

# Accuratesse en gebruik van schoolprestatie-indicatoren

R. Veenstra, L. Bleker en A. Knuver

## Samenvatting

De IST/RST-rapporten van de Inspectie van het Onderwijs en de kwaliteitsindicatoren van de Evaluatiegroep voor het Onderwijs in Groningen worden vergeleken met de standaarden die Visscher, Dijkstra, Karsten en Veenstra (2001) hebben opgesteld. Om de accuratesse van beide publicaties te bepalen, is gekeken naar de gegevens van 31 basisscholen in de stad Groningen. Van beide publicaties kan gezegd worden dat ze relatief goed scoren op accuratesse. De analyses laten echter wel verschillen zien in prestatieoordelen. Zo scoort ongeveer een zesde van de scholen bij de meerniveau-analyses van de Evaluatiegroep anders dan bij de (geaggregeerde) schoolniveau-analyses van de Inspectie. De bruikbaarheid van de informatie in beide publicaties lijkt voldoende te zijn voor schoolleiders en bestuurders. Geen van beide publicaties blijkt geschikt om zonder uitleg aan ouders of teamleden voor te leggen. Een minpunt van beide publicaties is dat geen inzicht gegeven wordt in de relatie tussen preces en uitkomsten.

## 1 Inleiding

Het verschijnen van publicaties met schoolprestatie-indicatoren valt goed te begrijpen vanuit een veranderende sturing in het publieke domein (zie ook Doolaard & Karstanje, 2001). Scholen hebben van de overheid meer autonomie gekregen. Die autonomie is samengegaan met controle op uitkomsten, waarbij van scholen wordt gevraagd rekenschap af te leggen van de bereikte resultaten. In de nota "Grenzeloos leren" bracht het ministerie van Onderwijs, Cultuur en Wetenschappen (2001) dat duidelijk onder woorden: de overheid geeft de richting aan, geeft scholen vervolgens de ruimte bij de uitvoering, maar vraagt hen wel om rekenschap af te leggen van het bereikte resultaat.

Via rapporten van bijvoorbeeld de Inspectie van het Onderwijs wordt rekenschap afgelegd over de prestaties van scholen, maar hoe accuraat en bruikbaar zijn die rapporten? En wat is de uiteindelijke impact van de rapporten? Na een uitgebreide weging van de voors en tegens van het openbaar maken van informatie over de kwaliteit van scholen die werd gepubliceerd in het boek "Het oog der natie: scholen op rapport. Standaarden voor de publicatie van schoolprestaties" (Dijkstra, Karsten, Veenstra, & Visscher, 2001), stelden de auteurs 13 standaarden voor. Deze standaarden betreffen de accuratesse, bruikbaarheid en zorgvuldigheid van schoolprestatie-indicatoren.

In onze bespreking gaan we in op de accuratesse en de bruikbaarheid van schoolprestatie-indicatoren, waarbij we een vergelijking maken tussen de rapporten van het Integraal en Regulier Schooltoezicht (IST/RST) van de Inspectie van het Onderwijs - de meest gezaghebbende publicatie van prestatiegegevens van basisscholen die op dit moment in Nederland beschikbaar is - en een schoolrapport met indicatoren die op een meerniveau-analyse gebaseerd zijn. Een dergelijk schoolrapport is sinds 1995 in gebruik in de stad Groningen. Vrijwel alle basisscholen in Groningen ontvangen in het kader van een kwaliteitsindicatorenproject, uitgevoerd door de Evaluatiegroep voor het Onderwijs in Groningen, jaarlijks een rapport waarin met behulp van een meerniveau-analyse prestatie-indicatoren voor de scores van leerlingen in de groepen 4 en 8 worden gepresenteerd, na controle voor het individuele leerlinggewicht (Bleker, Knuver, & Mulder, 2002). De gegevens van de Evaluatiegroep geven een gedetailleerd beeld van de prestaties van Groninger basisscholen en lenen zich goed voor een vergelijkende beoordeling van de accuratesse van prestatie-indicatoren. We gaan daarbij na, in hoeverre er verschil bestaat in de beelden van de scholen die uit beide benaderingen naar voren komen. We

leggen de IST/RST-rapporten en de Groninger rapporten langs de eerder genoemde standaarden voor publicatie van prestatiegegevens. Beide keren staat de vraag naar het accuraat meten van onderwijskwaliteit centraal. Vervolgens wordt ingegaan op het gebruik van de vergeleken rapporten door scholen en hun toezichhouders. Daarbij gaan we na of de rapporten voor scholen begrijpelijk zijn, of schoolleiders van oordeel zijn dat de rapporten een adequaat beeld geven van hun school, en in hoeverre (en waarvoor) de rapporten worden gebruikt en impact hebben.

Samenvattend: in dit artikel gaan we in op de accuratesse en bruikbaarheid van prestatie-indicatoren die bedoeld zijn voor scholen. De algemene probleemstelling voor dit artikel is: Wat is de accuratesse en bruikbaarheid van de prestatie-indicatoren voor basisscholen van respectievelijk de Inspectie van het Onderwijs en de Evaluatiegroep voor het Onderwijs in Groningen?

## 2 Accuratesse

Van de 39 Groningse basisscholen nemen 33 scholen deel aan het project “Kwaliteitsindicatoren in het Basisonderwijs”. De Inspectie heeft 37 scholen beoordeeld.<sup>1</sup> Voor 31 scholen zijn beide analysemethoden uitgevoerd. In deze paragraaf worden de op geaggregeerde gegevens gebaseerde rapporten van de Inspectie vergeleken met informatie die de Evaluatiegroep op individueel niveau verzamelde.

De vraag is of de vaak toegepaste analyse van geaggregeerde gegevens van leerlingen (gemiddelden op schoolniveau) volstaat, of dat indicatoren nodig zijn die gebaseerd zijn op een meerniveau-analyse (met scores van leerlingen binnen scholen). Verschillende onderzoeken laten zien dat beoordeling van scholen volgens beide methoden uitkomsten geeft die weliswaar redelijk samenhangen, maar die ook substantiële verschillen laten zien (Bosker, Béguin, & Rekers-Mombarg, 2001; Veenstra, Dijkstra, Peschar, & Snijders, 1998). Deze verschillen pakken voor individuele scholen soms ten onrechte voordeelig of nadelig uit, hetgeen vooral ongewenst is als de gegevens in de openbaarheid worden gebracht.

### 2.1 Beschikbare gegevens

Voor de vergelijking van de Inspectiegegevens met de gegevens van de Evaluatiegroep is gebruikgemaakt van drie gegevensbronnen.

*Gegevens van de Inspectie.* Er is gebruikgemaakt van een bestand van de Inspectie (2003; gegevens tot en met schooljaar 2001/2002) waarin op schoolniveau oordelen over de opbrengsten van het onderwijs beschikbaar zijn. De Inspectie hanteert de volgende regels bij het bepalen van haar oordeel over de eindopbrengsten:

- 1 Inventariseer het schoolgemiddelde op de eindtoets van de laatste drie jaar.
- 2 Corrigeer zonedig het schoolgemiddelde voor leerlingen die ten onrechte in de schoolscore zijn opgenomen. Dit geldt voor alle leerlingen die korter dan zes jaar op de school hebben gezeten, als het vermoeden bestaat dat er een selectieve invloed op het schoolgemiddelde uitgaat. Dit geldt eveneens voor alle leerlingen waarvan de school kan aantonen dat ze naar leerwegondersteunend- of praktijkonderwijs uitstromen. Dergelijke leerlingen worden ook wel WSNS-leerlingen genoemd.
- 3 Bepaal in welke van de scholengroepen de school jaarlijks valt. De scholengroep wordt zonedig gecorrigeerd als selectieve tussentijdse instroom en WSNS-leerlingen het gewichtpercentage vertekenen.
- 4 Vergelijk het jaarlijkse schoolgemiddelde met de Inspectienorm, met als range min of plus 2.5 standaardscorepunten rond het gemiddelde van de scholengroep.
- 5 Als de school ten minste één jaar binnen deze range scoort, zijn de resultaten rond het verwachte niveau. Als de school drie jaar achtereen onder de norm scoort, liggen de resultaten onder het verwachte niveau. Als de school drie jaar achtereen 2.5 punten boven het gemiddelde van de scholengroep scoort, liggen de resultaten boven het verwachte niveau (Inspectie van het Onderwijs, zonder jaartal).<sup>2</sup>

*Gegevens van de Evaluatiegroep.* Sinds 1995 verzamelt de Evaluatiegroep voor het Onderwijs in Groningen gegevens in het kader van het project “Kwaliteitsindicatoren in het Basisonderwijs”. Het betreft onder meer gegevens over de prestaties van leerlingen in

groep 8. De gemeente, besturen en scholen in Groningen ontvangen elk jaar een rapport waarin de (ontwikkeling in) prestaties en het sociaal-emotioneel functioneren van leerlingen in de groepen 4 en 8 worden weergegeven (zie Appendix en [www.abcg.nl/files/klokhuisrapport2002.pdf](http://www.abcg.nl/files/klokhuisrapport2002.pdf)). Voor de hier besproken analyses is gebruikgemaakt van Cito-eindtoetscores van de leerlingen van groep 8, en het leerlinggewicht. Het leerlinggewicht is gebruikt als schatting van het beginniveau. Vervolgens is de toegevoegde waarde berekend met behulp van meerniveau-analyse (MLwiN). Alle leerlingen waarvan een eindtoetscore bekend is, zijn in de analyse betrokken.<sup>3</sup> Deze waarden zijn vervolgens omgezet naar z-scores. Op grond van de z-scores zijn de scholen vervolgens ingedeeld in gemiddeld scorende scholen (maximaal .67 standaarddeviatie afwijkend van het gemiddelde van nul), onder- of bovengemiddeld scorende scholen (afwijking tussen .67 en 1.28 standaarddeviatie van het gemiddelde) en laag of hoog scorende scholen (afwijking van meer dan 1.28 standaarddeviatie van het gemiddelde). De indeling in groepen staat overzichtelijk in Tabel 1.

