

# Opbrengstgericht werken in het primair onderwijs: een effectieve weg naar onderwijsverbetering?

H. Blok, G. Ledoux en J. Roeleveld

## Samenvatting

Verbeterde leeropbrengsten voor taal en rekenen staan hoog op de politieke agenda. Opbrengstgericht werken wordt in dit verband als een belangrijk middel gepropageerd. We besteden vanuit twee invalshoeken aandacht aan opbrengstgericht werken: a. wat valt er onder te verstaan, en b. is er evidentie dat opbrengstgericht werken tot verhoogde leeropbrengsten leidt? De gehanteerde onderzoeksmethoden zijn begripsanalyse en een quick scan van de onderzoeksliteratuur. We concluderen ten eerste dat er verschillende opvattingen bestaan 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. Dat geldt zowel voor vormen waarbij het accent gelegd wordt op formatieve evaluatie, als voor summatieve evaluatievormen. 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.

**Kernwoorden:** opbrengstgericht werken, onderwijsverbetering, begripsanalyse, review

## 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 punt op de huidige 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* voor het primair onderwijs en *Beter Presteren* voor het voortgezet onderwijs (Ministerie van Onderwijs, Cultuur en Wetenschap, 2011a, 2011b). Opbrengstgericht werken – door scholen en leerkrachten – is een van de wegen waarlangs hogere leeropbrengsten 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 tweede 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?

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

- Wat wordt er verstaan over opbrengstgericht werken? Deze vraag beantwoorden we via een begripsanalyse.
- Is opbrengstgericht werken een effectieve aanpak om tot verbeterde leerresultaten te komen? Deze vraag beantwoorden we op basis van een quick scan van onderzoeksliteratuur.

Wij beperken ons in dit artikel tot opbrengstgericht werken in het primair onderwijs, hoewel de term ook in het voortgezet onderwijs een belangrijke rol speelt.

Toch menen wij dat sommige bevindingen ook 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 delen. 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 benoembare 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 passende term kunnen zijn.

De belangstelling voor opbrengstgerichtheid kent een voorgeschiedenis. In 1994 constateerde de Commissie Evaluatie Basisonderwijs – destijds ingesteld 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.” (Ministerie van Onderwijs, Cultuur en Wetenschap, 2008)

De in de brief aangewezen weg om dat te bereiken is een verbeterde evaluatiepraktijk: het hanteren van hoge standaarden, het via genormeerde toetsen 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 regering 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 invalshoek van de Inspectie van het Onderwijs stemt in grote lijnen overeen met 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. Deze cyclus bestaat uit een vaste sequentie: vastleggen van doelen en standaarden, het verzamelen van empirische evidentie over het gerealiseerde niveau via een vorm van toetsing (incl. vormen van observatie), het interpreteren van de toetsuitkomsten in het licht van de nagestreefde standaarden en het treffen van maatregelen om de standaarden alsnog te realiseren, indien zou blijken dat zulks nog niet gelukt is (Ledoux, Blok & Boogaard, 2009). Om te beoordelen in welke mate basisscholen opbrengstgericht

werken, hanteert de Inspectie een operationalisatie op basis van 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 2012-2013 stelde de Inspectie aan de hand van deze indicatoren vast dat slechts 26 procent van de basisscholen volledig opbrengstgericht werkt, dat wil zeggen op alle vijf criteria een voldoende scoort (Inspectie van het Onderwijs, 2014). Dit percentage is al enige jaren tamelijk stabiel.

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 relatief open opvatting van opbrengstgericht werken. *Opbrengstgericht werken betekent dat scholen of leraren regelmatig de vorderingen van hun leerlingen bepalen en de uitkomsten daarvan gebruiken bij de inrichting van volgende onderwijsactiviteiten.* Deze omschrijving laat open dat opbrengstgericht werken moet leiden tot verbeterde leeropbrengsten. Of dat effect er daadwerkelijk is, vormt geen onverbreekbaar element van opbrengstgericht werken. De omschrijving is óók open omdat er geen beperking is aangebracht tot de kernvakken Nederlandse taal en rekenen. Evenmin is in de omschrijving vastgelegd hoe de vorderingen moeten worden bepaald, via genormeerde toetsen of anderszins, via bijvoorbeeld methodegebonden toetsen, zelf gemaakte proefwerken of observatielijsten. Merk tot slot op dat opbrengstgericht werken een kenmerk kan zijn van de werkwijze van scholen, maar ook van leraren.

