

Effecten en neveneffecten van inspecties en prestatie-indicatoren in het onderwijs: Een overzicht van empirische studies

I. F. de Wolf en F. J. G. Janssens

Samenvatting

Dit artikel bevat een overzicht van empirische studies over effecten en neveneffecten van inspecties en prestatie-indicatoren in het onderwijs. Op basis van de besproken studies concluderen we dat causale effecten van inspectiebezoeken lastig zijn aan te tonen, en dat de bevindingen hierover uiteenlopen. De meeste studies concluderen dat inspectiebezoeken een positief effect hebben op schoolbeleid en gedrag van betrokkenen, maar in enkele studies naar effecten op leerprestaties wordt een klein negatief effect gerapporteerd. Een tweede conclusie is dat effecten van openbare prestatie-indicatoren positief lijken, maar geen invloed hebben op schoolkeuze-gedrag. Prestatie-indicatoren, en in iets mindere mate inspectiebezoeken, leiden daarentegen wel tot strategisch gedrag van scholen, zoals indicatorfixatie, 'window dressing' en andere vormen van 'gaming'.

1 Inleiding

Dit artikel geeft een overzicht van empirische studies naar effecten en neveneffecten van twee belangrijke controlemechanismen in het onderwijs, namelijk (1) inspecties en (2) 'accountability'-systemen met openbare prestatie-indicatoren. De Nederlandse situatie staat hierbij centraal, maar zowel het gehanteerde analysekader als de meeste besproken studies zijn relevant voor vergelijkbare (inspectie)organisaties en andere controlemechanismen in het onderwijs.

1.1 Waarom aandacht voor (neven-) effecten van controlemechanismen?

Het is om een aantal redenen relevant zicht te krijgen op effecten en neveneffecten van controlemechanismen in het (Nederlandse) onderwijs. Een eerste belangrijke reden betreft een reeks relatief recente, voornamelijk bui-

tenlandse, empirische studies op dit terrein. Uit deze studies, die verderop besproken worden, blijkt dat het onderwerp complex is, eenduidige conclusies vooralsnog ontbreken en er inmiddels wel een eerste aantal interessante onderzoeksresultaten zijn te melden.

Onderzoek naar (neven)effecten van controlemechanismen is daarnaast vanuit maatschappelijk perspectief relevant. Dit heeft alles van doen met de sterke toename van controlemechanismen (Power (1997) spreekt in dit kader van 'the audit society') en de hieraan gekoppelde discussie over de efficiency en effectiviteit hiervan. Zo wordt regelmatig de vraag gesteld of openbare controlemechanismen in het onderwijs wel werken (zie o.a. Leeuw, 2000; Wiggins & Tymms, 2000), wat het effect van externe controles op onderwijskwaliteit van scholen is (zie bijv. Rosenthal, 2004; Shaw, Newton, Aitkin & Darnell, 2003), wat de voor- en nadelen van alternatieve controlemechanismen met prestatie-indicatoren zijn (zie bijv. Canton & Webbink, 2004; Jacob, 2002; Kane & Staiger, 2001; Wiggins & Tymms, 2000) en welke omvang ongewenste neveneffecten van controlemechanismen hebben (zie Fitz-Gibbon, 1997; Jacob, 2002; Jacob & Levitt, 2003; Leeuw, 2000).

Tot slot is onderzoek naar (neven)effecten van toezicht van belang voor de overheid, in het bijzonder de beleidsmakers en toezichthouders. Zij kunnen de resultaten gebruiken om de kwaliteit en effectiviteit van controlemechanismen te versterken en maatregelen te nemen om eventuele neveneffecten tegen te gaan.

1.2 Onderzoeksvraag en opzet

De belangrijkste twee openbare controlemechanismen in het onderwijs zijn inspecties en accountability-systemen met openbare prestatie-indicatoren. Dit artikel bevat een inventarisatie van empirische studies naar deze twee controlemechanismen. De volgende on-

doelzoeksvraag staat hierbij centraal: Wat is er, in nationale en internationale studies, bekend over de effecten en neveneffecten van inspectiebezoeken en prestatie-indicatoren in het onderwijs?

2 Een analysekader

Alvorens we een antwoord geven op de vraag welke empirische evidentie er bestaat over (neven)effecten van inspectiebezoeken en prestatie-indicatoren, moeten we eerst bedenken naar welke effecten en neveneffecten we op zoek zijn. Kijken we bijvoorbeeld naar het effect op de kwaliteit van het onderwijs op een afzonderlijke school, naar effecten op het niveau van (wets)handhaving voor het gehele bestel of naar ongewenste gevolgen zoals strategisch gedrag van scholen? In deze paragraaf presenteren we een beknopt analysekader, bestaande uit beoogde effecten en mogelijke neveneffecten.

2.1 Beoogde effecten

Inspectiebezoeken hebben over het algemeen tot taak toezicht uit te oefenen op de kwaliteit van scholen, hierover te rapporteren en de kwaliteit te stimuleren (zie o.a. Macnab, 2004). In de Nederlandse Wet op het Onderwijstoezicht (2002) staan deze drie taken ook expliciet vermeld. Een beoogd effect is het “waarborgen” van de onderwijskwaliteit van scholen, oftewel zorgen dat zoveel mogelijk scholen voldoen aan een minimum niveau van onderwijskwaliteit. Een tweede beoogd effect betreft naleving van wet- en regelgeving; de onderwijsinspectie dient ervoor te zorgen dat scholen en onderwijsinstellingen zich aan de wet houden. Een derde beoogd effect van inspectietoezicht is kwaliteitsverbetering. Deze taak is zelfs expliciet in de Wet op het Onderwijstoezicht opgenomen, en een recente reconstructie van de beleidstheorie achter deze wet (Ehren, Leeuw, & Scheerens, 2005) laat ook zien dat inspectiebezoeken naar verwachting leiden tot verbetertrajecten op scholen, welke vervolgens resulteren in kwaliteitsverbetering en een stijging in toegevoegde waarde van scholen.

Accountability-systemen met prestatie-indicatoren beogen deels dezelfde effecten te

bewerkstelligen als inspectietoezicht. Ook in deze systemen gaat het om zoveel mogelijk waarborgen van onderwijskwaliteit en om kwaliteitsverbetering. Publicatie van prestatie-indicatoren leidt er namelijk toe dat (a) scholen zich gaan verantwoorden, (b) ouders beter geïnformeerd zijn en scholen gericht aan kunnen spreken op zwakke punten, (c) de prestatie-indicatoren gebruikt kunnen worden voor schoolkeuze (het zogenaamde “stemmen met de voeten”) en (d) scholen de indicatoren kunnen gebruiken voor ‘benchmarking’ oftewel inzicht verkrijgen in de (relatieve) positionering van de eigen school ten opzichte van andere scholen (zie o.a. Canton & Webbink, 2004; Dudok, 2004; Ehren et al., 2005). Aangenomen wordt dat dit schoolleiders en docenten prikkelt om ervoor te zorgen dat hun school “zo goed mogelijk scoort”.

In Nederland controleert de overheid de kwaliteit van scholen voornamelijk via het toezicht door de Inspectie van het Onderwijs. Daarnaast publiceert de inspectie kwaliteitsinformatie over scholen in de vorm van “kwaliteitskaarten”. Ook de media publiceren regelmatig overzichten van prestaties van scholen, de zogenaamde ‘league tables’. In de ons omringende landen zien we vergelijkbare systemen, al wordt niet overal even zwaar ingezet op publicatie van openbare prestatie-indicatoren door de overheid. In Noord-Amerika is dit laatste wel het geval, hier zijn inspecties weer een minder belangrijk controlemechanisme.