*Gegevens van de Citogroep.* Naast de hierboven genoemde gegevens, is bij het bepalen van de accuratesse ook gebruikgemaakt van direct via de Citogroep beschikbare gegevens. Dit zijn de ruwe gegevens waarop de Inspectieoordelen gebaseerd zijn. Het betreft per schooljaar de gemiddelde scores per school op de Eindtoets, de schoolscoregroep en de gemiddelde Eindtoetscore per schoolscoregroep.

## 2.2 Analyse

Bij de analyses is alleen gekeken naar de prestatiegegevens van de scholen; andere in-

formatie is buiten beschouwing gelaten. Bij de vergelijking van Inspectie- en Evaluatiegroepoordelen zijn de gegevens van de Inspectie als basis genomen, omdat deze niet jaarlijks worden verzameld. De prestatieoordelen van de Inspectie betreffen steeds het bezoekjaar en de twee jaren daarvoor. Voor deze drie jaren, waarop het Inspectieoordeel gebaseerd is, zijn ook de Evaluatiegroepgegevens bekeken. Bij de vergelijking van beide beoordelingen is de Inspectiemethode, waarbij het oordeel over de prestaties is gebaseerd op de scores van drie achtereenvolgende jaren, gehanteerd. Door op deze manier te werken, worden effecten van toevallige omstandigheden binnen een groep of schooljaar verkleind.

In Tabel 2 staan de betreffende oordelen naast elkaar weergegeven. Voor de overgrote meerderheid van scholen komen de oordelen van de Inspectie en de Evaluatiegroep overeen. Van een overeenkomst in oordeel is sprake wanneer het Evaluatiegroepoordeel als basis kan worden gebruikt voor het Inspectieoordeel. Dus: drie jaar achtereenvolgend onder of laag zou een Inspectieoordeel *onder niveau* moeten opleveren; twee jaar onder of laag en één keer midden zou een Inspectieoordeel *op niveau* moeten opleveren. Voor enkele scholen komen de oordelen niet met elkaar overeen. De eerste school in de tabel scoort bij de Evaluatiegroep gemiddeld, terwijl de Inspectie de opbrengst *boven niveau* beoordeelt. De drie scholen daaronder worden door de Evaluatiegroep als *boven niveau* beoordeeld, de Inspectie vindt echter dat de scholen *op niveau* scoren. Ook aan de onderkant van de tabel zijn nog enkele verschillen te vinden. Hier is een school die van de Inspectie het oordeel *op niveau* krijgt, terwijl de berekeningen van de Evaluatiegroep ople-

Tabel 1

*Betekenis van de waarderingscategorieën uit de schoolrapporten van de Groningse Evaluatiegroep*

Waardering	Indicatorwaarden (z-scores; $M = 0$ en $SD = 1$ )	Geschat percentage van het totaal aantal deelnemende scholen (uitgaande van normaalverdeling)
Laag	Waarde lager dan -1.28	10% van 33 scholen
Onder	Waarde tussen -1.28 en -.67	15% van 33 scholen
Midden	Waarde tussen -.67 en .67	50% van 33 scholen
Boven	Waarde tussen .67 en 1.28	15% van 33 scholen
Hoog	Waarde hoger dan 1.28	10% van 33 scholen

Tabel 2

Vergelijking van oordelen over scholen in Groningen door de Inspectie en door de Evaluatiegroep

	Inspectie	Evaluatiegroep	Overeenkomst
1 school	Boven niveau	3 jaren midden	-
1 school	Op niveau	3 jaren hoog	-
1 school	Op niveau	2 jaren hoog, 1 jaar boven	-
1 school	Op niveau	3 jaren boven	-
1 school	Op niveau	2 jaren hoog, 1 jaar midden	+
2 scholen	Op niveau	1 jaar hoog, 1 jaar boven, 1 jaar midden	+
5 scholen	Op niveau	1 jaar boven, 2 jaren midden	+
1 school	Op niveau	1 jaar hoog, 1 jaar midden, 1 jaar onder	+
1 school	Op niveau	1 jaar boven, 1 jaar midden, 1 jaar onder	+
6 scholen	Op niveau	3 jaren midden	+
3 scholen	Op niveau	2 jaren midden, 1 jaar onder	+
1 school	Op niveau	1 jaar midden, 1 jaar laag, 1 jaar geen gegevens	+
3 scholen	Op niveau	1 jaar midden, 2 jaren onder	+
2 scholen	Op niveau	1 jaar midden, 2 jaren laag	+
1 school	Op niveau	3 jaren onder	-
1 school	Onder niveau	1 jaar onder, 2 jaren laag	+

veren dat de school drie jaren achtereen *onder niveau* scoort.

Een van de oorzaken voor deze verschillen is dat de Inspectie een grote terughoudendheid betracht met het geven van positieve of negatieve oordelen over schoolprestaties. De uitspraken van de Inspectie werken sterk nivellerend: in de uiteindelijke oordelen zit nauwelijks nog variatie. Het oordeel boven- of onderpresteren velt de Inspectie alleen bij een relatief grote afwijking van de verwachte waarde. Dit heeft als voordeel dat de kans klein is dat scholen onterecht als zeer sterk of zeer zwak te boek zullen komen te staan. Gegeven dat verschillen tussen scholen meer met instroomkenmerken dan met kwaliteitsverschillen te maken hebben, is deze voorzichtigheid verstandig (Scheerens & Bosker, 1997; Veenstra, 1999). Als ondersteuning bij het schoolkeuzeproces van ouders is het ontbreken van verschillen echter weinig informatief. Door de manier van berekenen van de Evaluatiegroep scoort per definitie 50% van de scholen gemiddeld en de andere 50% boven- of ondergemiddeld. Het voordeel van deze werkwijze is dat de resul-

taten van scholen gespreid zijn en dat scholen die laag scoren, geprikkeld worden om zich te verbeteren. De keerzijde is echter dat ook als alle scholen absoluut voldoende scoren, er relatief gezien scholen zullen zijn die tot het onderste kwart behoren.

Een andere mogelijke oorzaak voor de gevonden verschillen is dat de Evaluatiegroepscores de uitkomsten zijn van meerniveau-analyses en die van de Inspectie van geaggregeerde gegevens. Om preciezer na te gaan in hoeverre de verschillende analysemethoden (geaggregeerde gegevens versus meerniveau-analyses) van de Inspectie en de Evaluatiegroep per schooljaar leiden tot dezelfde uitkomsten, is gebruikgemaakt van de gegevens van de Citogroep over de jaren 1999-2002. Deze gegevens vormen de feitelijke basis voor het oordeel van de Inspectie en de Evaluatiegroep. Op deze gegevens is zowel een keer de Inspectieprocedure toegepast, als de Evaluatiegroepprocedure. Volgens de procedure van de Inspectie is per schooljaar de gemiddelde Cito-eindtoetscore berekend (voor alle in het bestand aanwezige leerlingen) minus de verwachte waar-

Tabel 3

Vergelijking uitkomsten analyseprocedure Evaluatiegroep en Inspectie met behulp van hetzelfde bestand met Cito-eindtoetscores

