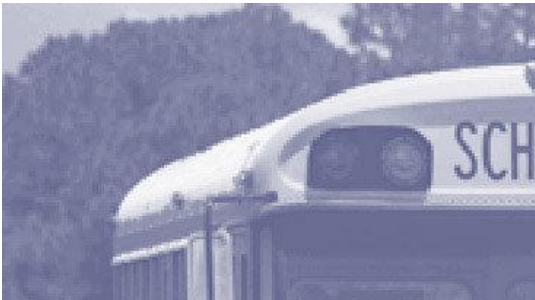




2012

Opbrengstgericht Werken en prestaties van leerlingen in het basisonderwijs



Bachelor afstudeeropdracht
van Marieke Abbes

Onderwijskunde, Universiteit Twente

Eerste Begeleider: Dr. Adrie Visscher

Tweede Begeleider: L. Staman Msc

8 mei 2012

Inhoud

1	Voorwoord	2
2	Algemene gegevens onderzoek	2
3	Samenvatting	2
4	Beschrijving van het onderzoek	3
4.1	Rationale	3
4.2	Conceptueel kader	3
4.2.1.	Opbrengstgericht werken en Focus	3
4.2.2.	Maat voor OGW	4
4.2.3.	Voorgaand onderzoek	4
4.2.4.	Moderatoren voor OGW	4
4.2.5.	Onderzoeksvragen	4
4.3	Relevantie	5
5	Onderzoeksontwerp	5
5.1	Type onderzoek	5
5.2	Procedure	5
5.3	Respondenten	6
5.4	Instrumenten	6
5.5	Data-analyse	6
6	Planning	7
6.1	Deadlines	7
6.2	Weekplanning	7
7	Referenties	8

1 Voorwoord

Dit document betreft het onderzoeksvoorstel ten behoeve van de bachelor afstudeerscriptie van ondergetekende. Het beschijft globaal waarom het onderzoek wordt uitgevoerd, welke concepten betrokken worden in de literatuurstudie, wat de onderzoeksvragen zijn en welke methode gebruikt zal worden. Tot slot is een planning van de uit te voeren activiteiten opgenomen met de bijbehorende deadlines. Het voorstel is opgesteld in samenspraak met de opdrachtgever, die tevens begeleider van het onderzoek is.

M. Abbes
3^e jaars student onderwijskunde

2 Algemene gegevens onderzoek

Onderzoeker	Marieke Abbes
Opleiding	Onderwijskunde
Vakgroep	Organisatie & Management
Datum	8 mei 2012
Eerste Begeleider	Dr. A.J. Visscher
Tweede Begeleider	L. Staman Msc.
Trefwoorden onderzoek	Opbrengstgericht werken, opbrengstgericht onderwijs, data driven, prestatiefeedback, leerlingprestaties, prestatiegerichtheid

Titel en ondertitel:

Opbrengstgericht Werken en prestaties van leerlingen in het basisonderwijs

Onderzoek naar de samenhang tussen de mate van opbrengstgericht werken van basisscholen, de prestatiegerichtheid van leerkrachten en de prestaties van leerlingen

3 Samenvatting

Het doel van dit onderzoek is een beredeneerde maat voor Opbrengstgericht werken op te stellen, en hiermee de relatie tussen de mate van Opbrengstgericht werken van scholen en de hoogte van hun leerlingprestaties te achterhalen. Ook zal de rol van prestatiegerichtheid van leerkrachten onderzocht worden, omdat er een sterke theoretische basis is voor een verband tussen deze variabele enerzijds en de mate van Opbrengstgericht werken en leerlingprestaties anderzijds. Het onderzoek bestaat uit een literatuurstudie en een correlatieel onderzoek. In de literatuurstudie zal onderzocht worden wat de uitkomst van voorgaand onderzoek zegt over de relatie tussen Opbrengstgericht werken en leerlingprestaties. In het correlatiele gedeelte zal gebruik worden gemaakt van reeds verzamelde data: interviews die de mate van OGW meten, toetsuitslagen voor de prestaties van leerlingen en een vragenlijst voor de prestatiegerichtheid van leerkrachten. Met deze data zullen meerdere regressie analyses uitgevoerd worden om de aard van het verband tussen de drie componenten in kaart te brengen.

4 Beschrijving van het onderzoek

4.1 Rationale

Dit onderzoek wordt uitgevoerd in het kader van de professionaliseringsinterventie Focus. Focus is een tweejarig scholingstraject dat erop gericht is een volledig team van leerkrachten, interne begeleiders en schoolleiders in het basisonderwijs professioneel te ontwikkelen op het gebied van Opbrengstgericht Werken. Met Opbrengstgericht werken (OGW) wordt bedoeld het systematisch en doelgericht werken aan het maximaliseren van onderwijsprestaties (Inspectie van het onderwijs, 2010). Van Opbrengstgericht werken wordt verwacht dat het een positief effect heeft op onderwijskwaliteit en leerprestaties. Daarom neemt de onderwijsinspectie sinds 2008/2009 de mate van opbrengstgericht werken van scholen mee in haar beoordeling van de kwaliteit van scholen. Mede om die reden zijn scholen erg geïnteresseerd in professionalisering op dit gebied, en hiertoe is het project Focus in 2008 gestart.

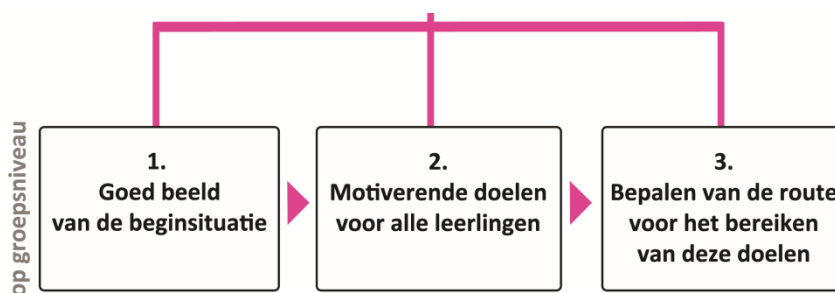
Het belang van dit onderzoek is te onderzoeken of er inderdaad een positief verband is tussen de mate van opbrengstgericht werken van een school en prestaties van leerlingen. Zijn de leerlingprestaties van scholen die meer opbrengstgericht werken ook daadwerkelijk hoger? Noodzakelijk voor het meten van de mate van opbrengstgericht werken van een school is een goede maat. Hiervoor bestaat al een maat die de inspectie hanteert, maar die is nogal globaal. In dit onderzoek wordt er een gedetailleerdere maat opgesteld door de inspectiemaat te combineren met de operationalisatie van het concept Opbrengstgericht werken die de onderzoekers van Focus hanteren. Met deze maat kan een uitspraak gedaan worden over de mate van Opbrengstgericht werken van de scholen die meededen aan de professionaliseringsinterventie Focus. Dan wordt onderzocht wat de aard van het verband is tussen de mate van Opbrengstgericht werken en de leerlingprestaties van die scholen, waarbij ook relatie met de variabele prestatiegerichtheid zal worden onderzocht.

4.2 Conceptueel kader

4.2.1. Opbrengstgericht werken en Focus

Er zijn verschillende interventies die erop gericht zijn scholen meer Opbrengstgericht te laten werken. Focus is hier een voorbeeld van. Focus is een tweejarig scholingstraject voor leerkrachten, schoolleiders en interne begeleiders in het basisonderwijs (Visscher, Peters & Staman, 2010). De cursus betreft scholing in het beter lezen en interpreteren van het leerlingvolgsysteem van de eigen school, het evalueren van leerling- en schoolprestaties, het stellen van goede doelen op verschillende niveaus en manieren om het onderwijs aan te passen op basis van al deze gegevens. De cursus richt zich specifiek op het rekenonderwijs, maar de aangeleerde technieken zijn in principe toe te passen op allerlei leergebieden. Van de cursus wordt verwacht dat ze de mate waarin scholen Opbrengstgericht werken vergroot, wat betekent dat zij beter gefundeerde beslissingen zullen nemen, en derhalve beter beoordeeld zullen worden door de onderwijsinspectie. Uiteindelijk moet dit ertoe leiden dat de onderwijskwaliteit en tenslotte de leerlingprestaties omhoog gaan.

Als uitgangspunt voor Focus wordt onderstaand model gehanteerd voor Opbrengstgericht werken (OGW), bestaande uit drie componenten, die grafisch zijn weergegeven in figuur 1.



Figuur 1. De drie componenten van Opbrengstgericht werken op groepsniveau. Overgenomen uit: “De effecten van het eerste jaar opbrengstgericht werken met Focus,” door L. Staman, A. J. Visscher en H. Luyten, 2012, paper gepresenteerd op de Onderwijs Research Dagen 2012, p. 3. Enschede: Universiteit Twente.

De eerste component is het krijgen van een goed beeld van de beginsituatie. In de cursus betekent dit dat er getraind wordt in het leren lezen en interpreteren van het leerlingvolgsysteem. De tweede component is het stellen van goede doelen. Deze doelen sluiten aan bij de principes van de goal setting theorie, wat erop neerkomt dat doelen uitdagend, haalbaar en evalueerbaar moeten zijn. De derde component tenslotte is het bepalen wat er moet gebeuren om het gestelde doel te behalen. De cursus geeft bijvoorbeeld informatie over de onderwijsactiviteiten van rekensterke scholen. In dit onderzoek gelden deze drie componenten voor OGW als uitgangspunt. Daarnaast wordt in het Focus project een onderscheid gemaakt tussen drie verschillende niveaus waarop OGW betrekking heeft: het schoolniveau, het groepsniveau en het leerlingniveau.