2.2 Ongewenste neveneffecten

Naast directe effecten, veroorzaken controlemechanismen ook neveneffecten. Neveneffecten zijn onbedoelde effecten die gewenst of ongewenst kunnen zijn. Ongewenste neveneffecten zijn het meest interessant, omdat ze in mindering gebracht moeten worden op de effecten en soms in staat zijn deze geheel teniet te doen (Leeuw, 2000). Inmiddels bestaan er vele algemene overzichten van ongewenste neveneffecten (Leeuw, 1996; 2000; Meyer & Gupta, 1996; Smith, 1998), alsook overzichten specifiek voor het onderwijs (Bosker & Scheerens, 1999; Fitz-Gibbon, 1997). In al deze overzichten worden weliswaar verschillende accenten gelegd, maar

bestaat er tegelijkertijd veel overlap in genoemde neveneffecten. Om deze reden wordt hieronder een algemeen overzicht voor inspectiebezoeken en prestatie-indicatoren gepresenteerd. We merken vast op dat we hierbij niet ingaan op de eventuele ernst van de neveneffecten; we verbinden hier dus geen waardering aan, maar volstaan met een eenvoudig overzicht op basis van de literatuur.

Een eerste groep ongewenste neveneffecten betreft “bewust strategisch gedrag” van scholen, oftewel ‘gaming’. De meest bekende vorm van bewust strategisch gedrag bij inspectiebezoeken is ‘window dressing’. Hieronder verstaan we het tot stand brengen van pro- en reactieve arrangementen in een school die alleen tot stand worden gebracht om beter beoordeeld te worden. Het gaat hier dus om maatregelen met een kortetermijneffect, die vooral gericht zijn op het “oppoetsen” van scholen. Soms gaat strategisch gedrag bij inspectiebezoeken en prestatie-indicatoren zo ver dat er sprake is van misrepresentatie of zelfs van fraude en bedrog (o.a. Smith, 1993). We moeten bij dit laatste denken aan het verstrekken van “valse” prestatie-indicatoren, het onterecht uitsluiten van leerlingen bij belangrijke toetsen (‘reshaping the testpool’) en het “helpen” van de leerlingen bij toetsen of examens. Ook het onterecht ziekmelden van bepaalde docenten tijdens het inspectiebezoek valt hieronder.

Strategisch gedrag van scholen kan ook onbewust zijn. Zo kunnen controlemechanismen onbedoeld een sturende werking hebben op de inhoud en organisatie van het onderwijs. Vaak gaat het hier om een (vaak onbewuste) eenzijdige nadruk op de aspecten waarop beoordeeld wordt. In de literatuur wordt in dit kader gesproken van ‘teaching to the test’ (of ‘teaching to inspection’), indicatorfixatie en ‘tunnel vision’. Een aardige theoretische uiteenzetting over teaching to the test is te vinden in Lazear (2004). Hij stelt eenvoudigweg dat alle scholen prikkels kennen om tot deze vorm van strategisch gedrag over te gaan, en beargumenteert dat er in het beleid tot reductie van dit verschijnsel een verschil gemaakt moet worden tussen scholen met moeilijk lerende leerlingen en scholen met makkelijk lerende leerlingen. Fitz-Gibbon (1997) gaat uitgebreider in op

indicatorfixatie en tunnelvision en wijst erop dat scholen als gevolg hiervan overmatig de nadruk leggen op examenprestaties, hetgeen weer leidt tot suboptimalisatie. Een ander mogelijk gevolg is ‘myopia’, dat wil zeggen concentratie op kortetermijnoplossingen ten koste van het langetermijnbeleid (Smith, 1993). Mogelijk leidt toezicht er ook toe dat scholen tot de middenmoot proberen te behoren (convergentie) en dat ze minder geneigd zijn met nieuwe en innovatieve methoden te experimenteren (ossificatie) (zie ook Hargreaves, 1995; Smith, 1993). Als deze neveneffecten maar vaak genoeg optreden, zal daarnaast sprake zijn van isomorfisme, dat wil zeggen dat alle scholen op elkaar gaan lijken. Een andere vorm van onbewust strategisch gedrag is formalisering, proceduralisering en een toenemende bureaucratisering.

Een tweede groep ongewenste neveneffecten heeft geen betrekking op strategisch gedrag, maar op andere (ongewenste) zaken die direct verband houden met controlemechanismen. Een veelgenoemd voorbeeld is het optreden van stress tijdens het inspectiebezoek. Deze stress wordt soms gezien als ongewenst neveneffect, omdat het druk en mogelijk ziekteverzuim veroorzaakt voor docenten en schoolleiders (zie o.a. Cullingford, 1997) en – in zijn meest extreme vorm – tevens de validiteit en de betrouwbaarheid van de beoordeling negatief kan beïnvloeden. Een ander mogelijk ongewenst neveneffect van de twee controlemechanismen is het ontstaan van een “zesjescultuur”. Hiermee doelen we op het feit dat scholen bij goede prestaties “achterover gaan leunen” en minder actief gaan werken aan verbetering van de onderwijskwaliteit. Janssens en Leeuw (2001) wijzen tot slot op nog een ander onbedoeld neveneffect op stelselniveau dat samenhangt met de publicatie van prestatie-indicatoren, namelijk de bijdrage die dit levert aan de marktwerking in het onderwijs, met alle voor- en nadelen die hieraan verbonden zijn.

Samenvattend kunnen we stellen dat er, in ieder geval in theorie, vele soorten ongewenste neveneffecten op kunnen treden. Vele hiervan hebben betrekking op strategisch gedrag van scholen, gedrag dat bewust of

Tabel 1

Neveneffecten van controlemechanismen, onderscheiden naar categorie en type

| Neveneffect (categorie en type) | Voorbeelden |
|-----------------------------------|---|
| Strategisch gedrag | |
| Bewust strategisch gedrag/gaming: | Window dressing Misrepresentatie Fraude en bedrog Orkestratie van peer review Reshaping the testpool |
| Onbewust strategisch gedrag | Formalisering en proceduralisering Teaching to the test Teaching to inspection Tunnel vision Indicatorfixatie Suboptimalisatie Myopia Convergentie Ossificatie Isomorfisme |
| Overige neveneffecten | |
| Stress docenten en schoolleiders | Toename (werk)druk Gezondheidsklachten, ziekte |
| Ontstaan zesjescultuur | Achteroverleunen goede scholen |
| Bijdrage aan marktwerking | |

onbewust kan zijn. Daarnaast bestaan er mogelijk neveneffecten als stress en het ontstaan van een zesjescultuur. In Tabel 1 staat een opsomming van de verschillende categorieën en typen neveneffecten, inclusief voorbeelden.

