



OPBRENGSTGERICHT WERKEN IN HET PRIMAIR ONDERWIJS: THEORIE EN PRAKTIJK

Discussiebijdrage, 6 november 2013

door Henk Blok, Guuske Ledoux en Jaap Roeleveld

Inhoudstabel

<i>Samenvatting</i>	2
<i>1 Inleiding</i>	3
<i>2 Begripsanalyse van opbrengstgericht werken</i>	4
<i>3 Hoe effectief is opbrengstgericht werken?</i>	7
<i>4 Twee gevalsstudies</i>	12
<i>5 Conclusies en discussie</i>	15
<i>Literatuur</i>	19
<i>Tabellen</i>	22
<i>Over de auteurs</i>	23

Kohnstamm Instituut

Plantage Muidersgracht 24

1018 TV Amsterdam

T 020 5251226

W www.kohnstammstituut.uva.nl

E az@kohnstamm.uva.nl

Samenvatting

Verbeterde leeropbrengsten voor taal en rekenen zijn een belangrijke doelstelling op de beleidsagenda van het Ministerie van Onderwijs, Cultuur en Wetenschap. Opbrengstgericht werken wordt allerwegen als een belangrijk middel beschouwd om dit doel te realiseren. In dit artikel wordt vanuit drie invalshoeken aandacht besteed aan opbrengstgericht werken: a. wat valt er onder te verstaan, b. is er empirische evidentie dat opbrengstgericht werken tot verhoogde leeropbrengsten leidt, en c. welke resultaten worden er gerealiseerd bij de implementatie van opbrengstgericht werken in de Nederlandse onderwijspraktijk? De gehanteerde onderzoeksmethoden zijn begripsanalyse, review van onderzoeksliteratuur en gevalbeschrijvingen. We concluderen ten eerste dat er geen consensus is over de betekenis van opbrengstgericht werken. Ten tweede, de effectiviteit van opbrengstgericht werken ten aanzien van verbeterde leeropbrengsten is blijkens het literatuuroverzicht aan twijfel onderhevig. Ten derde, de twee beschreven Nederlandse schoolverbeteringsprojecten – het project Streef: Gebruik Maken van Opbrengsten en het project Opbrengstgericht Leiderschap – hebben niet of nauwelijks tot hogere leeropbrengsten geleid. In de discussie wordt gesteld dat het verstandig is om de verwachtingen over de effectiviteit van opbrengstgericht werken naar een realistischer niveau terug te brengen. Ook wordt gepleit voor kleinschalige experimenten naar de werkzaamheid van nauw te omschrijven vormen van opbrengstgericht werken.

1 Inleiding

Opbrengstgericht werken is een term die in onderwijsland nog niet zo lang in zwang is, maar die al wel een belangrijke rol vervult in de beleidsdiscussie over onderwijsverbetering (Visscher & Ehren, 2011). Verbetering van het taal- en rekenonderwijs is een centraal oogmerk van de tegenwoordige beleidsagenda voor het primair onderwijs, zoals vastgelegd in de nota *Scholen voor morgen* (Ministerie van Onderwijs, Cultuur en Wetenschap, 2007) en de actieplannen *Basis voor presteren* (primair onderwijs) en *Beter Presteren* (voortgezet onderwijs) (Ministerie van Onderwijs, Cultuur en Wetenschap, 2011a). Opbrengstgericht werken – door scholen en leerkrachten – is een van de wegen waarlangs verbeteringen gerealiseerd moeten gaan worden.

Over nieuwe termen, dus ook over opbrengstgericht werken, zijn nuttige vragen te stellen (De Groot & Medendorp, 1986). Een belangrijke vraag is bijvoorbeeld of opbrengstgericht werken alleen maar een nieuwe naam is voor een al langer bestaande werkwijze, of dat er sprake is van een daadwerkelijk nieuwe werkwijze. Een andere vraag is of gebruikers van de term er allemaal hetzelfde mee bedoelen. Het is, zeker voor een nieuwe term, niet ongewoon dat er onduidelijkheid bestaat over de betekenis ervan. Die onduidelijkheid kan aanleiding geven tot werkwijzen die in naam gelijk zijn, maar in bedoeling of uitvoering verschillen.

Een andere invalshoek is welke effecten men mag verwachten van opbrengstgericht werken. Valt er van deze werkwijze inderdaad een prestatieverhoging te verwachten, en zo ja, onder welke voorwaarden? En, niet onbelangrijk, moet men bedacht zijn op mogelijke nadelige neveneffecten?

Een derde invalshoek is de vertaling van beleidsmatige uitgangspunten naar de praktijk. Hoe wordt dat gedaan, en zijn er al gunstige effecten bekend?

In dit artikel komen alle drie de invalshoeken aan bod via de volgende vragen.

- a. Wat wordt er verstaan over opbrengstgericht werken? Deze vraag beantwoorden we via een begripsanalyse.
- b. Is opbrengstgericht werken een effectieve aanpak om tot verbeterde leerresultaten te komen? Deze vraag beantwoorden we op basis van een review van onderzoeksliteratuur.
- c. Welke ervaringen zijn er inmiddels opgedaan met de bevordering van opbrengstgericht werken? Deze vraag beantwoorden we aan de hand van twee gevalsbeschrijvingen, het project *Streef: Gebruik maken van opbrengsten* (coördinatie: Rijksuniversiteit Groningen

en Cito) en het project *Opbrengstgericht Leiderschap* (coördinatie: Algemene Vereniging van Schoolleiders).

Vooraf nog dit: wij beperken ons in dit artikel in hoofdzaak tot opbrengstgericht werken in het primair onderwijs. Hoewel de term ook in het voortgezet onderwijs een belangrijke rol speelt, gaan we daar niet op in. Dat neemt niet weg dat sommige bevindingen relevant kunnen zijn voor opbrengstgericht werken in het voortgezet onderwijs. In de discussie komen we op dit punt terug.

2 Begripsanalyse van opbrengstgericht werken

Taalkundig gezien is opbrengstgericht werken een samenstelling, waarvan de betekenis bepaald wordt door de samenstellende leden. Het gaat om een manier van werken waarbij de opbrengst voorop staat, en niet iets anders als bijvoorbeeld werkprocessen of reductie van kosten. In combinatie met onderwijs is opbrengstgericht werken weinig anders dan een pleonasme, aangezien onderwijs per definitie op nader benoemde opbrengsten is gericht, bijvoorbeeld op de overdracht van kennis en vaardigheden. Tenzij men tot uitdrukking zou willen brengen dat de opbrengsten centraler zouden moeten staan, meer de nadruk zouden moeten krijgen en bijgevolg ook verhoogd zouden moeten worden. Taalkundig zou *opbrengstgericht* werken dan een voor de hand liggende term kunnen zijn.