4.2.2. Maat voor OGW

Om te kunnen meten of een cursus de mate van Opbrengstgericht werken verhoogt, en of OGW inderdaad zorgt voor beter onderwijs en hogere prestaties, is het nodig om de mate van OGW goed te kunnen meten. De ontwikkelaars van Focus hebben een gestructureerd interview gemaakt om de drie componenten van opbrengstgericht werken, zoals weergegeven in figuur 1, te operationaliseren. Deze operationalisatie is gedetailleerder dan de maat die de onderwijsinspectie gebruikt bij de jaarlijkse inspectie van basisscholen. Het interview biedt de mogelijkheid een nauwkeuriger maat te ontwikkelen, waarmee dan de mate van Opbrengstgericht werken van een school gemeten kan worden. Deze nauwkeurige maat kan vervolgens gebruikt worden voor verder onderzoek binnen Focus, maar wellicht ook in andere settings. In dit onderzoek zal het gebruikt worden om het verband te onderzoeken tussen de mate van opbrengstgericht werken van een controle en experimentele scholen en hun leerlingprestaties.

4.2.3. Voorgaand onderzoek

Voorafgaand aan het praktische deel van dit onderzoek, zal een literatuurstudie worden gedaan naar het verband tussen Opbrengstgericht werken en leerling prestaties. Het fenomeen onderwijsbeslissingen meer te baseren op verzamelde informatie over prestaties van leerlingen wordt in wetenschappelijke literatuur behandeld in artikelen over Data Driven Teaching, Data Driven Decision making en Assessment for Learning.

In de literatuurstudie zal worden ingegaan op eerder onderzoek naar het verband tussen de mate van OGW en leerling prestaties. Omdat de gebruikte termen nogal uiteenlopen en er verschillende manieren van onderzoeken zijn, kan niet zonder meer een conclusie worden getrokken over dit verband, maar zullen de manier van onderzoeken en de operationalisering van Opbrengstgericht werken vergeleken moeten worden.

4.2.4. Moderatoren voor OGW

Op basis van literatuur zijn sterke aanwijzingen dat er factoren zijn die verband houden met leerling prestaties, maar ook met de mate waarop scholen Opbrengstgericht werken. Deze factoren worden moderatoren voor OGW genoemd, omdat ervan verwacht wordt dat zij het effect van een training gericht op OGW versterken. Eén van deze moderatoren is prestatiegerichtheid. Van prestatiegerichtheid wordt verwacht dat er een sterk verband is met zowel OGW als leerling prestaties. De vraag is of dit inderdaad zo is, en wat precies de aard van dit verband is ten opzichte van leerling prestaties en Opbrengstgericht werken.

4.2.5. Onderzoeksvragen

Bovenstaande leidt tot de volgende onderzoeksvragen:

1. Literatuurstudie: wat is bekend over de relatie tussen OGW en leerprestaties (incl., hoe hard zijn de conclusies?).
2. Hoe ziet een beredeneerde maat voor Opbrengstgericht werken eruit?
3. Wat zeggen de verzamelde gegevens bij experimentele en controle scholen van Focus I over de relatie tussen de mate van opbrengstgericht werken van een school en hun leerlingprestaties?
4. Is er verschil in de mate van opbrengstgericht werken tussen experimentele en controle scholen?
5. Wat is bij de experimentele scholen de samenhang tussen enerzijds de prestatiegerichtheid van scholen en anderzijds de mate van OGW binnen die scholen en hun leerlingprestaties?

4.3 Relevantie

De onderwijsinspectie ziet Opbrengstgericht werken als belangrijk middel voor het verhogen van leerprestaties (Inspectie van het onderwijs, 2010). Deze ontwikkeling sluit aan bij internationale trends om meer meetgestuurd onderwijs te geven. Maar resulteert dit ook daadwerkelijk in beter onderwijs en leidt dit ook tot hogere leerprestaties? Voor onderzoek naar verbanden en effecten met betrekking tot Opbrengstgericht werken is een goede maat onontbeerlijk. De gedetailleerde maat voor OGW die voor dit onderzoek wordt opgesteld, kan gebruikt worden voor onderzoek binnen het Focus project, maar ook daarbuiten.

Vervolgens is het onderzoek een aanvulling op bestaand onderzoek naar OGW. Het onderzoek maakt inzichtelijk wat voor soort onderzoek er al gedaan is naar het verband tussen OGW en prestaties. Hierdoor kan worden ingeschat of er slechts sprake is van een verband, of dat causaliteit waarschijnlijk is. Als bijvoorbeeld experimenteel onderzoek ontbreekt, kan er geen uitspraak gedaan worden over effecten, maar slechts over verbanden.

5 Onderzoeksontwerp

5.1 Type onderzoek

Dit onderzoek bestaat uit een literatuurstudie en een praktisch onderzoek. In de literatuurstudie zal worden onderzocht wat er tot nu toe bekend is over het verband tussen OGW en leerling prestaties, en op welk soort onderzoek de conclusies berusten. Het praktische deel van het onderzoek is correlatieel. In tabel 1 wordt het type onderzoek per onderzoeksvraag weergegeven.

Tabel 1

Onderzoeksvragen met type onderzoek

Onderzoeksvraag	Type onderzoek
1. Wat zegt voorgaand onderzoek over de relatie tussen de mate van opbrengstgericht werken van een school en de prestaties van leerlingen?	Literatuurstudie
2. Hoe ziet een berekende maat voor Opbrengstgericht werken eruit?	Het combineren van gegevens
3. Wat zeggen de verzamelde gegevens bij experimentele en controle scholen van Focus I over de relatie tussen de mate van opbrengstgericht werken van een school en hun leerlingprestaties?	Samenhang onderzoek
4. Is er verschil in de mate van opbrengstgericht werken tussen experimentele en controle scholen?	Verskil onderzoek
5. Wat is bij de experimentele scholen de samenhang tussen enerzijds de prestatiegerichtheid van scholen en anderzijds de mate van OGW binnen die scholen en hun leerlingprestaties?	Samenhang onderzoek

5.2 Procedure

Voorafgaand aan het opstellen van dit onderzoeksvoorstel was er een oriëntatiefase. In deze fase zijn een aantal artikelen gelezen over Opbrengstgericht werken, vonden gesprekken plaats met de opdrachtgevers en is een cursusbijeenkomst van Focus bijgewoond. Daarnaast is de website van Focus bekeken en foldermateriaal over OGW en Focus. Ook zijn een aantal vragenlijsten ingevoerd in SPSS. Door deze activiteiten werden het onderwerp OGW, de inhoud van Focus, en de aanverwante begrippen duidelijk, en werd inzicht verkregen in de behoefte van de vakgroep en de aanwezige data.

Na goedkeuring van het voorstel zal allereerst een literatuurstudie worden uitgevoerd. Hierbij worden de volgende termen gebruikt: data-driven/data-based decision making/teaching, Assessment for

Learning en Opbrengstgericht werken/onderwijs, gecombineerd met outcomes, impact, effects, achievement en correlation (en de nederlandse woorden hiervoor). Het resultaat van de literatuurstudie wordt beschreven in het literatuurverslag.

Vervolgens wordt als voorbereiding op het praktische deel van het onderzoek een maat opgesteld voor opbrengstgericht werken. Dit zal gebeuren door de operationalisatie die voor Focus gebruikt wordt van de term OGW te vergelijken met de maat van de onderwijsinspectie. Ook de literatuurstudie kan inzicht geven hoe de maat het best kan worden samengesteld.

Dan zullen de ingevoerde interviews in SPSS bewerkt worden zodat ze aansluiten bij de opgestelde maat, zodat de mate van opbrengstgericht werken bij controle en experimentele scholen onderzocht kan worden. Ook zullen de toetsresultaten van leerlingen geschikt worden gemaakt voor analyses in SPSS, alsmede de vragenlijsten voor prestatiegerichtheid.

Vervolgens worden verschillende statistische analyses uitgevoerd in SPSS, waarna de resultaten worden beschreven in het onderzoeksverslag.

5.3 Respondenten

De respondenten zijn afkomstig van 43 experimentele en 43 controle scholen. De scholen zijn random aan de groepen toegewezen. Aan het onderzoek namen leerkrachten, Interne Begeleiders (IB'ers) en leerlingen deel.

De interviews zijn afgenomen onder de IB'ers van experimentele en controle scholen (N=86). De vragenlijsten voor prestatiegericht zijn afgenomen onder leerkrachten van experimentele scholen (N=600). De leerlingprestaties zijn afgenomen bij experimentele en controle scholen (N=?).

5.4 Instrumenten

Het praktische deel van het onderzoek bestaat uit het analyseren van reeds verzamelde data. Deze data zijn verzameld met de volgende instrumenten:

- Gestructureerde interviews, die de verschillende onderdelen van Opbrengstgericht werken meten. Deze interviews zijn afgenomen en ingevoerd in SPSS door Laura, zowel bij de experimentele als bij de controle scholen;
- Een vragenlijst waarmee prestatiegerichtheid is gemeten. De vragenlijsten zijn alleen afgenomen bij de experimentele scholen. De afgenomen vragenlijsten waren gedeeltelijk ingevoerd, en zijn gedeeltelijk door mij ingevuld.
- Toetsresultaten die leerling prestaties meten. De toetsen zijn afgenomen bij aanvang van de professionaliseringsinterventie, zowel bij de experimentele als bij de controle scholen. De toetsresultaten zijn reeds ingevoerd.

5.5 Data-analyse

In onderstaande tabel is beschreven hoe de data geanalyseerd zullen worden (tabel 2).