2.3 Concluderend

De beoogde effecten en de verwachte neveneffecten van inspectiebezoeken en openbare prestatie-indicatoren zijn centrale elementen voor ons conceptuele model. Kort samengevat komt het model erop neer dat we verwachten dat inspectiebezoeken een positief effect hebben op het waarborgniveau van onderwijskwaliteit, ze de naleving vergroten en leiden tot een verbetering van de kwaliteit van het onderwijs. Voor prestatie-indicatoren geldt dat ze ook leiden tot kwaliteitsverbetering en waarborgen wat betreft het minimumniveau van onderwijskwaliteit, zo is de verwachting. Een direct effect op naleving ligt hier minder voor de hand, tenzij de indicatoren expliciet betrekking hebben op naleving. Bij beide controlesystemen worden dezelfde of sterk vergelijkbare neveneffecten verwacht. Het gaat hier om bewust en onbewust strategisch gedrag (zoals window dressing, gaming, indicatorfixatie en teaching to the test) en het mogelijk ontstaan van een zesjescultuur. Op stelselniveau kan dit leiden tot steeds meer

gelijkvormigheid in het onderwijs (isomorfisme). Voor inspectiebezoeken geldt verder dat deze naar verwachting stress opleveren, terwijl publicatie van prestatie-indicatoren mogelijk bijdraagt aan verdere marktwerking in het onderwijs. In Tabel 2 staat het conceptuele model schematisch weergegeven.

Of er ook een empirische basis bestaat voor de verwachte effecten en neveneffecten, wordt duidelijk in de rest van dit artikel.

3 Onderzoek naar effecten

Empirische studies naar effecten van inspectiebezoeken en openbare prestatie-indicatoren zijn schaars, al is de laatste jaren wel sprake van forse groei. Hieronder geven we een overzicht van de belangrijkste onderzoeken op dit terrein.

3.1 Effecten van inspectiebezoeken

Onderzoek naar effecten van inspectiebezoeken is divers van aard, in die zin dat studies sterk verschillen in de manier waarop effecten gemeten worden. We kunnen hier onderscheid maken in effecten op de tevredenheid van docenten en schoolleiders, effecten op hun gedrag, effecten op beleid en effecten op leerprestaties.

Tabel 2

Instrumenten, beoogde effecten en mogelijke neveneffecten van inspectiebezoeken en prestatie-indicatoren

| Instrumenten | Beoogde effecten | Mogelijke neveneffecten |
|--------------------------------|--------------------------------|--|
| Inspectiebezoeken | Waarborg onderwijskwaliteit | Bewust strategisch gedrag van scholen: window dressing, teaching to inspection |
| | Garantie nalevingniveau | Onbewust strategisch gedrag van scholen; isomorfisme |
| | Verbetering onderwijskwaliteit | Stress Zesjescultuur |
| Openbare prestatie-indicatoren | Waarborg onderwijskwaliteit | Bewust strategisch gedrag van scholen: indicatorfraude, teaching to the test |
| | Verbetering onderwijskwaliteit | Onbewust strategisch gedrag van scholen; indicatorfixatie, suboptimalisatie |

Effecten op tevredenheid van docenten en schoolleiders

In satisfactieonderzoek staat tevredenheid met het toezicht centraal. Een grotere tevredenheid wordt hierbij veelal geïnterpreteerd als een groter gepercipieerd effect van het inspectiebezoek. Zo onderzochten Gray en Gardner (1999) 'the impact of school inspections' door scholen in Noord-Ierland te vragen naar hun ervaringen. Op basis van het feit dat de meeste scholen het inspectiebezoek als professioneel en ondersteunend ervaren, concluderen de onderzoekers dat inspectiebezoeken effectief zijn. Daarbij laten ze niet onvermeld dat de scholen op specifieke punten, zoals de benodigde tijdsinvestering, een stuk minder tevreden zijn. Vergelijkbare studies zijn in Engeland gedaan door Fitz-Gibbon en Stephenson-Forster (1996, 1999), maar zij komen tot iets andere resultaten. Volgens deze laatste auteurs is het nog altijd onduidelijk wat de precieze effecten van toezicht zijn, en zijn schoolleiders noch tevreden noch ontevreden over inspectiebezoeken.

In Nederland wordt relatief veel satisfactieonderzoek gedaan. Over het algemeen is dit type onderzoek niet direct bedoeld om zicht te krijgen op mogelijke effecten van toezicht, maar om verbeterpunten in kaart te brengen. Zo hebben Emmelot, Karsten, Ledoux en Vermeulen (2005) recentelijk een uitgebreide studie gedaan naar ervaringen met onderwijstoezicht. Uit deze studie komt naar voren dat een ruime meerderheid van de schoolleiders (ruim 80%) tevreden is over het inspectietoezicht en van mening is dat de bezoeken in positieve zin bijdragen aan de kwaliteit van scholen. Vergelijkbare bevindingen komen uit onderzoek gedaan door of in

opdracht van de Inspectie van het Onderwijs (Heliview, 1999; IvhO, 2004). Overigens geldt voor al deze studies dat er tevens een positieve samenhang gevonden wordt tussen het inspectieoordeel en de tevredenheid met het toezicht.

Effecten op gedrag van docenten en schoolleiders

De enkele studies waarin is onderzocht of inspectiebezoeken het didactisch handelen van docenten beïnvloeden, richten zich op de vraag of de beoordeling van lessen leidt tot veranderingen in dit didactisch handelen. Brimblecombe, Shaw en Ormston (1996) tonen aan dat 38% van de docenten van wie een les is beoordeeld, vlak na een inspectiebezoek geneigd is tot veranderingen. Dit betreft met name veranderingen in de instructie en klassenorganisatie. De geneigdheid tot verandering neemt toe naarmate de docent hogerop is geplaatst in de schoolorganisatie. In andere onderzoeken wordt meestal een iets sterkere relatie gevonden. Zo concludeert Chapman (2001, pp. 63-64) dat "about 50% of teachers agreed or strongly agreed that Ofsted inspection leads to changes in classroom practice" en dat "58% of teachers thought that Ofsted is a useful tool for school improvement." Zowel in de studie van Chapman (2001) als in de studies van Brimblecombe e.a. (1995, 1996) wordt geconcludeerd dat de relatie tussen docent en inspecteur sterk bijdraagt aan de mate waarin les- en inspectiebezoeken effect sorteren. Chapman (2001) suggereert dat *feedback* hier het sleutelbegrip is (p. 69). Hierbij dient overigens wel opgemerkt te worden dat deze studies betrekking hebben op de Engelse situatie waarin de inspectie tevens rapporteert op het niveau van

individuele docenten (in tegenstelling tot de Nederlandse inspectie).

Onderzoek naar effecten op gedrag van schoolleiders vindt, net als effecten op gedrag van docenten, over het algemeen plaats via vragenlijsten en interviews met betrokkenen. Ook uit deze studies, waarvan overzichten zijn te vinden in Learmonth (2000) en Ouston, Fidler en Earley (1997), blijkt dat inspectiebezoeken leiden tot gedragsveranderingen van een ruime meerderheid van schoolleiders. Over het algemeen rapporteren jongere, meer onervaren schoolleiders meer veranderingen als gevolg van inspectiebezoeken dan oudere, ervaren schoolleiders (zie ook Fitz-Gibbon & Stephenson-Forster, 1996).