School	Jaar Inspectie bezoek	1999		2000		2001		2002	
		Evaluatie- groep	Inspectie- tie	Evaluatie- groep	Inspectie- tie	Evaluatie- groep	Inspectie- tie	Evaluatie- groep	Inspectie- tie
1	98-99	-.21	-.19	.67	.51	1.31	1.06	-.66	-.46
2	98-99	-2.16	-2.43	-2.92	-1.72	-.41	-1.00	-.35	-1.21
3	98-99	-.04	-.01	1.19	1.16	.67	1.15	1.47	1.08
4	98-99	-1.26	-1.45	-.64	-.75	-.11	-.05	.75	1.18
5	98-99	.04	-.13	.96	.73	1.19	1.05	.59	.43
6	99-00	-.09	.03	-.27	.33	-2.09	-1.83	-2.10	-1.71
7	99-00	1.33	1.32	.31	.05	2.15	1.48	1.62	1.24
8	99-00	.71	.32	-.30	-.55	-.28	-.39	.57	.42
9	99-00	-.30	-.48	.01	-.29	1.19	.70	.25	-.01
10	99-00	-1.04	-.74	-1.24	-1.03	-.53	-.59	.26	.29
11	99-00	1.34	1.26	.44	.37	-.95	-1.01	-.49	-.22
12	99-00	.01	-.40	.08	.08	-.79	.21	.34	.73
13	99-00	.78	.61	-.38	-.66	-.62	-.36	.24	.21
14	99-00	.52	.31	-.24	-.32	-.85	-.84	.78	.86
15	99-00	.75	.62	1.09	.98	.22	.35	.93	.59
16	99-00	-.30	-.03	-.39	.10	-.81	-.06	-.78	-.07
17	99-00	-1.75	-1.64	-1.62	-1.69	-.14	.10	-1.97	-1.56
18	00-01	2.12	2.18	2.10	2.32	2.23	2.65	1.01	.76
19	00-01	-.59	.26	.71	2.28	-.37	.50	.89	1.67
20	00-01	.40	.55	-1.08	-1.40	-1.24	-1.25	-1.98	-1.90
21	00-01	-1.24	-1.15	.11	.02	-.62	-.59	-.30	-.13
22	00-01	.02	.19	-.45	-.97	.35	.28	-.06	.22
23	00-01	-.60	-.60	-1.27	-1.39	-.65	-.60	-1.15	-.84
24	00-01	-1.01	-1.14	-.34	-.58	-.69	-.34	-.01	.56
25	00-01	-.54	-.63	-.12	-.24	.94	.86	-.74	-.71
26	00-01	-1.50	-1.46	-.25	.16	-1.37	-1.50	-.32	-1.08
27	01-02	.02	.21	-.42	-.50	-.72	-1.55	-1.24	-1.54
28	01-02	.66	.98	.38	.02	.57	.64	1.14	1.00
29	01-02	-.20	-.18	.60	.58	.43	.34	.72	1.08
30	01-02	1.30	1.24	1.42	1.00	1.38	.94	.83	.61
31	01-02	1.31	1.29	1.81	1.47	.49	.22	-1.18	-1.82

*Noot.* Wanneer de uitkomsten van de beide analyseprocedures meer dan een halve standaarddeviatie verschillen, zijn de betreffende hokjes grijs gekleurd. Voor bijvoorbeeld school 19 leveren de verschillende analyseprocedures in 2002 een verschil in resultaat op van 0.78 standaarddeviatie. Volgens de meerniveau-analyse van de Evaluatiegroep scoort deze school 0.89 standaarddeviatie boven de verwachte waarde (Gronings gemiddelde, gecorrigeerd voor leerlinggewicht); volgens de analyse met geaggregeerde gegevens van de Inspectie scoort deze school 1.67 standaarddeviatie boven de verwachte waarde (gemiddelde van de schoolscoregroep waartoe deze school hoort).

de. De verwachte waarde is gelijk aan de gemiddelde Eindtoetscore voor de schoolscoregroep waarin de betreffende school valt.<sup>4</sup> De uitkomst is omgezet naar z-scores. In de procedure van de Evaluatiegroep is de schoolscore bepaald door een meerniveau-analyse uit te voeren met de Eindtoetscore als afhankelijke variabele en het leerlinggewicht als controlevariabele. De uitkomsten

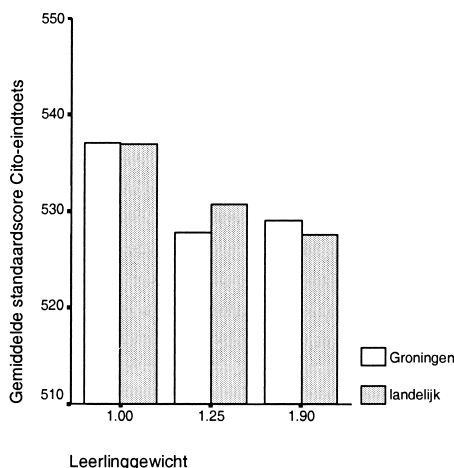
zijn omgezet naar z-scores. In Tabel 3 zijn de uitkomsten van beide procedures naast elkaar gezet. Ter verduidelijking zijn de hokjes waarin de resultaten meer dan een halve standaarddeviatie verschillen, grijs gekleurd.

Voor negen scholen leveren de analyseprocedures van de Inspectie en de Evaluatiegroep één of meer jaren niet dezelfde uitkomsten op bij het gestelde criterium van een

halve standaarddeviatie. Het mag duidelijk zijn dat een ander criterium weer andere resultaten zal opleveren. Als bijvoorbeeld het indelingscriterium van de Evaluatiegroep gehanteerd wordt, blijken voor 21 scholen de oordelen voor één of meer jaren te verschillen. Bij beide analysemethoden is de zogeheten toegevoegde waarde bepaald met behulp van leerlingachtergrondgegevens. Echter, de Evaluatiegroep gebruikt het leerlinggewicht op individueel niveau als controlevariabele in de meerniveau-analyse; de Inspectie gebruikt de schoolscoregroep als geaggregeerd referentiekader.

De schoolscoregroep is sinds 1999 een landelijk door de Citogroep gehanteerd indelingscriterium. De schoolscoregroep waartoe een school wordt gerekend, is gebaseerd op het percentage leerlingen met verschillende leerlinggewichten, waarbij ervan uit wordt gegaan dat allochtone kinderen slechter presteren dan autochtone achterstandsleerlingen, die op zich weer lager presteren dan autochtone leerlingen zonder achterstand. Bij het bepalen van dit criterium is uitgegaan van landelijk gemiddelde scores voor de onderscheiden leerlinggewichten, en is geen rekening gehouden met regionale verschillen in de bevolkingssamenstelling en eventuele effecten daarvan op leerlingprestaties. Voor de stad Groningen is bijvoorbeeld bekend dat autochtone achterstandsleerlingen slechter presteren dan allochtone leerlingen (Knuver & Mulder, 2003). De oorzaak hiervan is deels gelegen in de bevolkingssamenstelling van de stad Groningen: vergeleken met andere grote steden zijn laag presterende Turkse en Marokkaanse leerlingen ondervertegenwoordigd in de allochtone populatie (vgl. Driessen, Van Langen, & Vierke, 2002). Daarnaast is bekend dat autochtone achterstandsleerlingen in de noordelijke provincies gemiddeld lager scoren dan in andere delen van Nederland (Inspectie van het Onderwijs, 2002). In Figuur 1 staan ter illustratie de gemiddelde Cito-eindtoetscores voor de stad Groningen en landelijk weergegeven naar leerlinggewicht.

Uit de figuur wordt duidelijk dat de 1.25 (autochtone achterstands)leerlingen in de stad Groningen lager scoren dan 1.9 (allochtone achterstands)leerlingen. Bij beoordeling op



Figuur 1. De gemiddelde standaardscore in 2003 per leerlinggewicht op de Cito-eindtoets in de stad Groningen in vergelijking met landelijke gegevens uit het PRIMA-cohortonderzoek 2002 (bron: Mulder & Knuver, 2004).

een op landelijke gemiddelden gebaseerde norm worden Groningse scholen met veel 1.25-leerlingen onderschat en scholen met veel 1.90-leerlingen overschat. De gegevens in de figuur hebben betrekking op 2003. Ook in de vijf daaraan voorafgaande jaren scoorden 1.90-leerlingen in Groningen steeds een of twee punten hoger op de Cito-eindtoets dan 1.25-leerlingen.

### 2.3 Accuratessecriteria

In "Het oog der natie: scholen op rapport" staan verschillende standaarden voor accurate genoemd (Visscher et al., 2001). Voor enkele relevante standaarden zijn hier de scores in de publicaties van de Inspectie en Evaluatiegroep vergeleken. Voor de meeste standaarden is weinig verschil te constateren in de procedures van Inspectie en Evaluatiegroep: Standaard A1 stelt dat inzicht gegeven moet worden in de toegevoegde waarde van scholen. Beide beoordelingsprocedures voldoen in zekere mate aan dit criterium, hoewel het op verschillende manieren is uitgewerkt (zie hierboven). In beide gevallen is geen beginmeting voorhanden, maar deze wordt geschat met behulp van het leerlinggewicht. Een van de problemen met schattingen van prestaties op grond van het leerlinggewicht, ligt in het feit dat de 1.00-leerlingen een zeer

gevarieerde groep vormen (variërend van kinderen met ouders waarvan de één een mavo-diploma heeft en de ander geen diploma, tot kinderen met twee wetenschappelijk opgeleide ouders). Deze vertekening in het beoordelen van de toegevoegde waarde doet zich zowel bij de Inspectieprocedure als de Evaluatiegroepprocedure voor. Standaard A3 stelt dat de betrouwbaarheidsintervallen weergegeven zouden moeten worden om aan te geven dat er onzekerheidsmarges zijn in de berekening van de indicatoren. In geen van beide beoordelingsprocedures worden betrouwbaarheidsintervallen berekend en in de publicatie opgenomen. Wel wordt zowel door de Inspectie als door de Evaluatiegroep rekening gehouden met toevalligheden in de uiteindelijke beoordeling. De Evaluatiegroep wijst in haar inleiding op een aantal punten die bij de interpretatie in acht moeten worden genomen, onder meer onbetrouwbaarheid van de cijfers bij een klein aantal leerlingen. De Inspectie houdt rekening met onbetrouwbaarheid van de gegevens door scores van drie opeenvolgende jaren te bekijken én een grote marge aan te houden bij het toedelen van een positief of negatief oordeel. Standaard A4 stelt dat de prestatie-indicatoren gebaseerd moeten zijn op minimaal 10 leerlingen in de school. Voor beide analyses geldt dat dit zo is. Standaard A6 stelt dat de berekeningen controleerbaar moeten zijn. Globaal zijn de door Inspectie gehanteerde criteria voor het bepalen van de oordelen over de opbrengst beschreven in de Informatiebrochure voor scholen. De precieze criteria zijn beschreven in de interne notitie "Analyse en waardering van opbrengsten" (Inspectie van het Onderwijs, zonder jaartal). Van de door de Evaluatiegroep gehanteerde procedure is via internet een integrale beschrijving beschikbaar.