Tabel 2

Beschrijving van de data-analyse per deelvraag

Onderzoeksvraag	Data-analyse
1	Vergelijken van de conclusies uit verschillende onderzoeken en in kaart brengen op welk soort onderzoek de conclusies zijn gebaseerd.
2	Maat opstellen op basis van de interviews, de door de inspecties gebruikte maat en literatuur over de onderdelen van opbrengstgericht werken.
3	Enkelvoudige regressie analyse (op basis van de uit interviews opgestelde maat voor OGW en leerlingprestaties)

4	Onafhankelijke T-toets (gemiddelden vergelijken tussen e/c scholen met de opgestelde maat)
5	Meervoudige regressie analyse (maat voor OGW en vragenlijsten prestatiegerichtheid als voorspellers van leerlingprestaties)

6 Planning

6.1 Deadlines

Aangezien mijn tweede begeleider L. Staman per 11 juni met zwangerschapsverlof gaat, is het wenselijk dat dit onderzoek voor die datum is afgerond. Dit is echter gezien de studielast die de bachelor afstudeerhandleiding aangeeft wat krap. Daarom zal het literatuuronderzoek voor eind mei af worden gerond, zodat er nog een ruime week is waarin Laura kan helpen bij het opstellen van de maat voor OGW en voor de statistische analyses in SPSS.

Tabel 3

Deadlines afstudeeronderzoek

31 mei 2012	Deadline literatuuronderzoek (methode en resultaatendeel verslag af)
8 juni 2012	Laatste werkdag Laura (zwangerschapsverlof)
6 juli 2012	Deadline beide verslagen
16 juli 2012	Deadline groen licht vergadering
24 juli 2012	Deadline verzoek afstuderen naar BOZ
28 augustus 2012	Vergadering examencommissie
30 augustus 2012	Afstudeerdatum (Dit moet na 28 ^e i.v.m. vergadering examencommissie, maar voor 1 september i.v.m. toegang tot de Master)

6.2 Weekplanning

Tabel 4

Activiteiten per week

Week	Bijzonderheden	Activiteiten
20	do/vrij hemelvaart	Afmaken onderzoeksvoorstel Systematisch literatuur zoeken Lezen en literatuurmatrijs opstellen
21	kids vakantie Atelier 5 toetsen afnemen en evalueren	Schrijven literatuurverslag
22	maandag Pinksteren afsluiting Filosofie/reflectie	Maat opstellen voor OGW SPSS: OGW maat samenstellen Toetsresultaten voorbereiden voor analyse
23	afsluiting Atelier 5 vrijdag naar Zwitserland	Statische analyses SPSS uitvoeren en bespreken met Laura
24	Laura start verlof Geen andere vakken meer	Schrijven methodedeel onderzoek Schrijven resultaatendeel onderzoek
25		Schrijven inleiding, theoretisch kader, conclusie en discussie
26		Afmaken literatuurverslag: inleiding, conclusie, discussie
27		6 juli: beide verslagen af en mailen naar begeleiders
28		Begeleiders lezen verslagen
29	16 juli groenlichtvergadering	Eventueel aanpassen verslagen Verslag en afstudeerverzoek naar BOZ

7 Referenties

- Inspectie van het Onderwijs (2010). *Opbrengstgericht werken in het basisonderwijs*. Utrecht: Inspectie van het Onderwijs.
- Staman, L., Visscher, A. J. & Luyten, H. (2012). *De effecten van het eerste jaar opbrengstgericht werken met Focus*. Paper gepresenteerd tijdens de Onderwijs Research Dagen 2012. Enschede: Universiteit Twente.
- Visscher, A., Peters, M. & Staman, L. (2010). *Het Focus-project: opbrengstgericht werken op basis van prestatie-feedback*. Enschede: Universiteit Twente.

De Samenhang tussen Opbrengstgericht Werken op Basisscholen en Rekenprestaties van Leerlingen

Marieke Abbes
Universiteit Twente Enschede
6 augustus 2012

Samenvatting

Dit onderzoek is uitgevoerd in het kader van de professionaliseringsinterventie Focus. In 2010 is een eerste groep basisscholen gestart met de training Focus 1. Van deze groep werden vooraf een aantal gegevens verzameld door middel van vragenlijsten, interviews en toetsen. Deze gegevens vormen de basis van dit onderzoek. Voorafgaand aan het uitvoeren van statistische analyses is een scoreformulier ontwikkeld waarmee de mate van Opbrengstgericht Werken (OGW) per school kon worden gewaardeerd. Hiermee werd een nulmeting gedaan van de opbrengstgerichtheid van controle en experimentele scholen, op basis waarvan de samenhang werd onderzocht tussen de mate van OGW en leerling prestaties. Uit de resultaten blijkt dat controlescholen en experimentele scholen niet significant verschillen qua opbrengstgerichtheid. Met deze studie kon niet overtuigend worden aangetoond dat opbrengstgerichte scholen hogere leerling prestaties hebben.

1^e begeleider: Dr. A.J. Visscher

2^e begeleider: L. Staman MSc

De overheid verlangt van scholen dat zij meer opbrengstgericht gaan werken. De term Opbrengstgericht Werken (OGW) is redelijk nieuw, maar het principe niet. Het is gebaseerd op de principes van prestatiefeedback en goal setting. Van deze activiteiten is al langer bekend dat zij een gunstig effect hebben op leerprestaties. Ook in andere landen wordt veel aandacht besteed aan het gebruik van informatie over leerlingprestaties, maar vaak worden dan termen als “data-driven teaching” of “data-based decision making” gebruikt. Het basisprincipe is dat informatie over leerlingprestaties, vaak toetsresultaten, maar ook observaties of oudergesprekken, de basis vormen waarop beslissingen in het onderwijs worden genomen, wat uiteindelijk moet leiden tot beter onderwijs en betere leerlingprestaties. Ook de maatschappelijke ontwikkeling dat scholen verantwoording moeten afleggen over de kwaliteit van hun onderwijs, draagt bij aan de enorme aandacht voor deze werkwijze.

De Nederlandse onderwijsinspectie neemt sinds 2008/2009 de mate van opbrengstgericht werken van scholen mee in haar beoordeling van de kwaliteit van scholen. Opbrengstgericht werken wordt door de Inspectie van het Onderwijs omschreven als het systematisch en doelgericht werken aan het maximaliseren van de prestaties van leerlingen (OCW, 2011). Hierbij wordt gebruikt gemaakt van een vaste cyclus van meten van de beginsituatie, het stellen van doelen en het nemen van instructiemaatregelen om die doelen te bereiken. Voor het vaststellen van de beginsituatie is het de bedoeling dat beter gebruik wordt gemaakt van de informatie die leerlingvolgsystemen verschaffen. Omdat scholen vaak nog niet optimaal gebruik maken van hun leerlingvolgsysteem, en deze informatie niet gebruiken om hun onderwijs te verbeteren, investeert de Nederlandse overheid geld in professionaliseringsinterventies op dit gebied. Het project Focus is één van die interventies.

Focus is een tweejarige cursus voor basisscholen, opgezet door de Universiteit Twente, waarbij zowel leerkrachten als interne begeleiders en schoolleiders betrokken worden (Visscher, Peters & Staman, 2010). Scholen leren welke analyses ze met hun leerlingvolgsysteem kunnen uitvoeren en hoe ze de uitkomst van analyses moeten lezen en interpreteren. Tevens wordt in de cursus aandacht besteed aan het stellen van uitdagende, haalbare en SMART geformuleerde doelen op verschillende niveaus. Tenslotte leren cursisten hoe zij onderwijsaanpassingen kunnen doorvoeren. De cursus richt zich specifiek op het rekenonderwijs, maar de aangeleerde technieken zijn in principe ook toe te passen op andere vakken. Het doel van de cursus is het vergroten van de mate waarin scholen Opbrengstgericht werken, zodat zij beter beargumenteerde beslissingen zullen nemen, wat er uiteindelijk toe moet leiden dat de onderwijskwaliteit, en tenslotte de leerling prestaties omhoog gaan (Visscher, Peters & Staman, 2010).

Inmiddels is de derde groep scholen begonnen aan de training Focus, en heeft de eerste groep van 43 scholen de cursus afgerond. Om onderzoek te kunnen doen naar effecten van een dergelijke training is een meetinstrument nodig waarmee de mate van opbrengstgericht werken van een school kan worden vastgesteld. Met behulp van een dergelijk meetinstrument zal immers kunnen worden vastgesteld of er verschil is tussen experimentele en controle scholen in de mate waarin zij opbrengstgericht werken, of scholen meer opbrengstgericht gaan werken na het volgen van de cursus, en of de mate van opbrengstgericht werken beter onderwijs oplevert en hogere leerprestaties.

Dit onderzoek heeft zich beperkt tot de voormeting van de eerste groep controle scholen en experimentele scholen. De uitkomst van een gestructureerd interview onder die scholen op de nulmeting is gebruikt om een scoreformulier te ontwikkelen waarmee de mate van opbrengstgericht werken binnen de scholen kan worden gewaardeerd. Tevens is onderzocht of er een samenhang bestaat tussen de componenten van opbrengstgericht werken en de leerling prestaties van een school.

Conceptueel kader

In deze paragraaf worden de verschillende componenten van opbrengstgericht werken verder uitgediept, en zijn de resultaten van het literatuuronderzoek dat vooraf ging aan dit onderzoek samengevat. Dit literatuuronderzoek werd uitgevoerd om te onderzoeken hoe overtuigend de samenhang tussen Opbrengstgericht werken en leerprestaties al werd aangetoond in voorgaande studies. Ook is beschreven hoe de Onderwijsinspectie de opbrengstgerichtheid van een school op dit moment meet, en in hoeverre dit verbeterd zou kunnen worden.

Opbrengstgericht Werken

Voorafgaand aan dit onderzoek is een literatuurstudie gedaan naar de samenhang tussen opbrengstgericht werken en leerprestaties (Abbes, 2012). In deze literatuurstudie is eveneens de achtergrond van opbrengstgericht werken onderzocht. Hieronder volgt een korte samenvatting.

