



## UvA-DARE (Digital Academic Repository)

### Effects and side effects of inspections and accountability in education

de Wolff, I.; Janssens, F.

**Publication date**  
2005

[Link to publication](#)

#### **Citation for published version (APA):**

de Wolff, I., & Janssens, F. (2005). *Effects and side effects of inspections and accountability in education*. (Scholar Working Papers Series; No. WP53/05). Unknown Publisher.

#### **General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*Effects and side effects of inspections and  
accountability in education; an overview of  
empirical studies*

Inge F. de Wolf\*

University of Amsterdam / Netherlands' Education Inspectorate

Frans J.G. Janssens

University of Twente / Netherlands' Education Inspectorate

\* Research Institute Scholar, University of Amsterdam, Roetersstraat 11, 1018 WB  
Amsterdam, The Netherlands. Email: I.deWolf@owinsp.nl.

## **Abstract**

*This paper presents an overview of studies into effects and side effects of control mechanisms in education. We focus on effects and side effects of inspection visits and public performance indicators. A first conclusion is that the studies do not provide us with a clear answer to the question of whether inspection visits have positive causal effects on quality of schools. Results of studies on publications of public performance indicators are more unambiguous. They lead us to draw the conclusion that, although principals and teachers believe performance indicators are important, parents and pupils take very little notice of these indicators when choosing schools. A third conclusion concerns the occurrence of side effects of school inspections and other control mechanisms in education. A number of the studies discussed clearly refer to the existence of these side effects, like 'window dressing' and other types of 'gaming'.*

## ***1. Introduction***

This paper focuses on the effects and side effects of control mechanisms in education. The aim of the paper is to generate an overview of effects and side effects on the basis of empirical studies on this subject. We distinguish between two types of control mechanisms in education, namely (a) external evaluation by means of school inspections and (b) accountability systems including the publication of performance indicators about the quality of schools (for an overview of these and other educational control or evaluation systems in Europe, see Eurydice, 2004).

### ***1.1 Why focus on the (side) effects of control mechanisms in education?***

For a number of reasons it is relevant to acquire an insight into the effects and side effects of control mechanisms in education. An initial important reason concerns a series of relatively recent empirical studies in this field. This series of studies, which are discussed later in this paper, reveals that the subject is rather complex, that conclusions are ambiguous and that an initial number of interesting empirical findings have been made by now.

Research into the (side) effects of control mechanisms in education is also relevant from a policy perspective. This has everything to do with the considerable increase in supervision during the past two decades and the related discussion of the efficiency and effectiveness of control mechanisms. With regard to education, there exist large differences in control mechanisms between countries. Some countries, mostly European, have inspectorates while other countries regulate school quality by means of legislation or market mechanisms. As a result, the questions are often asked as to whether public control and accountability mechanisms actually work (see for example Leeuw, 2000; Wiggins & Tymms, 2000) and what the advantages and disadvantages of these mechanisms are (see for example Wiggins & Tymms, 2000; Kane & Staiger, 2001; Canton & Webbink, 2004).

Lastly, research into (side) effects of control mechanisms is important for assessors themselves. For example, Inspectorates of Education can use the conclusions to reinforce the

quality and effect of its method of working and to take concrete measures to counteract any side effects.

## ***1.2 Research question and structure***

This paper is a first step to acquire more insight into the effects and side effects of control mechanisms in education and comprises an inventory of the relevant empirical studies. The following question is central to the paper:

*What is known about the effects and side effects of control mechanisms in education?*

Before we examine the (side) effects of control mechanisms in education, we first have to ask ourselves which effects and side effects we are looking for.

## ***2. An analysis framework***

In this section we present an analysis framework for research into the effects and side effects of control mechanisms in education. We based the development of the framework on the two main control activities, which are school inspections and performance indicators.

### ***2.1 Intended effects***

#### *School inspections*

Most Inspectorates of Education have the statutory tasks to supervise, stimulate and report on quality of schools in their countries (see Macnab, 2004). For the Dutch inspectorate, these tasks have recently been examined in greater detail by Ehren, Leeuw and Scheerens (2005). On the basis of an analysis of documents and a series of interviews with the parties involved, Ehren et al. conclude that school inspections lead to an increase in the added value of individual schools. School inspections of the quality requirements, the use of an evaluation framework and the

principle of proportionality lead to schedules for improvements at schools, which result in quality improvements and an increase in the added value. In addition, school inspections have an indirect effect on the quality of education, namely via the public nature of the Inspectorate's assessments.

Next to improvement of the quality of education, inspectorates aim to guarantee a minimum level of education quality. School inspections may also have a positive effect on compliance with legislation and regulations at schools. However, there are important differences between countries in educational legislation as well as in the priority given at the compliance by inspectorates.

School visits are the most important instruments for Inspectorates. During these visits, that generally last a number of days and in which a school's strong and weak points are systematically vetted, the level of education quality and compliance with statutory regulations is assessed. A school is also informed about how the quality of education can be improved. By providing feedback on these findings to schools, but also by publishing the report on the school or institution, inspectorates expect it to be possible to influence the policy pursued at schools and, by doing so, contribute to an improvement in the quality of education at the school in question. In some countries, the findings of the inspection visits are also processed into public information and used for national reports (see also Macnab, 2004).

### *Performance indicators*

Accountability systems with (public) performance indicators are based on the ideas that (a) schools have to give account, (b) parents are informed more effectively and can challenge schools as regards weaknesses, (c) the performance indicators can be used by pupils and parents for school choice (so-called 'voting with their feet') and (d) the school can use the performance indicators as benchmark information. This information on quality is meant to bring about the active involvement of other stakeholders, particularly parents and pupils, in ensuring education quality at schools. It is also assumed that this public information on quality, or public performance indicators will, in itself, have a positive effect on the quality of education at schools. It in fact functions as an important stimulus for school principals and teachers to ensure that their school 'scores as well as possible' due to the

threat of ‘naming and shaming’ in the event of sub-standard scores and possible effects of such public information on the choices of pupils and parents (‘voting with their feet’). A more detailed description of these mechanisms can be found in Goldstein (2001), Kane and Staiger (2001) and Canton and Webbink (2004).

