

# 'De weging gewogen'

Een verkenning naar de beoordelingssystematiek van de opbrengsten in het basisonderwijs en de daaraan verbonden gevolgen voor de onderwijspraktijk

Herziene versie, september 2008

Herman Godlieb

## Inhoudsopgave

1	Samenvatting.....	3
2	Inleiding.....	5
3	Stellingen.....	7
4	Onderzoek.....	8
4.1	Opzet van het onderzoek.....	8
5	De resultaten.....	11
5.1	'Vergelijkbare' scholen.....	11
5.1.1	Schoolvergelijking opleidingsniveau ouders versus weging.....	11
5.1.2	Spreiding van de scores.....	14
5.1.3	Zeer zwakke scholen.....	18
5.1.4	Rol van het Cito.....	19
5.1.5	Kwaliteitsbeleid of machtsmisbruik?.....	20
5.2	Kleine scholen.....	20
5.3	Opbrengsten en intelligentie.....	23
5.3.1	Een laag 'ambitieniveau'.....	24
5.3.2	Onderwijsextrinsieke factoren.....	24
5.3.3	Beperkte invloed, kleine verschillen.....	25
5.4	Schadelijke neveneffecten.....	25
5.4.1	De verschraling van het onderwijsaanbod.....	26
5.4.2	Bedreiging effectiviteit en rendement.....	26
5.4.3	Bedreiging WSNS en passend onderwijs.....	26
5.4.4	Manipulatie toetsscores.....	26
5.4.5	Misbruik Cito-eindtoets.....	27
5.4.6	Groei van de administratieve last.....	28
5.4.7	Groei bureaucratie.....	28
5.4.8	Stagnerende innovaties.....	28
5.4.9	Beschadigd imago.....	29
5.4.10	De schandpaal.....	29
5.4.11	Bedreiging diversiteit.....	29
6	Conclusies.....	30
6.1	Opbrengstwaardering onbetrouwbaar.....	30
6.2	Zeer zwakke scholen.....	30
6.3	Manipulatie Cito-eindtoets.....	30
6.4	Positie kleine scholen.....	30
6.5	Beperkte invloed.....	30
6.6	Schadelijke neveneffecten van een te sterk gerichte opbrengstcultuur.....	31
7	Aanbevelingen.....	32
7.1	Een voorlopige betrouwbaardere systematiek.....	32
7.2	Kleine scholen.....	33
7.3	Een realistisch toezicht.....	33
7.4	Wetenschappelijk onderzoek.....	33
7.5	Stoppen opbrengstterreur.....	35
8	Tot slot.....	36
9	Geraadpleegde bronnen:.....	38

# 1 Samenvatting

De inspectie van het basisonderwijs vergelijkt opbrengsten van scholen onderling en verbindt hier oordelen aan. Om te bepalen in hoeverre de werkwijze van de inspectie betrouwbaar is, is een bescheiden onafhankelijk onderzoek uitgevoerd. Voor dit onderzoek hebben 57 scholen, van totaal 69 jaargroepen, gegevens aangeleverd. Het betreft de scores van de Cito-eindtoets van 1063 leerlingen en het opleidingsniveau van beide biologische ouders.

Een correctiesystematiek gebaseerd op het opleidingsniveau van beide ouders scoort aanmerkelijk beter dan de correctie van de inspectie met behulp van het GLG (Gemiddeld Leerling Gewicht).

Met behulp van drie voorbeeldscholen, met alleen ongewogen leerlingen, wordt aangetoond dat de inspectie geen scherp beeld heeft van de kenmerken van de leerlingpopulatie van een school.

Daardoor is de beoordelingssystematiek ten aanzien van de opbrengsten, zoals die door de inspectie van het basisonderwijs wordt gehanteerd, ondeugdelijk.

De kennis van de kenmerken van de leerlingpopulatie is volstrekt onvoldoende om betrouwbare uitspraken te kunnen doen over de opbrengsten en de kwaliteit van het onderwijs.

Door het gebrekkige inzicht in de kenmerken van de leerlingpopulatie zijn vooral scholen met een niet gemiddelde populatie het slachtoffer van een onjuiste beoordeling van de opbrengsten.

De werkwijze van de inspectie aangaande kleine scholen is statistisch niet te verdedigen. Kleine scholen hebben op statistische gronden een grotere kans om als zeer zwak te worden beoordeeld dan grotere scholen.

De onderzochte zeer zwakke scholen scores, in vergelijking met de vergelijkingsgroep, nagenoeg in overeenstemming met de kenmerken van de leerlingpopulatie.

Het publiekelijk aan de schandpaal nagelen van scholen op basis van een aantoonbaar gebrekkige systematiek is verwerpelijk en laakbaar.

De veronderstelde relatie tussen opbrengsten en onderwijskwaliteit is niet staande te houden. In plaats van kwaliteit meet de inspectie populatieverschillen, statistische fouten en toetsmanipulaties.

De opbrengstterreur in combinatie met een onbetrouwbare waarderingssystematiek heeft een destructieve uitwerking op de feitelijke onderwijskwaliteit. Het betreft verschraling van het onderwijsaanbod, aantasting van: het pedagogisch klimaat, de efficiency, het imago, de veranderingscapaciteit, de motivatie en de diversiteit.

Daarnaast wordt de ontwikkeling van passend onderwijs geblokkeerd.

Een niet te onderschatten bedreiging van de kwaliteit van het onderwijs tengevolge van de imagobeschadiging is het gegeven, dat talentvolle jonge mensen onder de huidige omstandigheden niet meer kiezen voor een onderwijsloopbaan.

Het huidige risicogerichte onderwijstoezicht is onvoldoende gefundeerd en dient te worden herzien. Opschorting van de werking van de Wet op het Onderwijstoezicht betekent automatisch een krachtige kwaliteitsimpuls.

De invloed van goed onderwijs op schoolsucces wordt schromelijk overschat. Een complex van onderwijsextrinsieke factoren hebben een bepalende invloed. Bij de beoordeling van de opbrengstgegevens wordt met deze factoren onvoldoende rekening gehouden.

De eindconclusie kan niet anders zijn dan dat de inspectie zich dient te onthouden van het beoordelen van opbrengsten van basisscholen op grond van de huidige systematiek. Een juridisch traject behoort tot de mogelijkheden.

Nader wetenschappelijk onderzoek moet de basis vormen voor een realistisch onderwijstoezicht. De achterdochtige en wantrouwende benadering moet plaats maken voor een klimaat gebaseerd op ruimte en vertrouwen voor de professional.

## 2 Inleiding

Deze verkenning is gebaseerd op hetgeen de laatste jaren over onderwijs en kwaliteit is geschreven, op de eigen ervaringen uit de dagelijkse onderwijspraktijk en op een bescheiden onafhankelijk onderzoek.

Door de ervaringen als werker in het veld, is het mogelijk om het gevoerde onderwijsbeleid en de daaruit voortvloeiende besluiten te toetsen aan de feitelijke uitwerking op de werkvloer.

Er is in de afgelopen decennia veel verbeterd, maar er zijn ook talrijke besluiten genomen (meestal met de beste bedoelingen) die een fnuikende uitwerking hebben gehad en hebben op de dagelijkse onderwijspraktijk.

In zijn algemeenheid kan worden gesteld dat veel maatregelen sterk zijn gericht op het verbeteren van een deelaspect, maar dat de samenhang met en de gevolgen ervan voor het totale systeem over het hoofd worden gezien.

Door de niet aflatende stroom negatieve publicaties over onderwijs is in ons land een vertekend beeld ontstaan over de kwaliteit van dat onderwijs. De ene dag blijkt uit onderzoek dat het rekenonderwijs niet deugt, een volgende dag levert het leesonderwijs te weinig op. Als werker in het onderwijs komt de vraag boven: "Wat doen we fout? Hoe heeft het zover kunnen komen?"

De inspectie legt sterk de nadruk op kwaliteit en opbrengsten en maakt lijstjes met zeer zwakke scholen. Uit de berichtgeving op de inspectie-site spreekt de grondgedachte, dat wanneer scholen maar beter hun best doen, opbrengsten kunnen worden verhoogd. Het onderwijsverslag van 2007 is er eveneens helder over. Scholen halen niet uit kinderen wat erin zit, staat er letterlijk te lezen.

Uit onderzoek blijkt dat scholen met een 'vergelijkbare' leerlingpopulatie heel verschillend kunnen scoren. Dit leidt tot de voor de hand liggende conclusie, dat de ene school het dus blijkbaar beter doet dan de andere.

De bedoeling van de overheid is om door middel van het be- en veroordelen van scholen, gevolgd door verscherpt toezicht en specifieke ondersteuning, de kwaliteit te verbeteren c.q. de opbrengsten te verhogen. Maar gaat dat ook lukken? Of is het onderwijs slachtoffer van overambitieuze verwachtingen. Waar liggen de grenzen van de maakbaarheid? Is het onderwijs een soort product dat oneindig verbeterd kan worden?

Binnen het onderwijs leeft het besef dat het altijd beter kan, maar er is in het veld ook een breed gevoel dat er onrechtmatigheden zitten in de beoordelingssystematiek, zoals die door de inspectie van het basisonderwijs wordt gehanteerd.

Deze verkenning en het daaraan gekoppelde onderzoek heeft niet het oogmerk aan te tonen dat er geen zwakke of zwakkere scholen bestaan. Hoofddoel is te onderzoeken of de gebruikte systematiek voldoende betrouwbaar is.

Een andere interessante vraag is of 'tegenvallende' opbrengsten inderdaad te wijten zijn aan zwak onderwijs of dat er andere factoren zijn die een grote(re) rol spelen.

Onder verantwoordelijkheid van staatssecretaris Netelenbos werd begin negentiger jaren het kwaliteitsbeleid serieus ter hand genomen. Vanuit de grondgedachte dat alles altijd beter kan, werd kwaliteitsbeleid een speerpunt. Vele brochures verschenen waar het woord kwaliteit tientallen malen in stond te lezen, zonder dat duidelijk werd gemaakt wat er nu eigenlijk mee werd bedoeld.<sup>1 1</sup>

Er bestond in het onderwijs de vrees dat ten aanzien van het begrip kwaliteit uitsluitend zou worden gekeken naar de hoogte van de output. Mede op basis van deze bezorgdheid is er in samenspraak met alle bij het onderwijs betrokken instanties en organisaties een uitgebreid toezichtskader ontwikkeld met een reeks indicatoren, die de basis zijn gaan vormen voor de beoordeling van de kwaliteit van scholen. Desondanks is de laatste jaren steeds sterker de nadruk komen te liggen op het onderdeel opbrengsten. Feitelijk is dit aspect het voornaamste onderdeel geworden voor het bepalen van risico's ten aanzien van de onderwijskwaliteit en het daarop gebaseerde toezichts-arrangement. Het breed opgezette indicatorensysteem is naar de achtergrond gedrongen. Ook is gebleken dat veel indicatoren nauwelijks relevant zijn voor dat, waar het blijkbaar écht om gaat: de opbrengsten.<sup>II</sup>

Doordat het toezicht zich momenteel in sterke mate richt op de opbrengsten, is de vrees die leefde bij de start van het kwaliteitsbeleid, uitgekomen.

Doordat de huidige systematiek voor scholen vérstrekkende gevolgen kan hebben, verdient de werkwijze een kritische analyse.

In een uitgave van de inspectie van het onderwijs met de titel 'Analyse en waardering van opbrengsten' wordt beschreven op welke wijze scores dienen te worden geïnterpreteerd en gewaardeerd. De toetsing van de werkwijze van de inspectie ten aanzien van met name de Cito-eindtoets, zoals beschreven in het aangehaalde document, is een belangrijk onderdeel van deze verkenning.<sup>III</sup>

Onderwijsmensen werken over het algemeen hard, met enthousiasme en passie. Veel van wat er zoal over het onderwijs wordt gezegd en geschreven laten de veldwerkers met een zekere gelatenheid langs zich heen glijden. Geen onbegrijpelijke reactie. Het gevaar bestaat echter dat de golf van negatieve aandacht het fundament van het onderwijs dreigt weg te spoelen.

De inbreng en het verweer uit het veld zullen krachtiger moeten klinken. De staatssecretaris van Onderwijs, Cultuur en Wetenschap mevr. Sharon Dijksma, nodigt het onderwijsveld nadrukkelijk uit deel te nemen aan de uitwerking van en de discussie over de Kwaliteitsagenda Primair Onderwijs. Deze verkenning hoopt hieraan een bijdrage te leveren.

Een bijzonder woord van dank is er voor de scholen die hebben meegewerkt aan het genoemde onderzoek door het beschikbaar stellen van relevante gegevens. Ook kan ir. Willem Godlieb niet onvermeld blijven voor het bewerken van de data.