Opbrengstgericht werken is gebaseerd op de principes van prestatiefeedback, gecombineerd met principes uit de Goal Setting Theorie (Locke & Latham, 2002). Prestatiefeedback is feedback die verkregen wordt uit prestaties van leerlingen. Leerlingen “vertellen” in feite met hun prestaties, in hoeverre het onderwijs effect heeft gehad en of verwachte doelen van leerkrachten bereikt zijn. Met hun prestaties op toetsen, maar ook door observeerbaar gedrag, geven leerlingen informatie over wat ze kunnen, wat ze lastig vinden en waar ze achterblijven vergeleken met andere leerlingen. Door nauwkeurige analyse van hun prestaties, bijvoorbeeld door het analyseren van toetsen op gedetailleerd niveau, of door het volgen van hun ontwikkelingsgroei, zouden leerkrachten beter moeten kunnen vaststellen wat het niveau van een leerling op dat moment is. Ook zouden ze hun onderwijs hiermee kunnen evalueren. Tegenwoordig maken leerlingvolgsystemen het mogelijk de vorderingen van een groep of van een school te vergelijken over leerjaren, tussen groepen onderling en zelfs tussen scholen op landelijk niveau. Dit is waardevolle informatie, en het juist leren lezen en interpreteren hiervan zou ervoor moeten zorgen dat beter gefundeerde onderwijsbeslissingen worden genomen. Dit geldt op groepsniveau, maar ook op het niveau van individuele leerlingen, en zelfs op school- of bovenschools niveau. In het literatuurverslag worden de onderdelen van feedback onderzocht vanuit een algemene feedback cyclus, zoals die in de technische en biologische wetenschappen is beschreven. Deze onderdelen worden gerelateerd aan feedback in het onderwijs, wat resulteert in de volgende schakels uit de feedback-cyclus: signalen, sensoren, signaalverwerkers en controllers.

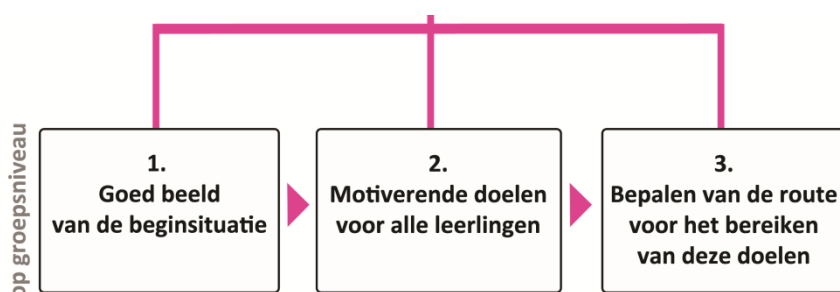
Het onderwijs wordt gezien als een systematisch proces, waarbij prestaties van leerlingen, maar ook observaties en oudergesprekken gezien kunnen worden als signalen die worden afgegeven. Deze signalen worden opgevangen door sensoren, de ogen en oren van een leerkracht. Deze kan hier al dan niet een leerlingvolgsysteem bij gebruiken. Een mechanisch systeem heeft vervolgens een signaalverwerker, in het onderwijs een leerkracht (of een interne begeleider of schoolleider), die de informatie verwerkt door deze te interpreteren en op waarde te schatten. Vervolgens bevat een feedbacksysteem altijd een controller, iets, en in het geval van het onderwijs iemand, die bijstelt op basis van de meting, zodat het systeem weer goed, of misschien zelfs beter, gaat werken. Dit kan een leerkracht zijn die zijn instructie bijstelt, een schoolleider die faciliteiten verbetert of leerkrachten bijstuurt, of een schoolbestuur dat organisatorische maatregelen neemt. In feite zijn al deze onderdelen schakels in de feedback cyclus van een schoolorganisatie.

Naast prestatiefeedback bevat opbrengstgericht werken ook nog een ander element, namelijk het stellen van goede doelen. Vanuit de theorie van de Goal Setting van Locke en Latham (2002) is bekend, dat het stellen van hogere, uitdagende doelen betere prestaties oplevert, mits de doelen haalbaar zijn en geaccepteerd worden door de betrokkenen.

Focus hanteert voor opbrengstgericht werken het model van Visscher & Ehren (2011). In dit model wordt opbrengstgericht werken op groepsniveau uitgesplitst in drie componenten, zoals te zien in figuur 1. Component 1 betreft het krijgen van een goed beeld van de beginsituatie. Hiervoor kan gebruik worden gemaakt van het leerlingvolgsysteem, een digitaal systeem waarmee de vorderingen van leerlingen zoals toetsresultaten en observaties, worden geregistreerd. Voor het analyseren van de toetsen worden verschillende leerlingvolgsystemen gebruikt. De meeste gebruikte systemen in Nederland zijn het computerprogramma LOVS van Cito en de administratiesystemen Parnassys, Esis en SchoolOAS (van Dotcom) (Meijer, Ledoux & Elshof, 2011). De systemen die in dit onderzoek door de scholen gebruikt werden, zijn Cito, Parnassys en Esis; Parnassys werd het meest gebruikt.

De tweede component heeft betrekking op het stellen van doelen. In Focus wordt aandacht besteed aan het gunstige effect van hoge doelen stellen. Tevens wordt geleerd dat doelen gesteld moeten worden volgens de SMART regel. Dit betekent dat doelen Specifiek, Meetbaar, Acceptabel, Realistisch en Tijdgebonden moeten zijn. Zo kan een doel worden gesteld in de vorm van een maximum percentage leerlingen dat een D- of E-score mag halen op de Cito eindtoets, maar het kan ook veel specifiek en uitdagender door te stellen dat leerling X over een half jaar een vijf punten hogere vaardigheidsscore haalt voor de rekentoets. Op die manier kan ook beter worden onderzocht waarom een doel wel of niet behaald is.

De derde component uit de OGW-cyclus van Visscher en Ehren (2011) is het nemen van instructiebeslissingen op basis van de vorige twee componenten. Het idee is, dat het juist inschatten van de beginsituatie, in combinatie met het stellen van uitdagende doelen, moet worden aangevuld met het kiezen van de juiste weg om de doelen te bereiken. De training geeft bijvoorbeeld informatie over de onderwijsactiviteiten van rekensterke scholen. In dit onderzoek gelden deze drie componenten voor OGW als uitgangspunt. Daarnaast wordt in het Focus project een onderscheid gemaakt tussen drie verschillende niveaus waarop OGW betrekking heeft: het schoolniveau, het groepsniveau en het leerlingniveau (Staman, Peters & Visscher, 2010).



Figuur 1. De drie componenten van Opbrengstgericht werken op groepsniveau. Overgenomen uit: “De effecten van het eerste jaar opbrengstgericht werken met Focus,” door L. Staman, A. J. Visscher en H. Luyten, 2012, paper gepresenteerd op de Onderwijs Research Dagen 2012, p. 3. Enschede: Universiteit Twente.

Effect van OGW op Leerprestaties

De enorme aandacht voor het gebruik van data voor onderwijsbeslissingen komt voort uit de No Child Left Behind Act uit de Verenigde Staten, beleid dat dateert uit 2001, en het Every Child Matters uit Engeland uit 2002. Dit waren initiatieven vanuit de overheid, die erop gericht waren scholen meer gebruik te laten maken van data bij het nemen van instructiebeslissingen. Enerzijds omdat het nemen van gefundeerde beslissingen zou leiden tot beter onderwijs, en anderzijds om meer inzicht te krijgen in de kwaliteit van scholen. Scholen moeten meer verantwoording afleggen over hun prestaties en over de onderwijsbeslissingen die zij nemen.

In de hieraan voorafgaande literatuurstudie is de samenhang tussen Opbrengstgericht werken en de prestaties van leerlingen in het basisonderwijs onderzocht. Omdat er slechts één correlatieel onderzoek werd gevonden op basis van de Nederlandse term, is verder gezocht naar vergelijkbaar onderzoek in het buitenland. Er is gezocht naar onderzoek naar effecten van het gebruik van data voor het aanpassen van instructie. Daartoe werd gezocht op de termen data-driven decision making en data-based teaching. Er werden twaalf studies gevonden, waarvan een aantal betrekking had op interventies die voortkwamen uit beleidsinitiatieven zoals in de vorige alinea genoemd. De studies gaan over onderzoek naar het effect van databenutting in het onderwijs. Goal setting is niet in de onderzoeken teruggevonden.

In de literatuurstudie is gezocht naar een effect van opbrengstgericht werken (data-driven teaching) op zowel leesprestaties van leerlingen als op rekenprestaties. De resultaten geven geen eenduidig antwoord op de vraag of opbrengstgericht werken ook daadwerkelijk hogere prestaties oplevert. Voor rekenprestaties is bij vijf van de onderzoeken sprake van een interventie, en in één onderzoek is er sprake van correlatieel onderzoek. Er werden positieve effecten gevonden in de experimentele studies van Carlson, Borman en Robinson (2011), Slavin, Cheung, Holmes, Madden en Chamberlain (2011) en Marsh, McCombs en Martorell (2010). Bij deze drie onderzoeken betreft de interventie training en/of coaching in databenutting, data-analyse en data-interpretatie. Slavin et al. vinden grotere effecten voor latere jaren in de interventie, waarin ook het onderwijs daadwerkelijk werd aangepast. Dit sloot aan bij de resultaten van Stecker (2005), die geen positief effect had gevonden, omdat scholen hun onderwijs niet aanpasten. Ook Jones en Krouse (1986) vonden geen positief effect, maar hier duurde de interventie slechts 8 weken. In een quasi-experimenteel, maar goed opgezet onderzoek van McCaffrey, Hamilton en Rand Education (2007) werd evenmin een positief

effect gevonden. Hier interpreteerden scholen de data niet juist, en pasten zij het onderwijs niet aan. De onderzoekers in dit experiment concluderen dat het data-analyse systeem vaak niet eens gebruikt werd. Data-driven teaching/ opbrengstgericht werken lijkt een positief effect te hebben op rekenprestaties, als niet alleen de data geanalyseerd worden, maar deze ook juist worden geïnterpreteerd, en het onderwijs wordt aangepast. Het aanleren van effectieve instructiemethoden vergroot het effect van een interventie. In het correlatieve onderzoek van de Inspectie (Inspectie van het onderwijs, 2010) was geen interventie, maar werd een positieve samenhang aangetoond tussen de mate waarin scholen opbrengstgericht werken en de prestaties van leerlingen.