In most countries, the government publishes performance indicators of schools, such as school report cards. In some countries, like the UK, Ireland and the Netherlands, this task is carried out by the Inspectorate (see Macnab, 2004). Next to the government, the press is an active player in publishing performance indicators. Their ‘league tables’ often consist of a selection and weighting of indicators making up a ranking of schools.

### ***2.3 Undesirable side effects***

In addition to direct effects, control systems have side effects. Side effects are unintentional effects which may be desirable or undesirable. Leeuw (2000) points out that undesirable side effects sometimes have the capacity to completely undo intended effects. In any event, they have to be deducted from the actual effects.

In the meantime, lots of studies have been carried out into possible undesirable side effects, in general (Smith, 1995; Meyer et al., 1994; Leeuw, 2000) or specifically in education (Fitz-Gibbon, 1997; Bosker and Scheerens, 1999). Although the accent in all these overviews is different, there exists considerable overlap in the side effects they refer to.

A first important group of undesirable side effects concerns the ‘intended strategic behaviour’ of schools, otherwise referred to as ‘gaming’. The most well-known form of intended strategic behaviour is ‘window dressing’. This means the creation of proactive and reactive arrangements which are generated simply and solely to be assessed more favourably by the supervisor. Sometimes, this form of strategic behaviour is so excessive that it constitutes a ‘misrepresentation’, or even ‘fraud’ and ‘deception’ (see for example, Smith, 1993). Examples of the latter are the drawing up of false documents, the wrongful exclusion of pupils from important tests because it is

feared that they will bring down the school average and the reporting in ill of teachers whose lessons a school would rather not have assessed.

Another more specific form of intended strategic behaviour sometimes occurs in the case of assessment systems in which assessors themselves are given an (external) incentive to ensure a positive assessment. The latter may, for example, be the case in the event of an assessment by colleagues (interview, 'peer review') as is implemented in many systems of higher education. It concerns assessing the quality of courses and research groups by a group of colleagues, but can also take the form of the allocation of research funding and the acceptance of scientific articles. Despite the fact that, in all these forms of 'peer review', a great deal is being done to ensure fair assessments (anonymity, etc.), numerous detrimental effects can occur at the same time. This may lead to excessively positive or negative assessments of certain colleagues. When this results in wrongful lack of negative final verdicts, we might speak of 'orchestration of peer review' (Leeuw, 2003).

A second important group of undesirable side effects is constituted by the various forms of 'unintended strategic behaviour'. With this we mean the unintended influencing of behaviour by the assessor and/or influencing the method of working used for the assessment. Examples include formalisation and 'proceduralisation' ('are the records in order', a 'tick & flick'-approach). Something that poses possibly even more of a threat is the (unintended) influencing of the content and organisation of education by the method of assessment. In effect this means a (usually unintended) one-sided emphasis on the elements that are assessed. In the literature, this is referred to as 'teaching to the test' or 'teaching to inspection'. An interesting theoretical explanation of this phenomenon can be found in Lazear (2004). He simply states that all schools are subject to stimuli to indulge in this form of strategic behaviour and argues that the policy to reduce this phenomenon has to make a distinction between schools with strong and weak pupils (between schools with 'high cost learners' and schools with 'efficient learners'). Related side effects include 'tunnel vision' and 'indicator fixation'. Fitz-Gibbon (1997) points out that this leads to an excessive emphasis on exam performance and 'sub-optimalisation'. This may also result in 'myopia'; a focus on short-term solutions at the expense of the long-term policy (Smith, 1993). School inspections can also result in



schools trying to stay in the middle bracket ('convergence') and in them being less inclined to experiment with new and innovative methods ('ossification') (see for example Smith, 1993). The more often this series of side effects occurs, the greater the chance of 'isomorphism', meaning that all schools will start to become alike.

A final group of undesirable side effects is unrelated to strategic behaviour but is related to other (undesirable) matters which are directly linked to control mechanisms. An example which is often referred to is the occurrence of stress during the assessment, particularly stress experienced by teachers and school principals during school inspections. Many researchers simply regard the occurrence of stress as undesirable (see for example Cullingford, 1997), but stress can also hinder the assessment and jeopardise the validity and reliability of the assessment. Janssens and Leeuw (2001) refer to another unintentional side effect at system level, a side effect that is linked to the publications of performance indicators. According to Leeuw and Janssens, these publications contribute to market forces in education. Students move from poorly functioning schools to properly functioning schools, especially the better students. An unintentional side effect might be that poorly functioning schools get worse and properly functioning schools improve even more.

We conclude by stating that, at least theoretically, a lot of different types of undesirable side effects can occur. Table 1 is a summary of the different categories and types of side effects.

- insert Table 1 about here -

#### ***2.4 Conclusion: elements for a conceptual model***

The intended effects and the possible side effects of school inspections and performance indicators are the main elements of our conceptual model (see Table 2). We assume that inspection visits and performance indicators ought to result in a guarantee with regard to school quality, a certain level of compliance and an improvement in educational quality. Intended and unintended strategic behaviour by schools are undesirable side effects, but other side effects are also possible. Whether

an empirical basis of effects and side effects indeed exists, is subject of the remaining part of this paper.

- *insert Table 2 about here* -

-

### **3. Studies into effects**

In this section, we present an overview of the most important empirical studies with regard to effects of control mechanisms in education. We distinguish between effects of inspection visits and effects of public performance indicators.