De belangstelling voor leeropbrengsten kent een voorgeschiedenis. In 1994 constateerde de Commissie Evaluatie Basisonderwijs – destijds aan het werk gezet om de Wet op het Basisonderwijs van 1985 te evalueren – dat de opbrengsten van het basisonderwijs, uitgedrukt in leerresultaten, op belangrijke onderdelen sterk verbeterd konden worden (Commissie Evaluatie Basisonderwijs, 1994). Het onderwijsaanbod was, gezien de beschikbare onderwijstijd, te omvangrijk om hoge opbrengsten op allerlei terreinen mogelijk te maken. Daarnaast constateerde de commissie dat er in het basisonderwijs te weinig sprake was van een opbrengstgerichte cultuur. Basisscholen zouden meer zorggericht dan opbrengstgericht te werk gaan. De constatering dat onderwijsopbrengsten beter zouden moeten, is sindsdien vaak herhaald, onder andere in het rapport *Tijd voor Onderwijs* (Commissie Dijsselbloem, 2008), in studies en adviezen van de Onderwijsraad en in de onderwijsverslagen die de Inspectie van het Onderwijs jaarlijks uitbrengt.

Vanaf 2008 verschijnt de term opbrengstgericht werken in ministeriële beleidsteksten. In het *Voorstel voor verbetering van de toets- en verantwoordingspraktijk in het*

basisonderwijs wordt aangegeven dat scholen meer en beter opbrengstgericht moeten gaan werken:

“De centrale ambitie van de Kwaliteitsagenda is het verbeteren van onderwijsresultaten voor taal en rekenen. Een goede evaluatiepraktijk op scholen is daar een belangrijk onderdeel van. Het gebruik van toetsen is immers essentieel om de kwaliteit van het onderwijs te bewaken en tijdig achterstanden bij leerlingen te onderkennen.” (Brief aan de Tweede Kamer, 16 juni 2008)

De in de brief aangewezen weg om dat te bereiken is een verbeterde evaluatiepraktijk: het hanteren van hoge standaarden, het nauwgezet volgen van de ontwikkeling van leerlingen en het aanpassen van het onderwijs aan de hand van toetsuitslagen. Deze trits wordt samengevat onder de noemer opbrengstgericht werken.

Ongeveer tezelfdertijd brengt de Onderwijsraad een advies uit over hoe scholen opbrengstgericht kunnen werken (Onderwijsraad, 2008). Opbrengstgerichtheid wordt als volgt omschreven:

“Met opbrengstgerichtheid doelt de raad op een kenmerk van een school waarbij concrete doelstellingen voor leerlingen worden geformuleerd op basis van de leerlingenpopulatie en waarbij alle betrokken partijen er gericht aan werken de gestelde doelen te bereiken.” (o.c., p. 13)

Deze omschrijving is in een aantal opzichten opener of breder dan wat het Ministerie van Onderwijs, Cultuur en Wetenschap voor ogen heeft: a. de expliciete koppeling aan taal en rekenen ontbreekt, b. er is geen sprake van verbeterde leeropbrengsten, en c. de verbeterde evaluatiepraktijk wordt niet expliciet aangeduid als de te bewandelen weg. In plaats daarvan benadrukt de Onderwijsraad het belang van een doelgerichte werkwijze en beveelt ze aan om werk te maken van een verbeterde samenwerking tussen de drie belangrijkste partijen: leerlingen en ouders, leraren en schoolleiders en schoolbesturen.

De door de Onderwijsraad geadviseerde verbreding van opbrengstgericht werken is door de nationale overheid niet overgenomen, zo blijkt uit het actieplan *Basis voor presteren* (Ministerie van Onderwijs, Cultuur en Wetenschap, 2011a). In dit plan neemt de bevordering van opbrengstgericht werken een centrale plaats in. Blijkens de uitwerking gaat het bij

opbrengstgericht werken om het realiseren van verbeterde leeropbrengsten voor taal en rekenen en om het versterken van leraarcompetenties. Leraren moeten hoge verwachtingen hanteren en goed inspelen op de verschillende behoeften van leerlingen. Daarbij moeten ze regelmatig toetsen afnemen en de uitkomsten daarvan analyseren en vertalen in een gedifferentieerd leerstofaanbod.

De Inspectie van het Onderwijs conformeert zich in grote lijnen aan de opvattingen van het Ministerie. Onder opbrengstgericht werken

“... verstaat de Inspectie het systematisch en doelgericht werken aan het maximaliseren van leerlingprestaties. Het gaat dan met name om het doorlopen van alle stappen van de evaluatieve cyclus.” (Inspectie van het Onderwijs, 2010)

Evenals in de opvatting van het Ministerie is hier een sterke verwevenheid te constateren tussen middel en doel. Maximalisering van leervorderingen moet bereikt worden door een integrale toepassing van de evaluatieve cyclus. Die verwevenheid blijkt ook uit de operationele definitie die de Inspectie heeft ontwikkeld. Om te beoordelen in welke mate basisscholen opbrengstgericht werken, hanteert de Inspectie vijf indicatoren. Deze hebben alle te maken met vorderingentoetsen en andere procedures voor evaluatie:

- a. De school gebruikt een samenhangend systeem van instrumenten en procedures voor het volgen van de prestaties en de ontwikkeling van leerlingen.
- b. De leraren volgen en analyseren de vorderingen van hun leerlingen systematisch.
- c. De school gaat de effecten van de zorg na.
- d. De school evalueert jaarlijks systematisch de kwaliteit van haar opbrengsten.
- e. De school evalueert regelmatig het onderwijsleerproces.

Voor het schooljaar 2011-2012 stelde de Inspectie aan de hand van deze indicatoren vast dat slechts 28 procent van de basisscholen volledig opbrengstgericht werkt, dat wil zeggen op alle vijf criteria een voldoende scoort (Inspectie van het Onderwijs, 2013).

Historisch gezien is opbrengstgericht werken dus gekoppeld aan tegenvallende leeropbrengsten of ook aan hogere ambities. Maar het voorgaande laat zien dat er tot op heden geen sprake is van eenduidige opvattingen over wat opbrengstgericht werken betekent. Wij benoemen een tweetal kwesties waarover de opvattingen divergeren. De eerste is of opbrengstgericht werken als een effectief middel wordt beschouwd om hogere leeropbrengsten te realiseren. Het Ministerie en de Inspectie menen van wel. De

Onderwijsraad is daarover minder uitgesproken. Zij bepleit het opbrengstgericht werken vanuit een ander perspectief: scholen moeten concrete doelstellingen formuleren voor hun leerlingenpopulatie en een samenwerking aangaan, onder andere met ouders, om de gestelde doelen te bereiken. De tweede divergentie betreft de vraag of opbrengstgericht werken verbonden is met de kernvakken Nederlandse taal en rekenen, dan wel een wijder bereik heeft. Een expliciete koppeling aan taal en rekenen treffen we alleen aan in opvattingen van het Ministerie, maar niet bij de Onderwijsraad of de Inspectie.