Nieuwe Pekela, augustus 2008

Herman Godlieb, directeur obs De Linde/Theo Thijssen te Nieuwe Pekela.

hermangodlieb@deweginggewogen.nl

---

<sup>1</sup> I verwijst naar de geraadpleegde bronnen

### 3 Stellingen

Deze verkenning is opgebouwd aan de hand van een viertal stellingen:

1. *Een belangrijk deel van de verschillen in opbrengsten tussen 'vergelijkbare' scholen wordt niet in hoofdzaak verklaard door verschillen in onderwijskwaliteit, maar door verschillen in de leerlingpopulatie.*
2. *Door onvoldoende rekening te houden met de schoolgrootte worden kleine scholen ten onrechte in de categorie zwakke scholen ingedeeld.*
3. *Vooraf onderwijsextrinsieke factoren bepalen het succes van onderwijs.*
4. *De sterke nadruk van de overheid op de opbrengsten veroorzaakt een reeks ongewenste effecten, die het onderwijs schade toebrengt.*

## 4 Onderzoek

Om met enige grond iets te kunnen zeggen over de geponeerde stellingen is er een bescheiden onafhankelijk onderzoek uitgevoerd. Voor dit onderzoek moest van een aanzienlijke groep schoolverlaters de Cito-eindtoets worden verzameld en het opleidingsniveau van beide biologische ouders. Het zou wenselijk zijn geweest van enkele tienduizenden leerlingen over deze gegevens te kunnen beschikken. Binnen de opzet van dit onderzoek is dat niet haalbaar gebleken en was het streven om van ten minste 1000 leerlingen gegevens te verzamelen.

### 4.1 Opzet van het onderzoek

In totaal zijn er enkele honderden scholen, verspreid over het land, via de mail benaderd met het verzoek gegevens aan te leveren over de jaren 2007 en 2008. In totaal hebben 57 scholen het invulformulier ingevuld geretourneerd van in totaal 76 jaargroepen met de gegevens van 1063 leerlingen. De gegevens van de leerlingen zijn derhalve niet afkomstig van een willekeurige groep scholen, maar van scholen die vrijwillig gegevens beschikbaar hebben gesteld. Omdat het onderzoek zich in beginsel sterk richt op de correlatie opleidingsniveau van de ouders en de schoolprestaties van de individuele leerling, is het niet echt wezenlijk van welke school de gegevens afkomstig zijn.

Scholen worden op basis van het percentage ongewogen leerlingen ingedeeld in schoolgroepen. In dit verband zijn de schoolgroepen 1 t/m 4 van belang.

Schoolgroep 1	100%	1.0 lln.
Schoolgroep 2	76 - 99%	1.0 lln
Schoolgroep 3	51 - 75%	1.0 lln
Schoolgroep 4	<50%	1.0 lln
	<25%	1.9 lln

In tabel 1 wordt de verdeling per schoolgroep aangegeven. Het merendeel behoort tot schoolgroep 2 en 3.

Tabel 1: Aantal jaargroepen per schoolgroep in het databestand

<b>schoolgroep</b>	<b>aantal</b>	<b>perc</b>
1	10	14%
2	29	42%
3	17	25%
4	13	19%

Het totale databestand valt gemiddeld net in schoolgroep 3. Uit tabel 2 blijkt dat 73% van de leerlingen valt in de categorie ongewogen leerlingen.

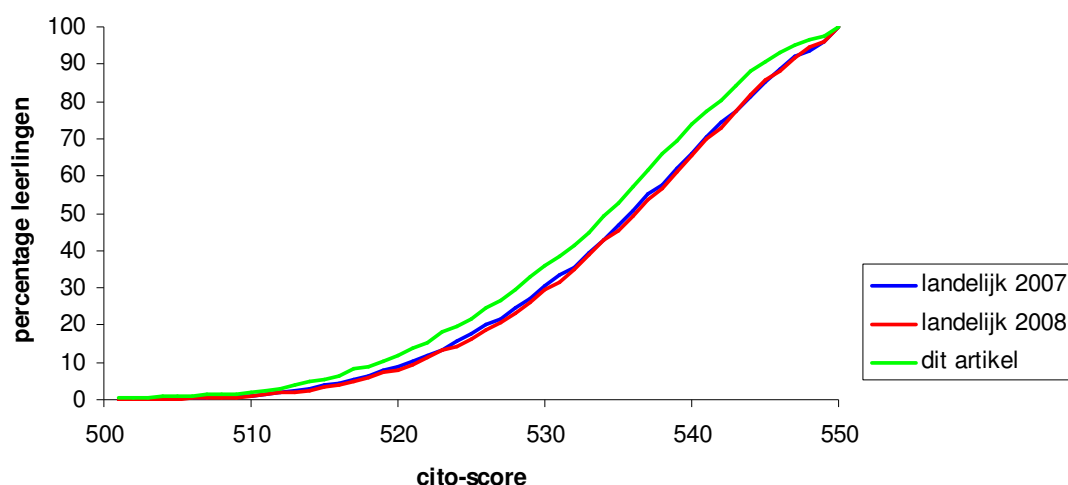


Tabel 2: Verdeling leerlingen per wegingscategorie en de daarbij behorende gemiddelde Cito-score.

weging	gem	aantal	perc
1	534,7	776	73%
1,25	527,3	222	21%
1,9	531,6	65	6%

Het percentage gewogen leerlingen verklaart dat de gemiddelde score van het totale bestand iets onder het landelijk gemiddelde ligt, daarnaast zijn leerlingen met een PRO-beschikking of met LWOO meegenomen. In het bestand zijn geen scholen opgenomen die door de inspectie zijn aangemerkt als zeer zwakke school.

Figuur 1 - Cumulatieve Cito-scores van dit artikel en de landelijke gegevens van 2007 en 2008.



Uit de vergelijking met de landelijke cijfers van het Cito uit 2007 en 2008 <sup>IV</sup> blijkt ook dat de gemiddelde score van het databestand enkele punten lager ligt. Ook zijn er naar verhouding weinig leerlingen met scores boven de 545. De lijn vertoont verder nagenoeg dezelfde curve, waaruit mag worden geconcludeerd dat het gebruikte databestand voldoende omvangrijk is.

De aanname is gerechtvaardigd dat het een verzameling leerlingen is van gemiddelde scholen.

Voor het bepalen van het opleidingsniveau is gebruik gemaakt van de gegevens uit de gewichtenregistratie van de school.

Middels een digitaal antwoordformulier hadden de respondenten de keuze uit totaal 17 categorieën om het opleidingsniveau aan te geven. Te weten; geen, vso, lbo, mavo, havo, vwo, mbo, hbo en wo.

Bij de acht genoemde opleidingen kon ook worden aangegeven of de opleiding al dan niet was afgesloten met een diploma.

Voor een doelmatige statistische verwerking is er voor gekozen bepaalde opleidingen te clusteren tot één categorie. In totaal is er sprake van 6 relevante categorieën: geen, lbo, mavo, mbo, havo en hbo.

De invulling van de categorieën is als volgt:

- geen - geen, vso, lbo zonder diploma
- lbo - lbo met diploma
- mavo - mavo met en zonder diploma
- mbo - mbo met en zonder diploma
- havo - havo met en zonder diploma
- hbo - vwo, hbo en wo

Naast het opleidingsniveau van beide biologische ouders en de standaardscore van de Cito-eindtoets is ook de wegingsfactor (oude weging) verzameld. Voor een deel van de leerlingen is de score van de Nederlandse Intelligentietest Onderwijsniveau (NIO) ook bekend.

Tabel 3 geeft een overzicht van de aantallen ouders per opleidingsniveau uit het databestand. In tabel 4 zijn de aantallen opleidingscombinaties weergegeven. De aantallen per klasse zijn over het algemeen voldoende groot.

Tabel 3- Aantallen ouders per opleidingsniveau

	<b>moeder</b>	<b>vader</b>
geen	83	66
lbo	257	328
mavo	166	89
mbo	299	270
havo	83	46
hbo	137	178

Tabel 4 - Aantallen oudercombinaties per opleidingsniveau

	<b>geen</b>	<b>lbo</b>	<b>mavo</b>	<b>mbo</b>	<b>havo</b>	<b>hbo</b>
geen	38					
lbo	26	154				
mavo	10	87	33			
mbo	10	91	43	148		
havo	1	34	23	27	10	
hbo	10	17	19	84	20	79

## 5 De resultaten

### 5.1 'Vergelijkbare' scholen

Stelling 1:

*Een belangrijk deel van de verschillen in opbrengsten tussen 'vergelijkbare' scholen wordt niet in hoofdzaak verklaard door verschillen in onderwijskwaliteit, maar door verschillen in de leerlingpopulatie.*

#### 5.1.1 Schoolvergelijking opleidingsniveau ouders versus weging

Cito-scores van basisscholen worden door de inspectie gecorrigeerd op instroom. Voor deze correctie wordt het Gemiddeld Leerling Gewicht (GLG) gebruikt. Het GLG is afgeleid van aantal en soort gewogen leerlingen, gerelateerd aan het opleidingsniveau van de hoogst opgeleide ouder.

Voor de volledigheid is het van belang vast te stellen dat de weginggegevens, waarop de GLG is gebaseerd, niet zijn verzameld met het oogmerk de populatie van een school nauwkeurig in beeld te brengen. De gegevens zijn verzameld met het doel additionele middelen voor de bestrijding van achterstanden te verdelen over scholen.<sup>2</sup>

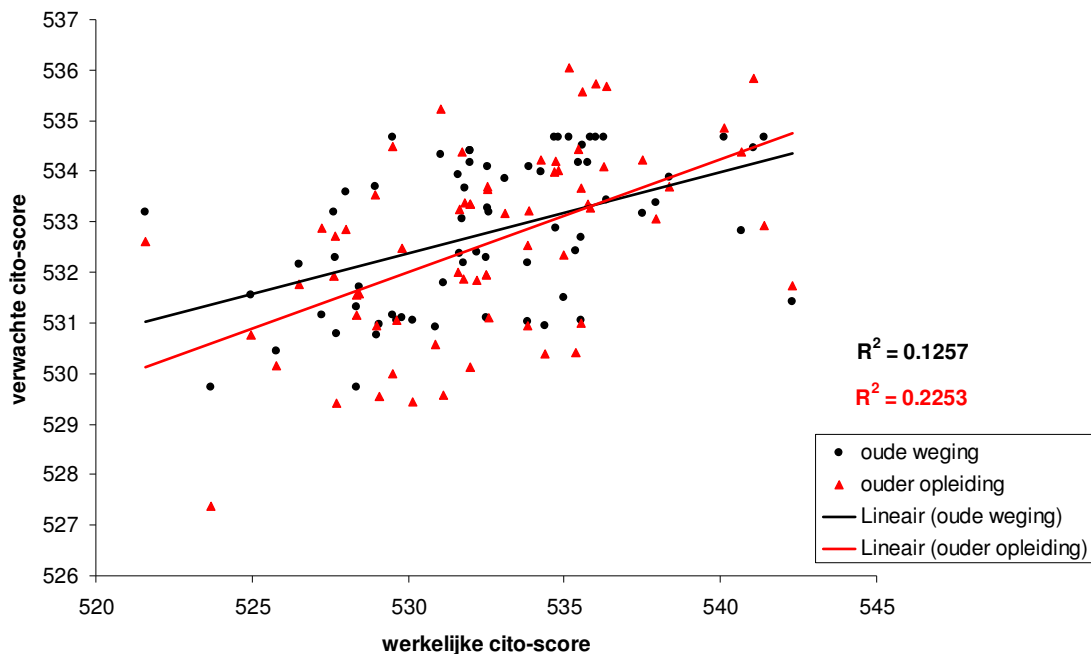
Daarnaast zijn er kanttekeningen te plaatsen bij de validiteit van de leerling-weging zelf. Scholen hebben er baat bij de weging in een bepaalde richting te buigen, omdat deze van invloed is op de hoogte van de bekostiging. Anderzijds spiegelen bepaalde groepen ouders hun opleidingsniveau vaak rooskleuriger voor, dan die in werkelijkheid is. Controle hierop is lastig.

---

<sup>2</sup> Doordat ten tijde van de herziening van de gewichtenregeling bekend was dat de weginggegevens ook worden gebruikt voor het bepalen van de kenmerken van de leerlingenpopulatie van een school, was het mogelijk geweest de wegingssystematiek hierop aan te passen. Hiertoe is niet besloten. Een gemiste kans. (\* 2, 6)

In figuur 2 wordt een vergelijking gemaakt tussen de weginggegevens (GLG - oude weging) en een nieuwe systematiek waaraan het opleidingsniveau van beide biologische ouders ten grondslag ligt. Voor elk van de 21 combinaties wordt de Cito-score gebruikt uit tabel 5. Daarnaast is voor kinderen met slechts één of geen ouderinformatie een 22<sup>ste</sup> categorie geformeerd met een gemiddelde Cito-score van 532.9. Deze categorie betreft 9% van het bestand.

Figuur 2 – Correlatie tussen verwachte scores en werkelijke gemiddelde Cito-score voor alle groepen uit het bestand op basis van de oude weging.