#### ***3.1 Effects of inspection visits***

There is a wide variety of research into the effects of inspection visits, in the sense that studies differ considerably as regards the way in which effects are measured. We make a distinction between the following approaches:

1. effects on the satisfaction of teachers and school principals
2. effects on the behaviour of teachers and school principals
3. effects on school improvement
4. effects on output, that is effects on pupils' educational achievements

#### ***Effects of inspection visits on satisfaction***

The central assumption of research on the satisfaction of teachers and school principals is that a stronger (perceived) effect leads to greater satisfaction with school inspections. For example, Gray and Gardner (1999) and Fitz-Gibbon and Stephenson-Forster (1999, 1996) investigated 'the impact of school inspections' by asking schools about their experiences. Gray and Gardner (1999) found that most schools experience the inspection visit as professional and supportive, so they concluded that these visits are effective. They also report that the schools are less satisfied about specific

points, such as the required time investment. According to Fitz-Gibbon and Stephenson-Forster (1999, 1996), however, it is still unclear what the exact effects of school inspections are. These authors found that school principals are neither satisfied nor dissatisfied with inspection visits.

Recently, Emmelot, Karsten, Ledoux and Vermeulen (2004) and Matthews and Sammons (2004) carried out comparable studies. On the basis of these results, we have to draw the conclusion that a significant majority (well over 80%) of the school principals are satisfied with the inspection visits and believe that the visits contribute positively to the quality of schools. The satisfaction with school inspections is, as might be expected, also positively influenced by the Inspectorate's judgement of a school (Emmelot et al., 2004; Matthews & Sammons, 2004).

#### *Effects of inspection visits on behaviour*

The few studies that investigate the effects of inspection visits on teachers' teaching strategies focus on the question of whether the assessment of lessons results in changes in teaching strategies of teachers. All these studies focus on school inspections in the UK, where the inspectorate (Ofsted) assess individual teachers. Brimblecombe, Shaw and Ormston (1996) reveal that 38% of the assessed teachers are inclined to implement changes shortly after an inspection visit, particularly in relation to the way they teach and organise their class. The tendency to change increases the higher up the teacher is within the organisation. Other investigations usually establish a slightly stronger relationship. For example, Chapman (2001, p. 63-64) concludes that 'about 50% of teachers agreed or strongly agreed that (Ofsted) inspection leads to changes in classroom practice' and that '58% of teachers thought that Ofsted is a useful tool for school improvement'. Both the study of Chapman (2001) and the studies by Brimblecombe et al. (1996, 1995) conclude that the relationship between teacher and inspector contributes significantly to the extent to which lesson and inspection visits have an effect. Chapman (2001) also suggest that feedback is a key factor (p. 69).

Interesting summaries of studies on effects on the behaviour of school principals can be found in Learmonth (2000) and Ouston, Fidler and Earley (1997). These authors show that inspection

visits result in changes in behaviour among a large majority of school principals. In general, younger, less experienced school principals are more likely to implement changes as a consequence of inspection visits than older experienced school principals (see also Fitz-Gibbon & Stephenson-Forster, 1996).

### *Effects of inspection visits on school improvement*

Another traditional method of researching the effects of inspection visits on education quality focuses on the relationship between the inspection visit and (changes in) school policy. The main issue in this type of research is the extent to which inspection visits lead to school improvement. Empirical analyses of the relationship between inspection visits and school improvement generally involves the use of questionnaires submitted to school principals and teachers. An interesting overview of these studies is included in the literature studies of Learmonth (2000) and Ouston et al. (1997). On the basis of these studies, one can conclude that inspection visits have positive effects on school improvement. However, few conclusions can be drawn as regards the type of school improvement and the extent and consequences thereof. Although almost all the discussed studies demonstrate that the management of schools usually implement policy changes, conclusions regarding the extent to which this happens vary from slightly to very much (Learmonth, 2000; Ouston et al., 1997). It is plausible that the extent to which inspection visits contribute to school improvement, is influenced by other matters such as context and culture (the latter is demonstrated by, among others, Chapman, 2001).

Frederiks (1996) and Jeliaskova and Westerheijden (2000) did not study the effect of inspection visits directly, but did carry out research into effects of review committees in university education. The central issue in both studies was how external evaluations lead to policy changes at Dutch universities. Frederiks interviewed the actors who were directly involved to establish which activities had resulted from the review judgements. He concludes that the visitation system contributes to new and/or changed policy at schools. Jeliaskova and Westerheijden (2000) nuance this conclusion and state that policy changes only occur in cases where there has not been enough

structural attention on quality insurance. When educational quality is a structural point of attention, reviews mainly result in 'incremental changes' (p. 26).

### *Effects of inspection visits on educational achievement*

Different researchers have examined whether the intended quality improvements by the English Inspectorate, Ofsted, are reflected in improved examination results at secondary schools. The researchers investigated the relationship between a visit by Ofsted and examination results. Although the analysis method varies from very straightforward (Cullingford & Daniels, 1999; Wilcox & Gray, 1996) to complex (Shaw, Newton, Aitkin & Darnell, 2003; Rosenthal, 2004) and a critical assessment can only really be made of the analysis method used in the latter study (Rosenthal, 2004), it is striking that all four studies referred to establish that Ofsted visits have a negative effect, at least in the short term (that is in the year of the inspection visit). Possible causes of this negative effect, according to the researchers, are stress and the need to prepare thoroughly for inspection visits. Unfortunately, possible effects in the longer term have not been investigated in these studies, nor have effects in other sectors of education.

In a study by Scalon (1999) school principals and teachers were simply asked whether inspection visits by Ofsted lead to higher student results. On low quality schools the high majority answered positive (60% of the teachers and 79% of the principals). On other schools, one third of the principals and one quarter of the teachers had the opinion that Ofsted-inspections lead to higher student results (p. 47).

### ***3.2 Effects of public performance indicators***

By the publication of performance indicators, inspectorates aim to stimulate school quality in a more indirect way. Most international research into the effects of these (public) performance indicators is limited to analyses relating to the reliability of league tables (see for example Kane & Staiger, 2001; Klein, Hamilton, McCaffrey & Stencher, 2000; Goldstein, 2001; Wiggins & Tymms,

2000). The main conclusion of these studies is that league tables are insufficiently reliable to have any effect at all. The authors refer, without exception, to the intended effects on school choice behaviour. Possible effects on the involvement of parents and pupils ('voice') and a more effective contribution by other stakeholders are often not mentioned. We discuss the main findings on both types of effects.