Op een van de standaarden verschillen beide procedures wezenlijk: Standaard A2 stelt dat de indicatoren zo veel mogelijk gebaseerd moeten zijn op meerniveau-analyses. Aan deze standaard voldoet alleen de analyseprocedure van de Evaluatiegroep. Door onder meer Veenstra e.a. (1998) en Bosker e.a. (2001) is met behulp van analyses op gegevens uit het voortgezet onderwijs beargumenteerd waarom de toegevoegde waarde

van scholen bepaald zou moeten worden op grond van gegevens over de prestaties die individuele leerlingen boeken (voor verdere informatie zie Snijders & Bosker, 1999). Uit hun berekeningen blijkt dat beoordelingen die gebaseerd zijn op geaggregeerde gegevens voor enkele scholen niet adequaat zijn. Uit de hierboven gepresenteerde analyses op Cito-eindtoetsscores blijkt eveneens dat meerniveau-analyses en analyses op geaggregeerde data voor een aantal scholen verschillende resultaten opleveren. Onafhankelijk van de uiteindelijk gehanteerde beslissingsregels voor het toekennen van een kwalificatie *onderpresteren* (zwak) of *overpresteren* (sterk) kunnen deze verschillen voor de betreffende scholen onterecht bezorgdheid of juist geruststelling bij scholen en ouders opleveren.

Bij het uitvoeren van deze vergelijking voor Groningse scholen moet nog opgemerkt worden dat de gevonden verschillen deels ook een effect kunnen zijn van het feit dat de meerniveau-analyse per definitie alleen Groningse prestatieverschillen tussen leerlinggewichten in beschouwing neemt, terwijl de Inspectieprocedure uitgaat van landelijk gemiddelde scores per leerlinggewicht.<sup>5</sup>

### 3 Gebruik

#### 3.1 Achtergrond

Kennen scholen de schoolrapporten en maken ze er gebruik van? Benutten scholen de rapporten om naar hun sterke en zwakke kanten te kijken? Of zijn er scholen waar het rapport in een la verdwijnt, omdat de school zich er niet in herkent of omdat de uitkomsten het team zouden demotiveren? Over het gebruik van kwaliteitskaarten is nog niet zo veel bekend. Als het gaat om ouders, lijken alleen sommige groepen interesse te tonen, zoals hoger opgeleiden. De invloed van prestatiegegevens op het schoolkeuzeprocess lijkt voorlopig echter niet groot (Dijkstra & Witziers, 2001). Voor wat het gebruik door scholen betreft, bieden Doolaard en Karstanje (2001) een eerste balans. Hun oordeel over de baat die scholen bij dergelijke publicaties hebben, is vooralsnog negatief. Vooral aansluiting bij ontwikkelingen binnen de school

en herkenbare, op de schoolpraktijk betrokken informatie is van belang. Publicaties met vooral uitkomsten - zoals hitlijsten in kranten - geven een school weinig aanknopingspunten voor verbetering, terwijl voor schoolverbetering inzicht in het proces binnen de school cruciaal is.

Over de tevredenheid bij schoolleiders en -besturen over de schoolrapporten van de Inspectie van het Onderwijs zijn enige gegevens bekend. De Inspectie heeft namelijk de opvattingen van scholen over IST/RST-rapporten onderzocht. Daaruit kwam naar voren dat scholen de rapporten van de Inspectie betrouwbaar en gezaghebbend vinden (Van de Grift, 2001). In de praktijk blijken scholen bovendien andere analyse-instrumenten steeds meer te laten aansluiten bij de eisen van de Inspectie. De Evaluatiegroep voor het Onderwijs deed ook onderzoek naar de mening van Groningse basisscholen over hun eigen schoolrapporten. De evaluatie liet zien dat scholen en besturen tevreden waren over de rapporten. Ook bleken ze de rapporten geregeld te gebruiken (Evaluatiegroep voor het Onderwijs, 2001).

In deze bijdrage presenteren we de resultaten van een vergelijkende beoordeling van het gebruik door scholen van de rapporten van de Inspectie en de Groningse Evaluatiegroep. De aanleiding daarvoor is in de eerste plaats dat we een bijdrage willen leveren aan de bezinning op de verdere ontwikkeling en optimalisering van schoolprestatie-indicatoren. Dat willen we doen door toepassing van de standaarden van Visscher e.a. (2001) op de rapporten van de Inspectie en de Evaluatiegroep voor het Onderwijs. We willen de standaarden over het gebruik van prestatiegegevens door scholen toetsen op praktische bruikbaarheid en relevantie. Ook willen we nagaan of het nodig is om die standaarden verder te verfijnen. Door beide publicaties tegen elkaar af te zetten en met de standaarden te vergelijken, willen we verder een beeld van de sterke en zwakke punten van beide rapporten schetsen.

### **3.2 Methode**

De standaarden uit Visscher e.a. (2001) zijn het uitgangspunt voor de evaluatie van het gebruik door scholen van de rapporten van

Evaluatiegroep en Inspectie (IST/RST).<sup>6</sup> Dit deel van ons onderzoek is in twee stappen uitgevoerd. De eerste stap is een analytische evaluatie die bestond uit inventarisatie en weging van de sterke en zwakke punten van beide publicaties door de onderzoekers. In de tweede stap zijn acht sleutelinformanten bevroegd op hun ervaringen met de publicaties van de Evaluatiegroep en de Inspectie. Deze stap heeft een dubbel doel. Deze raadpleging geeft een directe impressie van de ervaringen die gebruikers met beide publicaties hebben opgedaan. Daarmee kunnen de resultaten van de analytische evaluatie verder worden aangevuld en ingekleurd. Ook zijn de bevindingen van de gebruikers gebruikt voor validatie en weging van de resultaten van de analytische evaluatie.

#### *Raadpleging van gebruikers*

De vergelijkende vraagstelling maakte de selectie van sleutelinformanten eenvoudig: gesprekspartners die vanuit verschillende functies zoals schoolleider, toezichthouder of bevoegd gezag, in hun werk ervaringen hadden opgedaan met de rapporten van de Inspectie én de Groningse Evaluatiegroep. De groep informanten bestond daarmee uit schoolleiders en bestuurders van basisscholen in de stad Groningen die via een regulier, lokaal overlegplatform waren uitgenodigd voor deelname aan de raadpleging. De raadpleging vond plaats in de vorm van een groepsgesprek. Bij het gesprek waren zes directeuren aanwezig en twee bestuurders. De deelnemers vertegenwoordigden samen 28 Groningse scholen. Alle deelnemers aan het gesprek waren bekend met de schoolrapporten van de Evaluatiegroep voor het Onderwijs en met de IST/RST-rapporten van de Inspectie. Het gesprek verliep heel open. De verschillende deelnemers kennen elkaar al langer en zijn gewend om informatie uit te wisselen. De sfeer van het gesprek was collegiaal en niet competitief.

Ter voorbereiding op de raadpleging is de deelnemers een vragenlijst toegestuurd. Deze vragenlijst, die ook de leidraad was voor het groepsgesprek, bestond uit een opsomming van de vier onderwerpen die in de bespreking aan de orde zouden komen. Dat waren tevens de kernvragen uit de analytische evaluatie:



Hoe is de begrijpelijkheid van de rapporten? Geven de rapporten, naar de mening van de gebruikers, een adequaat beeld van de school? Maken de scholen gebruik van de informatie uit de rapporten? In hoeverre hebben de rapporten invloed op de school of op de context waarin de school opereert? Per onderwerp werd vervolgens een aantal mogelijke aspecten onderscheiden. De bespreking van de resultaten van de analytische evaluatie, waarin ook de uitkomsten van de raadpleging van gebruikers is verwerkt, is vanuit deze vier kernvragen opgezet.

#### *Standaarden voor gebruik door scholen*

De standaarden uit Visscher e.a. (2001) bevatten in de categorie *bruikbaarheid* overwegingen die van belang zijn voor gebruik door scholen. Het gaat dan om het belang van een adequate toelichting (B1), de eis dat informatie beschikbaar moet zijn over de functie die de gepubliceerde gegevens vervullen (B2), het belang van ruime verspreiding van kwaliteitsinformatie (B3) en de begrijpelijkheid daarvan (B4). Verder bieden ook standaarden over accuratesse van publieke schoolprestatiegegevens enige aanknopingspunten, met name waar het gaat om controleerbaarheid (A6) en zorgvuldige procedures (A7). Ten slotte is ook een van de eisen van zorgvuldigheid voor scholen relevant, te weten het vooraf kunnen inzien van de gegevens (Z2).