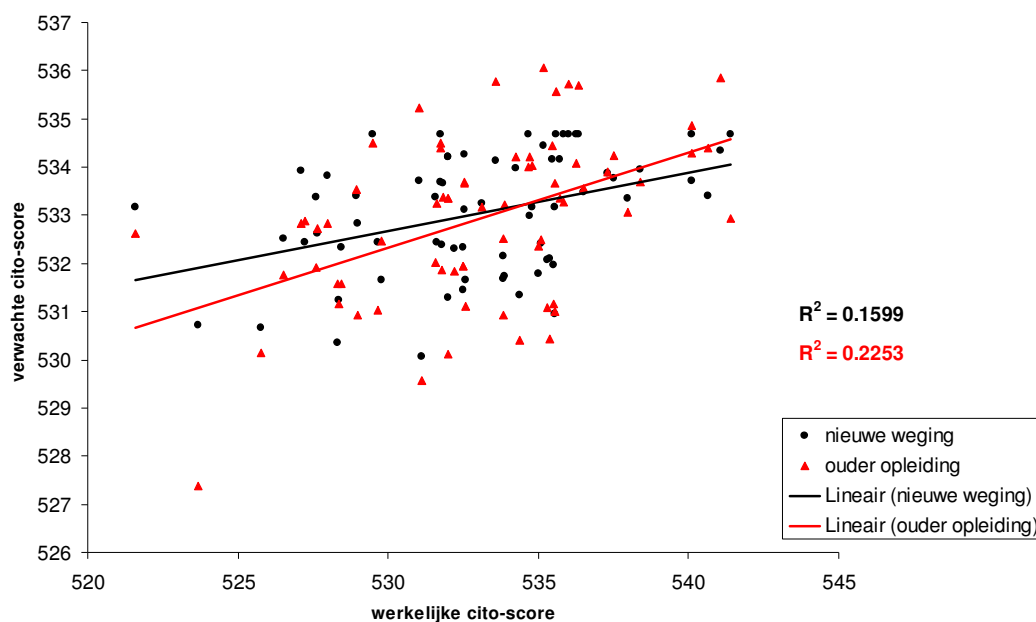


Elke zwarte stip komt overeen met een van de 69 jaargroepen uit het bestand. Voor de zwarte stippen is de verwachte waarde gebaseerd op de gemiddelde Cito-score van leerlingen met dezelfde weging. Voor de rode driehoekjes is de verwachte waarde gebaseerd op de gemiddelde score van leerlingen met dezelfde oudercombinatie uit het databestand.

De correlatie(R) voor de voorspelling gebaseerd op het opleidingsniveau van beide ouders is 0,475. Dit is aanzienlijk hoger dan de voorspelling gebaseerd op de oude weging (R=0,355).<sup>3</sup> Niet verwonderlijk, wanneer we bedenken dat de weging van de inspectie uitgaat van de hoogst genoten opleiding van één van de ouders én er vanaf mavo-niveau geen onderscheid meer wordt gemaakt.

<sup>3</sup> De exacte berekeningssystematiek van het GLG kon niet worden toegepast. De hoofdinspectie bleek niet in staat om de gegevens aan te leveren en het Cito is hiertoe niet bereid gebleken.

Figuur 2a – Correlatie tussen verwachte scores en werkelijke gemiddelde Cito-score voor alle groepen uit het bestand op basis van de nieuwe weging.

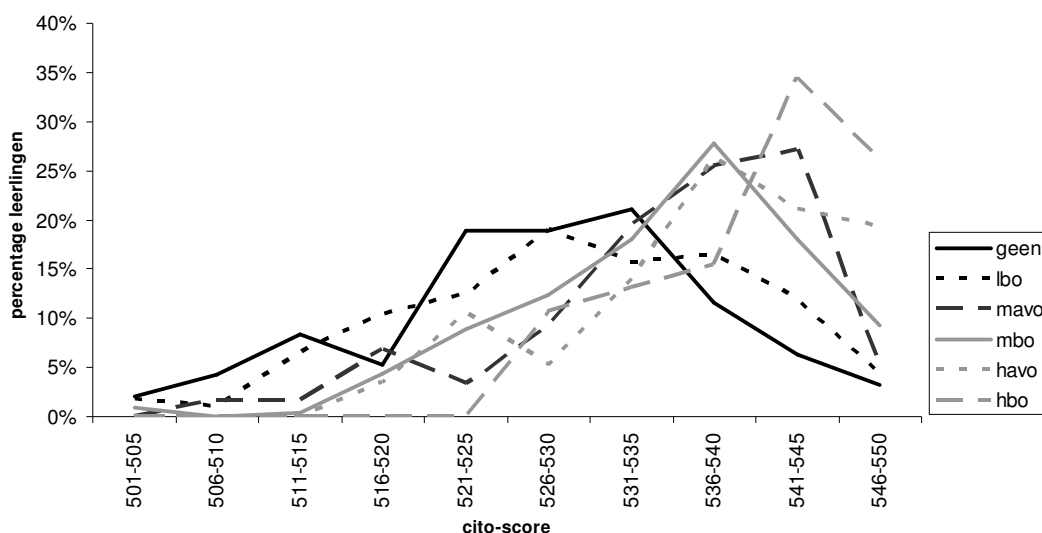


Voor de volledigheid is dezelfde vergelijking ook gemaakt op basis van de nieuwe gewichtenregeling. De correlatie(R) stijgt van 0,355 naar 0,400. Een lichte verbetering die wordt veroorzaakt doordat het aspect etniciteit in de nieuwe gewichtenregeling geen rol meer speelt. Van een vernieuwd GLG, waaraan de nieuwe weging ten grondslag ligt, hoeven we dan ook geen wonderen te verwachten. Ondanks dat het opleidingsniveau van ouders de beste voorspeller is van schoolsucces, is de correlatie in alle gevallen laag.

Ook zijn de opbrengstgegevens van verschillende jaargangen van dezelfde school vergeleken. De gemiddelde Cito-scores van deze jaargroepen liggen in extreme gevallen wel 5 punten uiteen, nadat er gecorrigeerd is voor het opleidingsniveau van beide ouders. In vergelijking met de door de inspectie gehanteerde marge van 4.6 punt tussen onvoldoende en goed is 5 punten een erg groot verschil. Het is volstrekt onwaarschijnlijk dat de kwaliteit van het onderwijs dermate fluctueert op een school, dat dergelijke grote verschillen er door verklaard kunnen worden.

## 5.1.2 Spreiding van de scores

Figuur 3 - Spreiding Cito-scores per categorie op basis van de laagst opgeleide ouder



Figuur 3 laat de spreiding zien van de Cito-scores per categorie op basis van het opleidingsniveau van de laagst opgeleide ouders. Vanzelfsprekend valt er uit op te maken dat het opleidingsniveau van de ouder van invloed is op de hoogte van de score. Er is evenwel sprake van een grote spreiding. Kinderen van zeer laag opgeleide ouders kunnen 550 scoren en omgekeerd kunnen kinderen van zeer hoog opgeleide ouders 525 scoren.

De Cito-toets discrimineert beter voor de zwakke en gemiddelde leerling dan voor de toppers. De spreiding van Cito-scores voor kinderen van de hbo categorie stopt abrupt. De Cito-scores leveren niet de informatie die noodzakelijk is voor de inspectie om te kunnen bepalen of scholen met veel hoog opgeleide ouders voldoende hoog scoren. Op zich geen tekortkoming van de eindtoets, omdat deze informatie gelet op het oogmerk van de toets, niet relevant is.

Tabel 5 - Gemiddelde Cito-scores per oudercombinatie.

	geen	lbo	mavo	mbo	havo	hbo
geen	529.5					
lbo	523.5	527.1				
mavo	526.3	531.0	533.3			
mbo	526.1	531.8	534.7	534.9		
havo	535.0	532.4	537.0	535.0	538.5	
hbo	533.2	534.8	536.6	536.1	540.4	540.5

Tabel 5 geeft een beeld van de gemiddelde Cito-scores betreffende de leerlingen van de diverse oudercombinaties. In het grijze gebied staan de scores van de leerlingen met oudercombinaties die door de inspectie worden aangemerkt als 1.25 leerlingen. Het is opmerkelijk te noemen dat de

combinaties geen-mavo en geen-mbo gemiddeld lager scoren dan de combinaties lbo-lbo, terwijl de weging voor deze categorie 1.0 is.

Het gegeven dat de geen-geen combinatie hoger scoort dan bijvoorbeeld de geen-lbo groep, kan worden verklaard doordat de geen-geen combinatie voor 39% bestaat uit 1.9 leerlingen.

Een ander opvallend gegeven is het verschijnsel dat de verticale kolommen minder variatie laten zien dan de horizontale rijen. Dit betekent dat de laagst opgeleide ouder meer invloed heeft op de resultaten van het kind dan de hoogst opgeleide ouder. De wegingsystematiek van de overheid gaat daarentegen juist uit van de hoogst opgeleide ouder.

Tabel 6 - Gemiddelde Cito-scores per categorie op basis van de hoogst opgeleide ouder

<b>categorie</b>	<b>cito-score</b>
geen	529.5
lbo	526.6
mavo	531.2
mbo	533.8
havo	534.9
hbo	537.8

Tabel 6 geeft een overzicht van de verschillen tussen gemiddelde Cito-scores per opleidingsniveau gerelateerd aan de hoogst opgeleide ouder. Het verschil tussen de score van mavo en bijvoorbeeld hbo bedraagt 6.6 Cito-punten.

Alle kinderen waarvan één ouder ten minste een mavo-opleiding heeft genoten, worden door het wegingsysteem in dezelfde categorie geplaatst (1.0 leerlingen).

Om aan te tonen dat scholen met een voor de inspectie vergelijkbare populatie enorm kunnen verschillen, vergelijken we een drietal fictief geconstrueerde scholen.

Het betreft de scholen A, B en C. De scholen hebben **geen** gewogen leerlingen en de groepsgrootte is steeds 25.

Leerlingen waarvan van één of geen ouder het opleidingsniveau bekend is, worden niet gebruikt bij het formeren van deze fictieve scholen.

**A** - Op school A hebben de ouders maximaal een mavo-opleiding. Het betreft de combinaties: lbo-mavo en mavo-mavo.

**B** - Op school B hebben de ouders minimaal een havo-opleiding. Het betreft de combinaties: havo-havo, hbo-havo en hbo-hbo.

**C** - School C is een gemiddelde school met leerlingen geformeerd uit alle oudercombinaties van 1.0 leerlingen (wit gekleurd in tabel 5).

De inspectie hanteert onder- en bovengrenzen voor het waarderen van opbrengsten. Deze grenzen zijn gebruikt om te bezien hoe groot de kansen van de scholen A, B en C zijn om ten aanzien van de opbrengsten als zwak, voldoende of goed te worden beoordeeld.

Ter verduidelijking:

Scoort een school 3 jaren achtereen gemiddeld onder de ondergrens, dan wordt deze school als *zwak* betiteld. Als de gemiddelde score 3 jaar achtereen boven de bovengrens ligt, wordt de school als *goed* beoordeeld.

De fictieve klassen zijn willekeurig gevuld met leerlingen uit het totale bestand van 1.0 leerlingen met de desbetreffende oudercombinaties. Totaal zijn er 3000 groepen per fictieve school geformeerd. Steeds 1000 voor elk afzonderlijk schooljaar.

De inspectie hanteert verschillende grenzen voor verschillende schoolgroepen.

Tabel 7: Onder- en bovengrenzen gehanteerd door de inspectie voor schoolgroep 1, 2 en 3 voor 2007.

	<b>ondergrens</b>	<b>bovengrens</b>
schoolgroep 1	535,1	539,7
schoolgroep 2	533,8	538,4
schoolgroep 3	531,9	536,5

Aangezien er uitsluitend 1.0 leerlingen op onze fictieve school zitten, moeten de grenzen die van toepassing zijn op schoolgroep 1 worden gehanteerd. Onder verwijzing naar de samenstelling van ons databestand (tabel 1), valt het gehele databestand nagenoeg op de grens van schoolgroep 2 en 3. Alhoewel slechts 1.0 leerlingen zijn opgenomen in de fictieve scholen, zijn toch ook de resultaten voor de marges van schoolgroep 2 en van schoolgroep 3 vermeld.

Doordat alleen leerlingen zijn gekozen van normale gemiddelde scholen, moeten de scholen in alle gevallen *voldoende* scores. De tabellen 8a, b en c geven een ander beeld.