Effecten op verbetering van schoolbeleid

Een andere onderzoekstraditie rond effecten van inspectiebezoeken op onderwijskwaliteit, richt zich op de relatie tussen het inspectiebezoek en veranderingen in het beleid van scholen. Hiervoor worden meestal schoolleiders en docenten ondervraagd. Een aardig overzicht van dit type onderzoek is opgenomen in de literatuurstudies van Learmonth (2000) en Ouston e.a. (1997). Op basis van deze studies wordt geconcludeerd dat inspectiebezoeken leiden tot verbeteringen in het beleid op scholen. Met betrekking tot het type beleidsverandering, de omvang en gevolgen ervan zijn echter weinig eenduidige conclusies te trekken. Alhoewel vrijwel alle besproken studies aantonen dat het management van scholen naar aanleiding van inspectiebezoeken veranderingen doorvoert, variëren conclusies over de mate waarin dit gebeurt van *enigszins* tot *zeer sterk* (Learmonth, 2000; Ouston et al., 1997). Mogelijk worden effecten van inspectiebezoeken op beleidsveranderingen ook sterk beïnvloed door andere zaken, zoals de context van en de cultuur op een school (zie ook Chapman, 2001). In de meeste onderzoeken wordt echter onvoldoende of geen rekening gehouden met dergelijke interveniërende variabelen.

In Nederland is nog weinig onderzoek gedaan naar de relatie tussen inspectietoezicht en het beleid van scholen. Een studie die hier wel op ingaat, is de eerder genoemde studie

naar ervaringen met het onderwijstoezicht (Emmelot et al., 2004). Uit dit onderzoek komt naar voren dat schoolleiders vinden dat de inspectiebezoeken in enige mate bijdragen aan verbetering van de kwaliteit van scholen. Schoolleiders in het basis- en voortgezet onderwijs zijn van mening dat het inspectierapport (rapport met de bevindingen van een inspectiebezoek) van nut is voor kwaliteitsverbetering, in het bijzonder wat betreft (a) de kwaliteitsdomeinen zoals gedefinieerd door de inspectie, (b) aanwijzingen voor leraren en (c) de wijze waarop kwaliteitszorg verbeterd kan worden (Emmelot et al., 2004, pp. 123 en 129).

Effecten op leerprestaties

In Engeland is sinds 1999 een reeks studies verschenen naar de effecten van inspectiebezoeken op de leerprestaties van leerlingen. Verschillende onderzoekers hebben getoetst of inspectiebezoeken van de Engelse inspectie, Ofsted, resulteren in een verbetering van de examenresultaten op middelbare scholen. De analysemethode varieert hierbij van zeer eenvoudig (Cullingford & Daniels, 1999; Wilcox & Gray, 1996) tot complex (Rosenthal, 2004; Shaw et al., 2003). Alhoewel alleen in de analysemethode van Rosenthal (2004) sprake is van een goede controlegroep, is het opvallend dat alle auteurs constateren dat er sprake is van een negatief effect van inspectiebezoeken op examenresultaten, althans in het jaar van het inspectiebezoek. Mogelijk zijn stress en de tijd die nodig is voor een grondige voorbereiding van het inspectiebezoek oorzaken van dit negatieve effect, aldus de auteurs. Eventuele effecten op de langere termijn zijn helaas niet onderzocht, effecten van inspectiebezoeken op leerprestaties in andere onderwijssectoren evenmin.

In 1999 is, in een onderzoek dat zich ook richtte op effecten van Ofsted-bezoeken, eenvoudigweg aan de betrokken schoolleiders en docenten gevraagd of inspectiebezoeken resulteren in betere leerprestaties van leerlingen (Scanlon, 1999). Met name op zwakke scholen antwoordde men instemmend: hier meende een ruime meerderheid (60% van de docenten en 79% van de schoolleiders) dat inspectiebezoeken tot betere leerresultaten

leiden. Van de overige scholen was een derde van de schoolleiders en een kwart van de docenten deze mening toegedaan (p. 47).

3.2 Effecten van openbare prestatie-indicatoren

Veel internationaal onderzoek naar (openbare) prestatie-indicatoren (de zogenaamde league tables en kwaliteitskaarten) beperkt zich tot analyses rond de betrouwbaarheid van ranglijsten (zie bijv. Goldstein, 2000; Kane & Staiger, 2001; Klein, Hamilton, McCaffrey, & Stencher, 2000; Oosterbeek & Webbink, 2001; Wiggins & Tymms, 2000). De centrale conclusie in deze studies is dat ranglijsten veelal onvoldoende betrouwbaar zijn om enig effect te kunnen sorteren. Ondanks deze vrij vérgaande conclusie, bestaat er tevens een reeks studies naar mogelijke effecten van prestatie-indicatoren. In deze paragraaf wordt een overzicht van deze studies gegeven, waarbij een onderscheid gemaakt wordt naar effecten op keuzegedrag, 'voice', schoolverbetering en leerprestaties.

Effecten van prestatie-indicatoren op keuzegedrag

Onderzoek naar het schoolkeuzegedrag van ouders en hun kinderen in Engeland en Frankrijk toont niet duidelijk aan dat het publiceren van schoolkwaliteitsgegevens van invloed is op schoolkeuzegedrag van ouders en leerlingen (voor een overzicht van deze studies, zie Karsten & Visscher, 2001). Ouders laten zich over het algemeen weinig gelegen liggen aan openbare prestatie-indicatoren. Volgens Janssens en Visscher (2004) hangt dit samen met de volgende factoren:

- Een behoorlijk deel van de ouders heeft om verschillende redenen geen toegang tot deze informatie, of is niet op de hoogte van het bestaan ervan.
- Het gebruik van de beschikbare informatie vereist de nodige expertise waarover de gemiddelde ouder niet beschikt, of de informatie is moeilijk naar de eigen situatie te vertalen.
- De keuzemogelijkheden zijn beperkt (ruimtelijke grenzen, een gering of weinig divers scholenaanbod, formele en materiële beperkingen).
- De onderwijsprestaties van scholen vor-

men slechts een van de factoren waarop ouders zich bij hun schoolkeuze baseren. Volgens Janssens en Visscher worden – in zowel Frankrijk als Engeland – de publicaties met name benut door ouders uit de hogere klassen en de middenklasse. Deze groepen ouders kennen het belang van onderwijs en investeren over het algemeen meer in de schoolkeuze van hun kinderen, zeker wanneer zij in sociaal-heterogene wijken wonen waar geen groot en divers scholenaanbod bestaat. Die investering strekt zich ook langer in de tijd uit: indien hun kinderen niet op de school van hun eerste keus terecht komen, doen deze ouders langer hun best om alsnog hun doelen te bereiken (zij weten 'how to work the system'). Binnen deze groep ouders worden in Engeland en Frankrijk, zeker wanneer er nog belangrijke formele en/of materiële beperkingen in het onderwijsaanbod zijn, ook zeer ongewenste neveneffecten signaleerd, zoals het ruilen van adressen in geval van beperkte schoolkeuzemogelijkheden, het zoeken van vreemde vakken (pakketten), het vragen om homogene klassen en het protesteren tegen schooltoewijzingen (Janssens & Visscher, 2004).