Wanneer we de door Visscher en collega's voorgestelde standaarden bezien vanuit de invalshoek van het gebruik door *scholen*, valt het op dat deze doorgaans niet rechtstreeks op dat perspectief zijn betrokken. De standaarden vragen vaak om een verdere uitwerking alvorens ze in de praktijk kunnen worden toegepast. We stellen daarom een specificatie van de standaarden voor. In deze bijdrage zullen we daarvoor enkele voorstellen doen als het gaat om het gebruik van kwaliteitsrapporten door scholen en de functie van schoolverbetering.

#### *Bruikbaarheid*

Begrijpelijkheid is wel de minste voorwaarde waaraan informatie over de kwaliteit van een school moet voldoen. Visscher e.a. (2001),

die deze standaard formuleren voor het gebruik door ouders, wijzen dan op de manier waarop de gegevens worden gepresenteerd, en de noodzaak van de toevoeging van een toelichting. De begrijpelijkheid van een publicatie hangt af van de duidelijkheid van de opgenomen gegevens, maar ook van de kans dat deze op verschillende manieren begrepen kunnen worden. Een ander aspect is de inspanning die lezers moeten doen om de gegevens te begrijpen: Zijn deze in één oogopslag correct te interpreteren, of moet er eerst uitgebreid kennis worden genomen van de manier waarop het materiaal gelezen moet worden? We vinden het dan ook verstandig de eis van *begrijpelijkheid* aan te vullen met die van een overzichtelijke en inzichtelijke presentatie van de gegevens; een aantrekkelijke en duidelijke vormgeving vergemakkelijkt het doornemen van schoolrapporten.

De analyse van de inhoud van de IST/RST-rapporten van basisscholen in de stad Groningen in de periode 1998-2002 laat zien dat deze doorgaans in duidelijke taal zijn opgesteld. Ook is de tekst doorgaans eenduidig en zonder al te veel moeite te interpreteren. De tekstuele presentatie van gegevens is weliswaar omvangrijk en is pas na enige inspanning te overzien, maar het schema van "bolletjes" geeft een overzichtelijke samenvatting van de gegevens. Een sterk punt is ook de passage waarin de "afspraken" met de school op een rij worden gezet. Hoewel het geen integrale samenvatting van het rapport is, geven ze voor de lezer op een eenvoudige manier inzicht in de verbeteringen die de Inspectie van de school vraagt. De rapporten van de Inspectie voldoen ook aan de eis dat de gegevens van een toelichting zijn voorzien. Ook het schoolrapport van de Groningse Evaluatiegroep is duidelijk opgeschreven. De nadruk op cijfermatige en grafische weergave van de gegevens betekent wel dat deze onderdelen van het rapport voor veel lezers zonder goede kennisname van de toelichting niet te begrijpen zullen zijn. Die toelichtingen worden gegeven, maar de duidelijkheid daarvan verschilt. Anderzijds leidt de opbouw vanuit tabellen en grafieken met toelichtende teksten tot een overzichtelijk rapport dat voor enigszins ervaren gebruikers snel inzicht geeft in uiteenlopende aspecten

van het onderwijs. Dat geldt nog meer voor de samenvatting die van elk schoolrapport beschikbaar is: veel informatie die tot een overzichtelijk, grafisch weergegeven geheel wordt teruggebracht. Een sterk punt van deze samenvattingen is ook dat de centrale conclusies kernachtig worden weergegeven.

De raadpleging van gebruikers weerspiegelt dit beeld. De informanten gaven aan dat de begrijpelijkheid van de gegevens afhangt van de geleidingen die de rapporten lezen. Schoolleiders waren van mening met beide rapporten zelf redelijk uit de voeten te kunnen, maar hielden voor hun teams een slag om de arm. Toch zouden de rapporten met enige uitleg ook voor deze groep wel te begrijpen zijn, zo dachten ze. Deze oordelen golden voor de rapporten van de Inspectie (“eenvoudig geschreven”) overigens wat minder dan voor die van de Evaluatiegroep; “Pittige stof”, zo oordeelden de directeuren over de laatste. Het belang van uitleg gold volgens de schoolleiders zeker voor ouders en de medezeggenschapsraad; de informanten gaven aan dat de begrijpelijkheid voor beide groepen zonder toelichting onvoldoende is. Afgezien daarvan gaven de directeuren aan, het van belang te vinden om, als de rapporten in het team of door de medezeggenschapsraad worden gebruikt, deze altijd mondeling toe te lichten “en van een context te voorzien”. Van de rapporten van de Inspectie waardeerden de gebruikers de uitvoerige beschrijvende informatie die deze verslagen geven: “die uitgebreidheid is een plus”. Hoewel de vormgeving van beide onderzochte rapporten nogal verschilt, gaven de informanten niettemin te kennen met beide tevreden te zijn. De wens om in de Inspectierapporten ook gegevens in tabelvorm op te nemen, helpt dit gegeven te interpreteren: met name een mix van gecomprimeerde, overzichtelijke gegevens zoals tabellen en meer gedetailleerde beschrijvende informatie, leek schoolleiders ideaal.

Visscher e.a. (2001) wijzen ook op het belang van duidelijkheid over de aard van de gegevens: Wat betekenen ze wel, en wat niet? Die duidelijkheid heeft betrekking op de betekenis van de gepresenteerde gegevens en op uitleg van wat dat zegt over de prestaties van een school, zoals dat het precieze aandeel

van de school onduidelijk is, en dat de gegevens betrekking hebben op waarschijnlijkheden. Het rapport van de Evaluatiegroep bevat dergelijke uitleg, maar het rapport van de Inspectie niet. Verder is duidelijkheid over het doel waarvoor de gegevens geproduceerd zijn belangrijk: zijn ze te gebruiken voor externe verantwoording, voor de keuze van een school, of voor de verbetering van het onderwijs? Dat geen van beide rapporten aan dit criterium voldoet, hoeft niet te verbazen; het verbinden van prestatiegegevens aan procesinformatie, zeker in periodieke rendementsrapporten, staat nog in de kinderschoenen.

#### *Adequaatheid*

Eén van de standaarden van Visscher e.a. (2001) luidt dat scholen het recht hebben de gegevens die voor de berekening van indicatoren zijn gebruikt vooraf in te zien. Aangezien scholen onevenredig geschaad kunnen worden door een onterecht negatief oordeel, draagt recht op inzage bij aan vermindering van de kans dat fouten in de publicaties terecht komen. Vanuit het belang van de school is wat te zeggen voor een ruimere formulering van deze standaard. Niet alleen de gegevens die als input voor de analyses gebruikt worden, zouden op juistheid gecontroleerd moeten kunnen worden, maar ook de conclusies die daarover in het rapport worden opgenomen. Dat een school niet het laatste woord heeft bij de bepaling van het oordeel behoeft geen betoog, maar kennisname van het conceptrapport draagt bij aan de publicatie van adequate, feitelijk juiste gegevens. De Inspectie volgt een dergelijke procedure.<sup>7</sup>

Naast de al genoemde eisen dat de berekeningen achter schoolprestatie-indicatoren controleerbaar moeten zijn, en dat de beoordeling van de prestaties van scholen helder en controleerbaar verloopt, kan daar vanuit de optiek van scholen nog het element van doorzichtigheid *voor de school* aan worden toegevoegd. Hebben schoolleiders zicht op de herkomst van de gepresenteerde kengetallen, en op hoe die gerelateerd zijn aan de voor hen waarneembare werkelijkheid van de school? Voldoen schoolprestatie-indicatoren aan het criterium van inhoudsvaliditeit, of is de school als gebruiker geheel aangewezen op het vertrouwen in de deskundigheid van de proe-

ducent van de prestatie-indicatoren?

Schoolrapporten die vooral uit uitgebreide, beschrijvende teksten zijn opgebouwd, voldoen doorgaans gemakkelijker aan deze eis. Naarmate rapporten echter meer op analyse van kwantitatieve gegevens berusten, is het criterium van transparante gegevens en oordelen een grotere opgave, zeker als daarvoor geavanceerde analysemethoden zijn gebruikt. De sterk kwantitatief georiënteerde rapporten van de Evaluatiegroep en de beschrijvend opgezette IST/RST-rapporten van de Inspectie illustreren dan zo ongeveer de uitersten die denkbaar zijn. Hoewel de Evaluatiegroep de analyseprocedure gedetailleerd publiceert, zal deze voor de meeste gebruikers niet begrijpelijk zijn. Voor de op Cito-scores gebaseerde beoordeling van opbrengsten door de Inspectie geldt tot op zekere hoogte hetzelfde. De gedetailleerde beschrijvingen uit de IST/RST-rapporten zijn daarentegen voor scholen goed te doorzien.