Met onze 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. In deze paragraaf besteden we aandacht aan drie initiatieven: een tweetal onderzoeken van de Inspectie van het Onderwijs en een tweetal onderzoeken naar schoolverbeteringsprojecten, respectievelijk *Streef* en *Opbrengstgericht Leiderschap*.

De beide onderzoeken van de Inspectie (Inspectie van het Onderwijs, 2009, 2010) berusten op het vergelijken van groepen scholen. In de eerste publicatie uit 2009 wordt onderscheid 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 achtereen 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 paragraaf 2). Voor beide groepen scholen is het percentage bepaald dat op de betreffende indicator voldoende scoort. Het verschil is uitgedrukt in een z-score. Uit de vergelijking komt naar voren dat er op slechts twee indicatoren 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. De effectgroottes bedragen omstreeks 0.30. Deze bivariate en correlatieve verbanden hebben evenwel weinig bewijskracht voor de stelling dat opbrengstgericht werken tot verbeterde leeropbrengsten leidt. In de eerste plaats is door

de dichotomisering (voldoende versus niet voldoende, representatief versus sterk) de gebruikte maat voor de samenhang erg grof. In de tweede plaats, en dat is nog relevanter, behoeft de waargenomen samenhang niet op causaliteit te berusten. Zeer waarschijnlijk is er sprake van een complex van factoren die in samenhang bepalen hoe effectief een school is, de zogenoemde effectketen.

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. Dit onderzoek naar het onderwijs rekenen-wiskunde vertoont echter vergelijkbare methodologische beperkingen als het eerdere onderzoek uit 2009.

De projecten *Streef* (Deunk & Doolaard, 2013; Doolaard, 2013) en *Opbrengstgericht Leiderschap* (Verbeek, Ledoux & Glaudé, 2012) hebben beide tot doel gehad basisscholen te ondersteunen bij de invoering van opbrengstgericht werken en tevens de effecten daarvan op de leeropbrengsten te bepalen. In *Streef* betrof de interventie een trainingstraject voor leerkrachten gedurende een jaar, gericht op een effectiever gebruik van het leerlingvolgsysteem van Cito. Per school (n = 40) heeft slechts een deel van de leraren de training gekregen; de niet-getrainde leraren vormden de controlegroep. Het project *Opbrengstgericht Leiderschap* heeft een langere looptijd gehad: gedurende drie schooljaren zijn enkele honderden scholen gefaciliteerd om de opbrengstgerichte werkwijze in te voeren. Activiteiten hadden plaats onder regie van de betrokken schoolbesturen, waardoor scholen veel ruimte kregen voor

een eigen uitwerking. Praktisch alle scholen hebben externe adviseurs of experts over de vloer gehad en alle schoolteams hebben studiemiddagen ingezet.

In beide gevallen 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 bij leerlingen. In het project *Opbrengstgericht Leiderschap* werd het schoolgemiddelde op de Eindtoets Basisonderwijs van Cito gedurende vier jaar gemonitord bij 526 deelnemende scholen. De uitkomst is dat de projectscholen gedurende deze jaren geen ontwikkeling laten zien, niet positief, niet negatief. Vanzelfsprekend roepen de uitkomsten vervolgvragen op, waarvan de belangrijkste is: ligt het aan een onvolkomen implementatie, of gaat van de werkwijze op zichzelf geen gunstige werking uit?