### *Effects of performance indicators on school choice behaviour*

Research on school choice behaviour in England and France does not make it abundantly clear that the publication of performance indicators matters. In fact, Karsten and Visscher (2001) conclude, on the basis of a number of international studies, that parents in general pay little attention to public performance indicators. According to Janssens and Visscher (2004) and Visscher (2004), this is related to the following factors:

- A significant number of the parents have no access to this information or are not aware of its existence.
- The use of the available information requires certain training, or the information is difficult to convert to people's individual situations.
- The choice options are limited (by spatial borders and other limitations).
- The performance indicators often reflect only one of the factors on which parents base their choice of school.

According to Janssens and Visscher, the publications are used, in both France and England, particularly by parents from the upper and middle classes. When the educational provision is subject to significant formal and/or material limitations, very undesirable side effects are also observed within this group of parents, such as the exchanging of addresses in the event of a limited choice of schools, searching for unusual subjects or subject combinations, requests for homogenous classes and protests against school allocations (Janssens and Visscher, 2004; Visscher 2004).

Teachers, on the other hand, appear to be more sensitive to the scores on the school report card, as demonstrated by Waterreus (2003). He reveals that, in Dutch secondary education, scores on these cards influence teacher mobility. Staff turnover is less at schools with good scores than at other secondary schools and these schools also attract more new teachers. It has to be pointed out, however, that the detected effects are minor.

#### *Effects of performance indicators on 'voice'*

In addition to the effect on choice of school, inspectorates intend the publication of performance indicators to increase the involvement of parents and pupils in the school and to provide them with instruments to communicate with the school as regards its strong and weak points. This enables parents (and other stakeholders) to contribute more effectively to (improving) the management of schools. In this context, Meijer (2004) refers to stimulating and improving the 'voice', but immediately attaches to this the conclusion that there is hardly empirical evidence of this possible effect of school report cards. In fact, according to the Dutch parents interviewed by Meijer, the performance indicators on the school report card are currently not, or scarcely, used and few parents are aware of the existence of this information.

#### *Effects of performance indicators on school improvement*

Performance indicators appear to have hardly any effects on 'school improvement' (see for example Janssens & Visscher, 2004; Doolaard & Karstanje, 2001). Neither in France nor in England do schools actually work on the quality of the education they provide as a result of the publications. However, there is evidence of 'window dressing' and a desire to achieve short-term effects. School principals, however, have a different opinion. The study by Emmelot et al. (2004), for example, shows that Dutch school principals regard the 'stimulating of schools to improve quality' as the most important effect of publishing reports and information on the quality of schools.

### *Effects of performance indicators on educational achievement*

A recent paper of Hanushek and Raymond (2004) shows that the introduction of the accountability system in the United States has a 'clear positive impact on student achievement'. By using both (a) variance between states in the year of introduction different elements of the accountability policy and (b) statistical methods to isolate the policy effects of other state effects, the authors prove that the introduction of state accountability has a positive impact of on student performance. Furthermore, interesting differences between subgroups are found; Hispanic and white students gain much more than black students. On the basis of these finding, Hanushek et al. conclude that the accountability system improves student achievement, but widens the black-white gap. Of the introduction of school report cards itself, one of the key elements of the accountability systems, they did not find significant effects on student performance.

### **3.3 Other effects**

The studies described above all focus on direct or indirect effects that improve the quality of education. We already stated that school inspections are also intended to guarantee a certain level of compliance with laws of education, but empirical studies into effects on compliance are scarce. Regarding this topic, we only found a single study by Matthews and Sammons (2004). This study consists of a combination of empirical data on a range of trends in education and the opinion of experts on the effects of school inspections. Both sub-studies draw similar conclusions, namely many effects of school inspections on English education including a positive effect on the level of compliance. Whether there is an actual causal relationship continues, however, to be unclear



because the study focuses on effects at system level and the methods used are not very suitable for demonstrating causal relationships.

### ***3.4 Methodological issues***

The focus of each of the studies discussed above is on the extent to which inspection visits contribute to an improvement in the quality of education. The –assumed- causality in this question is the most interesting but, at the same time, the most problematic aspect. Although a lot of studies establish a relationship between inspection visits and improvement in behaviour/policy/quality of schools, the question still remains as to whether there is actually a causal connection. It is in any event typical that the studies make almost no use of research designs which focus more strongly on exposing causal relationships, such as the use of control groups and more quasi-experimental approaches (see for example Shadish, Cook and Campbell et al., 2002; Cook, 2002; Oosterbeek, 2001).

Moreover, there are two other important methodological issues relating to the assessment of effects of inspection visits to which little systematic attention has been paid. These are (a) the distinction between two types of schools: low-performing and other schools and (b) the period of time during which (quality) changes might become visible. With regard to low-performing and other there is a major difference in the possible effects of school inspections. For example, certain intended effects are aimed primarily at (quality improvements of) low-performing schools (such as compliance with the guarantee level of education quality and the supervision of compliance). Moreover, most countries employ different methods of working with regard to school inspections of low-performing and other schools. In any case, the (literature) study by Learmonth (2000) demonstrates that as regards effects it is worth making a distinction between the two types of schools.

With regard to the period of time during which the changes can take place, and particularly the period of time during which quality improvements become visible, peculiar choices are sometimes made in the studies discussed. A distinction is hardly ever made between short-term and long-term

effects. Moreover, the studies which examine the effects of inspection visits on educational achievements (Rosenthal, 2004; Shaw et al., 2003; Wilcox and Gray, 1996) relate primarily to the year in which the inspection visit has taken place or the year immediately thereafter. The question is, of course, whether the intended effects of inspection visits on pupils' educational achievements will appear so quickly. In research into school effectiveness, a period of five years usually applies, but Learmonth (2000) and Matthews and Sammons (2004) point out that changes take place a lot earlier (even in the case of changes in student achievement). We plea for a distinction between short- and long-term effects. This also makes it easier to distinguish between different types of effects and enables researchers to test whether some effects, like changes in school improvement, are typical short-term effects, while other take place on a longer term, like effects on student achievement.