Tabel 8a: Kans op kwalificatie door inspectie voor school A, B en C voor de inspectienorm voor schoolgroep 1

	<b>zwak</b>	<b>voldoende</b>	<b>goed</b>
school A (mavo-)	82%	18%	0%
school B (havo+)	0%	93%	7%
school C (combi)	21%	79%	0%

Tabel 8b: Kans op kwalificatie door inspectie voor school A, B en C voor de inspectienorm voor schoolgroep 2

	<b>zwak</b>	<b>voldoende</b>	<b>goed</b>
school A (mavo-)	46%	54%	0%
school B (havo+)	0%	59%	41%
school C (combi)	4%	96%	0%



Tabel 8c: Kans op kwalificatie door inspectie voor school A, B en C voor de inspectienorm voor schoolgroep 3

	<b>zwak</b>	<b>voldoende</b>	<b>goed</b>
school A (mavo-)	6%	94%	0%
school B (havo+)	0%	8%	93%
school C (combi)	0%	100%	0%

De bovenstaande drie tabellen laten grote verschillende zien. Voor school C (Combi), berekend voor schoolgroep 3, laat tabel 8c een uitkomst zien, die in alle gevallen de verwachte uitkomst had moeten zijn. De vergelijking laat afwijkingen zien van minimaal 4% tot maximaal 82% om onterecht als zwak beoordeeld te worden ten aanzien van opbrengsten.

Voor de 1000 A-scholen (mavo-) valt de gemiddelde Cito-score in 46% van de gevallen 3 jaar achtereen onder de ondergrens voor schoolgroep 2 en zou in dat geval als *zwak* worden beoordeeld ten aanzien van het aspect opbrengsten. De A-scholen (mavo-) hebben geen enkele kans om ooit als *goed* te worden gekwalificeerd.

Voor de 1000 B-scholen (havo+) zijn de resultaten omgekeerd. Deze scholen hebben een grote kans om onterecht als *goed* te worden beoordeeld. Volgens de inspectie leveren de B-scholen kwaliteit, terwijl zij slechts 'geluk' hebben met de samenstelling van de schoolpopulatie.

De C-scholen (combi) in schoolgroep B hebben een kans van 4% om onterecht als *zwak* te worden beoordeeld op opbrengsten.

#### **Resumerend:**

Ongeacht met welke maat er wordt gemeten heeft School A (mavo-) een forse kans om onterecht als *zwak* te worden beoordeeld ten aanzien van de opbrengsten.

Zelfs bij de strengste norm heeft School B (havo+) een kans van 7% om onterecht als *goed* te worden beoordeeld ten aanzien van de opbrengsten.

School C (combi) met een gemiddelde populatie heeft geen enkele kans om onterecht als goed te worden beoordeeld ten aanzien van opbrengsten. Daarentegen is er wel een kans dat ze onterecht als zwak beoordeeld wordt.

De nieuwste werkwijze van de toezichthouder is gebaseerd op een risicoanalyse aangaande de opbrengsten, de jaarstukken en eventuele signalen.

In het geval school B (havo+) de schoolgids en het schoolplan tijdig indient, ziet men de inspectie in geen vier jaar. School A (mavo-) heeft een aanzienlijke kans dat de inspectie op de stoep staat. De inspectie zal vanuit de filosofie dat lage opbrengsten zijn te wijten aan zwak onderwijs op zoek gaan naar bewijzen voor deze filosofie. De inspecteur zal altijd op zaken stuiten die in de ogen van de inspectie niet 100% zijn. Alle lijstjes met indicatoren ten spijt, een inspecteur is ook maar een mens.

School B (havo+) zal tijdens het vierjaarlijkse schoolbezoek door de inspectie met hele andere ogen worden bekeken. De opbrengsten zijn immers van hoog niveau, dus moet het onderwijs wel goed zijn.

(Het zou buitengewoon interessant zijn te onderzoeken hoe de kwaliteit van een willekeurige school wordt beoordeeld in het geval de betreffende inspecteur niet op de hoogte zou zijn van de opbrengsten van deze school).

De stelling dat een belangrijk deel van de verschillen in opbrengsten tussen 'vergelijkbare' scholen niet in hoofdzaak wordt verklaard door verschillen in onderwijskwaliteit, maar door verschillen in de leerlingpopulatie, is hiermee aangetoond. De scholen A en B hebben in de optiek van de inspectie een volkomen vergelijkbare leerlingpopulatie. In werkelijkheid zijn de populaties onvergelijkbaar.

De basis van het huidige schooltoezicht, dat is gefundeerd op onderlinge schoolvergelijking, valt hiermee in duigen. De inspectie meet niet wat ze veronderstelt te meten.

### 5.1.3 Zeer zwakke scholen

Van enkele scholen, die door de inspectie als zeer zwak zijn beoordeeld, zijn de opbrengsten over een periode van drie schooljaren verzameld en gecorrigeerd voor het opleidingsniveau van de ouders. In vergelijking met de overige 76 groepen is het te verwachten dat de gecorrigeerde score van deze groepen onderaan bungelt. Dit blijkt geenszins het geval. De ene *zwakke* school scoort weliswaar één jaar laag, maar de twee daarop volgende jaren ligt de score op niveau. De andere *zwakke* school scoort alle jaren op niveau.

Naar het oordeel van de inspectie liggen de resultaten van de leerlingen op beide scholen **niet** op het niveau dat op grond van de kenmerken van de leerlingpopulatie mag worden verwacht.

De conclusie na vergelijking met de resultaten van de 1066 leerlingen uit dit onderzoek is, dat de scores van deze scholen wél in overeenstemming zijn met hetgeen mag worden verwacht.<sup>v</sup>

Van één van de zeer zwakke scholen is bekend dat de school zich de laatste drie jaren een slag in de rondte heeft gewerkt om het onderwijs te vormen naar de wensen van de inspectie. De inspectie toont zich hierover tevreden. Het is opmerkelijk dat deze 'verbeteringen' niet tot uiting komen in de opbrengsten. De groepen van gemiddeld 27 leerlingen scoren in de periode 2006-2008 respectievelijk -0,5 / +0,8 / +1,7 ten opzichte van de verwachte score gerelateerd aan de leerlingen uit het databestand gecorrigeerd naar opleidingsniveau van beide ouders. Ondanks een minimale stijging is de inspectie van oordeel dat de school onvoldoende opbrengsten realiseert.

Om eindelijk weer van de lijst zwakke scholen te worden geschrapt moet de school twee achtereenvolgende jaren boven de ondergrens scoren. Dit bleek niet te lukken, totdat de leerkracht van groep 8 ontdekte dat 1 leerling niet meegeteld hoefde te worden (korter dan vier jaar in ons land). Deze correctie bracht de school net boven de ondergrens. Dit enkele voorbeeld toont de rigiditeit van de systematiek. Eén enkele leerling die een dag eerder of later ons land binnenkomt, kan er toe leiden dat een school weer een jaar langer op de lijst met zeer zwakke scholen prijkt.

Dezelfde methode met de 1000 fictieve groepen, zoals die is gehanteerd voor de fictieve scholen A, B en C, is ook uitgevoerd voor de populatie van beide zwakke scholen. De kans dat scholen met een

vergelijkbare populatie op statistische gronden als *zwak* worden beoordeeld ten aanzien van het aspect opbrengsten is ongeveer 20%.

De opbrengsten van scholen met normale opbrengsten worden beoordeeld als onvoldoende en leiden er uiteindelijk mede toe dat deze scholen worden gekwalificeerd als zwakke school.

Vanuit deze constatering zou het omgekeerde ook mogelijk moeten zijn. Scholen die beduidend onder de verwachting scoren en waarvan de opbrengsten door de inspectie als voldoende worden beoordeeld. Ook dit blijkt het geval. Er zijn scholen in het bestand die opeenvolgende jaren beduidend lager scoren dan op basis van de vergelijking zou mogen worden verwacht. De inspectie beoordeelt de opbrengsten echter als voldoende. Er ontstaat een beeld van toeval en willekeur.

Voor scholen die worden bezocht door een niet gemiddelde leerlingpopulatie, bestaat een aanzienlijke kans dat de opbrengsten onjuist door de inspectie worden gewaardeerd.

#### 5.1.4 Rol van het Cito

Het Cito is de leverancier van de gegevens. De gegevens die het Cito levert zijn over het algemeen van een goede kwaliteit. Het Cito kan vanzelfsprekend niet verantwoordelijk worden gehouden voor de wijze waarop anderen de gegevens gebruiken.

In het artikel 'Scholen langs de meetlat' samengesteld door Norman Verhelst, Gerrit Staphorsius en Frans Kleintjes, verbonden aan de Cito-groep, wordt helder aangetoond dat het vergelijken van scholen een hachelijke onderneming is.<sup>VI</sup>

Het Cito maakt beslist geen geheim van deze wetenschap. In de toelichting van de rapportage van de Entreeoets 2008<sup>VII</sup> staat te lezen:

*Via landelijke gemiddelden is een directe vergelijking met andere scholen mogelijk. Het is echter de vraag of een dergelijke vergelijking zinvolle informatie oplevert. Dat zou zo zijn als we het op een school gegeven onderwijs als enige verklaring voor verschillen tussen schoolgemiddelden zouden kunnen beschouwen. Dat is echter niet zo. Verschillen tussen schoolgemiddelden worden namelijk mede bepaald door factoren waarop het onderwijs niet of nauwelijks invloed heeft, maar die wel van grote betekenis zijn voor de onderwijsresultaten. Als we die factoren allemaal zouden kennen en als we de invloed van al die factoren op het schoolgemiddelde zouden kunnen neutraliseren, dan zou een vergelijking met het landelijk gemiddelde of tussen schoolgemiddelden onderling gegevens kunnen opleveren voor conclusies over het gegeven onderwijs. Maar we kennen niet al die factoren en voor zover we ze kennen, zijn ze niet allemaal even gemakkelijk en feilloos meetbaar.*

Heldere taal, maar de inspectie houdt zich van de domme. Scholen worden wel onderling vergeleken en het complete risicogerichte toezicht is er nota bene op gebaseerd.

### 5.1.5 Kwaliteitsbeleid of machtsmisbruik?

Er mag worden verondersteld, dat er bij de inspectie twijfel moet bestaan over de betrouwbaarheid van de gehanteerde waarderingssystematiek. Vele wetenschappelijke publicaties hebben de afgelopen jaren duidelijk gemaakt dat het waarderen van scholen op basis van de huidige opbrengstgegevens niet kan.<sup>4</sup> Deze wetenschap **moet** tot de inspectie zijn doorgedrongen.

In het geval dat een waarderingssystematiek niet betrouwbaar is, is het verwerpelijk scholen op basis hiervan te veroordelen en deze veroordeling publiekelijk te verspreiden.

Er zijn wellicht andere beweegredenen van de overheid om vast te houden aan een volkomen onbetrouwbare systematiek. De huidige werkwijze is bijvoorbeeld bijzonder geschikt voor het verkrijgen van greep op het onderwijs. We spreken dan niet meer over kwaliteit, maar over macht. Of beter gezegd over misbruik van macht.

## 5.2 Kleine scholen

Stelling 2:

*Door onvoldoende rekening te houden met de schoolgrootte, worden kleine scholen ten onrechte in de categorie zwakke scholen ingedeeld.*

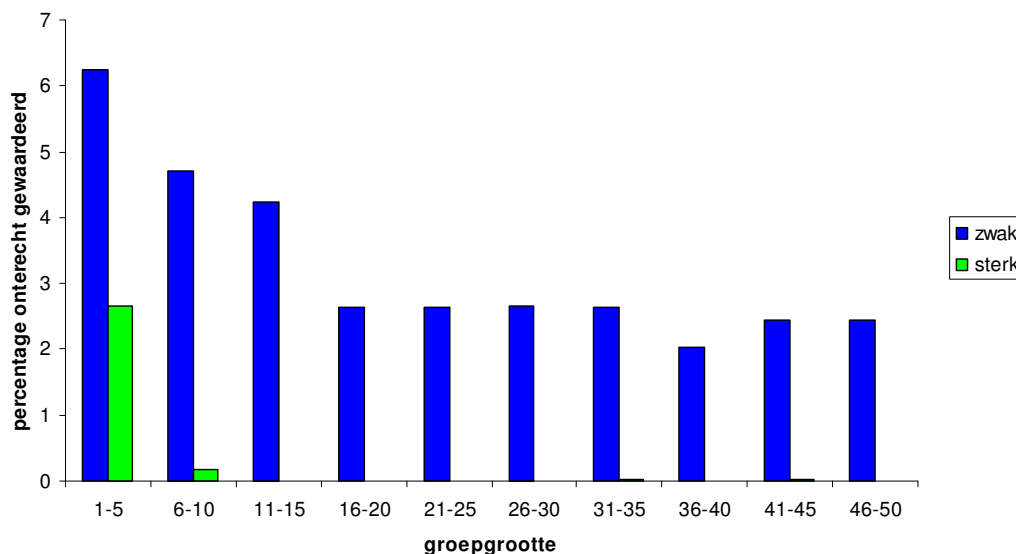
Met betrekking tot de beoordeling van kleine en zeer kleine scholen is het inspectie-document 'Analyse en waarderingen van opbrengsten Primair Onderwijs' niet eenduidig. In de eerste zin van de paragraaf 'Eindtoets in kleine en zeer kleine groepen' wordt gesteld dat in principe de omvang van een groep niet bepalend is voor een gefundeerde beoordeling. Daarentegen wordt in de tweede zin aangegeven dat voor groepen kleiner dan 10 leerlingen de extreem hoge of lage score van een enkele leerling een groot effect kan hebben op de gemiddelde schoolscore. Dit laatste is vanzelfsprekend het geval. De systematiek die wordt beschreven om een onjuiste kwalificatie van een kleine school te voorkomen, lijkt redelijk.