Ook voor lezen zijn de resultaten wisselend. Er wordt niet eenduidig aangetoond dat OGW effect heeft op prestaties, omdat de verklaring voor gevonden effecten niet enkel zijn toe te schrijven aan de interventie. Net als voor rekenen, geldt voor lezen dat de experimentele studies, alsmede de quasi-experimentele studie van McCaffrey en Hamilton alleen positieve effecten rapporteren als er ten minste sprake is van het leren interpreteren van data tijdens de interventie. Carlson, Borman & Robinson (2011) en Slavin, Cheung, Holmes, Madden, & Chamberlain (2011), bij wie op het moment van post-test nog geen sprake is van het hebben aangeleerd van effectieve instructiemethoden, vonden voor sommige groepen wel en voor sommige groepen geen significante effecten. Ronduit positieve effecten zijn gevonden in de quasi-experimentele studies (Bursuck, Smith, Munk, Mehlig & Perry, 2004; Abbott, Atwater, Lee & Edwards, 2011; Lai, McNaughton, Amituanai-Tolosa, Turner & Hsiao, 2009; McNaughton, Lai & Hsiao, 2012; Sexton, 2010). Bij deze studies kan sprake zijn van bias, bijvoorbeeld ten gevolge van het Hawthorne Effect, waardoor het zuivere effect overschat kan worden. Gevonden effecten kunnen namelijk deels of geheel het gevolg zijn van de sociale gevolgen van het enkele feit dat er een interventie is. Alleen al de extra aandacht die een interventie met zich meebrengt kan zorgen voor meer inspanning of een sterkere focus van leerlingen, leerkrachten of schoolleiders. Dit betekent dat elke interventie, onafhankelijk van de inhoud, hetzelfde effect zou kunnen hebben veroorzaakt. Ook kan het niet randomiseren van het onderzoek selectie-bias veroorzaken: in de experimentele groep zitten betere leerlingen of scholen dan in de controle groep, bijvoorbeeld omdat die meer openstaan voor deelname aan een interventie. Bij het ontbreken van een controle groep komen daar nog meer mogelijke vormen van bias bij, bijvoorbeeld maturation (de natuurlijke ontwikkeling van de capaciteiten van de leerlingen), een externe gebeurtenis (zoals meer aandacht voor schoolprestaties in de media), of eenvoudigweg het verschil in moeilijkheid tussen een pre- en een post-test. Ten slotte kenmerkten de quasi-experimentele studies zich door interventies die sterk gericht waren op zowel databenutting en interpretatie, alsmede op het aanleren van effectieve instructiemethoden.

De literatuurstudie leverde geen overtuigend bewijs dat het gebruik van data bij het nemen van instructiebeslissingen altijd en overal tot betere leerlingprestaties leidt. Wel lijkt het erop, dat een intensieve interventie die leerkrachten ook professionaliseert in het interpreteren van data en het kiezen van goede instructiemethoden, tot betere prestaties kan leiden. Voor lezen zijn er meer positieve resultaten dan voor rekenen, hoewel dit te maken kan hebben met de opzet van de studies (quasi-experimenteel). Ook kan het zo zijn, dat de interventies voor lezen meer gericht waren op het aanleren van effectieve instructiemethoden. Het kan zelfs zo zijn, dat er voor lezen meer duidelijkheid is over wat effectieve instructiemethoden zijn dan voor rekenen. Misschien weten leerkrachten wel dat hun leerlingen het niet goed doen op rekenen, maar vinden ze het moeilijker om hier effectieve instructiebeslissingen te nemen.

Het Meten van OGW

Uiteindelijk is het de bedoeling dat het effect van de professionaliseringsinterventie Focus op onderwijskwaliteit en leerling prestaties onderzocht gaat worden, maar dat valt buiten het bereik van dit onderzoek. Wel werd onderzocht hoe opbrengstgericht werken gemeten kan worden. Hiervoor is uitgegaan van de componenten van OGW zoals weergegeven in figuur 1. Tevens is de maat die de onderwijsinspectie hanteert bij de jaarlijkse inspectie van basisscholen bekeken.

In het onderwijsverslag 2010-2011 (Inspectie van het Onderwijs, 2012) zijn aan de hand van vijf indicatoren in kaart gebracht in hoeverre scholen opbrengstgericht werken. Dit zijn:

- Gebruik samenhangend systeem voor volgen leerlingen
- Systematisch volgen en analyseren vorderingen
- Evalueren effecten van de zorg

- Evalueren resultaten
- Evalueren onderwijsleerproces

Deze vijf indicatoren gaven een nogal grove waardering van opbrengstgerichtheid.

Omdat OGW op wordt opgenomen in de jaarlijkse inspectie van scholen, is door de inspectie een uitgebreider waarderingskader opgesteld, waarin ook andere kwaliteitsaspecten van opbrengstgericht werken zijn opgenomen. Dit waarderingskader werd gemaakt voor een onderzoek naar de samenhang tussen opbrengstgericht werken en leerling prestaties wat betreft rekenen. Bij dit onderzoek waren 166 scholen betrokken, die werden gewaardeerd op deze aspecten, waarbij ze per item voldoende of onvoldoende konden scoren. Op basis van analyses achteraf zijn vier schalen gevormd: kwaliteitszorg, afstemming, didactisch handelen en zorg. Onder “kwaliteitszorg” valt de evaluatie van leerresultaten (bijvoorbeeld door het uitvoeren van analyses met een leerlingvolgsysteem), het planmatig werken aan verbeteractiviteiten en een opbrengstgerichte schoolcultuur. “Afstemming” bestaat uit items waaruit blijkt of scholen hun onderwijs afstemmen op leerresultaten. “Didactisch handelen” betreft het al dan uitvoeren van effectieve didactiek, zoals duidelijke uitleg, feedback en hoge verwachtingen. “Zorg” tenslotte heeft betrekking op overdracht van leerling-gegevens en evaluatie van zorgplannen. De inspectie heeft hiermee een erg breed waarderingskader voor opbrengstgericht werken opgesteld, hoewel het niet overal even gedetailleerd is. Zo is het onderwerp “kwaliteitszorg” nog niet erg specifiek. Het uitvoeren van analyses wordt bevraagd met slechts één item, namelijk: “De school maakt trendanalyses van haar rekenen-wiskunde resultaten”. Hoewel er naast de trendanalyse veel meer soorten analyses mogelijk zijn, worden die niet genoemd. Hetzelfde geldt voor het bespreken van analyses. Er is één item dat luidt: “Alle partijen (leraren, directeur, intern begeleider) zijn betrokken bij de evaluatie van de resultaten voor rekenen-wiskunde.” Maar er wordt niet beoordeeld hoe ze besproken worden, met wie, of hoe vaak.

Voor dit onderzoek werd een instrument ontwikkeld dat met name het uitvoeren en bespreken van analyses gedetailleerder meet. De eerste onderzoeksvraag die werd opgesteld, moest een eenvoudig, duidelijk maar gedetailleerd scoreformulier opleveren, waarmee de mate van opbrengstgericht werken per school kon worden bepaald. Deze maat zou dan vervolgens gebruikt worden voor het beantwoorden voor de volgende onderzoeksvragen, die alle drie betrekking hadden op de mate van opbrengstgerichtheid van de scholen.

Onderzoeksvragen

Voor dit onderzoek zijn de volgende onderzoeksvragen opgesteld.

1. Hoe ziet een beredeneerde maat voor Opbrengstgericht werken eruit?
2. Is er voor aanvang van de interventie Focus verschil in de mate van opbrengstgericht werken tussen experimentele scholen en controle scholen?
3. Wat zeggen de verzamelde gegevens bij experimentele scholen en controle scholen van Focus I over de relatie tussen de mate van opbrengstgericht werken van een school en hun leerlingprestaties?

Onderzoeksvraag 1 leidde tot een scoreformulier, dat vervolgens als instrument werd ingezet bij het beantwoorden van de onderzoeksvragen 2 en 3.

Methode

Procedure

Voor het beantwoorden van de onderzoeksvragen is gebruik gemaakt van gestructureerde interviews, vragenlijsten en rekentoetsen die zijn afgenomen aan het begin van de training Focus I. Op basis van de gestructureerde interviews is allereerst een scoreformulier ontwikkeld, zodat de mate van opbrengstgericht werken per school kon worden uitgedrukt in een totaalscore. Dit scoreformulier werd daarmee een instrument voor de andere drie onderzoeksvragen. Met dit formulier kon namelijk worden gemeten of er verschil was in de mate waarin scholen opbrengstgericht werken tussen experimentele scholen en controle scholen bij aanvang van de interventie Focus. Vervolgens is met behulp van dit scoreformulier de samenhang onderzocht tussen opbrengstgericht werken en de prestaties van leerlingen op rekentoetsen. De respondenten, instrumenten en uitgevoerde data-analyses

zijn kort samengevat in tabel 1, en zijn verder toegelicht in de paragrafen Respondenten, Instrumenten en Data-analyse.