In de opvattingen van het Ministerie en de Inspectie zijn middel en doel sterk met elkaar verbonden: opbrengstgericht werken wordt beschouwd als het middel bij uitstek voor het realiseren van verhoogde leeropbrengsten. Vanuit wetenschappelijk perspectief is zo'n sterke koppeling wellicht niet zo verstandig. Een belangrijke reden voor een meer terughoudende opstelling is dat het niet zeker is dat de opbrengstgerichte werkwijze tot verhoogde leeropbrengsten zal leiden. Bovendien bestaat het risico dat andere middelen om tot hogere opbrengsten te komen, zoals leertijdverlenging, didactische verbeteringen of vergroting van ouderbetrokkenheid buiten beeld blijven.

In het vervolg van het artikel hanteren we een smalle opvatting van opbrengstgericht werken. *Opbrengstgericht werken betekent dat scholen en leraren regelmatig de vorderingen van hun leerlingen bepalen en de uitkomsten daarvan gebruiken bij de inrichting van volgende onderwijsactiviteiten.* Met deze opvatting zoeken we aansluiting bij het beleidsprogramma van het Ministerie, waarin een zwaar accent ligt op regelmatige evaluatie en op de benutting van evaluatiegegevens voor volgende activiteiten. Het is een specificatie die ook aansluit bij opvattingen van de Inspectie, die opbrengstgericht werken met name invult met toets- en evaluatieprocedures. Een vervelend nadeel van de hiervoor gegeven definitie is dat deze taalkundig gezien ongelukkig is; deze is veel specifieker dan men uit de samenstellende onderdelen 'opbrengst' en 'gericht' zou kunnen opmaken. Een alternatief zou kunnen zijn om over te stappen op een term als meetgestuurd onderwijs of toetsgestuurd onderwijs, maar het nadeel van dergelijke termen is dat ze buiten de huidige discussie staan omdat niemand ze gebruikt (maar zie Ledoux, Blok & Boogaard, 2009).

3 Hoe effectief is opbrengstgericht werken?

3.1 Nederlands onderzoek

Nederlands onderzoek naar de effectiviteit van opbrengstgericht werken is schaars. Wij kennen alleen onderzoek van de Inspectie (Inspectie van het Onderwijs, 2009, 2010). In de eerste publicatie uit 2009 wordt een vergelijking gemaakt tussen twee groepen basisscholen. De ene groep bestaat uit 2821 basisscholen, representatief voor de populatie van alle basisscholen. De andere groep wordt aangeduid als sterke scholen. Deze groep telt 209 basisscholen die drie jaar achtereenvolgende bovengemiddelde eindopbrengsten hebben gerealiseerd, blijkend uit het schoolgemiddelde op de Eindtoets Basisonderwijs van Cito. De twee groepen zijn onder meer vergeleken op de vijf indicatoren die de Inspectie gebruikt om te bepalen in hoeverre een school opbrengstgericht werkt (zie de vorige paragraaf). Voor beide groepen scholen is het percentage bepaald dat op de betreffende indicator voldoende scoort. Het verschil is uitgedrukt in een z-score. Tabel 1 laat zien dat er slechts in twee gevallen sprake is van een statistisch significant verschil tussen de twee groepen scholen. In de groep sterke scholen zijn vaker scholen aan te treffen die in voldoende mate de effecten van de leerlingenzorg nagaan of die jaarlijks de leeropbrengsten evalueren. Deze bivariante en correlatieve verbanden hebben evenwel weinig bewijskracht voor de stelling dat opbrengstgericht werken tot verbeterde leeropbrengsten leidt. Door de dichotomisering (voldoende versus niet voldoende, representatief versus sterk) is de gebruikte maat voor de samenhang bovendien erg grof. Relevanter is nog dat de samenhang niet op causaliteit hoeft te berusten en dat er zeer waarschijnlijk sprake is van een complex van factoren die in samenhang bepalen hoe effectief een school is.

-- HIER ONGEVEER TABEL 1 --

De tweede publicatie uit 2010 is gewijd aan de vormgeving en opbrengsten van het onderwijs in rekenen-wiskunde op een steekproef van 166 basisscholen. Aan de hand van het waarderingskader dat de Inspectie bij schoolbezoek hanteert, zijn de betrokken scholen ingedeeld in drie groepen: scholen die bij rekenen-wiskunde opbrengstgericht werken (23 procent), scholen die weinig opbrengstgericht tewerk gaan (16 procent) en een tussengroep, aangeduid als matig opbrengstgericht (61 procent). De eerste groep scholen bereikt hogere leeropbrengsten dan de beide andere groepen. Omgekeerd worden scholen met hoge leeropbrengsten positiever beoordeeld op een aantal aspecten van het didactische proces en de kwaliteitszorg. Het onderzoek naar het onderwijs rekenen-wiskunde vertoont vergelijkbare methodologische beperkingen als het eerdere onderzoek uit 2009.

Buitenlands onderzoek naar het belang van opbrengstgericht werken voor verbeterde leeropbrengsten is in veel ruimere mate beschikbaar dan Nederlands onderzoek. We beperken ons hier tot de hoofdlijnen, zoveel mogelijk op basis van overzichtsartikelen. In de bespreking sluiten we aan bij het bekende onderscheid tussen formatieve en summatieve toetsing. Formatieve toetsing of evaluatie is gericht op bijsturing van het onderwijsleerproces (*assessment for learning*), summatieve toetsing op evaluatie achteraf (*assessment of learning*).

3.2 De effectiviteit van formatieve toetsing

Vijftien jaar geleden publiceerden Black en Wiliam een omvattend overzicht van het onderzoek naar de effecten van toetsgebruik op de kwaliteit van het onderwijs (Black & Wiliam, 1998). Dit overzicht is gebaseerd op meer dan 250 afzonderlijke publicaties en heeft in de erop volgende jaren veel gezag verworven. Zo wordt er in latere literatuur veelvuldig naar verwezen. De review gaat met name in op effecten van *formative assessment* "... [which] is to be interpreted as encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (o.c., p. 8). In de review staan drie vragen centraal:

- a. Zijn er aanwijzingen dat formatieve toetsing leidt tot beter leren?
- b. Zijn er aanwijzingen dat de bestaande praktijk van toetsgebruik in de klas kan worden verbeterd?
- c. Is duidelijk welke maatregelen nodig zijn om tot een verbeterde formatieve toetsing te komen?