Juist omdat naar verhouding relatief veel kleine scholen worden gekwalificeerd als zwak, is het van belang te onderzoeken of dit effect niet mede wordt veroorzaakt door de gehanteerde systematiek. In het voorbeeld uit het document 'Analyse en waardering van opbrengsten', waar de scores over een periode van drie jaar worden beoordeeld, betreft het slechts 16 leerlingen. Een te bescheiden groep. Voor een school met 200 leerlingen levert een periode van drie jaren een vergelijkingsgroep op van ongeveer 75 leerlingen. Een factor vier groter.

---

<sup>4</sup> Een niet te veronachtzamen feit is het gegeven dat in veel wetenschappelijk onderzoek dezelfde gebrekkige populatiegegevens worden gebruikt als correctiefactor. De gevolgen voor de betrouwbaarheid van dergelijk onderzoek laat zich raden.

Figuur 4a - Onterecht zwak of goed gewaardeerde scholen in relatie tot groepsgrootte conform de inspectienormen van schoolgroep 2.



Er zijn uit het databestand willekeurig fictieve groepen geformeerd met uitsluitend 1.0 leerlingen. Voor 1000 scholen zijn telkens 3 fictieve jaargroepen geformeerd. In het geval de gemiddelde Cito-score in elk van deze 3 fictieve groepen onder de ondergrens van schoolgroep 2 valt, zijn de opbrengsten van deze fictieve school als *zwak* beoordeeld. Dezelfde methode is gehanteerd voor de bovengrens op basis van dezelfde normen. Deze methode is uitgevoerd voor groepen van 1 tot en met 50 leerlingen.

Aangezien er alleen leerlingen zijn gebruikt van gemiddelde scholen, zouden alle scholen in alle gevallen *voldoende* moeten scoren. Vanaf 10 leerlingen hanteert de inspectie de standaardmethode. In figuur 4a valt echter af te lezen dat voor scholen met 11 tot 15 leerlingen de kans op een onterechte waardering nog steeds 1,5% hoger is, dan voor grotere groepen.

Voor groepen met 1 tot 5 leerlingen is de kans dat de school onterecht *zwak* of *goed* wordt beoordeeld op opbrengsten ongeveer 6%. Zoals aangegeven hanteert de inspectie voor scholen met deze omvang een aangepaste methode, waarbij een gemiddelde score wordt berekend over meerdere jaren.

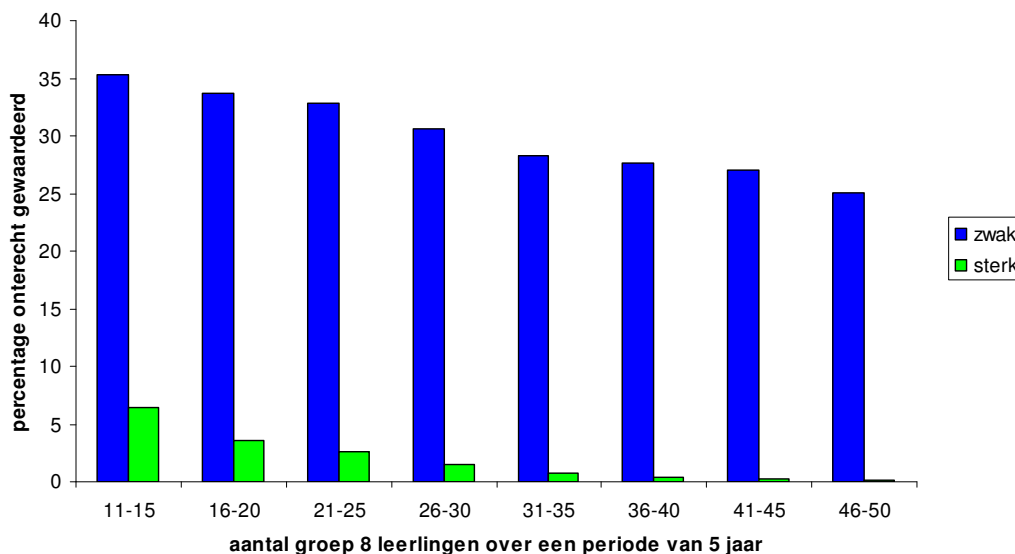
Figuur 4b geeft een beeld van de kans om onterecht als *zwak* te worden beoordeeld in het geval bij hele kleine scholen de aangepaste systematiek wordt gehanteerd. In de berekening worden de scores over een periode van 5 jaren gemeten. Als minimum is 11 schoolverlaters aangehouden per vijf jaar. Een school heeft geen bestaansrecht bij minder leerlingen. Bij 50 leerlingen of meer wordt de reguliere werkwijze gebruikt.

Er is geput uit de categorie 1.0 leerlingen die wederom zijn beoordeeld naar de normen van schoolgroep 2.

Uit de percentages onterechte beoordelingen als *zwak* blijkt, dat de aangepaste werkwijze van de inspectie voor hele kleine scholen ontoereikend is. Fluctuaties worden gedempt, maar er is ook sprake van een forse denkfout. Het gegeven dat ten aanzien van de reguliere werkwijze drie opeenvolgende

jaren onvoldoende gescoord moet worden, staat in schril contrast met het feit dat er in het geval van kleine scholen wordt gewerkt met één gemiddelde score die boven de norm moet liggen per waarderingsmoment. De kans om onterecht zwak te scoren, stijgt daardoor explosief. Een grote school kan twee maal laag scoren en éénmaal het 'geluk' hebben boven de norm te scoren, terwijl het gemiddelde van de deze school heel best onder de norm kan liggen. Deze geluksfactor hebben hele kleine scholen niet.

Figuur 4b - Onterecht zwak of goed gewaardeerde hele kleine scholen op basis van de aangepaste systematiek conform de inspectienormen van schoolgroep 2.



Ondanks de aangepaste systematiek voor hele kleine scholen hebben de hele kleine scholen gemiddeld een kans van 30% om ten onterechte als zwak te worden beoordeeld qua opbrengsten. Een ontoelaatbaar hoog percentage. Om in termen van de inspectie te spreken is de aanduiding 'zeer zwak' voor deze werkwijze een eufemistische kwalificatie.

Het is statistisch niet te verdedigen, dat de inspectie de marges van de normering niet aanpast aan de grootte van de groep.

Ook het Cito heeft jaren geleden al glashelder aangetoond, dat de systematiek van de inspectie met betrekking tot de kleine scholen statistisch niet deugt. Er is de statistische wetmatigheid die zegt dat gemiddelden, berekend op kleine steekproeven een grotere variabiliteit vertonen dan gemiddelden berekend op grote steekproeven.<sup>VIII</sup>

### 5.3 Opbrengsten en intelligentie

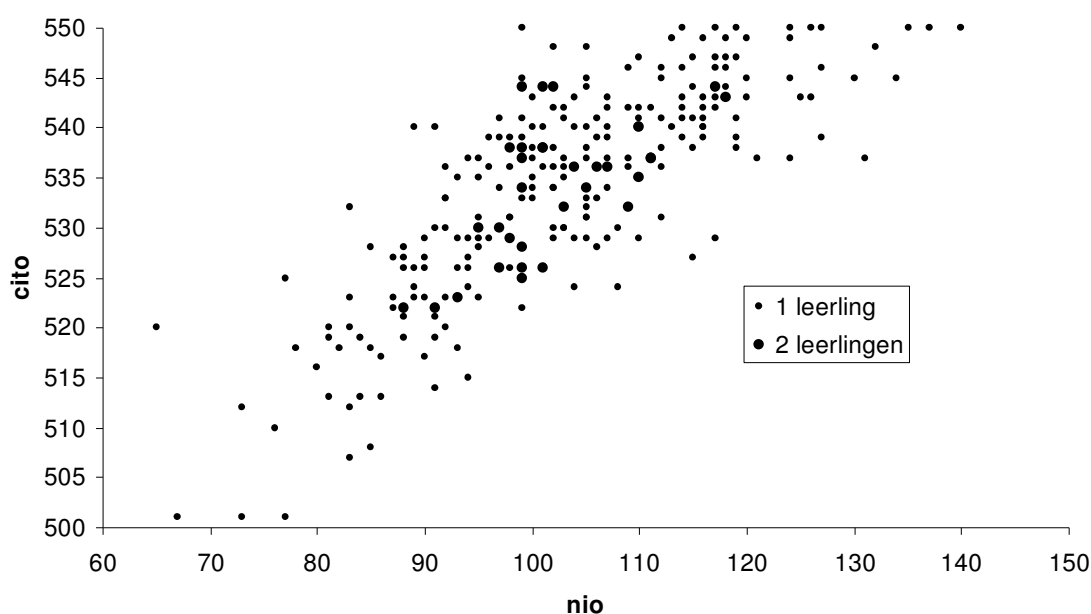
Stelling 3:

*Vooraf onderwijsextrinsieke factoren bepalen het succes van onderwijs.*

Van 274 leerlingen uit het databestand is naast het ouder-opleidingsniveau en de Cito-score ook de score van de Nederlandse Intelligentietest voor Onderwijsniveau (NIO) bekend.

Figuur 5 toont de correlatie tussen Cito en NIO. Het betreft twee hele verschillende toetsen. De Cito-eindtoets is een schoolvorderingentoets, die meet wat kinderen in 8 jaar onderwijs hebben geleerd. Het NIO is een intelligentietest, die bepaalt wat kinderen 'in huis hebben'.

Figuur 5, relatie Cito-score en de NIO



In figuur 5 staan de NIO en Cito-scores van 274 leerlingen. Elke stip in de grafiek komt overeen met 1 of 2 leerlingen. Er is een relatie tussen de scores van de Cito-eindtoets en de score van de NIO. Gemiddeld gesproken komen de uitslagen redelijk overeen. De correlatie(R) is 0,79. De verschillen in individuele gevallen zijn evenwel bijzonder groot.

De brochure van de NIO geeft aan, dat de uitslagen van beide soorten tests in het algemeen meer overeenkomen als het onderwijs goed is geweest.<sup>IX</sup> Echter tussen leerlingen in dezelfde groep met dezelfde gemeten intelligentie zijn er grote verschillen. Deze leerlingen hebben hetzelfde onderwijs genoten van dezelfde leerkrachten in dezelfde omstandigheden. De verschillen zijn dus niet te verklaren, doordat de ene leerling beter onderwijs heeft genoten dan de andere.

*De uitslagen op beide soorten tests komen in het algemeen óók meer overeen als de leerling met de goede instelling de school heeft bezocht.*

Deze opmerking, uit de brochure van de NIO, lijkt van groter belang.

De ene leerling met een gemeten intelligentie van 91 scoort 540 en een andere leerling uit dezelfde groep scoort met dezelfde intelligentie 519 op de Cito-eindtoets. Met een intelligentie van 100 scoort de ene leerling 544 en de andere 533, een jongen met een intelligentie van 115 komt niet verder dan 538 en een meisje met 105 behaalt op de eindtoets 548. Grote verschillen die niet te verklaren zijn op grond van intelligentie of de kwaliteit van het onderwijs.

Van enkele groepen leerlingen uit het databestand zijn naast de scores, ook de persoonlijkheidskenmerken en deels de niet gedeelde omgevingsfactoren bekend. Er blijkt een samenhang tussen de behaalde resultaten en de verschillen in de kindkenmerken en de milieufactoren.

Het ambitieniveau van kind en thuis lijken een aanzienlijke invloed te hebben op het schoolsucces.

Een gevolg hiervan kan zijn dat een school, die de pech heeft te zijn gesitueerd in een sociaal economisch zwakke omgeving en wordt bezocht door kinderen uit gezinnen met een laag ambitieniveau, het moeilijk krijgt wanneer het om opbrengsten gaat. Zelfs het beste leerkrachtenteam, werkend met de beste leermiddelen, zal hier maar heel beperkt succes kunnen boeken. Leerachterstanden zijn in de regel hardnekkig en vereisen een lange adem om ze om te buigen.

Het hardnekkige fenomeen laaggeletterdheid (analfabetisme) in Nederland is een ander illustratief voorbeeld van het opbotsen tegen grenzen.

### 5.3.1 Een laag 'ambitieniveau'

Een laag ambitieniveau is niet iets waar bewust voor wordt gekozen. Het is het gevolg van een complex samenspel van veel factoren, dat uiteindelijk tot een laag ambitieniveau leidt. De omstandigheden waarbinnen kinderen opgroeien zijn zeer verschillend. Beschikken de ouders bijvoorbeeld over een minimale opvoedkundige vaardigheid, is er een zekere taalrijkdom aanwezig, is er een goede basiscommunicatie, is er voldoende tijd en aandacht van de ouder, is er de overtuiging dat onderwijs (de school) van het allergrootste belang is, is er thuis ondersteuning van het leerproces, etc. etc.