Tabel 1

Respondenten, instrumenten en data-analyse per onderzoeksvraag

Vraag	Respondenten	N (con/exp.)	Instrumenten	Data-analyse
1	IB'ers	86 (43/43)	Gestructureerd interview	Maat opstellen op basis van de interviews, de door de inspecties gebruikte maat en literatuur over de onderdelen van opbrengstgericht werken.
2	IB'ers	84 (42/42)	Scoreformulier (opgesteld op basis van gestructureerd interview)	Onafhankelijke T-toets (gemiddelden vergelijken tussen e/c scholen met de opgestelde maat)
3	IB'ers	82 (40/42)	Scoreformulier (opgesteld op basis van gestructureerd interview)	Bivariate correlaties tussen leerlingprestaties en de totaalscores van scholen op de maat voor OGW (Pearson) en voor enkele aparte items (Spearman)
	Leerlingen (groep 1 t/m 7)	10.838 (4.751/6.087)	Rekentoetsen van Cito LOVS	

Respondenten

De respondenten voor dit onderzoek zijn afkomstig van 86 basisscholen, waarvan 43 scholen in de experimentele groep en 43 scholen in de controle groep van het Focus-project. De experimentele scholen zijn scholen die deelnemen aan het Focus-project, een training in opbrengstgericht werken. De scholen komen voornamelijk uit de regio Twente en Salland (provincie Overijssel). De controlescholen zijn verkregen door middel van Propensity Score Matching. Hiervoor is een logistische regressie met een binaire afhankelijke variabele gebruikt om de propensity scores te berekenen (Staman, Visscher en Luyten, 2012). Het doel van matching op basis van propensity scores is zoveel mogelijk gelijkheid tussen groepen te krijgen met betrekking tot de covariaten (Thoemmes & Kim, 2011). Voor het berekenen van de propensity scores is gebruik gemaakt van de volgende covariaten op schoolniveau: denominatie, SES, percentage autochtone leerlingen, toezichtsarrangement (zeer zwak, zwak of goed), geslacht (percentage vrouwen per school) en schoolgrootte (op basis van aantal fte). Scholen met propensity scores vergelijkbaar met de scores van experimentele scholen zijn telefonisch benaderd om als controleschool mee te werken aan het onderzoek.

Van deze 86 scholen namen leerkrachten, Interne Begeleiders (IB'ers) en leerlingen deel. De interviews zijn afgenomen onder de IB'ers van zowel de experimentele als de controle scholen (N=86). De toetsscores zijn verkregen bij de leerlingen van die scholen.

Interne begeleiders

Bij alle 86 scholen zijn gestructureerde interviews afgenomen bij één IB'er per school. Op basis van deze interviews zijn scores per school berekend voor de mate van opbrengstgericht werken. Er werden twee scholen uitgesloten bij de ontwikkeling van het scoreformulier en het verdere onderzoek, omdat deze scholen te sterk afweken van de andere scholen. Het betrof een school met voornamelijk asielzoekers, wat een vertekend beeld zou geven van de resultaten, en één school uit de controlegroep, omdat die ook teveel afweek van een doorsnee-school. Bovendien was er voor twee scholen uit de

controlegroep sprake van teveel missende waarden, waardoor er uiteindelijk 82 scholen overbleven waarvoor een totaalscore voor OGW werd berekend (42 experimentele/ 40 controle scholen).

Leerlingen

Van de 84 scholen werden van de groepen 1 tot en met 7 leerlingdata afgenomen in de vorm van een Cito-LOVS rekentoets. In totaal namen 10.838 leerlingen deel: 4751 in de controle groep en 6087 in de experimentele groep. De leerlingen waren afkomstig uit verschillende leerjaren. In tabel 2 is te zien van hoeveel leerlingen uit de verschillende leerjaren een toetsscore beschikbaar was.

Tabel 2

Aantal leerlingen per leerjaar waarvan data zijn afgenomen

Leerjaar	Aantal leerlingen		Totaal
	Controle groep	Experimentele groep	
1	662	897	1559
2	774	770	1544
3	718	1017	1735
4	692	1003	1695
5	727	997	1724
6	666	797	1463
7	512	606	1118
Totaal	4751	6087	10838

Het aantal leerlingen waarvan een toetsscore overbleef per leerjaar en per school liep sterk uiteen. Er is per leerjaar per school een gemiddelde Z-score berekend. Dit wordt verder uitgelegd bij de paragraaf Instrumenten. Een dergelijke gemiddelde Z-score per leerjaar per school werd alleen dan berekend, wanneer hiervoor van minimaal 5 leerlingen een toetsscore beschikbaar was. Hierdoor konden uiteindelijk kon voor 74 scholen een gemiddelde Z-score per leerjaar worden berekend. Een gemiddelde Z-score over alle leerjaren gezamenlijk per school werd verkregen door het gemiddelde uit te rekenen over alle gemiddelde Z-scores per leerjaar op een school, waarbij er minimaal voor 4 van de 7 groepen een score moest zijn. Hierdoor kon voor 68 scholen een gemiddelde Z-score per school worden berekend (33 controle scholen en 35 experimentele scholen). Geslacht en de leerlinggewichten (gebaseerd op het opleidingsniveau van de ouders) zijn niet meegenomen bij de analyses, omdat gegevens over het geslacht van de leerlingen bij 60 % ontbrak en gegevens over het leerlinggewicht bij 21 % van de leerlingen.

Instrumenten

De data voor dit onderzoek zijn verzameld met de volgende instrumenten:

- Een gestructureerd interview over de mate van OGW;
- Een scoreformulier voor de mate van OGW op basis van het interview;
- Gestandaardiseerde rekentoetsen;
- Een vragenlijst met betrekking tot prestatiegerichtheid.

De gebruikte instrumenten worden in de volgende paragrafen nader toegelicht.

Een gestructureerd interview over de mate van OGW

Hiermee zijn de verschillende onderdelen van Opbrengstgericht werken in kaart gebracht. Het interview bestaat zowel uit open, als uit gesloten vragen. Het eerste deel van het interview bestaat uit vragen omtrent het uitvoeren en bespreken van analyses van leerlingprestaties. Dergelijke analyses kunnen door scholen worden uitgevoerd met behulp van een leerlingvolgsysteem. De mate waarin scholen deze analyses uitvoeren is bevraagd met gesloten vragen. Er zijn zes verschillende analyses mogelijk:

- een dwarsdoorsnede van de school

- een trendanalyse van jaargroepen
- een trendanalyse van leerlingcohorten
- een analyse van de vaardigheidsgroei van leerlingen
- een groepsanalyse
- een analyse van de afzonderlijke categorieën uit een toets

Per analyse is voor elke jaargroep afzonderlijk gevraagd of de analyse nooit, incidenteel, jaarlijks of twee keer per jaar op de school van de betreffende IB'er wordt afgenomen. Op dezelfde wijze is gevraagd hoe vaak de uitgevoerde analyses besproken worden met de schoolleider, met de groepsleerkracht of in het schoolteam. Het tweede deel van het interview heeft een meer open karakter, doordat naast gesloten vragen ook wordt gevraagd naar voorbeelden. Dit tweede deel bevat onderwerpen als: het trekken van conclusies op basis van de analyses, het stellen van doelen, het nemen van maatregelen met betrekking tot onderwijsaanpassingen en de manier waarop de uitvoering van voorgenomen maatregelen bewaakt wordt. Zo is bijvoorbeeld gevraagd hoe doelen worden gesteld, wat voorbeelden van maatregelen zijn die worden genomen, en of die maatregelen betrekking hebben op alle leerlingen.

De interviews zijn afgenomen onder interne begeleiders van zowel de experimentele als de controle scholen door een onderzoeker van Focus. Vervolgens zijn de antwoorden op de vragen ingevoerd in SPSS. Alle vragen die gebruikt werden voor de analyses zijn numeriek gecodeerd. Zo werden de antwoorden voor het uitvoeren van analyses als volgt gecodeerd: 0=nooit, 1=1x per jaar, 2=2x per jaar en 3=3x per jaar. Voor het bespreken zijn de antwoorden als volgt gecodeerd: 1=2x per jaar, 2=1x per jaar, 3=incidenteel, 4=nooit. Later, bij het ontwerpen van een scoreformulier, zijn deze antwoorden gehercodeerd, zodat ze consistentere waren, en een opbouw van weinig naar vaak kregen. De meer open vragen zijn eveneens numeriek gecodeerd, zodat elk antwoord een aparte code kreeg. Bij de doelen zijn de antwoorden gesplitst in drie variabelen: doelen op schoolniveau, groepsniveau en leerlingniveau.

Een scoreformulier over de mate van OGW

Op basis van de ingevoerde interviewdata is in antwoord op onderzoeksvraag 1 een scoreformulier ontwikkeld. Met dit score formulier kan de mate van Opbrengstgerichtheid van een school in een numerieke waarde worden uitgedrukt. Hoe dit scoreformulier tot stand is gekomen wordt uitgebreid beschreven in de resultatensectie van dit verslag, aangezien dit tevens het antwoord op de eerste onderzoeksvraag gaf. Het uiteindelijke scoreformulier dat gebruikt is voor het beantwoorden van onderzoeksvraag 2, 3 en 4 bestaat uit een schaal van 20 items: het uitvoeren en bespreken van analyses ($\alpha=0,92$), en 4 losse items, waarvan 2 voor het stellen van doelen en 2 voor het aanpassen van het onderwijs.

Voor de eerste 20 items werden de scores uit het gestructureerde interviewbestand voor uitvoeren en bespreken van analyses gebruikt op basis van de volgende codering: 0=nooit, 1=incidenteel, 2=1x per jaar, 3=2x per jaar. Voor elke analyse is een gemiddelde score berekend door de onafgeronde scores van groep 1 tot en met groep 8 bij elkaar op te tellen en te delen door 8. Voor het bespreken gold, dat een score van 0 bij uitvoeren, direct zorgde voor een score van 0 bij bespreken. De onafgeronde scores voor de items 1.1 tot en met 1.20 zijn bij elkaar opgeteld en gedeeld door 20. Dit resulteerde dus op een onafgeronde, gemiddelde score per school voor de eerste component van OGW.