Het antwoord op de beide eerste vragen is bevestigend: a. regelmatige feedback aan leerlingen heeft een positief effect op hun vorderingen, en b. de praktijk in klaslokalen laat nog veel ruimte voor verbeteringen. De auteurs signaleren de volgende problemen: veel toetsmethoden versterken oppervlakkig leren, ze leiden tot competitie tussen leerlingen (in plaats van *personal improvement*) en ze zijn meer gericht op het afleggen van verantwoording dan op verbeteren van het leren. Kortom, leraren toetsen wel, maar ze gebruiken de verkeerde toetsen, of de goede toetsen op de verkeerde manier. Ze toetsen summatief, terwijl ze formatief zouden moeten toetsen.

Black en Wiliam staan in deze conclusies niet alleen. In een andere review, gewijd aan de effectiviteit van feedback, wordt eveneens een negatieve conclusie getrokken over de manier waarop leraren toetsen en andere meetinstrumenten gebruiken (Hattie & Timperley,

2007). We citeren: “Most current assessments provide minimal feedback, too often because they rely on recall and are used as external accountability thermometers rather than as feedback devices that are integral to the teaching and learning process” (p. 104). Hattie en Timperley merken op dat feedback in schoolklassen vaak persoonsgericht is in plaats van taakgericht en derhalve ineffectief. In hun model voor effectieve feedback staan drie vragen centraal: wat zijn de doelen, hoe ver ben ik al gevorderd, en wat moet ik doen om de doelen te realiseren. Effectieve feedback zou volgens dit model op alle drie de vragen antwoord moeten geven, zowel als het gaat om feedback aan leerlingen als om feedback aan leraren.

Vergelijkbare inzichten zijn te vinden in de conclusies die Shute (2008) verbindt aan haar review over de effectiviteit van formatieve feedback. Formatieve feedback moet aan bepaalde voorwaarden voldoen om het leren te kunnen ondersteunen: niet-veroordelend (*nonevaluative*), helpend (*supportive*), op het juiste tijdstip (*timely*) en precies of duidelijk (*specific*).

De derde vraag van Black en Wiliam, over welke maatregelen er nodig zijn om tot een verbeterde formatieve toetsing te komen, valt buiten het bestek van deze sectie. Toch is hun antwoord op zichzelf genomen interessant. Zij staan op het standpunt dat verbeteringen niet gemakkelijk te realiseren zijn en langlopende verbeteringstrajecten vergen, waarin experts en leraren intensief samenwerken. De ineffectieve toetsingsmethoden zijn diep geworteld, zowel in de schoolorganisatie als in de leraar. In latere artikelen gaan ze in op hun ervaringen, opgedaan in projecten om leraren te helpen hun toetspraktijk te verbeteren (Black e.a., 2004; Black & Wiliam, 2003; Wiliam, 2008). Deze ervaringen bevestigen dat veranderingen in de toetspraktijk veel tijd en intensieve begeleiding vergen.

3.3 De effectiviteit van summatieve toetsing

Summatieve toetsing vindt vaak plaats aan het eind van een cursus of onderwijsprogramma, maar ook wel tussentijds als een programma meerdere jaren beslaat. De uitkomsten worden in hoofdzaak voor twee doeleinden aangewend, voor het certificeren van leerlingen en voor evaluatie van het betreffende onderwijsonderdeel en – in het verlengde daarvan – voor de evaluatie van de leraar en de school. In het Nederlandse basisonderwijs is certificering strikt genomen niet aan de orde, al kunnen scores op de Eindtoets Basisonderwijs van Cito en vergelijkbare toetsen wel een rol spelen bij de beslissing hoe de schoolloopbaan verder zal verlopen. In deze sectie beperken we ons tot de evaluatieve functie: leidt de regelmatige

afname van summatieve toetsen en het gebruik daarvan met het oog op evaluatie van programma's en scholen tot onderwijsverbetering?

Over het antwoord op deze vraag wordt al decennialang gedebatteerd, ongeveer zo lang als dat er met dergelijke toetspraktijken ervaring is opgedaan. De discussie vindt de laatste jaren plaats onder de noemer van *standards-based education*, en eerder onder *outcome-based education*. In de Verenigde Staten, en overigens ook elders, zijn veel scholen onderworpen aan zogenoemde *accountability* programma's. Zulke programma's eisen dat scholen jaarlijks toetsen afnemen en over de resultaten daarvan rapporteren. Vaak zijn de af te nemen toetsen precies voorgeschreven en zijn aan de toetsen ook standaarden verbonden. Het gebruik van standaarden maakt het mogelijk een uitspraak te doen over goede en slechte scholen. Een slechte school is bijvoorbeeld een school waar relatief weinig leerlingen de standaard hebben gehaald. Het Amerikaanse programma *No Child Left Behind* (NCLB) uit 2002 is een bekend voorbeeld van een *accountability* programma. NCLB verplicht de Amerikaanse staten tot het instellen van een verplicht toetsprogramma, jaarlijks af te nemen aan alle leerlingen in het basisonderwijs en het voortgezet onderwijs (*grade 3 – 8*). Verplichte toetsing moet leiden tot een hoger onderwijspeil, onder andere door de identificatie van leerlingen en scholen die onder de maat (*substandard*) presteren. Om scholen in beweging te krijgen, zijn aan negatieve uitkomsten sancties verbonden.

De effectiviteit van de *standards-based reform movement* is omstreden. In de discussie gaat veel aandacht uit naar onbedoelde negatieve bijeffecten (Darling-Hammond, 2004; De Wolf & Janssens, 2007; Ingram, Louis & Schroeder, 2004; Kornhaber, 2004; McGill-Frantzen & Allington, 2006; Mehrens, 2002; Shepard, 2003; Wang, Beckett & Brown, 2006). We vatten ze samen onder vijf noemers.

a. Summatieve toetsing kan leraren en scholen tot onzuiver handelen brengen. De kans daarop is groter, naarmate de belangen voor de school groter zijn. Leraren kunnen bijvoorbeeld de zwakste leerlingen buiten de toetsing houden. Om verzekerd te blijven van betrouwbare uitkomsten zijn controlemaatregelen nodig. Ehren en Swanborn (2012) laten zien dat onzuiver handelen in eigen land voorkomt bij de afname van de Eindtoets Basisonderwijs van Cito.

b. Summatieve toetsing kan leiden tot verarming van het curriculum. De kans daarop is groter naarmate de meetinstrumenten een slechtere afspiegeling vormen van het leerplan. In dat geval laten leraren zich verleiden tot een overaccentuering van de getoetste kennis en vaardigheden (*teaching to the test*).