Een leerling die na 8 jaar de basisschool verlaat heeft 7500 uur onderwijs gevolgd. In diezelfde periode hebben de ouders ruim 35.000 uren de tijd gehad in hun kroost te investeren.

### 5.3.2 Onderwijsextrinsieke factoren

Prof. dr. Jelle Jolles noemt 7 onderwijsextrinsieke factoren die bepalend zijn voor de efficiency van het leerproces en de daaraan gekoppelde resultaten. Jolles onderscheidt cognitieve neuropsychologische, psychobiologische, ontwikkelingsbiologische, sociodemografische, klinisch neuropsychologische en cognitief gerontologische factoren.<sup>x</sup>

Het onderwijs, maar ook de inspectie, zouden rekening moeten houden met dit complex van factoren.

Op de afdeling Biologische Psychologie van de Vrije Universiteit Amsterdam (VU) wordt sinds 1991 onderzoek gedaan naar cognitieve vaardigheden en factoren die van invloed zijn op de cognitieve



vaardigheden. Bij dit onderzoek wordt gebruik gemaakt van het Nederlands Tweelingen Register. Studies bij tweelingen suggereren dat de ontwikkeling van mentale vaardigheden wordt beïnvloed door een complex samenspel van erfelijkheid en omgeving. Vervolgonderzoek naar dit complexe samenspel is inmiddels gestart. Het doel is nauwkeuriger te gaan uitzoeken hoe omgeving en genetische aanleg de mentale ontwikkeling beïnvloeden.

### 5.3.3 Beperkte invloed, kleine verschillen

In de literatuur wordt algemeen aangenomen dat de school maximaal 20% van invloed is op het schoolsucces. Doordat in ons land de opleidingseisen van de leerkrachten en de hoogte van de bekostiging voor alle basisscholen nagenoeg gelijk zijn, mag worden verondersteld dat de verschillen tussen de sterkste en zwakste scholen slechts met enkele procenten van invloed is op het schoolsucces. Deze kleine verschillen zijn slechts enkele Cito-punten per leerling. Uit figuur 5 blijkt dat een verschil van 20 Cito-punten voor leerlingen met een gelijke intelligentie mogelijk is en uit figuur 3 blijkt dat leerlingen met ouders met hetzelfde opleidingsniveau tientallen punten uiteenlopen. Het is dus onmogelijk om het kleine effect van slechts enkele punten tussen goede en slechte scholen te meten terwijl allerlei andere invloeden op schoolsucces niet gemeten worden.

Daarnaast is het mogelijk de uitslag van de Cito-toets door manipulaties met 8 punten te laten stijgen (zie paragraaf 7.4). Het mag duidelijk zijn dat met de huidige systematiek de toegevoegde waarde van een school en de daaraan veronderstelde kwaliteit onmogelijk te bepalen is. Binnen het spectrum van de andere beïnvloedingscomponenten zijn de gevolgen van kwaliteitsverschillen veel te klein om met de huidige ruwe systematiek te kunnen meten.

## 5.4 Schadelijke neveneffecten

Stelling 4:

*De sterke nadruk van de overheid op de opbrengsten veroorzaakt een reeks ongewenste effecten, die het onderwijs in zijn geheel schade toebrengt.*

Binnen het basisonderwijs is een reeks ongewenste effecten te constateren tengevolge van de bovenmatige nadruk op de opbrengsten.

Voor geen enkele school is het een aanlokkelijk idee om opgenomen te worden in de lijst van zeer zwakke scholen. Heel Nederland weet dan, dat je je werk blijkbaar niet goed doet. Het is dan ook voor de hand liggend dat scholen op allerlei terreinen acties ondernemen om uit de gevarezone te blijven. Hieronder in willekeurige volgorde een opsomming van de diverse schadelijke neveneffecten en andere onbedoelde nadelige gevolgen.<sup>XI</sup>

#### 5.4.1 De verschraling van het onderwijsaanbod

Op veel scholen wordt er gaandeweg steeds meer tijd besteed aan lezen, rekenen en taal. De opbrengsten voor deze vakgebieden zullen wellicht iets stijgen, al kan er ook sprake zijn van tijdelijke schijnopbrengsten. De mogelijkheden om te werken aan een brede ontwikkeling van het kind worden er door beperkt. Deze sterke fixatie op de opbrengsten van een beperkt deel van het onderwijspakket, ondergraaft de geest van de Wet op het Primair Onderwijs.

#### 5.4.2 Bedreiging effectiviteit en rendement

Er zijn veel kinderen met hele specifieke talenten, die soms niet op het gebied van rekenen en taal liggen. Het is van belang juist deze specifieke talenten optimaal te ontwikkelen. De bovenmatige nadruk op enkele ontwikkelingsaspecten benadeelt juist deze groep kinderen. Het bedreigt eveneens de effectiviteit en het rendement van het onderwijs. Bepaalde economische wetten gelden ook voor het onderwijsproces. De beschikbare tijd kan maar eenmaal worden besteed. In een krampachtige poging aan de gestelde eisen te voldoen, besteden leerkrachten naar verhouding onevenredig veel onderwijstijd aan enkele leerlingen met een minimaal of niet te meten resultaat.

#### 5.4.3 Bedreiging WSNS en passend onderwijs

In groepen groeit, onder druk van de prestatiedwang, een sterk competitief klimaat. Kinderen met een meer praktische aanleg en leerlingen met leermoeilijkheden voelen zich daar minder thuis. Een aangetast welbevinden van deze groep leerlingen belemmert hun ontwikkeling en een grotere groep kinderen zal dientengevolge aangewezen zijn op speciaal basisonderwijs.

Voor het opkrikken van de Cito-scores kan het ook 'gunstig' zijn deze categorie leerlingen te verwijzen c.q. niet toe te laten. Hiermee worden de uitgangspunten van WSNS ondergraven. Dit bovenmatig competitieve klimaat maakt een succesvolle ontwikkeling van passend onderwijs bij voorbaat kansloos.

#### 5.4.4 Manipulatie toetsscores

De inspectie onderzocht op verzoek van de politiek de wijze waarop de Cito-eindtoets wordt afgenomen. <sup>xii</sup> Een onderdeel van dit onderzoek betrof de vraag of de toets overeenkomstig de aanwijzingen wordt afgenomen. Uit het onderzoek komt naar voren dat 94,5% van de scholen zich houdt aan de aanwijzingen uit de handleiding. Ten aanzien van deze uitkomsten zijn enkele opmerkingen te maken.

- Het merendeel van de scholen heeft geen echte behoefte om de afname te manipuleren. Bij een bepaalde categorie scholen zal daartoe een grotere neiging bestaan. Het onderzoek houdt hier geen rekening mee.

- Scholen die hebben afgeweken van de aanwijzingen zullen niet direct de behoefte voelen hier openheid van zaken over te geven.
- In veel gevallen hoeft de directeur helemaal niet op de hoogte te zijn of er geheel conform de aanwijzingen is gewerkt en zal hij of zij veronderstellen dat dit het geval is geweest.

Het beeld dat uit het aangehaalde onderzoek naar voren komt, lijkt te rooskleurig. De Cito-eindtoets is op een uiterst eenvoudige wijze te manipuleren. Het gaat dan om training van de toetsaanpak, het voorlezen van een deel van een tekst, het samen 'voorbepreken' van enkele opgaven, het verlengen van de tijd, het gebruik van kladpapier etc.

Bij een combinatie van meerdere minder subtiele manipulaties is onbedoeld proefondervindelijk vastgesteld dat de score wel met gemiddeld 8 punten kan worden opgeschroefd. Van een school uit het onderzoek is vastgesteld dat door het voorlezen van de tekst en het oprekken van de tijd er onverwacht grote effecten zijn waar te nemen. Subtiele en nagenoeg onmerkbare manipulaties zullen de score met enkele punten laten stijgen.

De verhoging van de scores op scholen waar de inspectie op bezoek is geweest<sup>xiii</sup>, alsmede de gemeten hogere opbrengsten na interventie op zwakke scholen, zijn waarschijnlijk in belangrijke mate toe te schrijven aan subtiele manipulaties van de afnamecondities van de toets.

Onder de huidige omstandigheden kan deze handelwijze de scholen niet kwalijk worden genomen. Een als onrechtvaardig ervaren handelwijze, lokt onreglementair gedrag uit. Jammer is het, dat het de opvatting van de overheid versterkt dat leerlingen beter gaan presteren nadat de inspectie op bezoek is geweest. Daarnaast is het betreurenswaardig dat de waarde van de Cito-eindtoets, als hulpmiddel bij de schoolkeuze, door de manipulaties wordt aangetast.<sup>xiv</sup>

#### 5.4.5 Misbruik Cito-eindtoets

De Cito-eindtoets is, als hulpmiddel voor de bepaling van de schoolkeuze, een valide instrument.

Met betrekking tot het gebruik van deze toets om de opbrengsten van het onderwijs te bepalen, is het goed te bedenken dat deze toets daar niet voor is ontwikkeld.

Om te dienen als hulpmiddel voor de schoolkeuze moet de toets een stabiel niveau hebben en voldoende discriminerend zijn. Om het niveau stabiel te houden, in vergelijking met toetsen uit voorgaande jaren, zullen in bepaalde situaties opgaven en/of de normen worden aangepast.

Als bij wijze van spreken het onderwijs de onwaarschijnlijke en extreme prestatie zou leveren alle leerlingen boven de 540 te laten scoren, past het Cito het niveau en de normen aan, zodat er weer een bruikbaar instrument ontstaat om onderscheid te maken. Op basis van het model zullen er altijd scholen 'onder de maat' blijven presteren. Zelfs in de situatie van het bovenstaande voorbeeld. Dit is inherent aan de systematiek.<sup>xv</sup>

#### 5.4.6 Groei van de administratieve last

Van scholen wordt inmiddels verwacht dat ze al hun handelen kunnen aantonen met protocollen, projectplannen, handelings- en uitvoeringsplannen, gespreksverslagen, evaluaties en onderzoeken.

Niet wat er feitelijk in de klas gebeurt staat centraal, maar het feit of alles goed op papier staat. Hoe het feitelijk in de klas gaat, komt er door in de verdrukking en onttrekt zich grotendeels aan het gezichtsveld van de inspectie. Het onderwijs zou er mee gediend zijn, als al deze kostbare menskracht ook daadwerkelijk aan onderwijs wordt besteed.

Er zijn parallellen te trekken met de kwaliteit van bijvoorbeeld de ouderenzorg. Grote instellingen werken volgens keurige protocollen, behalen prachtige keurmerken, de leiding verdient vorstelijk en ondertussen schaamt het personeel zich voor de kwaliteit van de dagelijkse zorg.

#### 5.4.7 Groei bureaucratie

De overheid heeft aangekondigd de eerste verantwoordelijkheid voor de kwaliteit van het onderwijs bij de besturen te leggen. Schoolbesturen zullen deze verantwoordelijkheid serieus nemen.

Er is echter sprake van het afschuiven van de last.<sup>XVI</sup>

Decennia lang heeft de overheid geprobeerd met honderden regels, wetten, projecten en een stortvloed aan gele circulaires het onderwijs te verbeteren. Zonder noemenswaardig resultaat. Nu wordt de bal bij de besturen gelegd. Het gevolg is dat er door al die besturen een machinerie wordt opgetuigd van kleine 'Zoetermeertjes', waarbij de omvang van het oude Ministerie van Onderwijs verbleekt. Deze ontwikkeling zal een verdergaande bureaucratisering veroorzaken met allerlei systemen van beheren en beheersen, zonder enig bewijs dat de kwaliteit hierdoor verbetert. Het tegenovergestelde is aannemelijker, omdat er nog meer onderwijsgeld wordt onttrokken aan het primaire proces.

#### 5.4.8 Stagnerende innovaties

De huidige beoordelingssystematiek maakt leerkrachten onzeker en in een bepaald opzicht angstig. In dit klimaat gedijen experimenten en innovaties niet. Om het onderwijs goed te laten aansluiten bij de zich steeds wijzigende omstandigheden, heeft de professional vertrouwen, ruimte en vrijheid nodig om het onderwijs te optimaliseren en aan te passen aan de snel wijzigende omstandigheden.

Het geïnstitutionaliseerde wantrouwen ten opzichte van de professional, treffend beschreven in het rapport 'Bewijzen van goede dienstverlening' door de Wetenschappelijke Raad voor het Regeringsbeleid, zal plaats moeten maken voor een geheel andere cultuur.<sup>XVII</sup>

#### 5.4.9 Beschadigd imago

De meest bepalende factor voor de kwaliteit van het onderwijs is de kwaliteit van de leerkracht (Stevens).