De overige 4 items werden gewaardeerd met 0, 1 of 2 punten, afhankelijk van de kwaliteit van het antwoord. In bijlage 2 is te zien, wanneer een school 0, 1 of 2 punten scoorde voor een antwoord. Deze beoordelingscriteria zijn opgesteld in overleg met de begeleiders van dit onderzoek, waarvan één van hen ook het interview had afgenomen. Een voorbeeld van een vraag is:

		0	1	2
2.1	Naar aanleiding van analyses worden prestatiedoelen op schoolniveau gesteld (2 punten wanneer deze schoolprestatiedoelen (mede) uitgedrukt worden in vaardigheidsscores)			

Gestandaardiseerde rekentoetsen.

Voor het waarderen van de rekenprestaties van de leerlingen van de scholen is gebruik gemaakt van de resultaten op de Cito-LOVS rekentoetsen, die op de betreffende scholen elk half jaar worden afgenomen. Voor dit onderzoek is gebruik gemaakt van de toetsgegevens van de laatste beschikbare afname vóór de start of aan het begin de interventie Focus; de afname data liggen tussen maart 2009 en april 2010. Er zijn leerlingdata van zowel experimentele als controle scholen, van de groepen 1 tot en met 7. Het betreft gestandaardiseerde toetsen van het CITO. De resultaten van de leerlingen op de toetsen zijn tijdens een schoolbezoek verzameld en geëxporteerd uit de verschillende leerlingvolgsystemen van de scholen. Deze leerlingdata per school zijn met behulp van SPSS samengevoegd tot één bestand. Voor het berekenen van de leerling prestaties is gebruik gemaakt van vaardigheidsscores. De vaardigheidsscores en de bijbehorende normen geven de beste indicaties voor het ontwikkelingsniveau van leerlingen (Staman, Visscher & Luyten, 2012). Oude toetsen (waarbij nog niet met vaardigheidsscores gewerkt wordt) zijn omgeschaald, zodat ze vergeleken konden worden met de scores op nieuwe toetsen.

Omdat vaardigheidsscores per leerjaar evolueren, en een gemiddelde vaardigheidsscore over meerdere leerjaren daarom geen betrouwbare indicator voor prestaties is, zijn deze scores daarna per leerjaar omgeschaald tot Z-scores. Zo kon een inschatting worden gemaakt van de leerling prestaties van een school ten opzichte van het gemiddelde van alle deelnemende scholen voor een bepaald leerjaar. Tenslotte is per school de gemiddelde Z-score per school berekend over de groepen 1 tot en met 7. Voor de analyses is zowel gebruik gemaakt van de gemiddelde Z-score per school per leerjaar, als van de totale gemiddelde Z-score per school. Deze Z-scores per school zijn gebruikt als inschatting van de leerling prestaties per school.

Data-analyse

De data zijn geanalyseerd met behulp van statistische toetsen uit SPSS. Met behulp van factor analyse en Cronbach's Alpha werd beoordeeld hoe de items voor het scoreformulier met elkaar samenhangen, en of het onderscheid in meerdere niveaus (school/groep/leerling) terug te vinden was in de data. Ook werd de consistentie van de subschalen "doelen" en "aanpassen onderwijs" beoordeeld, en werd onderzocht waarom items niet correleerden met andere items uit een subschaal. Uiteindelijk is besloten één grote schaal te vormen voor het uitvoeren en bespreken van analyses, en daarnaast vier losse items (twee voor doelen; twee voor aanpassen onderwijs).

Vervolgens zijn de correlaties berekend tussen de verschillende onderdelen van dit scoreformulier en leerlingprestaties. De leerlingprestaties bestaan uit een totale gemiddelde Z-score per school, en gemiddelde Z-scores per school per leerjaar. Voor correlaties met de schaal "Uitvoeren en Bespreken van analyses" werd Pearson's correlatiemaat gebruikt, aangezien de totaalscore een intervalwaarde was en deze ook normaal verdeeld bleek. Voor correlaties met de overige vier losse items werd Spearman's Rho gebruikt als correlatiemaat, aangezien dit waarden op ordinaal niveau betrof. Er is geen regressie-analyse uitgevoerd, omdat de correlatie tussen de mate van opbrengstgericht werken van de scholen en de Z-scores van de leerlingen te laag bleek en een regressie analyse dan niet zinvol is.

Resultaten

Ontwikkeling van een Scoreformulier

Bij het ontwikkelen van een scoreformulier bleken er een aantal obstakels. Allereerst konden niet alle onderdelen van opbrengstgericht werken helemaal gedekt worden vanuit het gestructureerde interview. Dit had te maken met het open karakter van veel vragen, waardoor het niet terecht was om een antwoord uit het interview aan een score te verbinden. Ook was het lastig te beoordelen welk antwoord beter was met betrekking tot opbrengstgerichtheid. Een voorbeeld hiervan is het bewaken of voorgenomen maatregelen ook worden uitgevoerd. Respondenten antwoordden op de vraag hoe dit bewaakt werd met antwoorden als: "1x per jaar door observatie door de schoolleider", of "2x per jaar door observatie door een Interne Begeleider met behulp van het observatie-instrument Kijkwijzer", of "door Intercollegiale Consultatie". Voor een aantal vragen is in overleg met de begeleiders, die tevens

projectleiders van Focus zijn, een opzet bedacht met een 0, 1 en 2 punten systeem. Het idee was, dat scholen telkens 0, 1 of 2 punten voor een item kregen, afhankelijk van hoe opbrengstgericht het antwoord op het item was. De punten voor alle items zouden bij elkaar opgeteld worden, zodat een totale score voor opbrengstgerichtheid werd verkregen. Het voordeel was dat dit een overzichtelijk, inzichtelijk en beknopt scoreformulier zou opleveren, zodat in de toekomst op een efficiënte manier opbrengstgerichtheid gemeten kon worden. De verwachting was dat de beschikbare data, zodra hiermee het scoreformulier zou zijn ingevuld, een samenhangende schaal zouden laten zien, met waarschijnlijk ook nog een onderscheid tussen de verschillende niveaus (school/groep en leerling). Uit de volgende alinea's blijkt, dat dit niet het geval was.

Onderdelen van een scoreformulier

Het te ontwikkelen scoreformulier zou alle drie componenten van opbrengstgericht werken, zoals beschreven in het conceptueel kader, volledig moeten omvatten. Helaas bleek dit niet mogelijk op basis van de beschikbare interviewdata. De eerste component, het in kaart brengen van de beginsituatie, kwam overeen met wat de Inspectie omschrijft als evaluatie van leerresultaten. De inspectie bekijkt of scholen trendanalyses maken van rekenresultaten, en of alle partijen betrokken zijn bij de evaluatie van de resultaten voor rekenen (twee items). Het Focus-interview vraagt gedetailleerder naar dit aspect, door per mogelijke analyse (zes in totaal), per leerjaar te vragen hoe vaak deze analyse wordt uitgevoerd, hoe vaak die wordt besproken en met wie (schoolleider, in het team, met de groepsleerkracht of anders). Het scoreformulier kan hiermee een veel gedetailleerder beeld geven van de mate waarin scholen analyses gebruiken. Er werden in totaal 20 items opgesteld over het uitvoeren en bespreken van analyses op basis van de interviewgegevens (bijlage 1). Helaas moest de eerste component van OGW zich hiertoe beperken. Het interpreteren van de analyses, wat ook een onderdeel is van het vaststellen van de beginsituatie, is wel bevraagd in het interview, maar deze informatie kon niet gebruikt worden voor het toekennen van een score. Zo werd er bijvoorbeeld gevraagd of er zorgsignalen worden afgeleid, maar op deze vraag antwoordden slechts drie scholen negatief, zodat de vraag niet erg onderscheidend is. Vervolgens werd om een voorbeeld gevraagd van een zorgsignaal dat was afgeleid naar aanleiding van de laatste toetsafname in juni. Als hier minder goede, of weinig zorgsignalen werden genoemd door een respondent, kon door de vraagstelling niet met zekerheid worden besloten dat de kwaliteit van het afleiden van zorgsignalen op die school dus onvoldoende was. Het ging immers slechts om een voorbeeld. Ook was in het interview gevraagd welk soort diagnoses/oorzaken voor tegenvallende prestaties in juni waren vastgesteld naar aanleiding van de LVS-toetsen voor rekenen. Hier werden verschillende oorzaken en diagnoses genoemd, maar dit betekende niet dat een andere (betere) diagnose zeker niet was gesteld. Bovendien kon niet worden beoordeeld of een diagnose van toepassing was; de informatie uit het interview over de diagnose stond los van de data die daaraan ten grondslag lagen, dus kon geen uitspraak gedaan worden over de geschiktheid van de genoemde diagnoses. Bovendien werden geen argumenten genoemd op basis waarvan een diagnose werd gesteld. Op deze manier was een groot deel van het interview niet bruikbaar voor het scoreformulier.

Dit had tot gevolg dat ook de tweede component, het stellen van doelen, niet volledig kon worden omvat met de interviewdata. Zo gaf het interview geen inzicht of de doelen hoog en uitdagend waren. Ook kon niet met zekerheid worden vastgesteld of de doelen tijdgebonden waren. Wel kon worden bepaald of er doelen op groepsniveau werden gesteld voor alle leerlingen en of dit soms, altijd of vaak gebeurde. Tevens kon worden vastgesteld of doelen op schoolniveau werden gesteld, en of deze doelen werden uitgedrukt in vaardigheidsscores. Het gebruik van vaardigheidsscores in de doelen werd door de projectleiders van Focus beoordeeld als een betere manier van doelen stellen, omdat vaardigheidsscores heel specifiek zijn en nauwkeuriger aansluiten bij het prestatieniveau van de leerling.