c. Summatieve toetsing is onbillijk tegenover scholen met achterstandsleerlingen. Die onbillijkheid ontstaat doordat schoolvorderingen niet alleen afhankelijk zijn van de kwaliteit van de school, maar bijvoorbeeld ook van de thuissituatie van leerlingen. Door middel van het schatten van de zogenoemde toegevoegde waarde kan hiervoor worden gecorrigeerd, maar aan deze handelwijze kleven haken en ogen (Amrein-Beardsley, 2008; Visscher, 2001). Laatstgenoemde auteur onderscheidt technisch-analytische problemen, bruikbaarheidsproblemen en ethische en maatschappelijke problemen.

d. Summatieve toetsing brengt risico's met zich mee voor zwakke leerlingen. Scholen kunnen ertoe besluiten zwakke leerlingen te laten zitten, naar andere scholen te verwijzen of niet toe te laten om daarmee het schoolgemiddelde positief te beïnvloeden. Dit risico geldt in het bijzonder de leerlingen die de gestelde standaard niet halen en leerlingen voor wie de instructietaal niet de eerste taal is (zie voor deze laatste groep: Solórzano, 2008).

e. Summatieve toetsing kan leiden tot onderstimulering van getalenteerde leerlingen. Zulke leerlingen halen de standaard toch wel en leraren kunnen daardoor in de verleiding komen hun aandacht te richten op leerlingen die 'op de wip' zitten, de wip van net wel of net niet de standaard halen.

4 Twee gevalsstudies

4.1 Project Streef: Gebruik maken van opbrengsten

In het schooljaar 2010-2011 is door de Rijksuniversiteit Groningen en enkele andere organisaties een interventie uitgevoerd op 40 basisscholen. De interventie betrof een trainingstraject voor leerkrachten. Het doel was leerkrachten opbrengstgericht te laten werken door ze te leren effectief gebruik te maken van het leerlingvolgsysteem van Cito. De onderzoeksopzet omvatte diverse voor- en nametingen, onder andere van de analysevaardigheden van leraren, hun attitude ten opzichte van opbrengstgericht werken en de leeropbrengsten van hun leerlingen. Per school heeft slechts een deel van de leraren de training gekregen. De niet-getrainde leraren vormden de controlegroep. Door de training aan wisselende bouwen aan te bieden, werd een balanceringsproces naar leerjaren nagestreefd (Doolaard, 2013).

De training omvatte negen bijeenkomsten, verspreid over het schooljaar, deels binnen de eigen school, deels schooloverstijgend. Tijdens twee bijeenkomsten zijn ook lesobservaties uitgevoerd om leerkrachten te ondersteunen bij het toepassen van differentiatie. Tussendoor

hebben leraren huiswerkopdrachten gekregen, waaronder het analyseren van toetsresultaten van de eigen leerlingen. De geschatte duur van de training beliep minimaal 40 uur (20 voor de bijeenkomsten, 20 voor huiswerkopdrachten).

De evaluatie van de eventuele interventie-effecten is gebaseerd op een contrastvergelijking tussen de ontwikkeling (verschil voor- en nameting) in de controlegroep en in de interventiegroep. Uit vragenlijstonderzoek onder de leerkrachten komt naar voren dat er op slechts een van de schalen sprake is van een significant verschillende ontwikkeling. Vergeleken met ongetrainde leraren reflecteren getrainde leraren meer op hun professionele handelen; ze kijken met name vaker terug naar hoe ze instructie hebben gegeven. Op de andere schalen – het zijn er twaalf, waaronder schalen voor onderwijskundig handelen en een schaal voor de attitude ten opzichte van opbrengstgericht werken – bleek het verschil tussen beide groepen leraren niet significant te zijn. Of de analysevaardigheden van de leerkrachten gedurende het schooljaar verbeterd zijn, kon niet worden vastgesteld; te weinig leerkrachten hebben twee of meer analyserapporten ingeleverd. Ook leeropbrengsten bij leerlingen zijn bij het effectonderzoek betrokken. Gebruik is gemaakt van uitkomsten op toetsen uit het leerlingvolgsysteem van Cito. Aan het eind van het interventiejaar is voor vier van de 25 bij de evaluatie betrokken toetsen vastgesteld dat er sprake is van een significant verschil, één maal in het voordeel van de controlegroep, drie maal in het voordeel van de interventiegroep. Een vergelijking een jaar later, gericht op het vaststellen van een retentie-effect laat zien dat er tussen beide groepen geen enkel significant verschil meer bestaat.

Al met al zijn de effecten van de interventie mager: geen of bijna geen significante effecten op gedrag en houding van leerkrachten, geen of bijna geen significante effecten op leeropbrengsten van leerlingen. De onderzoeker (Doolaard, o.c.) voert hiervoor twee verklaringen aan. De eerste is dat leraren nog onvoldoende samenwerken met hun collega's in het schoolteam. Leraren zouden zich te veel op de eigen werksituatie richten en het bevorderen van leeropbrengsten onvoldoende als een kwestie van *team work* zien. Een andere verklaring kan zijn dat leraren de stap van een toetsanalyse naar het opstellen en uitvoeren van een groepsplan nog onvoldoende beheersen. Die stap stelt hoge eisen aan de organisatorische en didactische bekwaamheden van leerkrachten, waaraan in het trainingstraject volgens de onderzoekster wellicht nog te weinig aandacht is besteed.

4.2 Project Opbrengstgericht Leiderschap

Onder auspiciën van de PO-Raad heeft de Algemene Vereniging van Schoolleiders gedurende drie schooljaren uitvoering gegeven aan het project *Opbrengstgericht Leiderschap*. Oogmerk was om onder schoolbesturen en scholen de opbrengstgerichte werkwijze te propageren.

Daartoe zijn er 79 pilots gestart, verspreid over het hele land. Tussen april 2009 en juli 2012 zijn deze pilots in de gelegenheid gesteld ervaring op te doen met opbrengstgericht werken. Een pilot bestond veelal uit een schoolbestuur met één of meer bijbehorende scholen. Doel van de pilots was dat schoolleiders en schoolbesturen ervaring zouden opdoen met opbrengstgericht werken en daarover kennis zouden verzamelen. Vragen die in de pilots aan de orde konden komen, waren onder andere: “Wat is opbrengstgericht werken precies, hoe kan dat vanuit de verschillende verantwoordelijkheden van bestuurder en schoolleider worden ingevuld, welke instrumenten en hulpmiddelen zijn beschikbaar, hoe wordt een goede analyse uitgevoerd, hoe wordt rekening gehouden met tal van andere, vaak moeilijk meetbare ontwikkelingen, ...” (geciteerd uit het projectplan, versie 11, maart 2009). Het is – daarnaast – een expliciet oogmerk van het project geweest, dat verworven kennis en inzichten beschikbaar worden gesteld aan besturen en scholen buiten de pilots. Het project is onderwerp geweest van een onafhankelijke evaluatie, die de bron vormt voor de navolgende beschrijving (Verbeek, Ledoux & Glaudé, 2012).