De enige ons bekende Nederlandse studies waarin de empirische relatie tussen de kwaliteitskaart en keuzegedrag van ouders centraal staat, zijn studies van Van Bommel (2004) en Dronkers (1999). Zij hebben beide onderzocht in welke mate (veranderingen in) leerlingaantallen op v.o.-scholen samenhangen met de publicatie van prestatie-indicatoren. Dronkers (1999) vindt in een aantal gevallen een (kleine) positieve correlatie tussen de schoolcijfers in "Trouw" en de verandering in leerlingaantallen in het jaar erop, maar concludeert wat snel dat hier sprake is van een causaal effect. Hij gaat bijvoorbeeld voorbij aan het feit dat de kwaliteit van de school zelf van invloed kan zijn op de leerlingaanmeldingen. Van Bommel (2004) gaat dieper in op mogelijke oorzaken van veranderingen in leerlingaantallen en toetst mogelijke effecten van de prestatie-indicatoren (in dit geval de kwaliteitskaart) los van andere mogelijke verbanden. In een enkel geval lijkt er sprake van enig effect, maar dit is in omvang zeer gering. Op basis van deze studie moeten we dan ook concluderen dat er

ook in Nederland niet of nauwelijks sprake is van “stemmen met de voeten”, althans niet in het voortgezet onderwijs. Ook Emmelot e.a. (2004) en Meijer (2004) komen middels een reeks gesprekken met ouders en schoolleiders tot de conclusie dat de kwaliteitskaart nauwelijks effect heeft op de schoolkeuze. Zo is een ruime meerderheid van de schoolleiders van mening dat openbare kwaliteitsinformatie niet merkbaar is in de aanmelding van nieuwe leerlingen, al vinden ze vrijwel allemaal dat dit voor de schoolkeuze nuttige informatie is. Mogelijk is de onbekendheid van ouders met deze prestatie-informatie hier mede debet aan (zie ook Dudok, 2004; Meijer, 2004).

Waterreus (2003) laat verder zien dat publicatie van prestatie-indicatoren, de kwaliteitskaart, wel een klein effect heeft op docentmobiliteit in het voortgezet onderwijs. Scholen met goede oordelen op de kwaliteitskaart kennen een grotere netto instroom van docenten. Het personeelsverloop is hier niet alleen minder groot dan op andere v.o.-scholen, er stromen bij vacatures ook makkelijker nieuwe docenten binnen.

Effecten van prestatie-indicatoren op voice

Hierboven hebben we gezien dat prestatie-indicatoren de betrokkenheid van ouders en leerlingen bij de school mogelijk vergroten en ze instrumenten in handen geven om de school aan te spreken op sterke en zwakke punten. Hierdoor kunnen ouders en leerlingen effectiever bijdragen aan (verbetering van) de kwaliteit van scholen. Op dit moment ontbreekt echter de empirische evidentie voor dit mogelijke effect. Sterker nog, volgens de geïnterviewde ouders in de studie van Meijer (2004) wordt de kwaliteitskaart hier niet of nauwelijks voor gebruikt, en zijn ook weinig ouders op de hoogte van het bestaan van deze informatie.

Effecten van prestatie-indicatoren op schoolverbetering

Effecten van prestatie-indicatoren op schoolverbetering lijken nauwelijks op te treden (zie o.a. Janssens & Visscher, 2004). Noch in Frankrijk, noch in Engeland zien we dat scholen op grond van de publicaties daadwerkelijk aan de kwaliteit van hun onderwijs

werken. Doolaard en Karstanje (2001) zijn, op basis van een analyse van de Nederlandse situatie, ook pessimistisch over hetgeen scholen met de kwaliteitskaart in het kader van schoolverbetering kunnen doen. Opvallend is dat de schoolleiders een andere mening zijn toegedaan. Zo blijkt uit de studie van Emmelot e.a. (2004) dat schoolleiders het prikkelen van scholen tot verbeteren van de kwaliteit als belangrijkste effect zien van het publiek maken van rapporten en kwaliteitsinformatie.

Effecten van prestatie-indicatoren op leerprestaties

Gezien het bovenstaande, is het op zijn minst opvallend dat Jacob (2002) aantoonde dat de introductie van accountability-systemen met prestatie-indicatoren leidt tot verbetering van de leerprestaties op Amerikaanse scholen. Hij doet dit door fraai gebruik te maken van het feit dat deze systemen niet in alle staten tegelijk zijn ingevoerd. In staten waar deze controlesystemen eerder zijn ingevoerd, blijken de leerlingprestaties sterker te zijn genomen dan in staten waar een dergelijk systeem nog geen praktijk was. Een vergelijkbaar onderzoeksdesign is gehanteerd door Hanushek en Raymond (2004), die ook concluderen dat accountability-systemen tot betere leerprestaties op scholen leiden. Zij tonen echter aan dat dit het gevolg is van een set aan beleidsmaatregelen en niet van de publicatie van de prestatie-indicatoren zelf.

3.3 Methodologische kwesties

In elk van bovengenoemde studies draait het om de vraag in welke mate inspectiebezoeken en prestatie-indicatoren bijdragen aan een verbetering van de kwaliteit van het onderwijs. De (veronderstelde) causaliteit in deze vraag is het meest interessant, maar tegelijkertijd het meest problematisch om te toetsen. Zo is het maar de vraag of een gevonden relatie tussen inspectiebezoeken en verbetering van beleid op scholen daadwerkelijk causaal van aard is. Typerend is in ieder geval dat er bij de meeste studies geen gebruikgemaakt wordt van onderzoeksdesigns die zich specifiek richten op het blootleggen van causale relaties, zoals het gebruik van controlegroepen en meer quasi-

experimenteel onderzoek (zie bijv. Cook, 2002; Oosterbeek, 2001; Shadish, Cook, & Campbell, 2002).

Verder zijn er twee andere belangrijke methodologische kwesties rond de toetsing van effecten van toezicht waaraan weinig systematisch aandacht besteed wordt. Dit betreft (a) het onderscheid tussen sterke en zwakke scholen en (b) de termijn waarop mogelijk (kwaliteits)veranderingen zichtbaar worden. Met betrekking tot sterke en zwakke scholen verschillen in de meeste landen zowel de werkwijze als de beoogde effecten. De (literatuur)studie van Learmonth (2000) toont ook aan dat het de moeite waard is bij effecten onderscheid te maken tussen sterke en zwakke scholen. Desondanks wordt er in de meeste studies weinig systematisch nagedacht over mogelijke verschillen, en wordt bij de empirische toetsing vrijwel nooit gedifferentieerd tussen sterke en zwakke scholen.

Wat betreft de veranderingstermijn, hier met name de termijn waarop kwaliteitsverbeteringen zichtbaar worden, zijn in de besproken studies soms wonderlijke keuzes gemaakt. Zo hebben de studies waarin de effecten van inspectiebezoeken op leerprestaties zijn onderzocht (Rosenthal, 2004; Shaw et al., 2003; Wilcox & Gray, 2003), vrijwel allemaal betrekking op het jaar waarin het inspectiebezoek heeft plaatsgevonden, of het jaar dat hier direct op volgt. Het is natuurlijk de vraag of inspectiebezoeken op zo'n korte termijn überhaupt wel effect kunnen sorteren op leerprestaties. Learmonth (2000) en Matthews en Sammons (2004) gaan uit van wat langere veranderingstermijnen, met name voor de betere scholen. Een onderscheid tussen korte- en langetermijneffecten zien we bijna nergens.

4 Studies naar ongewenste neveneffecten van toezicht

In deze paragraaf wordt een overzicht gepresenteerd van studies naar neveneffecten, waarbij we onderscheid maken tussen onderzoek naar neveneffecten van inspectiebezoeken en neveneffecten van prestatie-indicatoren.

