007) Macropolitics caught up in micropolitics: the case of the policy on quality control in Flanders (Belgium). *Journal of Educational Policy* 22: 471-491.
- Klassen RM, Bong M, Usher EL, et al. (2009) Exploring the Validity of a Teachers' Self-Efficacy Scale in Five Countries. *Contemporary Educational Psychology* 34: 67-76.
- Kogan M and Maden M. (1999) An evaluation of evaluators: The Ofsted system of school inspection. In: Cullingford C (ed) *An inspector calls: Ofsted and its effect on school standards*. London: Kogan Page, 9-32.
- Kunter M, Frenzel A, Nagy G, et al. (2011) Teacher enthusiasm: Dimensionality and context specificity. *Contemporary Educational Psychology* 36: 289-301.
- Learmonth J. (2000) *Inspection. What's in it for schools?*, London/New York: Routledge Falmer.
- Lowe G. (1998) Inspection and change in the classroom: rhetoric and reality? In: Earley P (ed) *School Improvement After Inspection? School and LEA Responses*. London: Paul Chapman Publishing, 97-109.
- Macbeath J. (2006) *School Inspection and Self-Evaluation. Working with the new Relationship*, New York/London: Routledge.
- Macnab D. (2004) Hearts, minds and external supervision of schools: direction and development. *Educational Review* 56: 53-64.
- Martin S. (2005) Evaluation, Inspection and the Improvement Agenda Contrasting Fortunes in an Era of Evidence-Based Policy-Making. *Evaluation* 11: 496-504.
- Matthews P and Sammons P. (2004) Improvement through inspection. An evaluation of the impact of Ofsted's work. London: Ofsted.
- Matthews P and Smith G. (1995) OfSTED: inspecting schools and improvement through inspection. *Cambridge Journal of Education* 25: 23-34.
- McCrone T, Rudd P, Blenkinsop S, et al. (2007) Evaluation of the Impact of Section 5 Inspections. Slough: National Foundation for Educational Research.
- McNamara G and O'Hara J. (2006) Workable compromise or pointless exercise? School-based evaluation in the Irish context. *Educational Management, Administration and Leadership* 34: 564-582.
- Murphy KR and Davidshofer CO. (1988) *Psychological testing: Principles and applications*, Englewood Cliffs, New Jersey: Prentice-Hall.
- Nicolaidou M and Ainscow M. (2005) Understanding failing schools: perspectives from the inside. *School Effectiveness and School Improvement* 16: 229-248.
- OECD. (2013) Synergies for better learning. An international perspective on evaluation and assessment. *OECD Reviews of Evaluation and Assessment in Education*. Paris: OECD.
- Onderwijsinspectie. (2013) *Onderwijs Spiegel 2013 [Education Mirror]*. Brussel: Onderwijsinspectie / Vlaams Ministerie van Onderwijs en Vorming, 114.
- Ouston J and Davies J. (1998) OFSTED and afterwards? Schools' responses to inspection. In: Earley P (ed) *School improvement after inspection? School and LEA responses*. London: Paul Chapman Publishing, 13-24.
- Perryman J. (2006) Panoptic performativity and school inspection regimes: disciplinary mechanisms and life under special measures. *Journal of Education Policy* 21: 147-161.
- Perryman J. (2009) Inspection and the fabrication of professional and performative processes. *Journal of Education Policy* 24: 611-631.
- Plowright D. (2007) Self-evaluation and Ofsted inspection. Developing an integrative model of school improvement. *Educational Management Administration & Leadership* 35: 373-393.
- Rossi PH, Lipsey MW and Freeman HE. (1999) *Evaluation: A systematic approach*, Thousand Oaks, CA: Sage Publications, Inc.
- Sandbrook I. (1996) *Making sense of primary inspection*, Buckingham, UK/Bristol, USA: Open University Press.
- Scanlon M. (1999) The impact of Ofsted inspections. Slough, UK: National Foundation for Educational Research.
- Standaert R. (2001) *Inspectorates of Education in Europe. A critical analysis*, Leuven, Belgium: Acco.
- Stoll L and Fink D. (1996) *Changing our schools*, Buckingham, UK: Open University Press.

- Thoonen, E. E. J., Slegers, P. J. C., Peetsma, T. T. D., & Oort, F. J. (2011). Can Teachers Motivate Students to Learn? *Educational Studies*, 37(3), 345-360.
- Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783-805.
- Van Bruggen JC. (2010) Inspectorates of Education in Europe; some Comparative Remarks about their Tasks and Work. Brussels: SICI.
- Vanotterdijk R. (2008) (Gedifferentieerd) doorlichten: leren dansen op een slappe koord [Inspection (differentiated approach): learning to show one's paces]. *Tijdschrift voor Onderwijsrecht en Onderwijsbeleid* 2007-08: 291-314.
- Visscher AJ. (2002) A framework for studying school performance feedback systems. In: Visscher AJ and Coe R (eds) *School improvement through performance feedback* Lisse, The Netherlands: Swets & Zeitlinger.
- Vlaamse Overheid. (2013). Vlaams Onderwijs in Cijfers 2012-2013 [Flemish Education in Figures 2012-2013]. <http://www.ond.vlaanderen.be/onderwijsstatistieken>
- Wilcox B and Gray J. (1995) Reactions to inspection: A study of three variants. In: Gray J and Wilcox B (eds) *Good school, bad school: evaluating performance and encouraging improvement*. Buckingham, UK/Philadelphia, USA: Open University Press, 149-166.
- Wilcox B and Gray J. (1996) *Inspecting schools: Holding schools to account and helping schools to improve*, Buckingham, UK: Open University Press.
- Woods P and Jeffrey B. (1998) Choosing positions: living the contradictions of OFSTED. *British Journal of Sociology of Education* 19: 547-570.
- Yeung SYS. (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.