Ook aan de gebruikers van beide rapporten zijn deze punten voorgelegd. Over het geheel genomen beoordeelden de geraadpleegde informanten beide rapporten als adequaat. Zo gaven de schoolleiders aan, hun school te herkennen in het beeld dat de rapporten daarvan geven. Ze wezen er bijvoorbeeld op dat ze het knap vonden wat de inspecteurs na een bezoek van drie dagen over een school konden melden. Onder meer de uitgebreidheid van de rapporten van de Inspectie speelt daarbij een rol. In de woorden van een directeur: "De diepte van het IST doet recht aan de school en draagt bij aan de herkenbaarheid van het beeld." Daarbij hoeft het niet te gaan over informatie die nog niet aan de school bekend was: "Het signaleert dingen die je al wist. De meerwaarde zit in de toelichting, en in de beleidsmatige aangrijpingspunten om er vat op te krijgen en er iets mee te doen."

Een aspect van de adequaatheid van de gegevens is ook de vraag of het beeld van de school voldoende specifiek is, en of er onderwerpen worden gemist. Eén van de punten waarvoor de schoolleiders aandacht vroegen, is het niveau waarop de gepubliceerde gegevens betrekking hebben. Zo kwam als sterk punt van de IST-rapporten en voor een deel van de rapporten van de Evaluatiegroep naar voren dat aan afzonderlijke locaties recht

wordt gedaan. Rapporten over scholen, in plaats van over vestigingen, sluiten volgens de directies onvoldoende aan bij de uiteenlopende praktijken waarvan op de locaties sprake kan zijn. Daarnaast pleitte een aantal informanten voor een rapportage die naar de groepen binnen de school is uitgewerkt; soms kunnen er aanzienlijke verschillen bestaan tussen leraren of situaties in klassen, die wegvallen in het algemene beeld van de school. Wanneer daartoe aanleiding bestaat, besteden de rapporten van de Inspectie daar overigens aandacht aan. Verder vonden gebruikers dat de sociaal-emotionele ontwikkeling in de Inspectierapporten onvoldoende aan bod komt. Informatie daarover, zoals maten voor welbevinden, sociale integratie en pesten, is wel opgenomen in de rapporten van de Evaluatiegroep. Ook voor de creatieve ontwikkeling van leerlingen, zoals muzikale vorming, bestond volgens de schoolleiders weinig aandacht. Tevens misten schoolleiders oordelen over de rol van leerkrachten, en vonden ze meer aandacht voor het personeelsbeleid van belang. Over de IST/RST-rapporten gaf een aantal schoolleiders aan, duidelijke beoordelingscriteria te missen, "zodat je weet waar de Inspectie op let". Ondanks de beschikbaarheid van toelichtingen op de onderzoeken van de Inspectie, slaagt die er blijkbaar onvoldoende in om de scholen daar afdoende mee te bereiken.

De scholen gaven aan, tevreden te zijn over de juistheid van de informatie en de mogelijkheden om deze in te zien en zonodig te corrigeren. Dit geldt zelfs in zodanige mate dat enkele schoolleiders aangaven dat het oordeel van de Inspectie tot op zekere hoogte "maakbaar" is. In de woorden van één van hen: "Je kunt een onvoldoende ombuigen tot een voldoende." Zo kan volgens schoolleiders een onvoldoende op een indicator soms worden afgewend door snel maatregelen te nemen. Ook waren de informanten van mening dat de Inspectie via sommige indicatoren oordelen vormt over het onderwijsproces, terwijl het empirisch bewijs ontbreekt dat sommige processen beter of slechter zijn. Soms kan, zo gaven enkele directeuren aan, op een indicator onvoldoende worden gescoord, zonder dat eerder is aangetoond dat een bepaalde methode of manier van handelen ongeschikt is.

### *Gebruik en impact*

Als laatste gaan we in op de bruikbaarheid en impact van beide schoolrapporten. Wanneer we afgaan op de ervaringen van de sleutel-informanten, springt vooral de “spiegelfunctie” in het oog. De rapporten zetten nog eens op een rij wat de schoolleiders al dachten te weten. Het gebruik van verschillende bronnen geeft een breed beeld van de school. De uitgebreide en systematische wijze waarop dat gebeurt (Inspectie) en de vergelijking met andere scholen (Evaluatiegroep) ondersteunen en versterken dat zelfbeeld; ze functioneren zo als een check op het beeld dat directeuren zelf van de situatie hebben. De gebruikers vonden beide rapporten dan ook nuttige instrumenten.

Wel lijkt de impact te veranderen. De informanten waren van mening dat schoolrapporten een belangrijke rol hebben gespeeld bij een omslag in de aandacht voor kwaliteit. Ze waren een belangrijke prikkel voor meer aandacht voor de prestaties van de school. Inmiddels is die verandering gerealiseerd, zo gaf een aantal gebruikers aan, en neemt het belang van publieke kwaliteitsbeoordelingen af. In de woorden van een directeur: “De rapporten hadden een functie als omslag, om kritisch naar jezelf te kijken. In die zin verliezen ze nu hun functie. Nu missen we zonder deze rapporten niet zoveel.” Ook de rol die de gebruikers voor de rapporten van de Inspectie en die van de Evaluatiegroep te midden van andere informatiebronnen zien, wijst daarop. Scholen beschikken nu meestal over een leerlingvolgsysteem dat feedback geeft over de prestaties van leerlingen. Ook gebruiken scholen veelal diagnose-instrumenten en sterkte/zwakte-analyses, waaraan ze informatie ontleen die eenvoudig naar de schoolpraktijk vertaald kan worden. Zoals een van de schoolleiders het uitdrukte: “Schooldiagnose-instrumenten geven gemakkelijke handvatten voor verbetering.” De beschikbaarheid van andere informatiebronnen en de beperkte bruikbaarheid voor schoolverbetering leiden er dus toe dat schoolleiders de rapporten van de Inspectie en de Evaluatiegroep weliswaar waardevol vinden, maar het gewicht daarvan relativeren.

Dat ligt anders voor bestuurders, die aangaven de rapporten van groot belang te vin-

den. Zonder deze rapporten zijn ze voor hun oordeel vooral afhankelijk van de directeur en vormen ze hun beeld op basis van werkbezoeken, gesprekken met schoolleiders of toevallige persoonlijke contacten. Met de rapporten van de Inspectie en de Evaluatiegroep hebben besturen onafhankelijke gegevens voorhanden. Dat wordt nog versterkt door het beschikbaar komen van meerjarige informatie, zoals in het rapport van de Evaluatiegroep dat cijfers over een periode van negen jaar op een rij zet. Een bestuurder gaf aan dat deze “trends een goed beeld geven van de ontwikkeling van de school” en een nuttig instrument zijn om met de school in gesprek te gaan over de kwaliteit van het onderwijs.

Hoewel de impact van schoolprestatie-indicatoren geen criterium is voor de beoordeling van de kwaliteit van schoolrapporten, is het nuttig om inzicht te hebben in de mate waarin, en de manier waarop scholen de rapporten gebruiken. Als scholen deze niet of nauwelijks gebruiken, rijst de vraag of de opzet voldoende is toegesneden op de wensen van de school of de capaciteit om ze te gebruiken. Vanuit het oogpunt van de school en schoolverbetering is het dan ook belangrijk dat producenten van prestatie-indicatoren het gebruik dat scholen van de rapporten maken, regelmatig evalueren.

Van een schoolexterne impact als gevolg van de publicatie van prestatiegegevens merken de schoolleiders en bestuurders weinig. Zo zijn ouders, volgens de informanten, nauwelijks op de hoogte van publicaties over de kwaliteit van scholen. Wel speelt de informatie binnen de school een rol. Met name aan de rapporten van de Inspectie wordt binnen de school veel gezag gehecht. Ze fungeren onder meer als “definitie” van de situatie, met name ook in situaties waarin veranderingen worden nagestreefd. Zo vertelden schoolleiders en bestuurders een negatief Inspectieoordeel als “breekijzer” te gebruiken om veranderingen in de school door te zetten. Een nadeel van de rapporten van de Inspectie was volgens de gebruikers dat ze onregelmatig verschijnen, en de manier van rapporteren binnen korte tijd steeds verandert. Om dezelfde reden kenden de schoolleiders invloed toe aan de rapporten van de Evaluatiegroep.

Deze verschijnen elk jaar in steeds dezelfde vorm, zodat de informatie uit deze rapporten een plaats krijgt bij de beeldbepaling over de school en een rol kan spelen bij de bepaling van beleid.

Het belangrijkste vonden de schoolleiders echter dat de rapporten over de prestaties van de school maar weinig aanknopingspunten geven om tot verbetering te komen. Om van belang te kunnen zijn voor verbetering van het onderwijs was het volgens hen nodig dat de rapporten inzicht geven in wat er veranderd moet worden om de uitkomsten te verbeteren. De rapporten geven geen of onvoldoende inzicht in de relatie tussen de prestaties en het proces in de school. Dat bevestigt nog eens de analyse van Doolaard en Karstanje (2001), die wijzen op de beperkte bruikbaarheid van huidige vormen van prestatie-informatie voor schoolverbetering.

#### 4 Conclusie

In dit artikel zijn de accuratesse en bruikbaarheid van schoolprestatie-indicatoren onderzocht. Wanneer de indicatoren van de Inspectie met die van de Evaluatiegroep worden vergeleken, blijkt allereerst dat de Inspectie een grote terughoudendheid betracht in het geven van positieve of negatieve oordelen over schoolprestaties. Van de 31 in Groningen beoordeelde scholen scoren er slechts twee *onder of ruim boven niveau*. De overige 29 scholen scoren allemaal *op niveau*. De Evaluatiegroep hanteert andere criteria voor de indeling en laat 50% van de scholen in de midden categorie vallen. De overige scholen scoren boven- of onder gemiddeld. Per definitie geven de uitkomsten van de Evaluatiegroep dus meer spreiding te zien.