#### ***4. Studies into side effects***

The eventual effect of control mechanisms is also influenced by unintentional side effects. Fortunately, more and more specific research is being carried out into possible side effects of school inspections and performance indicators. It has already been pointed out above that this primarily concerns intended strategic behaviour (such as gaming, window dressing and deception) and unintended strategic behaviour (such as 'proceduralisation', teaching to inspection, tunnel vision and myopia).

##### ***4.1 Side effects of school inspections***

There exists some empirical evidence of strategic behaviour during inspection visits, especially on 'window dressing' and 'gaming the inspection'. Case, Case and Catling (2000), for example, conclude that teachers tend to put on an act during inspection visits. Fitz-Gibbon and Stephenson (1996) demonstrate, via a questionnaire-based investigation among school principals, that a great deal of investments have been made in preparations for inspection visits, amounting to an average

of almost 1,200 English pounds per school. These investments appear to pay off because the researchers actually detected a positive link between the amount of money that is spent on preparing and the assessments by the Inspectorate.

In addition, inspection visits may also indulge in unintended strategic behaviour, such as ‘proceduralism’, ‘teaching to inspection’, ‘tunnel vision’, ‘bureaucratisation’, ‘sub-optimisation’ and ‘isomorphism’ (see Table 3). In 1995, Hargreaves reached the conclusion that ‘inspection will result in an inhibition or diversity and innovation’ and many have subsequently agreed. However, we did not find proper empirical evidence on these types of side effects. There are only indications, such as the fact that it has transpired that recently developed instruments for school self-evaluation and quality assurance often become copies of instruments of inspectorates or written with the inspection frameworks in mind.

A series of other studies on side effects focus on the unintentional effects of inspection visits on the frame of mind and behaviour of the actors involved, particularly stress. For example, there exists a series of booklets by practical experts with lively descriptions of inspection visits and typical behaviour of inspectors. In general, they tend to point towards the conclusion that inspection visits primarily have a detrimental effect on schools because they cause stress among teachers and management teams.

Empirical research into stress and other side effects generally uses satisfaction questionnaires. These are implicitly based on the assumption that a greater degree of dissatisfaction implies the existence of undesirable side effects. As stated above, these studies show that there is general satisfaction about inspection visits, but they also demonstrate that the visit implies a considerable time investment and this usually leads to stress among teachers and school principals (see for example Learmonth, 2000; Cullingford, 1999; and Fitz-Gibbon et al., 1996, 1999).

#### ***4.1 Side effects of performance indicators***

Most research on side effects focus on side effects of performance indicators, in particularly strategic behaviour of schools. Recently, Jacob and Levitt (2003) demonstrated that fraud is

committed with regard to test scores at no fewer than 4 to 5% of the primary school classes in Chicago each year. Jacob and Levitt reach this conclusion on the basis of an empirical analysis of fluctuations in test scores over the years and unconventional answer patterns in pupils' tests. In order to implement the latter, unique approach, the authors developed an algorithm with which they were able to identify three types of unconventional answer patterns, namely (1) blocks of identical answers within a single class, (2) unconventional correlation patterns between answers of pupils in a single class and (3) unconventional answer patterns of individual pupils (such as correct answers to very difficult and incorrect answers to very easy questions). Jacob and Levitt also demonstrate that the schools at which there are unconventional answer patterns, are also the schools at which the fluctuations in test scores have been high over the years. With regard to the ultimate identification of the fraudulent classes (the 4 to 5%), the authors applied fairly strict norms, i.e. the figures only pertain to schools with both a lot of unconventional answer patterns and seriously deviant test scores in the year in question. According to the authors, the 4 to 5% is therefore an underestimation of the actual extent of the fraud, certainly if we consider that the authors only studied a couple of forms of fraud affecting pupils' tests. The authors also conclude that fraudulent behaviour is largely the effect of public performance indicators and that the fraud among teachers increases significantly the greater the incentives are to generate favourable pupil results.

Publicising student achievement or test scores not only results in fraud, but also has another side effect, namely the exclusion of larger numbers of pupils from tests. Four relatively recent working papers (Jacob, 2002; Figlio and Getzler, 2002; Cullen and Reback, 2002; Hanushek and Raymond, 2004) demonstrate convincingly that the introduction of the public accountability mechanisms, which are part of the 'No Child Left Behind Act' (an American program of educational reforms introduced in 2001), has led to an increase in the number of pupils not participating in the relevant tests. Although the studies differ as regards the methodology used and relate to different states, the findings and conclusions clearly correspond. The studies not only show that 'school accountability' results in the exclusion of more pupils from tests (referred to as 'reshaping the test pool' by Figlio and Getzler, 2002) but also that this phenomenon occurs more

often at weak schools and at schools with a lot of disadvantaged pupils (the so-called 'high-poverty schools').

Moreover, Jacob (2002) demonstrates that the publication of test scores has led to an increase in the number of repeaters in the years before the tests. In addition, there is a suspicion that the publication of test scores results in higher dropout rates (see for example Figlio and Getzler, 2002), albeit that the latter has – to our knowledge – never properly been empirically demonstrated.

A recent study by Canton and Webbink (2004) also examines possible intended strategic behaviour by schools. The authors conclude, partly on the basis of the above-mentioned international research, that (intended) strategic behaviour by schools (*gaming*) is a reality in case of performance incentives such as accountability (p. 36). In addition to this conclusion, Canton and Webbink (2004, p. 40) also link strategic behaviour with the possible unreliability of quality measurements:

*'With regard to stimulating the achievements of schools via accountability, there is a dilemma. On the one hand, league tables can provide a powerful incentive on the basis of school achievements for improvements in the quality of education. On the other hand, recent studies show that even league tables that have been constructed on the basis of 'state of the art' techniques are unreliable. Detrimental stimuli are just waiting to happen.'*

Scientific research into unintended strategic behaviour has focused mainly on 'teaching to the test'.