### 3.2 Buitenlands onderzoek

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 overzichtartikelen die de afgelopen 15 tot 20 jaar verschenen zijn. De hier te bespreken literatuur is voornamelijk langs drie wegen verzameld: via *searches* en citatie-analyses in *Web of Science*, via zoekopdrachten op het web in Google en andere zoekmachines, en via de sneeuwbal methode. Gebruikte zoektermen zijn onder meer: standards-based testing, data-driven/data-based decision making, formative assessment/evaluation, summative assesment/evaluation, assessment for learning, accountability systems, high stakes testing, review, meta-analysis. Het resultaat van alle spuurwerk is wellicht niet een complete verzameling van alle beschikbare reviews vanaf 1995. Maar de aanwezigheid van veel geciteerde artikelen die qua invalshoek en verschijningsdatum een zekere verscheidenheid laten zien, bieden toch enige garantie dat de verzamelde literatuur een representatief beeld geeft. Verder is nog op te merken dat in de verzamelde internationale literatuur een zwaar accent ligt op effecten van

toetsgebruik, een beperking die overigens in overeenstemming is met de essentiële plaats van toetsgebruik binnen de opbrengstgerichte werkwijze.

In de bespreking houden we het bekende onderscheid aan 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*). Beide vormen van evaluatie kunnen geschaard worden onder opbrengstgericht werken, zij het op een verschillende manier. Bij formatieve evaluatie wordt veelal gestreefd naar verbeterde leeropbrengsten bij dezelfde leerlingen wier kennis en vaardigheden vastgesteld zijn, summatieve evaluatie biedt met name mogelijkheden om bij volgende leerlingcohorten naar hogere opbrengsten te streven.

#### *De effectiviteit van formatieve toetsing*

Ruim 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). 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 toetsen 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 latere reviews 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. 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 voorwaarden voldoen om het leren te kunnen ondersteunen: niet-veroordelend (*nonevaluative*), helpend (*supportive*), op het juiste tijdstip (*timely*) en precies (*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 ineffektieve 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, Harrison, Lee, Marshall & Wiliam, 2004; Black & Wiliam, 2003; Wiliam, 2008). Deze ervaringen bevestigen dat veranderingen in

de toetspraktijk veel tijd en intensieve begeleiding vergen (zie voor een uitwerking: Sluismans, Joosten ten Brinke & Van der Vleuten, 2013).

De meta-analyse van Faber en Visscher (2014) berust op vijftien studies, waarvan er negen gepubliceerd zijn na het omvattende overzicht van Black en Wiliam uit 1998. In deze meta-analyse wordt aandacht geschonken aan twee vragen: de vraag naar de effectiviteit van leerlingvolgsystemen (leidt het gebruik ervan tot hogere leeropbrengsten?) en de vraag naar de factoren die van invloed zijn op het doelmatige gebruik van leerlingvolgsystemen. De *overall* effectgrootte die met het gebruik van een leerlingvolgsysteem samenhangt, bleek 0.07 (s.e. 0.02) te bedragen, een betrekkelijk bescheiden grootte. Een opvallende, maar moeilijk te duiden bevinding is dat de grootste effecten gevonden zijn in de oudste studies, de studies van voor de eeuwwisseling. Een andere bevinding is dat de effectgroottes sterker zijn in studies met kleinere steekproeven (kleiner dan 1000 leerlingen) dan in grotere steekproeven. Ook deze uitkomst is moeilijk te duiden. De uitkomst kan duiden op een schaalproblematiek, in die zin dat de effectiviteit grotendeels verloren gaat bij opschaling van de interventie. Maar evenzeer is mogelijk dat de uitkomst een gevolg is van *publication bias*: kleinschaliger onderzoek dat niet tot verwerping van de nulhypothese leidt, loopt een grotere kans in 'de la van ongepubliceerd onderzoek' te verdwijnen dan kleinschalig onderzoek dat wèl tot verwerping van de nulhypothese leidt (Dickersin, 1990). Wat betreft de effectiviteitsbevorderende factoren, de tweede vraag van Faber en Visscher, blijkt een hoge feedbackfrequentie – minimaal maandelijks – van belang te zijn. Wat ook bevorderend werkt, is als toetsuitkomsten rechtstreeks verbonden zijn aan instructie-adviezen. Maar wat de meta-analyse van Faber en Visscher vooral ook laat zien, is dat er een grote variëteit is in toetsgebruik en opbrengstgericht werken en dat zeker niet alle uitwerkingen tot verbeterde leeropbrengsten bij leerlingen leiden. Veel hangt waarschijnlijk af van de inbedding van leerlingvolgsystemen in de evaluatieve cyclus als geheel.