Bij twee scholen in Groningen velt de Inspectie wel een positief (*boven niveau*) of negatief (*onder niveau*) oordeel. Deze twee uitspraken worden slechts ten dele bevestigd door de gegevens van de Evaluatiegroep. De school die volgens de Inspectie *onder niveau* presteert, blijkt volgens de Evaluatiegroep ook het laagst te presteren. Daarbij moet worden aangetekend dat er nog een andere school uit Groningen overwegend laag scoort, maar die wordt door de Inspectie niet

als onderpresteerder aangemerkt. De enige school in Groningen die volgens de Inspectie *ruim boven niveau* presteert, scoort volgens de Evaluatiegroep in alle jaren in de midden categorie. In dit individuele geval wijkt het oordeel gebaseerd op meerniveau analyse (Evaluatiegroep) duidelijk af van de analyse op schoolniveau (Inspectie). Omgekeerd is er juist een school die volgens de Evaluatiegroep in alle jaren hoog scoort, maar die volgens de Inspectie toch *op niveau* presteert. Ook enkele andere scholen behoren volgens de Evaluatiegroep overwegend tot de categorie boven of hoog, maar worden daarvoor door de Inspectie niet beloond. Uit deze vergelijking valt dus op te maken dat er verschil is in classificatie tussen de schoolniveau- en de meerniveau-analyse.

In de bespreking van die analyses wordt ook de vraag opgeroepen hoe geschikt de schoolscoregroepen nu werkelijk zijn als referentiekader voor de prestaties van scholen. Die schoolscoregroepen blijken in ieder geval ongeschikt voor Groningen. Daar blijken autochtone achterstandsleerlingen (1.25) namelijk slechter te presteren dan allochtone achterstandsleerlingen (1.9). De heterogeniteit in de groep 1.0-leerlingen brengt verder met zich mee dat het leerlinggewicht wellicht de beste, maar voor de 1.0-groep geen adequate controlevariabele is.

Bij zowel de Inspectie als bij de Evaluatiegroep is er dus het streven om de toegevoegde waarde van scholen te berekenen. Daarbij treden echter twee verschillen op: de Inspectie gebruikt een correctie op groepsniveau, terwijl de Evaluatiegroep leerlingkenmerken op individueel niveau gebruikt. Het gebruik van de meerniveau-analyse maakt het vervolgens mogelijk daarvoor ook de adequate analyse uit te voeren. Onze analyse laat zien dat de werkwijze van de Inspectie tot conclusies kan leiden die verschillen van een directe, op individueel niveau uitgevoerde correctie voor de kenmerken van de leerlingpopulatie van een school.

Ouders, scholen, overheid en de publieke opinie zijn de mogelijke gebruikers van schoolprestatie-indicatoren. In dit artikel is de bruikbaarheid van de indicatoren van de Inspectie en de Evaluatiegroep voor scholen nagegaan. Daarvoor hebben we de standaard

den voor publicatie van schoolprestatiegegevens uit “Het oog der natie: scholen op rapport” (Dijkstra et al., 2001) geanalyseerd en naast de rapporten van de Inspectie en de Evaluatiegroep gelegd. Ook hebben we een groep gebruikers gevraagd naar de ervaringen die ze in de praktijk met beide publicaties hebben opgedaan. De geraadpleegde gebruikers waren van mening dat de rapporten van beide instanties tamelijk toegankelijk zijn. Ook vonden de informanten de rapporten van beide instanties adequaat; het beeld dat in de rapporten van scholen wordt geschetst, fungeert in veel gevallen vooral als controle of “autorisatie” van het eigen beeld. Onder meer de gedetailleerde en systematische rapportage van de Inspectie, en de mogelijkheid in de rapporten van de Evaluatiegroep om de eigen school met andere scholen en met zichzelf in de tijd te vergelijken, speelden daarbij een rol. De impact van de schoolprestatie-indicatoren, in het bijzonder die van de Inspectie, beschouwden de gebruikers als groot. Een tekort van beide publicaties is dat de rapporten geen inzicht geven in de relatie tussen proces en uitkomsten.

De toepassing van de door Visscher e.a. (2001) voorgestelde standaarden maakt duidelijk dat uitwerking voor de verschillende gebruikersgroepen zinvol lijkt. De formuleringen van Visscher en collega’s zijn met name toegesneden op gebruik door ouders. We stellen voor, in de standaarden expliciet tot uitdrukking te brengen dat publicaties voor verschillende gebruikersgroepen - ouders, scholen, besturen, overheid en samenleving - om uiteenlopende uitwerkingen vragen. Dat betekent dat de standaarden uit “Het oog der natie: scholen op rapport” worden uitgebreid met verwijzingen naar de relevante gebruikersgroepen. In deze bijdrage hebben we daarvoor op verschillende plaatsen enkele voorstellen gedaan als het gaat om het gebruik van kwaliteitsrapporten door scholen en de functie van schoolverbetering. We formuleren deze voorstellen voor een verdere uitwerking van de standaarden ten slotte nog eens expliciet. De standaarden voor begrijpelijkheid en bereik van Visscher e.a. (2001) zijn toegesneden op ouders en niet op andere gebruikers. Een meer nauwkeurige formulering van beide standaarden luidt:

B3 Schoolprestatie-indicatoren moeten zó worden verspreid dat overheid, scholen en met name ouders zo veel mogelijk worden bereikt.

B4 De wijze waarop de indicatoren worden gepresenteerd, evenals de toelichting daarop, dient wat begrijpelijkheid betreft te zijn afgestemd op de doelgroep (beleidsmedewerkers, bestuurders, schoolleiders, leerkrachten en ouders).

Toegespitst op het gebruik door scholen betekent Standaard B4 dat de gegevens voor bestuurders, schoolleiders en leerkrachten duidelijk, voor één uitleg vatbaar, zonder veel inspanning correct te interpreteren, en overzichtelijk en aantrekkelijk gepresenteerd moeten zijn.

Ten aanzien van de eis van zorgvuldige procedures en berekeningen stellen we voor, het recht van scholen op inzage van gegevens uit te breiden:

Z2 Om eventuele fouten in de publicaties te voorkomen, hebben scholen het recht om de gegevens die voor berekening van de indicatoren zijn gebruikt, en de conclusies die daarop worden gebaseerd, vóór publicatie in te zien.

Bezien vanuit het oogpunt van schoolverbetering zou er wat voor te zeggen zijn hierbij ook het element van herkenbaarheid te betrekken. Het is belangrijk dat de signalering van problemen en het afleiden van diagnoses op een dusdanige manier verloopt dat het draagvlak onder de direct betrokkenen wordt bevorderd. Het hoeft dan geen toelichting dat praktijken van ‘naming and shaming’, zoals lijsten in kranten met koppen als “de zwakste scholen van Nederland”, ontoereikend zijn voor realisering van schoolverbetering. Met het oog op schoolverbetering en een goede afstemming van publieke schoolprestatie-indicatoren op de wensen van scholen formuleren we ten slotte als *aanbeveling* dat producenten van schoolrapporten gehouden zijn tot regelmatige raadpleging van de gebruikers, met als doel de gepubliceerde schoolprestatie-informatie te evalueren, verbeteren en af te stemmen op de behoeften van scholen voor verbetering van het onderwijsproces.<sup>8</sup>

Ter afsluiting: onder invloed van de door de overheid gestimuleerde vergroting van de autonomie van scholen is sinds eind jaren ne-

gentig van de vorige eeuw de 'accountability'-eis veelal vertaald in de eis van het openbaar maken van schoolresultaten. Op deze wijze zouden gebruikers en financiers van het Nederlandse onderwijs op de hoogte worden gesteld van de resultaten van de onderwijsinvesteringen. Ouders zouden tot een betere schoolkeuze voor hun kinderen kunnen komen en het bevoegd gezag zou de benodigde beleidsinformatie verkrijgen. Vooral nog lijkt de interesse van ouders in de publicaties over schoolresultaten bescheiden, en aanwijzingen ontbreken dat ze van grote invloed zijn op schoolkeuze (Dijkstra & Witziers, 2001). Bevoegd gezag, Inspectie en schoolleiding zijn wel gebaat bij relevante informatie over de effecten van hun inspanningen. Zeker nu de Inspectie in het kader van de Wet op het Onderwijstoezicht van scholen vraagt om zelf informatie te geven over de onderwijskwaliteit, is het voor de schoolleiding van belang om gedegen geïnformeerd te zijn over accuratesse en bruikbaarheid van bestaande instrumenten.