Different scientists have tried to put this phenomenon into operation and quantify it. In almost all cases, they opt to demonstrate discrepancies between learning outcomes in subjects which are and subjects which are not tested. The fact that this discrepancy has clearly increased with the introduction of public performance indicators in a number of American states is, according to Jacob (2002), proof of the existence of this undesirable side effect. Sturman (2003) chose to investigate the phenomenon of 'teaching to the test' using a questionnaire submitted to teachers. On the basis of this survey, Sturman concludes that 'teaching to the test' indeed takes place but that teachers interpret this in totally different ways. The author does not find any additional proof of the effectiveness of the strategies used, like practising tests or focussing on the central test subjects.

A study by Wiggins and Tymms (2000) shows that there are several other forms of unintended strategic behaviour. The authors used a questionnaire submitted to English and Scottish principals and teachers to investigate the extent to which publication of public performance indicators results in undesirable side effects. In regions in which public performance indicators were published, the researchers claim there was also:

- (a) a greater conflict between aims and targets,
- (b) a concentration on targets at the expense of other important objectives,
- (c) a narrowing of the effect on the curriculum,
- (d) a concentration of resources on borderline children, and
- (e) an increasing blame culture.

Although one has to wonder whether these established effects can all be attributed to the publications of public performance indicators, the three middle findings do refer to unintended strategic behaviour by schools.

## ***5. Conclusion and discussion***

### ***5.1 Conclusion***

The overview of studies into effects and side effects of control mechanisms in education presented above enables us to draw a number of conclusions. The main conclusion is that the studies do not provide us with a clear answer to the question of whether school inspections and performance indicators have causal effects. It is not only methodologically difficult to demonstrate causal effects, the methodology used also appears to have a strongly determinative effect on conclusions concerning the extent and direction of the effects. On the basis of research into the satisfaction of the parties involved in the inspection visit, we can conclude that inspection visits probably are effective. After all, almost all these studies show that on average around 80% of the parties directly involved (teachers, school principals) are generally satisfied with school inspections and are of the opinion that these visits are useful. According to the various studies, behaviour and changes in

school improvement are influenced less frequently. Nevertheless, these effects seem to occur at between 50 and 70% of the schools. However, these changes do not lead to an improvement in the educational achievements, in any case not in the short term at English secondary schools. In contrast, some researchers even found a negative effect of school inspections on student achievement.

Research into effects of the publications of public performance indicators is currently limited to research into the effects on choice behaviour of parents and pupils (and teachers in a few cases). The results of these studies are more unambiguous. They lead us, in any case, to draw the conclusion that, when choosing schools, parents and pupils take little notice of published performance indicators such as the school report card. Whether this takes place consciously or is due to unfamiliarity with the information on quality, is not known. On the other hand, principals and teachers have the opinion that performance indicators are important and there exists some evidence that accountability has a positive impact on student achievement. To put it differently, there exist no evidence that school report cards matter, but there exist some evidence that other elements of the accountability system lead to positive results on school improvement and student achievement.

A third conclusion concerns the occurrence of side effects of control mechanisms in education. A number of the studies discussed clearly refer to the existence of these side effects, particularly side effects of performance indicators. Most studies on these side effects prove that strategic behaviour of schools exist, like 'window dressing', 'teaching to the test', cheating in connection with test scores and 'reshaping the test pool'. On the basis of studies on side effects of school inspections, we have to conclude that the occurrence of stress is a clear side effect of inspection visits. Unfortunately, very little research has, as yet, been carried out on strategic behaviour of school inspections, like 'teaching to inspection', 'suboptimisation' and 'isomorphism'. From a theoretical point of view, the occurrence of these side effects is plausible, but lacks empirical evidence.

## *5.2 Discussion*

The problem with the existing studies into the effects and side effects of control mechanisms is that (a) the findings vary quite considerably, (b) the research methodology varies substantially and is not always of high quality and (c) the findings appear to be linked closely to the research methodology used. In addition, a lot of intended (side) effects have not or have scarcely been assessed. The recent increase in research into (side) effects is therefore a very welcome development.

More research into the (side) effects of school inspections and performance indicators is recommended but, at the same time, requires improved methodology. Exposing a causal relationship, which is in effect our intention, makes considerable demands of the research design. We suggest to use research designs which are tailored specifically for the assessment of causal relationships. Ideally, inspectorates use a traditional experimental design, including a set of totally randomised school inspections, in order to study the effect and side effects of school inspections. When this is not possible, other types of variation in inspections are suggested, like for example random variation in types or numbers of inspections and variation in inspection arrangements between geographical regions. Similar methods can be used in order to measure effects of performance indicators, as for example the variation in systems and/or indicators between regions and in time. In addition, we recommend, with regard to the design interpretation, that researchers reflect properly on which (side) effects are going to be assessed, whether strong or weak schools are primarily going to be involved and on which time period the (side) effects should reasonably become evident.



## ***Literature***

Bosker, R., & Scheerens, J. (1999) Openbare prestatiegegevens van scholen; nuttigheid en validiteit. In: *Pedagogische Studiën*, 76, 61-73.

[Public performance data on schools; usefulness and validity. In: *Educational Studies*, 76]

Brimblecombe, N., Ormston, M., & Shaw, M. (1995) Teachers' perceptions of school inspections: a stressful experience. In: *Cambridge Journal of Education*, 25(1), 53-62.

Brimblecombe, N., Shaw, M., & Ormston, M. (1996) Teachers' intention to change practice as a result of Ofsted school inspections. In: *Educational Management & Administration*, 24(4), 339-354.

Canton, E., & Webbink, D. (2004) *Prestatieprikkelers in het Nederlandse Onderwijs: Wat kunnen we leren van recente buitenlandse ervaringen?* Den Haag: Centraal Planbureau.

[*Performance incentives in Dutch education: What can we learn from recent experiences abroad?*]

Carpenter, B., & Stoneham, C. (1994) Inspection effectiveness: an analysis of an Ofsted inspection. In: *British Journal of Special Education*, 21(2), 70-72.