De werkwijze in het project is tamelijk open geweest. Pilots hebben zelf in documenten kunnen aangeven welke accenten men wilde leggen en welke wegen men daarbij dacht te volgen. Uit een analyse van de opgestelde plannen en een bevraging van de deelnemers komt een viertal ambities naar voren die zowel onder schoolbesturen als onder schoolleiders op een bijna volledige (90 procent of meer) instemming konden rekenen: a. verbeteren van de leeropbrengsten, b. leren analyseren en interpreteren van leerresultaten door schoolleiding en leerkrachten, c. leerkrachten duidelijk maken welke taken opbrengstgericht werken omvat, en c. de implementatie van een resultaatgerichte cultuur. De pilots hebben uiteenlopende instrumenten ingezet om hun doelen te bereiken, waaronder pilotoverstijgende uitwisselingsbijeenkomsten, collegiale-consultatiekringen, aanschaf of ontwikkeling van managementinformatiesystemen, uitbreiding of vervanging van leerlingvolgsystemen, cursussen in het gebruik van zulke systemen en begeleiding bij het opzetten en uitvoeren van schoolverbeterplannen. Praktisch alle scholen hebben externe adviseurs of experts over de vloer gehad. Alle schoolteams hebben studiemiddagen ingezet om de invoering van opbrengstgericht werken te bevorderen.

In de eindmonitor tonen zowel schoolbestuurders als schoolleiders zich in ruime mate positief over de bereikte resultaten. Schoolbestuurders waarderen de mate van implementatie van opbrengstgericht werken gemiddeld met een 7.6 (op een schaal van 1-10), schoolleiders met een gemiddelde van 7.0. Een ruime meerderheid van de schoolleiders (71 procent) meent dat opbrengstgericht werken een vast element in de schoolcultuur is geworden. Ook schoolbestuurders zijn hierover in ruime meerderheid positief. Gevraagd naar wat men zelf als het grootste succes beschouwt, valt vaak het woord *cultuuromslag*. Daarmee wordt bedoeld: er is een bewustwording in gang gezet, de nadruk ligt op leeropbrengsten en er is een gezamenlijke betrokkenheid en gedeelde verantwoordelijkheid ontstaan.

De cijfers over de effectiviteit van het project in termen van leeropbrengsten vertellen een ander verhaal. Van 526 bij het project betrokken basisscholen is de gemiddelde schoolscore op de Eindtoets Basisonderwijs van Cito gedurende vier jaar gevolgd en vergeleken met de landelijke cijfers. Dit heeft geleid tot twee bevindingen. De eerste is dat de projectscholen tussen 2009 en 2012 geen ontwikkeling laten zien, niet positief, niet negatief. De tweede is dat de schoolgemiddelden van de projectscholen in geen enkel genoemd schooljaar afwijken van de landelijke cijfers. Wat er mogelijkwijs ook is veranderd ten aanzien van het opbrengstgerichte werken op de projectscholen, het heeft niet tot verbeterde eindopbrengsten geleid, althans niet binnen de door het onderzoek bestreken termijn van drie jaar.

5 Conclusies en discussie

5.1 Conclusies

Opbrengstgericht werken is een term die de laatste jaren duidelijke opgang heeft gemaakt. De term brengt – globaal gesproken – tot uitdrukking dat scholen en leraren regelmatig toetsen moeten afnemen en de uitkomsten daarvan kritisch moeten analyseren en gebruiken voor onderwijsverbetering. De verwachting is dat deze werkwijze tot verbeterde leeropbrengsten leidt. De term wordt in steeds bredere kringen gebruikt, vanaf het Ministerie van Onderwijs, Cultuur en Wetenschap die de term voor het eerst in 2008 in een beleidstekst opnam tot aan schoolbegeleidingsdiensten die maatwerktrajecten opbrengstgericht werken aanbieden en scholen die in schoolplannen vastleggen dat hun aanpak opbrengstgericht is. Dat verschillende gebruikers van de term er een verschillende inhoud aan geven, lijkt de consensus over het belang van opbrengstgericht werken niet te deren. Daardoor is

opbrengstgericht werken een nimbusterm geworden, een term met een prachtige en verblindende stralenkrans. Een nimbusterm bergt ook een gevaar in zich, zoals A. D. de Groot, aan wie we deze kwalificatie ontlenu, zo treffend beschreef (De Groot & Medendorp, 1986). Dit gevaar valt aan te duiden als begrips- of zinsverduistering. Het is een manier van omgaan met taal waarmee verhullende of vervagende effecten kunnen worden bereikt. Om dit gevaar te bezweren is het noodzakelijk dat naar een meer eenduidige of precieze definitie wordt toegewerkt. In het eerste deel van dit artikel is een overzicht gegeven van de verschillende interpretaties. Dat heeft geleid tot het voorstel de betekenis van opbrengstgericht werken als volgt vast te leggen: *opbrengstgericht werken betekent dat scholen en leraren regelmatig de vorderingen van hun leerlingen bepalen en de uitkomsten daarvan gebruiken bij de inrichting van volgende onderwijsactiviteiten*. Een voordeel van deze, in vergelijking met andere interpretaties tamelijk smalle betekenisomschrijving, is dat deze aansluiting vindt bij het beleidsprogramma van het Ministerie, waarin de invoering van opbrengstgericht werken wordt gepropageerd.

Dat opbrengstgericht werken tot beter onderwijs of tot verbeterde leeropbrengsten leidt of zou moeten leiden, is *niet* in de voorgestelde definitie opgenomen. Een belangrijke reden daarvoor is dat dit verband aan twijfel onderhevig is, zoals de conclusie is van het tweede deel van het artikel waarin we een review van de onderzoeksliteratuur presenteren. In het literatuuroverzicht maken we onderscheid tussen formatieve toetsing (gericht op bijsturing van het onderwijsleerproces) en summatieve toetsing (achteraf vaststellen welke de vorderingen zijn geweest voor certificerings- of evaluatiedoeleinden). Van formatieve toetsing is vastgesteld dat deze bij kan dragen aan verbeterde leerresultaten. Van belang daarbij is vooral dat leerlingen inhoudelijke feedback krijgen, feedback die taakgericht is en die leerlingen aanwijzingen geeft over wat ze in hun denken of gedrag moeten veranderen. Tegelijkertijd blijkt uit onderzoek dat de toetspraktijk in klaslokalen nog veel te wensen overlaat. Leraren bepalen weliswaar regelmatig de vorderingen van hun leerlingen, maar hun feedback aan leerlingen is vaak persoonsgericht in plaats van taakgericht. Met andere woorden, ze informeren hun leerlingen over hoe goed ze iets doen, maar veelal niet over hoe ze zich kunnen verbeteren.