## Noten

- 1 De zes scholen in Groningen die niet deelnemen, zijn een vrije school, een katholieke Jenaplanschool, twee reformatorische en twee gereformeerd vrijgemaakte scholen. In het Inspectiebestand ontbreken de gegevens van twee scholen. De reden is onbekend; op de website van de Inspectie van het Onderwijs staat bij deze twee scholen aangegeven dat er nog geen rapport beschikbaar is.
- 2 De standaarddeviatie op schoolniveau is voor de Cito-eindtoets 5.3 (Uiterwijk & Theunissen, 1997). Een afwijking van 2.5 standaard-scorepunt komt dus overeen met een gestandaardiseerde afwijking van een halve standaarddeviatie ten opzichte van de scholengroep waartoe een school behoort.
- 3 Volgens de richtlijnen van het Cito doen leerlingen die naar het speciaal onderwijs worden verwezen en leerlingen die de Nederlandse taal onvoldoende beheersen en korter dan vier jaar in Nederland verblijven, niet mee aan de Cito-eindtoets. Uit controle op de gegevens blijkt dat in 1% van de gevallen leerlingen toch hebben deelgenomen aan de eindtoets.

- 4 De indeling van scholen in schoolscoregroepen gebeurt sinds 1999; de Citogroep heeft zeven schooltypen onderscheiden naar samenstelling van de leerlingpopulatie: schoolscoregroep 1 betreft scholen met een populatie vrijwel geheel bestaand uit autochtone leerlingen met hoog opgeleide ouders, schoolscoregroep 7 betreft scholen met meer dan 75% allochtone leerlingen.
- 5 Een nieuwe ontwikkeling in de Rapportage Eindtoets Basisonderwijs (zie [www.citogroep.nl](http://www.citogroep.nl)) is dat scholen zelf per leerling leerlinggewichten kunnen invoeren. Dit biedt wellicht op termijn de mogelijkheid om landelijke scores te bepalen met behulp van meerniveau-analyses.
- 6 Het toezicht op de scholen en het daarmee verbonden systeem van rapportage is sinds 2003 georganiseerd via Jaarlijkse Onderzoeken (JO) en Periodieke Kwaliteitsonderzoeken (PKO). De overgang naar dit systeem, dat in de plaats kwam van de eerdere IST/RST-rapporten, was met name ingegeven door een streven naar toezicht dat meer dan voorheen is gebaseerd op proportionaliteit (intensivering van het toezicht al naar gelang de gerealiseerde onderwijskwaliteit daarom vraagt). (Voor meer informatie zie Inspectie van het Onderwijs, 2003.) Onverlet de verschillen tussen de IST/RST- en JO/PKO-onderzoeken zijn de inhoudelijke overeenkomsten tussen beide systemen zodanig, dat de op IST/RST gebaseerde gegevens die we in dit artikel presenteren ook voor de JO/PKO-rapporten relevant zijn. Nadat de Inspectie en scholen voldoende ervaring met de nieuwe vormen van rapportage hebben opgedaan, zou herhaling van de hier gepresenteerde analyses te overwegen zijn.
- 7 De rapporten van de Evaluatiegroep zijn niet openbaar, zodat deze eis hier minder speelt. Mocht een school fouten in de gegevens vermoeden, dan wordt nagegaan of dit het geval is en waar de fout zit, en wordt deze in overleg met de Evaluatiegroep hersteld en in het volgende rapport verwerkt.
- 8 De Evaluatiegroep, bijvoorbeeld, raadpleegt de gebruikers van de rapporten drie maal per jaar.

## Literatuur

- Bleker, L. A. W., Knuver, J. W. M., & Mulder, J. J. (2002). *Schoolrapport 2002 voor de Klokhuis-school*. Groningen: Evaluatiegroep voor het Onderwijs in Groningen.
- Bosker, R., Béguin, A., & Rekers-Mombarg, L. (2001). Hoe meten we de prestaties van een school? 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. 121-135). Assen: Van Gorcum.
- Dijkstra, A. B., Karsten, S., Veenstra, R., & Visscher, A. J. (Reds.) (2001). *Het oog der natie: scholen op rapport. Standaarden voor de publicatie van schoolprestaties*. Assen: Van Gorcum.
- Dijkstra, A. B., & Witziers, B. (2001). Kwaliteit van scholen en keuzeprocessen in het onderwijs. 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. 139-154). Assen: Van Gorcum.
- Doolaard, 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.
- Driessen, G., Langen, A. van, & Vierke, H. (2002). *Basisonderwijs: veldwerkverslag, leerlinggegevens en ouder vragenlijsten. Basisrapportage PRIMA-cohortonderzoek. Vierde meting 2000-2001*. Nijmegen: ITS.
- Evaluatiegroep voor het Onderwijs in Groningen (2001). *Overzicht Evaluatie Schoolrapportage Kwaliteitsindicatoren (interne publicatie)*. Groningen: ABCG.
- Griff, W. van de. (2001). Evaluatie van de kwaliteit van het basisonderwijs door de Inspectie. 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. 219-235). Assen: Van Gorcum.
- Inspectie van het Onderwijs. (2002). *Onderwijsverslag 2001*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs. (2003). *Toezichtkader Primair Onderwijs. Inhoud en werkwijze van het Inspectietoezicht conform de WOT*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs. (zonder jaartal). *Analyse en waardering van opbrengsten*. Interne notitie. Zonder plaats.
- Knuver, J. W. M., & Mulder, J. J. (2003). *Evaluatie basisonderwijs gemeente Groningen 2002*. Groningen: Evaluatiegroep voor het Onderwijs in Groningen.
- Ministerie van onderwijs, Cultuur en Wetenschappen. (2001). *Grenzeloos leren. Een verkenning naar onderwijs en onderzoek in 2010*. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschappen.
- Mulder, J. J., & Knuver, J. W. M. (2004). *Evaluatie basisonderwijs gemeente Groningen 2003*. Groningen: Evaluatiegroep voor het Onderwijs in Groningen.
- Scheerens, J. & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford etc.: Pergamon.
- Snijders, T. A. B. & Bosker, R. J. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. Newbury Park, CA: Sage.
- Uiterwijk, J. J., & Theunissen, T. J. J. M. (1997) *Verantwoording Eindtoets Basisonderwijs 1994*. Arnhem: Cito.
- Veenstra, R. (1999). *Leerlingen - klassen - scholen. Prestaties en vorderingen in het voortgezet onderwijs*. Dissertatie, Rijksuniversiteit Groningen. Amsterdam: Thesis Publishers.
- Veenstra, R., Dijkstra, A. B., Peschar, J. L., & Snijders, T. A. B. (1998). Scholen op rapport. Een reactie op het Trouw-onderzoek naar schoolprestaties. *Pedagogische Studiën*, 75, 121-134.
- Visscher, A. J., Dijkstra, A. B., Karsten, S., & Veenstra, R. (2001). Standaarden en aanbevelingen voor de publicatie van schoolprestatie-indicatoren. 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. 239-250). Assen: Van Gorcum.

Manuscript aanvaard: 27 augustus 2004



## Auteurs

**René Veenstra** is als onderzoeker verbonden aan de Vakgroep Sociologie van de Rijksuniversiteit Groningen.

**Lianne Bleker** en **Anja Knuver** zijn als onderzoekers verbonden aan de Evaluatiegroep voor het Onderwijs in Groningen.

*Correspondentieadres:* R. Veenstra, Sociologie, Faculteit PPSW, Rijksuniversiteit Groningen, Grote Rozenstraat 31, 9712 TG Groningen, e-mail: D.R.Veenstra@ppsw.rug.nl

## Abstract

### **Accuracy and usefulness of two publications of school achievement indicators**

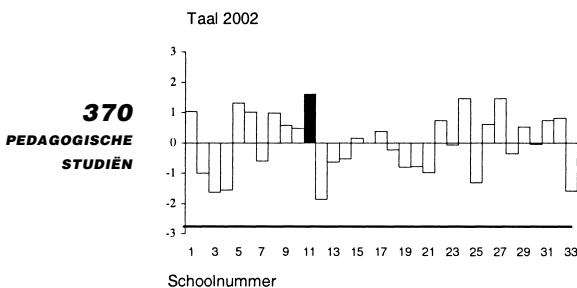
Accuracy and usefulness of two publications (produced by the Inspectorate of Education and the Educational Evaluation Group Groningen) have been judged from standards developed by Visscher, Dijkstra, Karsten, and Veenstra (2001). Accuracy has been checked by analyses on underlying data sets for both publications. Both publications appear to score relatively well on accuracy. However, for about a quarter of the schools the outcomes differ between the Inspectorate (using aggregated information) and the Evaluation Group (using multilevel analyses). To obtain insight into the usefulness of the two publications a small scale meeting was organized to gather information from school leaders and local authorities. Both publications seem to contain clear, recognizable and relevant information for school leaders and local authorities. However, both reports are inappropriate for informing parents or teachers without an extended clarification from the school leader. A weak point is also the lack of recommendations on the relation between process and outcomes.

## Appendix

### **Voorbeelden van gegevens uit het schoolrapport**

In de eerste grafiek staan de indicatoren van de 33 deelnemende Groningse scholen naast elkaar gepresenteerd. Deze indicatoren zijn de afwijkingen van het Gronings gemiddelde, berekend aan de hand van de Cito-eindtoetsscores. Het staafje van de betreffende school is ingekleurd. In de tweede grafiek staat de ontwikkeling van de scores op de verschillende onderdelen van de Cito-eindtoets voor één van de scholen weergegeven.

Harry Potterschool



Klokhuischool