De negatieve beeldvorming, in combinatie met een naar verhouding karige beloning, maakt het onderwijs niet aantrekkelijk voor jonge getalenteerde mensen. Dit klimaat, mede veroorzaakt door de handelwijze van een overheidsinstantie, is schadelijk voor het onderwijs. De beste leerkrachten stromen uit en talentvolle jonge mensen worden niet binnengehaald. Het vormt de allergrootste bedreiging voor de kwaliteit van het onderwijs.

#### 5.4.10 De schandpaal

Mede onder druk van de media worden zogenaamde 'zeer zwakke scholen' ernstig gestigmatiseerd. Het onderwijs op deze scholen is blijkbaar zo slecht, dat het gerechtvaardigd is om deze scholen jarenlang publiekelijk aan de schandpaal te (laten) nagelen. Voor teams moet dit een hele vernederende ervaring zijn. Waarschijnlijk is de strategie van 'blaming and shaming' de meest effectieve manier om een schoolteam te beschadigen en te demotiveren.<sup>XVIII XIX</sup>

#### 5.4.11 Bedreiging diversiteit

De verschillende vernieuwingscholen in ons land werken veelal vanuit een ander gezichtspunt en vanuit andere prioriteiten. Deze diversiteit is van grote waarde. De bovenmatige nadruk op slechts een deel van de onderwijsinhoud bedreigt deze diversiteit. Het toezicht wordt misbruikt als breekijzer om iedereen in hetzelfde keurslijf te dwingen. Op die manier ontstaat er een eenheidsworst, die op termijn tot economische schade zal leiden. Tevens staat niet vast dat deze indirect door de overheid afgedwongen werkwijze beter is dan de werkwijzen, zoals die door vernieuwingscholen worden gehanteerd. Er zijn vele wegen die naar Rome leiden.<sup>XX</sup>

## **6 Conclusies**

### **6.1 Opbrengstwaardering onbetrouwbaar**

De inspectie heeft een onvolledig en zeer beperkt beeld van de kenmerken van de leerlingpopulatie van een school. De correctie van de opbrengsten met behulp van het GLG is ontoereikend.

Zonder het hebben van een scherp beeld van de kenmerken van de leerlingpopulatie is het niet mogelijk de opbrengsten nauwkeurig en objectief te waarderen.

Scholen met een niet gemiddelde leerlingpopulatie lopen een grotere kans onjuist te worden beoordeeld dan scholen met een meer gemiddelde populatie.

Op basis van de huidige systematiek dient de inspectie zich te onthouden van het waarderen van opbrengsten.

### **6.2 Zeer zwakke scholen**

De onderzochte 'zeer zwakke' scholen scoren, in tegenstelling tot het oordeel van de inspectie, voldoende ten aanzien van het aspect opbrengsten.

### **6.3 Manipulatie Cito-eindtoets**

Onmerkbaar manipulaties van de toetsomstandigheden van de Cito-eindtoets beïnvloeden de score met twee tot drie punten. Minder subtiele manipulaties verhogen de score met minimaal 8 tot maximaal 15 punten.

### **6.4 Positie kleine scholen**

In het geval van kleine scholen is de beschreven procedure onbetrouwbaar. Statistische basisregels worden geschonden. Een kleine school heeft statistisch een grotere kans als zeer zwak te worden bestempeld dan een grote school. Hiermee is tevens verklaard dat kleine scholen zijn oververtegenwoordigd op de lijst zeer zwakke scholen. Voor de beoordeling van kleinere scholen moet een bredere marge worden toegepast.

### **6.5 Beperkte invloed**

De invloed van de kwaliteit van een school op het schoolsucces wordt schromelijk overschat. Kinderen met een gelijke intelligentie, die acht jaar lang hetzelfde onderwijs hebben genoten, waar identieke eisen aan zijn gesteld, scoren uiteindelijk heel verschillend op de Cito-eindtoets. Het zijn factoren die buiten de school liggen, die een grote invloed hebben op het schoolsucces. Juist de ervaringen uit de praktijk wijzen overduidelijk in de richting van een grote invloed van het ambitieniveau van kind en/of ouders op het schoolsucces.

## 6.6 Schadelijke neveneffecten van een te sterk gerichte opbrengstcultuur

Het huidige toezichtskader, met de sterke nadruk op bepaalde (eenvoudig te meten) opbrengsten, is een bedreiging voor de kwaliteit van het onderwijs.

Er zijn schadelijke gevolgen op diverse terreinen:

- Vakgebieden als muziek, handvaardigheid, dramatische expressie, mondelinge taalvaardigheid komen in de verdrukking door de te sterke nadruk op lezen, taal en rekenen.
- Het pedagogisch klimaat verhardt door de groei van een sterke eenzijdige competitieve prestatiedrang. Steeds minder kinderen zullen zich thuis voelen op de reguliere basisschool. De verwijzing zal verder stijgen.
- Toetsen worden onder druk van de belangen gemanipuleerd.
- De gegevens van de Cito-eindtoets worden misbruikt door de inspectie. Altijd zullen er scholen in de optiek van de inspectie onder de maat blijven presteren.
- Er wordt op scholen onevenredig veel tijd besteed aan allerlei registratie, die leidt tot minder menskracht voor het primaire onderwijsproces.
- Besturen zullen voor het opzetten van hun kwaliteitsbeleid een steeds grotere hap nemen uit de beschikbare budgetten.
- De afrekencultuur heeft een verlamme invloed op innovaties.
- Het systematisch afbreken van het imago van het onderwijs leidt tot een gebrek aan vakbekwame nieuwe leerkrachten.
- De plaatsing op de lijst van zeer zwakke scholen veroorzaakt angst en demotivatie binnen het onderwijs.
- De diversiteit binnen het Nederlandse basisonderwijs wordt ondergraven.
- Leerkrachten getroosten zich onder de druk enorm veel inspanningen in een poging om resultaten uit kinderen te halen, die er eenvoudigweg niet in zitten. Dit is verspilling van onderwijstijd die ten koste gaat van de overige leerlingen.

## 7 Aanbevelingen

### 7.1 Een voorlopige betrouwbaardere systematiek

Het terugbrengen van de opbrengstgerichtheid tot redelijk proporties verdient de absolute voorkeur. In het geval de politiek vasthoudt aan de beoordeling van basisscholen op basis van de 'opbrengsten', is aanpassing van de systematiek een vereiste.

Een 'betrouwbaardere' wijze van het beoordelen van opbrengsten zou kunnen bestaan uit de volgende werkwijze:

- **De Cito-eindtoets (of vergelijkbare toets)**

De toets wordt onder toezicht van een gecommitteerde afgenomen.

- **Opleidingsniveau ouders**

Het Cito reserveert op het scoreformulier een hoekje waar de school informatie over het opleidingsniveau van beide biologische ouders verschaft. De validiteit van deze gegevens is niet absoluut vast te stellen, maar hetzelfde geldt voor de gegevens die ten grondslag liggen aan de weging. De bron is in beide gevallen dezelfde.

- **Kindinformatie**

De school geeft op het scoreformulier enige kindinformatie in de vorm van bijvoorbeeld een vijfpunt-schaal. Onderdelen die opgenomen zouden kunnen worden zijn: houding t.o.v. de school, doorzettingsvermogen, motivatie, concentratie, werkhouding en thuistaal. Het betreft subjectieve gegevens, maar wel gegevens verstrekt door een professional die het kind goed kent.

- **De intelligentietoets**

Alle leerlingen maken een intelligentietoets.

De uitslag van de Cito-eindtoets in combinatie met de uitslag van de intelligentietest, het opleidingsniveau van beide ouders en de kindinformatie maken het mogelijk voor groepen **indicatief** iets te zeggen over de toegevoegde waarde van het onderwijs. Al deze informatie samen zal er toe leiden dat opbrengsten reëler kunnen worden geïnterpreteerd en dat de inspectie in staat is een onderbouwd advies te geven.<sup>5</sup> Deze voorgestelde werkwijze zal er ook toe leiden dat scholen met veel kansrijke leerlingen, die nu nog tevreden zijn met 'hoge' opbrengsten, aan het jasje kunnen worden getrokken, omdat er meer uit te halen valt.<sup>XXI</sup>

In die situaties waar het niveau van de opbrengsten sterk afwijkt van wat verwacht zou mogen worden, doet de inspectie onderzoek op schoolniveau naar het ambitieniveau van de

---

<sup>5</sup> Het is betreurenswaardig dat de component advisering en stimulering als onderdeel van een schoolbezoek compleet uit het toezicht is geschrapt.



leerlingpopulatie op deze scholen. Hier zullen handzame instrumenten voor moeten worden ontwikkeld.

Ziet de inspectie geen kans de systematiek structureel te verbeteren, dan zal de inspectie zich volledig moeten onthouden van het waarden van opbrengsten van basisscholen.

## 7.2 Kleine scholen

In zijn algemeenheid is te stellen dat de waardering van hele kleine scholen per definitie minder betrouwbaar is dan van grote scholen. Het zou aanleiding kunnen zijn voor de inspectie om alle hele kleine scholen, ongeacht de opbrengsten, frequenter te bezoeken. Vooral ook op basis van het gegeven, dat het effect van bijvoorbeeld een minder getalenteerde leerkracht op een hele kleine school aanmerkelijk groter is dan op een grote school.

## 7.3 Een realistisch toezicht

Het opzetten van een realistisch en wetenschappelijk verantwoord toezicht is dringend noodzakelijk, maar de ontwikkeling ervan zal enige jaren vergen.

Het zou gunstig zijn voor de kwaliteitsontwikkeling van het basisonderwijs, om in afwachting van een verantwoorde en gefundeerde vorm van onderwijstoezicht, de werking van de Wet op het Onderwijstoezicht op te schorten.

In de tussentijd richt het toezicht zich niet meer op documenten, maar op het onderwijs zelf.

Alle groepen van alle scholen worden door de inspectie eens per twee jaar bezocht. Het toezicht richt zich op de kwaliteit van de instructie, het effectieve gebruik van de leertijd en de afstemming van het onderwijs op de mogelijkheden van het kind.<sup>xxii</sup>

De inspecteur wordt tijdens elk schoolbezoek geassisteerd door een ervaren schooldirecteur. Het bezoek resulteert in een verslag. Het verslag met eventuele aanbevelingen en de reactie van de school daarop zijn publiek toegankelijk.

Het budget van de inspectie wordt dienovereenkomstig verhoogd.<sup>6 xxiii</sup>

## 7.4 Wetenschappelijk onderzoek

Een leger wetenschappers heeft de laatste 15 jaar aangetoond, dat het meten van effectiviteit en rendement van het onderwijs geen kwestie is van het analyseren van toetscores. Meijnen, Bosker, Dronkers, Veenstra, Leseman zijn enkele gerenommeerde onderzoekers die op dit terrein veel onderzoek verricht(t)en. Veenstra concludeert in 'Leerlingen – klassen – scholen'<sup>xxiv</sup>:

*.....om scholen eerlijk met elkaar te vergelijken moet de controle voor instroom uitgebreid zijn en is een multiniveau analyse nodig (tot op heden wordt in alle publicaties van prestatiegegevens niet voldaan aan beide eisen). Wanneer die methode wordt gebruikt dan blijkt dat prestaties en vorderingen veel meer samenhangen met kenmerken van leerlingen en gezinnen dan met de kwaliteit van een school.*

---

<sup>6</sup> Het heeft er alle schijn van dat de huidige inferieure kwaliteit van de inspectie mede is veroorzaakt door ondoordachte bezuinigingen van het vorige kabinet.

In 'Onderwijs en Ongelijkheid' schrijft Leseman<sup>xxv</sup>:

*.....het onderwijs kan geen actief tegenwicht bieden tegen dit complexe samenspel dat in de voorschoolse jaren al tot verschillen heeft geleid. Het onderwijs conformeert zich aan hetgeen binnenkomt en reproduceert jaarlijks de verschillen die de kinderen, individueel, bij binnenkomst meebrengen. Omdat wat ze gemiddeld per groep en dus per school, meebrengen sterk wordt beïnvloed door het educatieve niveau thuis, wordt daarmee weer eens bevestigd wat de aanleiding is geweest voor het onderwijsvoorrangsbeleid: er zijn meer en minder bedeelde scholen (Meijnen en Riemersma 1992). Deze constatering is niet geruststellend wanneer gehoopt wordt op de compenserende kwaliteit van het onderwijs.*

Dat de invloed van onderwijs op het schoolsucces minder groot is dan vaak wordt aangenomen, mag nooit als excuus dienen om achterover te leunen. Scholen zullen steeds alles op alles moeten zetten om het maximale te bereiken.

Onafhankelijk wetenschappelijk onderzoek naar de samenhang van alle beïnvloedingsaspecten van schoolsucces verschaft ons een scherper inzicht, waar met succes in geïnvesteerd dient te worden. Deels is dit onderzoek uitgevoerd en wordt onderwijs en beleidsmakers de weg gewezen. In de praktijk lijkt het er echter op of al deze waardevolle kennis alleen maar circuleert in het wetenschappelijke circuit. Een gevolg is in dit verband, dat politiek en overheid halsstarrig blijven vasthouden aan een controle- en beoordelingsmechanisme, waarvan vaststaat dat het een onbetrouwbaar mechanisme is, waarvan op geen enkele wijze wetenschappelijk is aangetoond of het mechanisme bevordert wat het beoogt.