### *De effectiviteit van summatieve toetsing*

Summatieve toetsing vindt plaats aan het eind van een cursus of onderwijsprogramma, maar ook wel tussentijds als een programma meer jaren beslaat. De uitkomsten kunnen in hoofdzaak voor twee doeleinden worden 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. In deze sectie beperken we ons daarom tot de evaluatieve functie: leidt de regelmatige afname van summatieve toetsen met het oog op de evaluatie van het geboden onderwijs tot onderwijsverbetering? Rond deze vraag bestaat veel literatuur. We groeperen deze onder twee noemers: *standards-based* en *data-driven reform*.

*Standards-based* schoolverbetering, in de Verenigde Staten een alom bekend fenomeen, onderwerpt scholen 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 (*grade 3 – 8*). Verplichte toetsing moet leiden tot een hoger onderwijspeil, onder andere door de identificatie van leerlingen en scholen die onder de standaard presteren. Om scholen in beweging te krijgen zijn aan negatieve uitkomsten sancties verbonden. Op dezelfde leest zijn inmiddels ook programma's ontwikkeld voor *teacher evaluation*, waarbij kan worden opgemerkt dat zulke programma's bij leraren weerstand kunnen oproepen (zie onder andere het tijdschrift *Teachers College Record*,

jaargang 116 nummer 1, een themanummer met de omineuze titel *High-stakes teacher evaluation: high cost – big losses*).

De effectiviteit van de *standards-based* schoolverbetering is omstreden. In de discussie gaat veel aandacht uit naar onbedoelde negatieve bijeffecten (Amrein-Beardsley, 2008; Darling-Hammond, 2004; De Wolf & Janssens, 2007; Ingram, Louis & Schroeder, 2004; Kornhaber, 2004; McGill-Frantzen & Allington, 2006; Mehrens, 2002; Shepard, 2002; Solórzano, 2008; Visscher, 2001; Wang, Beckett & Brown, 2006). Summatieve toetsing kan leraren en scholen tot onzuiver handelen brengen, het kan leiden tot verarming van het leerplan, het kan onbillijk zijn tegenover scholen met veel achterstandsleerlingen en het kan negatief uitpakken voor zwakke leerlingen. Deze negatieve effecten treden vaker op naarmate de belangen toeneemen (onder *high-stake conditions*).

Het boek *Leading schools in a data-rich world* (Earl & Katz, 2006) heeft de schijnwerpers gezet op *data-driven* schoolverbetering, in het Nederlands ook wel aangeduid als schoolprestatiefeedback (Visscher, Dijkstra & Karsten, 2012). De centrale gedachte hierachter is dat vorderingengegevens van leerlingen, maar ook allerlei andere gegevens zoals uitkomsten van oudertevredenheidspeilingen, benut kunnen worden voor onderwijsverbetering (Schildkamp, Ehren & Lai, 2012). Een verschil met het vorige onderzoeksdomein is dat er veelal geen sprake is van sancties, waardoor scholen niet onder grote druk worden gezet. Dat verkleint de kans op negatieve bijeffecten. Literatuur over data-gestuurde schoolverbetering is ontegenzeggelijk van belang voor de discussie over de effectiviteit van opbrengstgericht werken. Maar tegelijkertijd moeten we vaststellen dat de werkzaamheid van *data-driven* schoolverbetering nog verre van overtuigend is aangetoond (Turner & Coburn, 2012); zie voor voorbeelden van onderzoek met wisselende uitkomsten Campbell & Levin (2009); McNaughton, Lai & Hsiao (2012); Schildkamp, Vanhoof, Van Petegem & Visscher (2012); Slavin, Cheung, Holmes, Madden & Chamberlain (2013).