4.1 Neveneffecten van inspectiebezoeken

Er bestaat enige empirische evidentie voor bewust strategisch gedrag ten tijde van inspectiebezoeken, met name window dressing. Zo concluderen Case, Case en Catling (2000) dat docenten tijdens het inspectiebezoek een toneelstukje opvoeren. Fitz-Gibbon en Stephenson-Forster (1996) tonen middels een vragenlijstonderzoek onder schoolleiders aan dat er ook fors geïnvesteerd wordt in de voorbereiding van inspectiebezoeken, gemiddeld bijna £ 1200 per school. Deze investering lijkt te lonen; de onderzoekers vonden namelijk een positief verband tussen het bedrag dat besteed wordt aan de voorbereiding en de beoordeling door de inspectie.

Studies naar andere neveneffecten concentreren zich over het algemeen rond onbedoelde effecten van inspectiebezoeken op de gemoedstoestand en het gedrag van de betrokken actoren (met name stress). Zo bestaat er een reeks publicaties van ervaringsdeskundigen met levendige omschrijvingen van inspectiebezoeken en gedragingen van inspecteurs. Over het algemeen werken ze toe naar de conclusie dat inspectiebezoek voornamelijk schade toebrengt aan scholen doordat ze stress veroorzaken bij docenten en de schoolleiding. In meer empirisch onderzoek naar stress wordt meestal gebruikgemaakt van tevredenheidsvragenlijsten. Uit deze studies blijkt dat het inspectiebezoek een forse tijdsinvestering betekent, en over het algemeen leidt tot stress bij docenten en schoolleiders (zie bijv. Cullingford, 1997; Fitz-Gibbon & Stephenson-Forster, 1996, 1999; Learmonth, 2000).

4.2 Neveneffecten van prestatie-indicatoren

Recentelijk hebben Jacob en Levitt (2003) aangetoond dat bij minstens 4 tot 5% van de basisschoolklassen in Chicago jaarlijks sprake is van fraude met toetsscores. Jacob en Levitt komen tot deze conclusie op basis van een empirische analyse van fluctuaties in toetsscores over de jaren heen, en ongebruikelijke antwoordpatronen in toetsen. Voor deze laatste, originele, aanpak hebben de auteurs een algoritme ontwikkeld waarmee ze drie typen ongebruikelijke antwoordpatro-

nen konden identificeren, namelijk (1) blokken identieke antwoorden binnen een klas, (2) ongebruikelijke correlatiepatronen tussen antwoorden van leerlingen in een klas en (3) ongebruikelijke antwoordpatronen van individuele leerlingen (zoals het juist beantwoorden van zeer moeilijke, en het foutief beantwoorden van zeer makkelijke vragen). Jacob en Levitt tonen verder aan dat de scholen waar sprake is van ongebruikelijke antwoordpatronen, ook de scholen zijn waar de fluctuaties in toetsscores over de jaren heen hoog zijn. Bij de uiteindelijke identificatie van de frauderende klassen (de 4 tot 5%) hebben de auteurs vrij strenge normen gehanteerd; dit betrof alleen scholen met zowel veel ongebruikelijk antwoordpatronen als sterk afwijkende toetsscores in het betreffende jaar. Volgens de auteurs is de gevonden 4 tot 5% dan ook een onderschatting van de werkelijke omvang van de fraude, zeker als we hierbij bedenken dat de auteurs slechts enkele vormen van fraude met toetsen hebben bestudeerd (zo hebben ze bijvoorbeeld niet kunnen toetsen of leerlingen meer tijd kregen voor de toets, of dat zwakke leerlingen werden uitgesloten van toetsdeelname). De auteurs laten verder zien dat verschillen in beleid samenhangen met verschillen in de omvang van de fraude, waarbij geldt: hoe groter de prikkels, des te meer fraude met toetsscores.

Het publiceren van leerprestaties of toetsscores van leerlingen heeft ook een ander neveneffect, namelijk uitsluiting van leerlingen bij toetsdeelname. In een drietal (NBER) 'working papers' (Cullen & Reback, 2002; Figlio & Getzler, 2002; Jacob, 2002) wordt op overtuigende wijze aangetoond dat, met de introductie van de publieke verantwoordingsmechanismen die deel uitmaken van de No Child Left Behind Act in de Verenigde Staten, grotere groepen leerlingen niet deelnemen aan de toetsen dan daarvoor het geval was. Alhoewel de studies verschillen in de gehanteerde methodologie en op verschillende staten betrekking hebben, komen de bevindingen en conclusies duidelijk overeen. De studies tonen niet alleen aan dat 'school accountability' resulteert in het uitsluiten van meer leerlingen van toetsdeelname ('reshaping the testpool', aldus Figlio en Getzler, 2002), maar dat dit verschijnsel

ook sterker optreedt op zwakke scholen en op scholen met veel achterstandsleerlingen (de zogenaamde 'high-poverty schools'). Voorts laat Jacob (2002) zien dat, door openbare publicatie van toetsscores, het aantal zittendblijvers in de jaren voor de afname van toetsen toeneemt. Daarnaast bestaat het vermoeden dat openbare publicatie van toetsscores leidt tot meer schooluitval (zoals o.a. wordt gesuggereerd door Figlio & Getzler, 2002), al is dit laatste, voor zover ons bekend, nooit goed empirisch onderzocht.

In een recente studie van Canton en Webbink (2004) wordt ook ingegaan op mogelijk bewust strategisch gedrag van scholen. De auteurs concluderen, mede op basis van bovengenoemd internationaal onderzoek, dat (bewust) strategisch gedrag van scholen (gaming) realiteit is in geval van prestatieprikkels en accountability (p. 36). Daarnaast leggen Canton en Webbink (2004) een link tussen strategisch gedrag en onbetrouwbaarheid van kwaliteitsmaten. In hun conclusie stellen ze dat, ondanks het feit dat ranglijsten een prikkel kunnen zijn voor verbetering van de onderwijskwaliteit, perverse prikkels op de loer liggen als gevolg van het feit dat veel ranglijsten onbetrouwbaar zijn (p. 40).

In Nederland is nog weinig empirisch onderzoek gedaan naar bewust strategisch gedrag van scholen. Wel concludeert Visser (2001) dat Amsterdamse basisscholen op een onwenselijke manier Cito-toetsscores positief beïnvloeden. Hij baseert deze conclusie op verschillende gesprekken met betrokkenen, maar onderbouwt deze verder niet.

Naast bewust strategisch gedrag leidt de publicatie van prestatie-indicatoren mogelijk tot onbewust strategisch gedrag, zoals teaching to the test. Verschillende wetenschappers hebben inmiddels geprobeerd dit verschijnsel te operationaliseren en te kwantificeren. In vrijwel alle gevallen wordt hierbij gekozen voor het aantonen van discrepanties tussen leerresultaten van vakken die wel, en vakken die niet met gestandaardiseerde toetsen worden getoetst. Het feit dat deze discrepantie duidelijk is toegenomen met de introductie van openbare prestatie-indicatoren in enkele Amerikaanse staten, is volgens Jacob (2002) het bewijs van het optreden van dit ongewenste neveneffect. Sturman

(2003) heeft ervoor gekozen het verschijnsel *teaching to the test* te onderzoeken middels een vragenlijst onder docenten. Op basis van dit survey-onderzoek concludeert Sturman dat er inderdaad sprake is van *teaching to the test*, maar dat docenten hier totaal verschillende dingen onder verstaan.