De effectiviteit van summatieve toetsing is omstreden. In de Verenigde Staten is veel ervaring opgedaan met summatieve toetsingsprogramma's, aan de uitkomsten waarvan voor leraren en scholen consequenties zijn verbonden. In zulke programma's worden leraren en scholen met tegenvallende toetsuitslagen onder druk gezet, bijvoorbeeld via openbaarmaking

van de resultaten. In de discussie over de effectiviteit van zulke programma's ligt het accent op negatieve bijeffecten, waaronder onzuiver handelen van scholen en leraren en verarming van het curriculum tot alleen die onderwerpen die worden getoetst. Uit Nederlands onderzoek van de Inspectie van het Onderwijs komt naar voren dat de mate waarin scholen opbrengstgericht werken geen of alleen een zwakke samenhang vertoont met de hoogte van de leeropbrengsten.

In het derde deel van het artikel beschrijven we twee schoolverbeteringsprojecten die te beschouwen zijn als pogingen om het opbrengstgerichte werken in de schoolpraktijk te integreren, het project *Streef: Gebruik Maken van Opbrengsten (Streef)* en het project *Opbrengstgericht Leiderschap (OL)*. Qua uitvoering zijn de projecten niet helemaal vergelijkbaar. Het project Streef is voor scholen een wellicht dwingender keurslijf geweest dan het OL-project, met minder mogelijkheden voor maatwerk. Het Streef-project heeft ingezet op leerkrachtcompetenties rond de analyse van toetsuitslagen en de vertaling daarvan in groepsplannen. In het OL-project heeft daarentegen een sterkere nadruk gelegen op de kwaliteit van het management en op managementinformatiesystemen. Voor beide projecten geldt dat de inspanningen *niet* hebben geleid tot aantoonbaar hogere leeropbrengsten. Ze leveren dus geen steun aan de veronderstelling dat opbrengstgericht werken tot hogere leeropbrengsten leidt.

5.2 Discussie

Het huidige onderwijsbeleid is gebaseerd op de veronderstelling dat opbrengstgericht werken een min of meer duidelijk te identificeren werkwijze van scholen en leraren is, die bij een juiste toepassing tot verbeterde leeropbrengsten leidt. Die verwachting, zo is in het voorafgaande betoogd, is aan twijfel onderhevig. Risico's in de huidige situatie zijn misbegrip en onbedoelde neveneffecten. Misbegrip kan ontstaan wanneer onder de noemer van opbrengstgericht werken een zwaar accent wordt gelegd op de verantwoordingsfunctie in plaats van op onderwijsverbetering. Onbedoelde neveneffecten kunnen zich voor gaan doen als aan hoge leeropbrengsten specifieke belangen (*high stakes*) worden verbonden, bijvoorbeeld in de sfeer van beloningen of straffen voor leraren en scholen (bonussen, bijzondere kwalificaties) of wanneer scholen naar hun leeropbrengsten zouden worden gerangordend zoals in *league tables*. Zulke belangen kunnen scholen en leraren verleiden tot onzuivere toetspraktijken, zoals in paragraaf 3.3 geschetst. Dat zulke praktijken niet alleen in

de Verenigde Staten maar ook in Nederland voorkomen, wordt beschreven door Ehren en Swanborn (2012).

De vraag is wat verstandige stappen zijn om uit de problemen te geraken. Wij bespreken er twee. De eerste is dat de hoge verwachtingen rond opbrengstgericht werken naar een realistischer niveau worden teruggebracht. Er moet ruimte komen voor nuchtere vragen aangaande de effectiviteit ervan en opbrengstgericht werken moet niet de norm worden die scholen en leraren dwingend wordt opgelegd. Terugschroeven van de verwachtingen betekent overigens niet – het zij expliciet gesteld – dat aan de waarde van hoge leeropbrengsten wordt afgedaan. Primaire taak voor het onderwijs is en blijft om leerlingen optimale ontwikkelingskansen te bieden. Een tweede stap is dat er gevarieerde en kleinschalige experimenten worden gedaan rond opbrengstgericht werken (Young & Kim, 2010). Vragen die in zulk onderzoek aan de orde gesteld moeten worden, betreffen allereerst de effectiviteit van verschillende uitwerkingen van opbrengstgericht werken. Factoren die in zulk onderzoek een rol zouden kunnen spelen zijn onder andere de wijze waarop leerresultaten worden bepaald, het type toetsen, de manier waarop de uitkomsten worden teruggekoppeld naar leerlingen, het uitvoeren van een kwaliteitscyclus (zoals *plan-do-check-act*), het hanteren van meer of minder gedifferentieerde onderwijsdoelen en het bieden van onderwijs-op-maat. Wanneer eenmaal kennis is opgedaan over effectieve elementen, doemen vragen op over de implementatie en borging. Hoe kunnen scholen en leerkrachten worden ondersteund bij de invoering van effectief gebleken vormen van opbrengstgericht werken, welke rol is weggelegd voor initiële scholing en nascholing? De genoemde kleinschaligheid is om twee redenen van belang. Kleinschaligheid draagt bij aan een getrouwe uitvoering van de onderzochte varianten (*treatment fidelity*). Kleinschaligheid is ook een kwestie van onderzoeksethiek; het gaat niet aan om grote groepen scholen, leraren en leerlingen te betrekken bij interventies waarvan de werkzaamheid nog onbekend is.

Tot slot, wat betekenen de hier getrokken conclusies die van toepassing zijn op het basisonderwijs, voor het voortgezet onderwijs? Wij vermoeden dat ze grotendeels ook daar geldig zijn. De conceptuele onduidelijkheid over opbrengstgericht werken is in elk geval niet sectorgebonden. En ook de twijfels over de effectiviteit van opbrengstgericht werken hebben niet exclusief betrekking op het primair onderwijs. Een deel van het in paragraaf 3 gerefereerde onderzoek is gedaan op Amerikaanse *high schools*. Een verdere overeenkomst is dat in de ogen van de overheid ook scholen voor voortgezet onderwijs opbrengstgericht moeten gaan werken (Ministerie van Onderwijs, Cultuur en Wetenschap, 2011b), zonder dat

er sprake is van *evidence-based* voorbeelden, waaruit blijkt dat deze aanpak tot verbeterde leeropbrengsten leidt. Een verschil tussen beide sectoren is er wel wat betreft de gangbare evaluatiepraktijk op scholen. Waar basisscholen al veelvuldig gebruik maken van genormeerde en geëquivalenteerde toetsen van het Cito, is dat bij scholen in het voortgezet onderwijs veel minder het geval. Dat heeft tot gevolg dat een betrouwbare vastlegging van leervorderingen in het voortgezet onderwijs veel minder makkelijk te realiseren is. Om die reden lijkt het ons verstandig om parallel aan de hiervoor bepleite kleinschalige experimenten rond opbrengstgericht werken in het basisonderwijs, zulke experimenten ook in het voortgezet onderwijs op te zetten.