## 4 Conclusies en discussie

### 4.1 Conclusies

Opbrengstgericht werken is een term die de laatste jaren opgang heeft gemaakt. De term brengt – globaal gesproken – tot uitdrukking dat scholen 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 inmiddels breed 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 verblindende stralenkrans. Een nimbusterm bergt ook een gevaar in zich, zoals A. D. de Groot, aan wie we deze kwalificatie ontlenen, 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 open 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, ontbreekt in de voorgestelde definitie. Een belangrijke reden daarvoor is dat dit verband nog aan twijfel onderhevig is, zoals de conclusie is van het tweede deel van het artikel waarin we een quick scan van de onderzoeksliteratuur presenteren. In het literatuuroverzicht maakten we onderscheid tussen formatieve en summatieve toetsing. Beide vormen lijken bruikbare uitwerkingen van opbrengstgericht werken. Van formatieve toetsing is vastgesteld dat deze kan bijdragen 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 naar deze vorm van toetsing 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 omstreven. In de Verenigde Staten is veel ervaring opgedaan met summatieve toetsingsprogramma's, aan de uitkomsten waarvan sancties verbonden kunnen worden. In zulke programma's worden 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. Onderzoek rond het thema *data-driven* schoolverbetering, waarbij aan lagere leeropbrengsten veelal geen sancties worden verbonden, levert wisselvallige uitkomsten op: soms is er sprake van daadwerkelijk verbeterde leeropbrengsten, maar soms ook niet.

#### 4.2 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 door onduidelijkheid over de functies van toetsen (formatief, summatief, beide?) en over de bruikbaarheid van de beschikbare toetsen. Onbedoelde neveneffecten kunnen zich voordoen als aan leeropbrengsten sancties of beloningen worden verbonden, of wanneer scholen naar hun leeropbrengsten zouden worden gerangordend zoals in *league tables*. Zulke consequenties kunnen scholen en leraren verleiden tot onzuivere toetspraktijken, zoals in paragraaf 3.3 is 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 vervolgstappen zijn, gegeven de positieve connotatie die aan opbrengstgericht werken kleeft. 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 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 scholen is en blijft om leerlingen optimale ontwikkelingskansen te bieden. Een tweede stap is dat er gevarieerd en kleinschalig onderzoek – met een experimentele opzet – wordt gedaan rond opbrengstgericht werken (Young & Kim, 2010) om een beter beeld te krijgen van de meerwaarde in de Nederlandse context. 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 onderwijsop-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. 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 (Heemskerk e.a., 2014). 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 geëquivalerde 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 (EVAAS). *Educational Researcher*, 37, 65-75.
- Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 9-21.
- Black, P. & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5, 7-75.
- Black, P. & William, D. (2003). 'In praise of educational research': formative assessment. *British Educational Research Journal*, 29, 623-637.
- Campbell, C., & Levin, B. (2009). Using data to support educational improvement. *Educational Assessment, Evaluation and Accountability*, 21, 47-65.
- 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.
- Deunk, M., & Doolaard, S. (2013). *Attitude en handelen van basisschoolleerkrachten met betrekking tot het verbeteren en borgen van leerlingresultaten*. Groningen: GION, Gronings Instituut voor Onderzoek van het Onderwijs.
- 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.
- Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association*, 263, 1385-1389.