Uit een studie van Wiggins en Tymms (2000) blijkt dat er sprake is van meer vormen van onbewust strategisch gedrag. De auteurs hebben, middels een vragenlijst-onderzoek onder Engelse en Schotse schoolhoofden en docenten, onderzocht in hoeverre publicatie van openbare prestatie-indicatoren resulteert in ongewenste neveneffecten. Volgens de onderzoekers was er in regio's waar prestatie-indicatoren gepubliceerd worden vaker sprake van:

- a toegenomen spanning tussen doelen en streefcijfers;
- b concentratie op streefcijfers ten koste van andere doelstellingen;
- c versmalling van het curriculum;
- d concentratie van ondersteuning bij de slechtste leerlingen; en
- e toename van de "verwijf"-cultuur.

Alhoewel het in deze studie natuurlijk de vraag is of de gevonden effecten in z'n geheel zijn toe te schrijven aan de publicatie van openbare prestatie-indicatoren, wijzen de bevindingen wel op onbewust strategisch gedrag van scholen.

5 Conclusie en nabeschuiving

5.1 Conclusie

Het hierboven gepresenteerde overzicht van studies naar effecten en neveneffecten van toezicht stelt ons in staat een aantal conclusies te trekken. Een eerste conclusie is dat inspectiebezoeken volgens de meeste studies een positief effect sorteren, maar dat er in enkele Engelse studies wel kleine negatieve effecten op leerprestaties gevonden worden. Voor al het onderzoek naar effecten van inspectiebezoeken geldt dat de omvang van de gevonden effecten sterk afhankelijk is van de gehanteerde onderzoeksmethode. In satisfactieonderzoek (onderzoek naar de tevredenheid van de betrokkenen met het inspectiebezoek) wordt over het algemeen aangetoond

dat ruim 80% van de direct betrokkenen (docenten, schoolleiders) van mening is dat inspectiebezoeken effect hebben. Wordt echter onderzoek gedaan naar gedragsbeïnvloeding en beleidsveranderingen, dan blijken deze effecten op zo'n 50 tot 70% van de scholen voor te komen. En wanneer gekeken wordt naar effecten op leerprestaties, dan blijken deze in ieder geval op de korte termijn op Engelse middelbare scholen licht negatief te zijn. Opvallend is verder dat vrijwel al het onderzoek naar effecten van toezicht zich richt op effecten op kwaliteitsverbetering en niet op andere beoogde effecten, zoals de waarborgfunctie en het naleven van wet- en regelgeving.

Op basis van onderzoek naar effecten van de publicatie van openbare prestatie-indicatoren moeten we concluderen dat er signalen zijn dat dit een positief effect op de kwaliteit van scholen heeft, maar dat tegelijkertijd blijkt dat er onduidelijkheid bestaat over mogelijke mechanismen die hieraan ten grondslag liggen. Sterker nog, enkele voor de hand liggende mechanismen worden door empirisch onderzoek niet ondersteund. Zo is er bijvoorbeeld niet of nauwelijks sprake van een relatie tussen openbare prestatie-indicatoren en schoolkeuzegedrag. Het is onbekend of dit met name ligt aan het feit dat ouders en leerlingen onbekend zijn met openbare prestatie-indicatoren of dat ze zich bewust weinig gelegen laten liggen aan deze indicatoren.

Een derde conclusie betreft het optreden van neveneffecten van de onderzochte controlemechanismen. We moeten op basis van verschillende studies concluderen dat er inderdaad een aantal neveneffecten bestaat, maar dat er tegelijkertijd een groep verwachte neveneffecten bestaat waarnaar nog nauwelijks onderzoek is gedaan. Op dit moment bestaat er met name empirisch bewijs voor bewust strategisch gedrag, zoals het beïnvloeden van toetsscores (middels het uitsluiten van leerlingen van toetsdeelname en het helpen van leerlingen bij toetsen) en window dressing. Onderzoek naar onbewust strategisch gedrag (indicatorfixatie, *teaching to the test*, etc.) is daarentegen schaarser van aard. Het optreden van deze neveneffecten ligt theoretisch voor de hand, maar empirische evidentie ontbreekt veelal. Eenzelfde

gebrek aan bewijs geldt voor veel andere veronderstelde neveneffecten, zoals het ontstaan van een zesjescultuur, gelijkvormigheid of isomorfisme, formalisering, versterking van de marktwerking en bureaucratisering. Een ongewenst neveneffect dat wel weer door onderzoeksbevindingen wordt ondersteund, is het optreden van stress tijdens inspectiebezoeken.

5.2 Nabeschuiving

Problematisch bij de hierboven gepresenteerde studies is dat (a) de bevindingen nogal uiteenlopen, (b) de onderzoeksmethodologie sterk varieert en niet altijd een schoonheidsprijs verdient en (c) de bevindingen sterk lijken samen te hangen met de gehanteerde methodologie. Ook zijn er veel beoogde (neven)effecten niet of nog nauwelijks getoetst. De recente toename in onderzoek naar effecten en neveneffecten van toezicht is dan ook toe te juichen, maar vereist wel een passende methodologie. Juist omdat we feitelijk een causale relatie willen blootleggen, worden er hoge eisen gesteld aan het onderzoeksdesign. Het ligt dan ook voor de hand onderzoeksdesigns te gebruiken die specifiek zijn toegesneden op toetsing van causale relaties. Zo is het bijvoorbeeld raadzaam goede controlegroepen te kiezen en/of quasi-experimentele designs toe te passen. We kunnen hierbij bijvoorbeeld denken aan het onderzoeken van effecten van inspectiebezoeken bij een aselechte groep scholen waar extra of juist geen inspectietoezicht heeft plaatsgevonden, of aan toetsing van verschillen tussen scholen waar als gevolg van toevallige verschillen in regio of tijd (nog) geen prestatie-indicatoren zijn gepubliceerd. Tot slot is het aan te bevelen om bij de invulling van het design goed na te denken over welke (neven)effecten getoetst gaan worden, of dit met name sterke of zwakke scholen betreft, en wat de termijn is waarop de (neven)effecten redelijkerwijs zichtbaar worden.

Een tweede aandachtspunt betreft de mogelijke invloed van contextfactoren op de verschillende bevindingen en de gevolgen die dit kan hebben voor de generaliseerbaarheid hiervan. Dit behoeft extra aandacht, omdat de besproken empirische studies betrekking hebben op verschillende landen. We

zijn daarom relatief voorzichtig geweest met onze conclusies en hebben deze alleen getrokken indien er in meerdere landen vergelijkbare resultaten zijn gevonden. Waar dit niet het geval is, spreken we slechts van een indicatie. We willen hieraan toevoegen dat controlesystemen als inspectiebezoeken en publicatie van prestatie-indicatoren in verschillende landen sterk op elkaar lijken (zie bijv. Macnab, 2004) en er theoretisch weinig reden is om aan te nemen dat er grote verschillen bestaan in (neven)effecten tussen verschillende landen.