Literatuur

- Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, 37(2), 65-75.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 9-21.
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5, 7-75.
- Black, P. & Wiliam, D. (2003). 'In praise of educational research': formative assessment. *British Educational research Journal*, 29, 623-637.
- Commissie Dijsselbloem (2008). *Tijd voor onderwijs*. Den Haag: Tweede Kamer der Staten-Generaal.
- Commissie Evaluatie Basisonderwijs (1994). *Zicht op kwaliteit. Evaluatie van het basisonderwijs* (Eindrapport). Utrecht: Inspectie van het Onderwijs.
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, 106, 1047-1085.
- De Groot, A. D. & Medendorp, F. L. (1986). *Term, begrip, theorie; inleiding tot de signifiante begripsanalyse*. Meppel/Amsterdam: Boom.
- De Wolf, I. F., & Janssens, F. J. G. (2007). Effects and side effects of inspections and accountability in education: an overview of empirical studies. *Oxford Review of Education*, 33, 379-396.

- Doolaard, S. (2013). *Effecten van het trainings- en begeleidingstraject 'Streef: Gebruik maken van opbrengsten'*. Groningen: GION, Gronings Instituut voor Onderzoek van het Onderwijs.
- Ehren, M. C. M., & Swanborn, M. S. L. (2012). Strategic data use of schools in accountability systems. *School Effectiveness and School Improvement*, 23, 257-280.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112.
- Ingram, D., Louis, K. S. & Schroeder, R.G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record*, 106, 1258-1287.
- Inspectie van het Onderwijs (2006). *Kwaliteit onderwijs in de vier grote steden in de periode 2003-2005*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2009). *De sterke basisschool; definitie en kenmerken*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2010). *Opbrengstgericht werken in het basisonderwijs; een onderzoek naar opbrengstgericht werken bij rekenen-wiskunde in het basisonderwijs*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2013). *De staat van het onderwijs; onderwijsverslag 2011/2012*. Utrecht: Inspectie van het Onderwijs.
- Kornhaber, M. L. (2004). Appropriate and inappropriate forms of testing, assessment, and accountability. *Educational Policy*, 18, 45-70.
- Ledoux, G., Blok, H., & Boogaard, M. (2009). *Opbrengstgericht werken; over de waarde van meetgestuurd onderwijs*. Amsterdam: SCO-Kohnstamm Instituut.
- McGill-Frantzen, A., & Allington, R. (2006). Contamination of current accountability systems. *Phi Delta Kappan*, 87(10), 762-766.
- Mehrens, W. A. (2002). Consequences of assessment: What is the evidence? In G. Tindal (Ed.), *Large-scale assessment programs for all students: validity, technical adequacy, and implementation* (pp. 149-177). Mahwah, N.J.: Erlbaum.
- Ministerie van Onderwijs, Cultuur en Wetenschap (2007).
- Ministerie van Onderwijs, Cultuur en Wetenschap (2011a). *Actieplan 'Basis voor presteren'*. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap.

- Ministerie van Onderwijs, Cultuur en Wetenschap (2011b). *Actieplan Beter Presteren: opbrengstgericht en ambitieus*. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap.
- Onderwijsraad (2008). *Partners in onderwijsopbrengst*. Den Haag: Onderwijsraad.
- Shepard, L. A. (2002). The hazards of high-stakes testing. *Issues in Science and Technology*, 19(2), 53-58.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153-189.
- Solórzano, R. W. (2008). High stakes testing: Issues, implications, and remedies for English language learners. *Review of Educational Research*, 78, 260-329.
- Verbeek, F., Ledoux, G., & Glaudé, M. (2012). *Op weg naar opbrengstgericht leiderschap*. Amsterdam: Kohnstamm Instituut.
- Visscher, A. (2001). Publieke schoolprestatie-indicatoren: de problemen op een rij. In A. B. Dijkstra, S. Karsten, R. Veenstra & A. J. Visscher (Red.), *Het oog der natie: scholen op rapport; standaarden voor de publicatie van schoolprestaties* (pp. 54-61). Assen: Koninklijke Van Gorcum.
- Visscher, A. & Ehren, M. (2011). *De eenvoud en complexiteit van opbrengstgericht werken* (analyse in opdracht van de Kenniskamer van het Ministerie van Onderwijs, Cultuur en Wetenschap). Opgehaald op 6 maart 2013, van <http://www.rijksoverheid.nl/bestanden/documenten-en-publicaties/rapporten/2011/07/13/de-eenvoud-en-complexiteit-van-opbrengstgericht-werken/visscher-ehren-eenvoud-en-complexiteit-van-opbrengstgericht-werken-def-1-7-11.pdf>.
- Wang, L., Beckett, G. H., & Brown, L. (2006). Controversies of standardized assessment in school accountability reform: A critical synthesis of multidisciplinary research evidence. *Applied Measurement in Education*, 19, 305-328.
- Wiliam, D. (2008). Changing classroom practice. *Educational Leadership*, 65(4), 36-42.
- Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: a literature review. *Education Policy Analysis Archives*, 18 (19). Opgehaald op 6 maart 2013, van <http://epaa.asu.edu/ojs/article/view/809>.

Tabel 1

Percentage 'voldoende' op indicatoren voor opbrengstgericht werken; tevens de z-score als aanduiding voor de grootte van het verschil; bron: Inspectie van het Onderwijs (2009)

Indicator	Percentage scholen met een voldoende beoordeling			Z-score
	Representatieve			
	groep scholen (N = 2.821)	Sterke scholen (N = 209)		
a. De school gebruikt een samenhangend systeem van instrumenten en procedures voor het volgen van de prestaties en de ontwikkeling van leerlingen.	92	94	0.06	
b. De leraren volgen en analyseren de vorderingen van hun leerlingen systematisch.	93	97	0.15	
c. De school gaat de effecten van de zorg na.	60	73	0.28*	
d. De school evalueert jaarlijks systematisch de kwaliteit van haar opbrengsten.	56	72	0.32*	
e. De school evalueert regelmatig het onderwijsleerproces.	54	53	- 0.02	

* $p < .05$

Over de auteurs

Henk Blok is als senior onderzoeker verbonden aan het Kohnstamm Instituut.

Guuske Ledoux is wetenschappelijk directeur van het Kohnstamm Instituut.

Jaap Roeleveld is als senior onderzoeker verbonden aan het Kohnstamm Instituut.