Case, P., Case, S., & Catling, S. (2000) Please show you're working; a critical assessment of the impact of Ofsted inspection on primary teachers. In: *British journal of sociology of education*, 21(4), 605-621.

Chapman, C. (2001) Changing classrooms through inspections. In: *School leadership and management*, 21(1), 59-73.

Cook, T.D. (2002) Randomized Experiments in Educational Policy Research: a critical examination of the reasons the educational evaluation community has offered for not doing them. In: *Educational Evaluation and Policy Analysis*, 24(3), 175-199.

Cullen, J.B., & Reback, R. (2002) *Tinkering towards accolades: school gaming under a performance accountability system*. Working paper. University of Michigan.

Cullingford, C., & Daniels, S. (1996) *Inspecting Her Majesty's Inspectors: should social science and social policy cohere?* Paper presented at the European Conference on Educational Research, Seville, Spain.

Cullingford, C. (1999) *An inspector calls; Ofsted and its effect on school standards*. Londen: Kogan Page.

Doolaard, S., & Karstanje, P. (2001) Gebruik van publieke prestatie-indicatoren voor schoolverbetering. In: Dijkstra, A.B., Karsten, S., Veenstra, R., & Visscher, A.J. (Eds.), *Het oog der natie: scholen op rapport. Standaarden voor de publicatie van schoolprestaties* (pp.155-172) Assen: Van Gorkum.

[Use of public performance indicators for school improvement. In Dijkstra et al., *The eye of the nation: schools on report*]

Ehren, M., Leeuw, F.L., & Scheerens, J. (2005) On the impact of the Dutch Educational Supervision Act. Analyzing assumptions concerning the inspection of primary education. In: *American Journal of Evaluation*, 26(1), 60-76.

Emmelot, Y., Karsten, S., Ledoux, G., & Vermeulen, A. (2004) *Ervaringen met het vernieuwde onderwijstoezicht*. Amsterdam: SCO/Kohnstamm Instituut.

[*Experiences with the new-style education school inspections*. Report by the SCO Kohnstamm Institute for Educational Research, Amsterdam]

Eurydice (2004) *Evaluation of Schools providing Compulsory Education in Europe*. Brussels: Eurydice Internet (<http://www.eurydice.org>)

Figlio, D.N., & Getzler, L.S. (2002) *Accountability, Ability and Disability: gaming the system*. NBER Working Paper.

Fitz-Gibbon, C.T., & Stephenson-Forster, N.J. (1996) *Inspecting her Majesty's Inspectors: should social science and social policy cohere?* Paper presented at the European Conference on Educational Research, Spain.

Fitz-Gibbon, C.T. (1997) *Feasibility studies for a national of value-added indicators*. Londen: SCAA.

Fitz-Gibbon, C.T., & Stephenson-Forster, N.J. (1999) Is Ofsted helpful? In: Cullingford, C., *An Inspector calls*. Londen: Kogan Page Limited.

Frederiks, M.M.H. (1996) *Beslissen over kwaliteit. Verklaring van besluitvorming over het gebruik van onderwijsvisitatierapporten in Nederlandse universiteiten*. Dissertatie, Universiteit Twente. [Deciding on quality. Statement on the decision process with regard to the use of education review reports in Dutch universities. Doctoral thesis, University of Twente]

Goldstein, H., & Spiegelhalter, D.J. (1996) League tables and their limitations: statistical issues in comparisons of institutional performance. In: *Journal of the Royal Statistical Society, Soc A*, 159(3), 385-443.

Goldstein, H. (2001) League tables and schooling. In: *Science in Parliament*, 58(2), 4-5.

Gray, C., & Gardner, J. (1999) The impact of school inspections. In: *Oxford Review of Education*, 25(4), 455-468.

Hanushek, E.A., & Raymond, M.E. (2004) *Does school accountability lead to improved student performance?* NBER-working paper.

Hargraves, D.H. (1995) Inspection and School Improvement. In: *Cambridge Journal of Education*, 25(1), 53-61.

Jacob, B.A. (2002) *Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago public schools*. NBER working paper.

Jacob, B.A., & Levitt, S.D. (2003) Rotten Apples, An Investigation of the Prevalence and Predictors of Teacher Cheating. In: *The Quarterly Journal of Economics*, 118(3), 843-877.

Janssens, F.J.G. & Leeuw, F.L. (2001) *Scholen maken verschil. Maar het ene verschil is het andere niet*. Utrecht: Inspectie van het Onderwijs.

[Schools do make a difference. But there are various types of differences. Utrecht: Inspectorate of Education]

Janssens, F.J.G., & Visscher, A.J. (2004) Naar een kwaliteitskaart voor het primair onderwijs. In: *Pedagogische Studiën*, 81, 371-383.

[Towards a school report card for primary education. ]

Jeliazkova, M., & Westerheijden, D. (2000) *Het zichtbare eindresultaat; doorwerking van onderwijsvisitaties in hogeronderwijsinstellingen*. Den Haag: Algemene Rekenkamer.  
[The visible final result; effects of educational visitations in institutes of higher education.  
Netherlands Court of Audit]

Kane, T.J., & Staiger, D.O. (2001) *Improving school accountability measures*. NBER working paper.

Karsten, S., & Visscher, A. (2001) Ervaringen met het openbaar maken van schoolprestaties in Engeland en Frankrijk. In: Dijkstra, A.B., Karsten, S., Veenstra, R., & Visscher, A.J. (Eds.), *Het oog der natie: scholen op rapport. Standaarden voor de publicatie van schoolprestaties* (pp. 36-53) Assen: Van Gorkum.  
[Experiences with publishing school performances in England and France. In: Dijkstra et al., *The eye of the nation: schools on report*]

Klein, S.P., Hamilton, L.S., McCaffrey, D.F., & Stencher, B.M. (2000) *What do test scores in Texas tell us?* Issue paper, Rand Education.

Learmonth, J. (2000) *Inspection. What's in it for schools?* London/New York: Routledge/Falmer.