De verhouding tussen enerzijds de onderwijsextrinsieke factoren, die het opbrengstniveau beïnvloeden en anderzijds de onderwijskwaliteit (of het gebrek eraan) moet worden vastgesteld, zodat realistische verwachtingen kunnen worden geformuleerd.

Het is raadzaam op basis van wat inmiddels bekend is niet al te overspannen verwachtingen te wekken. De huidige kwaliteit van het basisonderwijs maakt het onwaarschijnlijk, dat er momenteel sprake is van een sterke onderbenutting van talent of van 'grote stille reserves aan leerkapitaal'.<sup>xxvi</sup>

Voor bepaalde categorieën kinderen zal op termijn een verbetering van het schoolsucces mogelijk zijn. De door de staatssecretaris te hulp geroepen Michael Fullan toont dit bijvoorbeeld aan. Het gaat dan wel om kinderen, die ontplooiing hebben gemist door met name een gebrekkige beheersing van de taal. Het is niet te verwachten, dat de door Fullan gepropageerde werkwijze tot hetzelfde resultaat zal leiden in bepaalde armoede-cumulatiegebieden in Nederland. De aard van de problematiek en de achtergronden zijn heel verschillend. Hét kenmerk van deze achterstanden is de hardnekkigheid ervan. De intensieve taalstimuleringsprojecten van de laatste dertig jaar hebben dat duidelijk gemaakt. Mocht er al sprake zijn van onderbenutting van talent, dan moet die gezocht worden bij de betere leerlingen. Er van uitgaande dat het ene kind zich sneller ontwikkelt dan het andere, moeten de onderlinge verschillen steeds groter worden. Het Nederlandse basisonderwijs is echter momenteel niet in staat de onderlinge verschillen tussen de kinderen te vergroten. Blijkbaar lukt het door middel van een grote inzet de zwakkere leerlingen erbij te houden, maar remmen we daardoor tegelijkertijd

de betere af. Onderbenutting van gemiddeld en bovengemiddeld talent in het basisonderwijs is het gevolg.

## **7.5 Stoppen opbrengsterreur**

Het is van groot belang dat scholen hoge eisen stellen en de ambitie hebben zonder vooroordeel alles uit kinderen te halen wat erin zit. Het stellen van irreële doelen is echter in alle situaties schadelijk en heeft zonder uitzondering mislukking en frustratie tot gevolg. Het is wellicht usance dat politici op gebrekkige gronden ambitieuze doelen formuleren. Maar irreële doelen worden niet gehaald en aan het eind van de rit krijgt het onderwijs door politiek en inspectie de schuld in de schoenen geschoven.

Er is een eenvoudige en lucratieve methode om de opbrengsterreur een halt toe te roepen. Een zogenaamde zeer zwakke school, waarvan de opbrengsten door de inspectie als onvoldoende zijn beoordeeld, klaagt de inspectie aan wegens smaad. De school zal zonder twijfel door de rechter in het gelijk worden gesteld. Er is in dit land namelijk geen ter zake deskundige wetenschapper te vinden, die de werkwijze van de inspectie kan en wil onderbouwen. Daarentegen is het voor de school een kleinigheid om wetenschappers te bewegen om als getuige deskundige te worden gehoord. Zij zullen de werkwijze van de inspectie ten aanzien van de waardering van opbrengsten ontmaskeren als volkomen onbetrouwbaar.

Het lucratieve deel van de actie is de mogelijkheid van het schoolbestuur, maar ook voor de aan de school verbonden leerkrachten, een schadeclaim in te dienen tengevolge van de geleden immateriële en wellicht materiële schade. Overheid en inspectie zullen na de gerechtelijke uitspraak snel tot inkeer komen.

Nog beter zou het zijn, wanneer politici en inspectie beseffen dat ze zichzelf bij de neus nemen. Voorwaarde is, dat deze voornamelijk alfa-opgeleiden zich laten leiden door relevante cijfers en niet door mooie verhalen.

Het is de hoogste tijd om het onderwijstoezicht voor het basisonderwijs drastisch te herzien.

## 8 Tot slot

Een inspecteur maakte de vergelijking arts-patiënt versus inspectie-school. De overeenkomst is dat zowel het menselijk lichaam als onderwijs complex is. Ondanks die complexiteit wordt van een arts verwacht dat hij een diagnose stelt en overgaat tot behandeling. Hetzelfde geldt voor de inspectie. Een mooie vergelijking, maar hij gaat mank.

Artsen stellen hun diagnose op basis van uitgebreid onderzoek. Wanneer de oorzaak onduidelijk blijft, wordt steeds diepgaander (en duurder) onderzoek uitgevoerd, totdat een betrouwbare diagnose gesteld kan worden. De diagnosemethode van de inspectie is niet uitgebreid getest op bekende gevallen. Dit is ook onmogelijk, want er bestaan geen overduidelijke sterke en zwakke scholen waarop de diagnostische aanpak van de inspectie getest kan worden. Daarnaast is er geen duidelijkheid over het foutpositief percentage (*Type I error*) van de methode van de inspectie. Gelet op de resultaten uit dit rapport, de onvolkomenheden die door de wetenschap worden aangedragen, publicaties van het Cito en het eigen onderzoek, is dit foutpositief percentage ongetwijfeld groter dan enkele procenten. In vergelijking met slechts 1.5% zwakke scholen die op dit moment op de zwarte lijst staan, leidt dit onvermijdelijk tot de conclusie, dat vrijwel alle scholen waarvan de inspectie de opbrengsten als onvoldoende aanmerkt in werkelijkheid voldoende scores.

In het geval een arts overgaat tot het verstrekken van bijvoorbeeld medicatie, betreft het altijd goedgekeurde medicijnen, waarvan de werking in alle gevallen wetenschappelijk is aangetoond. Mocht een arts besluiten de huidige werkwijze van de inspectie te hanteren voor zijn of haar medisch handelen, dan wordt hij of zij zonder twijfel als kwakzalver uit het ambt gezet. Iedereen krijgt hetzelfde medicijn toegediend (waarvan de werking niet vaststaat), de bijwerkingen erger zijn dan de kwaal en gezonde patiënten worden gedwongen een medicijn te slikken.

De huidige situatie kon vooral ontstaan en zich versterken door het repressieve karakter van het optreden van de inspectie. De term 'opbrengstterreur' is wat zwaar. De inspectie doet haar werk met de beste bedoelingen, maar is toch niet geheel misplaatst.

Het onderwijs zal een strijd moeten voeren om het juk van dit ontaarde toezicht af te werpen. Het betreft dan niet alleen het basisonderwijs, maar ook het speciaal basisonderwijs, het speciaal onderwijs en het voortgezet onderwijs. Allen kampen met een vergelijkbare problematiek.

Tijdens het uitwerken van deze verkenning drong zich steeds sterker de vraag op hoe het mogelijk is, dat politiek en overheid beleid ontwikkelen en besluiten nemen op basis van niet bewezen feiten. Besluiten die ook nog eens een averechts effect hebben.

De uitgave 'De tien plagen van de staat'- De bedrijfsmatige overheid gewogen -, onder redactie van Frank Ankersmit en Leo Klinkers, geeft een reeks plausibele antwoorden op bovengestelde vraag.<sup>xxvii</sup> In de tiende plaag behandelt Klinkers het gebrek aan politiek vakmanschap. Een citaat over het 'Afrekenen op het resultaat' wil ik u niet onthouden.

*Met de komst van de - door Jouke de Vries - genoemde 'managementstaat' zijn we ondergedompeld in een hiervoor al even geduide logica die voor het bedrijfsleven normaal is, maar die wezensvreemd is aan de politiek. Het neoliberalisme bracht ons de bedrijfsmatige concepten van producenten, klanten, managers, productieafspraken, prestatiecontracten, resultaatverplichtingen, input-, output- en outcomesturing, shared services, registraties, protocollen en zo meer. Dat is een kwaal die de geloofwaardigheid van de democratische rechtsstaat ten gronde aantast. De staat is geen bedrijf, zijn ambtenaren zijn geen werknemers en de burgers geen klanten.*

*In de context van dat jargon heeft ook iemand bedacht 'ik reken je af op je resultaat'. Ik zal daarover kort zijn. Afrekenen doe je in een winkel of in het criminele circuit, niet tussen mensen die hun ambt uitoefenen in dienstbaarheid aan staat en samenleving.*

## 9 Geraadpleegde bronnen:

---

- <sup>I</sup> Netelenbos, Tineke – Een impuls voor het basisonderwijs, OCW Den Haag mei 1995
- <sup>II</sup> Grift van der, Wim- Ontwikkelingen in de kwaliteit van het basisonderwijs – Basisschool Management, september 2007
- <sup>III</sup> Inspectie van het Onderwijs – Analyse en waarderingen van opbrengsten primair onderwijs, Utrecht 6 april 2007
- <sup>IV</sup> Cito – Terugblik en resultaten Eindtoets basisonderwijs 2008, Arnhem februari 2008
- <sup>V</sup> Vos de, Teije – Meten is niet álles weten, Zorg Primair mei 2007
- <sup>VI</sup> Verhelst, Norman; Staphorsius, Gerrit; Kleintjes, Frans – Scholen langs de meetlat, Cito-groep Arnhem november 2003
- <sup>VII</sup> Cito – Rapportage Entreetoets 2008, Arnhem mei 2008
- <sup>VIII</sup> Verhelst, Norman; Staphorsius, Gerrit; Kleintjes, Frans – Scholen langs de meetlat, Cito-groep Arnhem november 2003
- <sup>IX</sup> Dijk van, Henk en Tellegen, Peter – Ouderbrochure Nederlandse Intelligentietest voor onderwijsniveau, Boom test uitgevers Amsterdam.
- <sup>X</sup> Goorhuis-Brouwer, Sienke; Essen van, Mineke; Levering Bas; Werf van der, Greetje (red.) – Mythes in het onderwijs, Amsterdam 2008
- <sup>XI</sup> Engelen, Ewald – Weg met de toetsterreur! Op naar een zachtmoedige meritocratie, 2008
- <sup>XII</sup> Inspectie van het Onderwijs – Eindtoets in het basisonderwijs (Een onderzoek naar leerlingen die niet meedoen en/of niet meetellen, Utrecht.
- <sup>XIII</sup> Webbing, Dinand; Luginbuhl, Rob; Timmerhuis, Jaqueline - Do school inspections improve primary school performance?, CPB Discussion Paper 83
- <sup>XIV</sup> Staphorsius, Gerrit – Optimalisering van het gebruik van onafhankelijke toetsen in een leerling- en onderwijsvolgsysteem: ‘strategische’ bijdrage aan duurzame onderwijsontwikkeling..., Cito maart 2008
- <sup>XV</sup> Glopper de, Kees – Barbertje moet hangen, Groningen maart/mei 2008
- <sup>XVI</sup> Sikkes, Robert – Inspectie geïnspecteerd, Het Onderwijsblad 7 januari 2006

- 
- <sup>xvii</sup> Wetenschappelijke Raad voor het Regeringsbeleid – Bewijzen van goede dienstverlening, Amsterdam 2004
- <sup>xviii</sup> Douma, Lisette – Friese basisscholen aan de schandpaal, Het Onderwijsblad 7 februari 2007
- <sup>xix</sup> Oosterbeek, Hessel en Webbink, Dinand – Een economische kijk op publieke indicatoren van schoolkwaliteit, februari 2001
- <sup>xx</sup> Samenwerkingsverband van Organisaties voor Onderwijsvernieuwing – Elke school is er één, Valthe september 2007
- <sup>xxi</sup> Jungbluth, Paul – Meer professionaliteit én betere kansen door verstandig toetsgebruik (Prestaties en toezicht in het onderwijs), Maastricht maart 2008
- <sup>xxii</sup> Grift van der, Wim- Ontwikkelingen in de kwaliteit van het basisonderwijs – Basisschool Management, september 2007
- <sup>xxiii</sup> Sikkens, Robert – Inspectie geïnspecteerd, Het Onderwijsblad 7 januari 2006
- <sup>xxiv</sup> Veenstra, Rene - Leerlingen - klassen - scholen, Prestaties en vorderingen van leerlingen in het voortgezet onderwijs, Amsterdam 1999
- <sup>xxv</sup> Karstens, Sjoerd; Slegers, Peter - Onderwijs en ongelijkheid: grenzen aan de maakbaarheid, Amsterdam november 2005
- <sup>xxvi</sup> Ministerie OCW – Scholen voor morgen (Samen op weg naar duurzame kwaliteit in het primair onderwijs), Den Haag 28 november 2007
- <sup>xxvii</sup> Ankersmit, Frank; Klinkers, Leo (red.) – De tien plagen van de staat – De bedrijfsmatige overheid gewogen, Den Haag / Groningen, januari 2008