Omdat in dit artikel twee controlemechanismen centraal staan (inspectiebezoeken en de prestatie-indicatoren), rijst de vraag of we ook iets vergelijkends kunnen zeggen over deze mechanismen. Welk systeem is nu beter, dat wil zeggen wel systeem sorteert de meeste effecten en kent de minste neveneffecten? Wij wagen ons niet aan een antwoord op deze vraag, omdat er nog veel puzzelstukjes ontbreken en het ondoenlijk is de verschillende effecten en neveneffecten te wegen. Wel mag duidelijk zijn dat inspectietoezicht in ieder geval lijkt te resulteren in kwaliteitsverbetering van scholen, al leidt dit niet tot hogere leerresultaten en is er tevens sprake van window dressing en stress. Publicatie van prestatie-indicatoren lijkt wel weer tot een verbetering van leerresultaten te leiden, al kan dit (mede) een gevolg zijn van bijkomend strategisch gedrag van scholen (indicatorfixatie, reshaping the testpool, en fraude bij toetsen). Meer onderzoek naar de causaliteit en onderliggende mechanismen blijft echter nodig.

Literatuur

- Bemmel, M. van. (2004). *De invloed van de kwaliteitskaart op de leerlinginstroom in het voortgezet onderwijs*. Doctoraalscriptie, Universiteit van Amsterdam.
- Bosker, R., & Scheerens, J. (1999). Openbare prestatiegegevens van scholen; nuttigheid en validiteit. *Pedagogische Studiën*, 76, 61-73.
- Brimblecombe, N., Shaw, M., & Ormston, M. (1996). Teachers' intention to change practice as a result of Ofsted school inspections. *Educational Management & Administration*, 24(4), 339-354.

- Brimblecombe, N., Ormston, M., & Shaw, M. (1995). Teachers' perceptions of school inspections: A stressful experience. *Cambridge Journal of Education*, 25(1), 53-62.
- Canton, E., & Webbink, D. (2004). *Prestatieprijkeles in het Nederlandse Onderwijs: Wat kunnen we leren van recente buitenlandse ervaringen?* Den Haag: Centraal Planbureau.
- Carpenter, B., & Stoneham, C. (1994). Inspection effectiveness: An analysis of an Ofsted inspection. *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. *British Journal of Sociology of Education*, 21(4), 605-621.
- Chapman, C. (2001). Changing classrooms through inspections. *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. *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. (Ed.). (1999). *An inspector calls: Ofsted and its effect on school standards*. Londen: Kogan Page.
- Doolgaard, S., & Karstanje, P. (2001). Gebruik van publieke prestatie-indicatoren voor schoolverbetering. In A. B. Dijkstra, S. Karsten, R. Veenstra, & A. J. Visscher (Reds.), *Het oog der natie: scholen op rapport. Standaarden voor de publicatie van schoolprestaties* (pp. 155-172). Assen: Van Gorcum.
- Dronkers, J. (1999). Veranderen leerlingaantallen in het voortgezet onderwijs in het schooljaar 1998-1999 door de publicatie van inspectiegegevens en de berekening van het schoolcijfer door Trouw in oktober 1997? Een nadere analyse. *Tijdschrift voor Onderwijsresearch*, 24, 63-66.
- Dudok, W. (2004). *Effecten van de publicatie van schoolprestaties*. Doctoraalscriptie, Radboud Universiteit Nijmegen.
- 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. *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.
- 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, Seville, Spain.
- Fitz-Gibbon, C. T., & Stephenson-Forster, N. J. (1999). Is Ofsted helpful? In C. Cullingford (Ed.), *An Inspector calls: Ofsted and its effect on school standards*. Londen: Kogan Page.
- Fitz-Gibbon, C.T. (1997). *Feasibility studies for a national system of value-added indicators*. Londen: SCAA.
- Goldstein, H., & Spiegelhalter, D. J. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society, Soc A*, 159(3), 385-443.
- Gray, C., & Gardner, J. (1999). The impact of school inspections. *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. *Cambridge Journal of Education*, 25(1), 53-61.
- Jacob, B., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, 118(3), 843-877.
- Jacob, B. A. (2002). *Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools*. NBER Working Paper.
- Janssens, F. J. G., & Leeuw, F. L. (2001). *Scholen maken verschil. Maar het ene verschil is het*

- andere niet. Utrecht: Inspectie van het Onderwijs.
- Janssens, F. J. G., & Visscher, A. J. (2004). Naar een kwaliteitskaart voor het primair onderwijs. *Pedagogische Studiën*, 81(5), 371-383.
- 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 A. B. Dijkstra, S. Karsten, R. Veenstra, & A. J. Visscher (Reds.), *Het oog der natie: scholen op rapport. Standaarden voor de publicatie van schoolprestaties* (pp. 36-53). Assen: Van Gorcum.
- 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 (Red.), *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.
- Macnab, D. (2004). Hearts, minds and external supervision of schools: Direction and development. *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.
- Meyer, M. W., & Gupta, V. (1994). The performance paradox. *Research in Organizational Behavior*, 16, 309-369.
- Oosterbeek, H., & Webbink, D. (2001). Risico's van indicatoren van schoolkwaliteit. In A. B. Dijkstra, S. Karsten, R. Veenstra, & A. J. Visscher (Reds.), *Het oog der natie: scholen op rapport. Standaarden voor de publicatie van schoolprestaties* (pp. 111-120). Assen: Van Gorcum.
- Oosterbeek, H. (2001). *Voortschrijdend inzicht*. Oratie, Universiteit van Amsterdam.
- Ouston, J., Fidler, B., & Earley, P. (1997). What do schools do after Ofsted school inspections, or before? *School Leadership & Management*, 17(1), 95-104.
- Power, M. (1997). *The audit society*. Oxford: Oxford University Press.
- Rosenthal, L. (2004). Do school inspections improve school quality? Ofsted inspections and school examination results in the UK. *Economics of Education Review*, 23, 143-151.
- Scalon, M. (1999). *The impact of Ofsted inspections*. Slough: 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? *British Educational Research Journal*, 29(1), 63-75.
- Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18, 277-310.
- Smith, P. (1993). Outcome-related performance indicators and organizational control in the public sector. *British Journal of Management*, 4, 135-151.
- Sturman, L. (2003). Teaching to the test: Science or intuition? *Educational Research*, 45(3), 261-273.
- Visser, M. (2001). Cito-toets in een dubbelrol: de kwaliteit van het Amsterdamse basisonderwijs. *Didaktief & School*, april 2001, 38-40.
- Visscher, A. J. (2004). De effecten van schoolprestatiepublicaties in andere landen. *Pedagogische Studiën*, 81, 384-396
- Waterreus, I. (2003). *Lessons in teacher pay: Studies on incentives and the labor market for teachers*. Dissertatie, 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 (ECER), Edinburgh.
- Wilcox, B., & Gray, J. (1996). *Inspecting schools:*

Holding schools to account and helping schools to improve. Buckingham: Open University Press.

Manuscript aanvaard: 9 oktober 2005

Auteurs

Dr. Inge de Wolf is werkzaam als adjunct-inspecteur/coördinerend onderzoeker bij de Inspectie van het Onderwijs en tevens als gast-onderzoeker verbonden aan onderzoeksinstituut Scholar van de Universiteit van Amsterdam.

Prof. dr. Frans Janssens is Coördinerend Inspecteur en programmaleider Onderzoek en Ontwikkeling van de Inspectie van het Onderwijs. Daarnaast is hij bijzonder hoogleraar Onderwijstoezicht aan de Universiteit Twente.

Correspondentieadres: Inge de Wolf, Inspectie van het Onderwijs, Postbus 2730, 3500 CS Utrecht, I.deWolf@owinsp.nl

Abstract

Effects and side effects of inspections and accountability in education: An overview of empirical studies

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”.