Lazear, E.P. (2004) *Speeding, Tax Fraud, and Teaching to the Test*. Paper Stanford University.

Leeuw, F.L. (2000) Onbedoelde neveneffecten van outputsturing, controle en toezicht? In: Raad voor Maatschappelijke Ontwikkeling, *Aansprekend burgerschap; de relatie tussen de organisatie van het publieke domein en de verantwoordelijkheid van de burgers* (pp. 151-171) Den Haag: Raad voor Maatschappelijke Ontwikkeling.  
[Unintentional side effects of output manipulation, control and school inspections? In: *Responsible citizenship; the relationship between the organisation of the public domain and the responsibility of the citizens*]

Leeuw, F.L. (2003) *Challenges for national quality assurance agencies: a view from the evaluation world*. Paper presented at the ENQA-conference, March.

Macnab, D. (2004) Hearts, minds and external supervision of schools: direction and development. In: *Educational Review*, 56(1), 53-64.

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

Meijer, A. (2004) *Vreemde ogen dwingen. De betekenis van internet voor maatschappelijke controle in de publieke sector*. Den Haag: Boom Juridische uitgevers.  
[The authority of external forces. The significance of the Internet for social control in the public sector]

Meyer, M.W., & Gupta, V. (1994) The performance paradox. In: *Research in Organizational Behavior*, 16, 309-369.

Oosterbeek, H. (2001) *Voortschrijdend inzicht*. Oratie Universiteit van Amsterdam.  
[Advancing insight. Lecture at the University of Amsterdam]

Ouston, J., Fidler, B., & Earley, P. (1997) What do Schools do after Ofsted school inspections, or before? In: *School leadership & Management*, 17(1), 95-104.

Rosenthal, L. (2004) Do school inspections improve school quality? Ofsted inspections and school examination results in the UK. In: *Economics of Education Review*, 23, 143-151.

Scalon, M. (1999) *The impact of Ofsted inspections*. National Foundation of Educational Research.

Shadish, D.T., Cook, W. R., & Campbell, T.D. (2002) *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.

Shaw, I., Newton, D.P., Aitkin, M., & Darnell, R. (2003) Do Ofsted inspections of secondary education make a difference tot GCSE results? In: *British Educational Research Journal*, 29(1), 63-75.

Smith, P. (1993) Outcome-related performance indicators and organizational control in the public sector. In: *British Journal of Management*, 4, 135-151.

Smith, P. (1995) On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18, 277-310

Sturman, L. (2003) Teaching to the test: science or intuition? In: *Educational Research*, 45(3), 261-273.

Visscher, A.J. (2004) De effecten van schoolprestatie-publicaties in andere landen. In: *Pedagogische Studiën*, 81, 384-396.

[The effects of school performance publications in other countries.]

Waterreus, I. (2003) *Lessons in teacher pay; studies on incentives and the labor market for teachers*. Doctoral Thesis, Universiteit van Amsterdam.

Wiggins, A., & Tymms, P. (2000) *Dysfunctional effects of public performance indicator systems: a comparison between English and Scottish Primary Schools*. Paper presented at the European Conference on Educational Research, Edinburgh.

Wilcox, B., & Gray, J. (1996) *Inspecting schools: Holding schools to account and helping schools to improve*. Buckingham: Open University Press.

## **TABLES**

*Table 1. Side effects of control mechanisms in education, differentiated into category and type*

<b>SIDE EFFECT (category and type)</b>	<b>EXAMPLES</b>
<b><i>A. Intended strategic behaviour</i></b>	
- strategic behaviour / gaming	<ul style="list-style-type: none"> <li>- window dressing</li> <li>- misrepresentation</li> <li>- fraud and deception</li> <li>- orchestration of peer review</li> </ul>
- side effect at system level	<ul style="list-style-type: none"> <li>- reduced reliability and validity of the assessment</li> </ul>
<b><i>B. Unintended strategic behaviour</i></b>	
- unintended strategic behaviour	<ul style="list-style-type: none"> <li>- formalisation and proceduralisation</li> <li>- teaching to the test / to inspection</li> <li>- tunnel vision</li> <li>- indicator fixation</li> <li>- sub-optimisation</li> <li>- myopia</li> <li>- convergence</li> <li>- ossification</li> </ul>
- side effect at system level	<ul style="list-style-type: none"> <li>- bureaucratisation</li> <li>- isomorphism</li> <li>- sub-optimisation</li> </ul>
<b><i>C. Other side effects</i></b>	
- stress felt by teachers and school principals	<ul style="list-style-type: none"> <li>- reduced support for the assessment</li> <li>- reduced reliability and validity of the assessment</li> </ul>

**Table 2. Instruments, intended effects and possible side effects of school inspections and performance indicators**

<i>Instruments</i>	<i>Intended effects</i>	<i>Possible side effects</i>
Inspection visits	<ul style="list-style-type: none"> <li>- guarantee education quality</li> <li>- guarantee level of compliance</li> </ul>	<ul style="list-style-type: none"> <li>- intended strategic behaviour by schools</li> <li>- unintended strategic behaviour by schools</li> </ul>
Public performance indicators	<ul style="list-style-type: none"> <li>- improve education quality</li> </ul>	<ul style="list-style-type: none"> <li>- other side effects</li> </ul>



### *Notes on Contributors*

Dr. Inge de Wolf works as a researcher at the research institute Scholar of the University of Amsterdam and a coordinating researcher at the Inspectorate of Education in the Netherlands. Her research concentrates on evaluation methodologies and effects and side effects of educational evaluation systems. During the last years, she developed a risk based inspection method for the Dutch Inspectorate. Her recent research at the University of Amsterdam focuses on effects of inspection visits on school improvement and on strategic behaviour ('gaming') of schools.

Prof. Dr. Frans Janssens is professor at the University of Twente and head of the Research & Development Department of the Netherlands' Education Inspectorate. His research program concentrates on 'Inspections and accountability of schools'. During the last decennium, he developed and implemented highly standardized methods and instruments of school inspections. He published on these methods in various articles and is now working on a book about the methodology of school inspections.