- Doolaard, S. (2013). *Effecten van het trainings- en begeleidingstraject 'Streef: Gebruik maken van opbrengsten'*. Groningen: GION, Gronings Instituut voor Onderzoek van het Onderwijs.
- Earl, L. M., & Katz, S. (2006). *Leading schools in a data-rich world. Harnessing data for school improvement*. Thousand Oaks, CA: Corwin Press.
- 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.
- Faber, J. M., & Visscher, A. J. (2014). *Digitale leerlingvolgsystemen: een review van de effecten op leerprestaties*. Enschede: Universiteit Twente.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112.
- Heemskerk, I., Verbeek, F., Kuiper, E., Oomens, M., Van de Linden, J., & Hilbrink, E. (2014). *Opbrengstgericht werken in het voortgezet onderwijs. Beeld en beleid*. Amsterdam: Kohnstamm Instituut.
- 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 (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 (2014). *De staat van het onderwijs; onderwijsverslag 2012/2013*. 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.
- McNaughton, S., Lai, M. K., & Hsiao, S. (2012). Testing the effectiveness of an intervention model based on data use: A replication series across clusters of schools. *School Effectiveness and School Improvement*, 23, 203-228.
- 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, NJ: Erlbaum.
- Ministerie van Onderwijs, Cultuur en Wetenschap (2007). *Scholen voor morgen*. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap.
- Ministerie van Onderwijs, Cultuur en Wetenschap (2008). *Brief van de staatssecretaris van Onderwijs, Cultuur en Wetenschap aan de voorzitter van de Tweede Kamer der Staten-Generaal, 16 juni 2008* (dossier 31 293, nr. 17). Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap.
- 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.
- Schildkamp, K., Ehren, M., & Lai, M. K. (2012). Editorial article for the special issue on data-based decision making around the world: from policy to practice to results. *School Effectiveness and School Improvement*, 23, 123-131.
- Schildkamp, K., Vanhoof, J., Van Petegem, P., & Visscher, A. (2012). The use of school self-evaluation results in the Netherlands and Flanders. *British Educational Research Journal*, 38, 125-152.
- 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.

- Slavin, R. E., Cheung, A., Holmes, G., Madden, N. A., & Chamberlain, A. (2013). Effects of a data-driven district reform model on state assessment outcomes. *American Educational Research Journal*, 50, 371-396.
- Sluijsmans, D., Joosten ten Brinke, D., & Van der Vleuten, C. (2013). *Toetsen met leerwaarde; een reviewstudie naar de effectieve kenmerken van formatief toetsen*. Maastricht: Universiteit Maastricht.
- Solórzano, R. W. (2008). High stakes testing: Issues, implications, and remedies for English language learners. *Review of Educational Research*, 78, 260-329.
- Turner, E. O., & Coburn, C. E. (2012). Interventions to promote data use: An introduction. *Teachers College Record*, 114, 1-13.
- 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., Dijkstra, A. B., & Karsten, S. (2012). Schoolprestatiefeedback als dé sleutel tot schoolverbetering? In A. B. Dijkstra & F. J. G. Janssens (Red.), *Om de kwaliteit van het onderwijs* (pp. 149-169). Den Haag: Boom Lemma.
- 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 30 januari 2014, 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.
- William, 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 30 januari 2014, van <http://epaa.asu.edu/ojs/article/view/809>.

## Auteurs

Alle drie de auteurs zijn werkzaam bij het Kohnstamm Instituut van de Universiteit van Amsterdam: **Henk Blok** en **Jaap Roeleveld** als senior onderzoeker, **Guuske Ledoux** als senior onderzoeker en wetenschappelijk directeur.

*Correspondentieadres:* Dr. H. Blok, Kohnstamm Instituut, Postbus 94208, 1090 GE Amsterdam, [h.blok@uva.nl](mailto:h.blok@uva.nl)

## Abstract

### Does data-driven teaching lead to better learning outcomes?

Better learning outcomes for language and math education is a central aim of the agenda of the Education Department in the Netherlands. The implementation of data-driven teaching (in Dutch: *opbrengstgericht werken*) has been chosen as the vehicle to realize this goal. In this article we evaluate this policy from two different points of view: a. is there consensus on the meaning of data-driven teaching and the constituent elements and b. is there sound empirical evidence that data-driven teaching leads to raised learning outcomes? The answers to both questions are negative: data-driven teaching has been given different interpretations by different parties concerned, and the effectiveness of data-driven teaching as a means to raise national standards does not withstand critical scrutiny. It is recommended to set up small-scale experiments to learn more about the effectiveness of specific operationalizations of data-driven teaching.