De laatste component, het aanpassen van het onderwijs, kon ook niet volledig gedekt worden met de interviewdata. Doordat voorbeelden werden genoemd van onderwijsaanpassingen, kon niet goed worden vastgesteld of scholen bepaalde aanpassingen wel of niet deden. Wel kon uit het interview worden afgeleid hoe vaak besluiten tot onderwijsaanpassingen betrekking hadden op alle leerlingen. Ook kon uit het interview worden afgeleid, of, op welke manier en hoe regelmatig scholen bewaken of voorgenomen maatregelen daadwerkelijk worden uitgevoerd.

Tot slot was in het gecodeerde interview bestand te lezen of de evaluatieve cyclus (plan-do-check-act) volledig werd doorlopen op een school.

Proefformulier

Na het selecteren van bruikbare vragen uit het interview werd vanuit de gecodeerde antwoorden uit het interviewbestand een proef-scoreformulier opgesteld. Dit formulier bestond uit 26 items (zie bijlage 1). Het score formulier werd zoveel mogelijk opgesteld conform de beschrijving van de componenten van OGW. Dit resulteerde in de hoofdonderdelen:

- het uitvoeren en bespreken van analyses (= het vaststellen van de beginsituatie)
- het stellen van doelen (=doelen stellen)
- het aanpassen van het onderwijs (=maatregelen)
- het volledig doorlopen van de OGW-cyclus (=de volledige cyclus inclusief monitoring en evaluatie)

Ook de verschillende niveaus waarop OGW betrekking kan hebben, werden aanvankelijk in het scoreformulier verwerkt. Het gaat dan om OGW op schoolniveau, op groepsniveau en op leerlingniveau (zie bijlage 1). Op basis van gesprekken met de begeleiders van het Focus-project , die inzicht hadden in het gebruik van de analyses door scholen, werd bepaald welke items bij welk niveau hoorden. Zo behoorde het bespreken van de dwarsdoorsnede met de schoolleider typisch bij OGW op schoolniveau, terwijl het bespreken van de categorieën-analyse met de leerkracht juist bij OGW op leerling niveau zou behoren. Na het coderen van de data zou moeten blijken, of factor analyse inderdaad dit onderscheid kon blootleggen.

Dit resulteerde in een formulier bestaande uit de volgende items: 20 items voor het vaststellen van de beginsituatie, waarvan 9 op schoolniveau, 9 op groepsniveau, en 2 op leerlingniveau; 2 items voor het stellen van doelen, waarvan 1 op schoolniveau en 1 op groepsniveau; 3 items voor het aanpassen van het onderwijs, waarbij geen onderscheid kon worden gemaakt tussen de verschillende niveaus en tot slot 1 item voor het volledig doorlopen van de cyclus. Eventueel zou bij dit laatste item nog het onderscheid tussen school- en groepsniveau kunnen worden aangebracht.

Na het opstellen van de 26 items werd een normering vastgesteld. Scholen kregen per item een score van 0, 1 of 2 punten, afhankelijk van of een antwoord hoorde bij een matig opbrengstgerichte school, een redelijk opbrengstgerichte school of een goede opbrengstgerichte school. Een voorbeeld van een item is:

HET UITVOEREN EN BESPREKEN VAN ANALYSES				
Schoolniveau				
Dwarsdoorsnede		0	1	2
1.1	Dwarsdoorsnede wordt jaarlijks uitgevoerd (2 punten als de analyse twee keer per jaar wordt uitgevoerd)			

Een school kreeg dus 2 punten als deze analyse 2 keer per jaar werd uitgevoerd, 1 punt wanneer die 1 keer per jaar werd uitgevoerd, en 0 punten wanneer deze minder dan 1x per jaar werd uitgevoerd. De scores werden berekend op basis van de data uit het interviewbestand. Hierin stonden voor het uitvoeren van de dwarsdoorsnede 4 mogelijke antwoorden, namelijk nooit, incidenteel, 1x per jaar of 2x per jaar. Dit werd dan per leerjaar beantwoord. Voor de score op het formulier werden “nooit” en “incidenteel” omgecodeerd tot score 0, “1x per jaar” werd score 1 en “2x per jaar” werd score 2. Daarna is het gemiddelde genomen van de groepen 1 tot en met 8, waarbij er maximaal 2 ontbrekende waarden mochten voorkomen per berekening. Dit gemiddelde is afgerond, zodat voor item 1.1 altijd een score 0, 1 of 2 werd verkregen, zelfs als er voor een groep (bijvoorbeeld een groep 1 of groep 8) geen antwoord was ingevuld. Op dezelfde manier is de score voor het uitvoeren van de andere vijf mogelijke analyses berekend.

Voor het bespreken van analyses was de werkwijze hetzelfde, maar met de volgende aanvulling. Als op uitvoeren van een analyse 0 punten werd gescoord, werd een code 98 of 99 (missing values) voor de score voor het bespreken van die analyse automatisch 0. Het is namelijk logisch dat als een analyse niet werd uitgevoerd op een school, deze ook niet werd besproken. Scholen vullen dan “niet van toepassing” in, of helemaal niets. Dit is op deze manier aangepast, omdat anders veel waarden zouden ontbreken, en de analyses beperkt zouden blijven.

Voor de 3 items behorend bij het stellen van doelen is dezelfde scoring toegepast: 0,1 of 2 punten. Maar omdat deze vragen een meer open karakter hadden, werd de waarde van een antwoord gewaardeerd met 0, 1 of 2 punten. Zo werd het stellen van doelen op schoolniveau gewaardeerd met 1 punt wanneer dit niet, en met 2 punten wanneer dit wel in vaardigheidsscores werd uitgedrukt.

Ook voor het aanpassen van het onderwijs werd op basis van persoonlijke, subjectieve inschatting bepaald welk antwoord 0, 1 of 2 punten verdiende, waarbij het van belang was of aanpassingen voor alle leerlingen golden, of de aanpassingen specifieke pedagogische/didactische maatregelen betrof, en of de IB'er of schoolleider regelmatig observeerde om de uitvoering te bewaken. In bijlage 1 staat de precieze formulering van de scoretoekenning.

Op het laatste item, het volledig doorlopen van de evaluatie OGW-cyclus, werd een iets afwijkende scoring toegepast: voor dit item werden 0 of 5 punten toegekend. Als scholen de cyclus volledig zouden doorlopen, en dus ook tussentijds het behalen van doelen zouden evalueren, dan zou dit meer gewicht moeten hebben dan bijvoorbeeld of één bepaalde analyse wel of niet wordt uitgevoerd. Daarom is gekozen voor 5 punten als scholen de evaluatieve cyclus volledig doorlopen.

Uitkomst data-analyse van het proefformulier

Om te beoordelen of het scoreformulier een goed, samenhangend formulier was, zijn alle controle en experimentele scholen gescoord. De schalen bleken op één item na op alle items te variëren. Alleen op het laatste item, het volledig doorlopen van de OGW-cyclus, bleek dat alle scholen 0 punten scoorden. Dit item is daarom niet meegenomen in de analyses.

Allereerst werd een totaalscore berekend voor een aantal voorbeeldscholen, zodat een inschatting gemaakt kon worden of het formulier in deze vorm goede, slechte en matige opbrengstgerichte scholen kon onderscheiden. De interviewer had op basis van haar inzicht bepaald wat 2 goede, 2 gemiddelde en 2 matig opbrengstgerichte scholen waren. De scoring liet dit ook zien: scholen die door de interviewer van Focus weinig opbrengstgericht werden geacht behaalden de laagste scores, middelmatige scholen behaalden de middenscores, en de scholen die waren opgevallen als erg opbrengstgericht kregen de hoogste scores.

Vervolgens werd een factoranalyse uitgevoerd, om te achterhalen of er meerdere factoren te onderscheiden waren in de variabele opbrengstgericht werken, en of de verschillende niveaus te herleiden waren uit de data. Dit bleek niet het geval. De factoranalyse gaf meerdere factoren, maar de items laadden niet duidelijk op slechts een enkele factor, waardoor geen duidelijk onderscheid tussen de factoren konden worden afgeleid. Vervolgens werd de interne consistentie van de gehele schaal beoordeeld met Cronbach's Alpha. Voor het complete formulier was de totale betrouwbaarheid groot ($\alpha=0,89$ op basis van 25 items), maar de samenhang tussen items onderling was matig tot slecht. Vooral de samenhang tussen de twee items over doelen hingen matig samen met het totaal (0,182 en 0,206) en de kwaliteit van de aanpassingen van het onderwijs hing zelfs negatief samen met het totaal (-0,155). Het verwijderen van items uit de tweede en derde component zou een hogere Alpha geven. Ook waren de totaalscores niet normaal verdeeld. Daarom is besloten het formulier niet als één schaal te gebruiken, maar als drie schalen op basis van de verschillende componenten van OGW. De interne consistentie van de eerste component was goed ($\alpha=0,91$ op basis van 20 items), die van de tweede component was redelijk (0,67 op basis van 2 items), maar de twee items samen werden niet als een valide schaal gezien voor het stellen van doelen, aangezien het zo'n beperkt deel van doelen stellen omvatte. De interne consistentie van de derde component was slecht (0,134 op basis van 3 items). De items van de derde component, het aanpassen van het onderwijs, vertoonden weinig samenhang, waarbij vooral het tweede item afweek. Dit item moest ofwel een afzonderlijk item vormen, of worden verwijderd van het formulier. Na bestudering van dit item is besloten het te verwijderen. Er was namelijk een indeling gemaakt van goede en slechte onderwijsaanpassingen, waarbij erop werd gelet of de aanpassingen specifiek waren, en of het inhoudelijke (pedagogische of didactische), dan wel oppervlakkige (materialen betreffend) of vaag geformuleerde aanpassingen betrof. Vervolgens werden

respectievelijk een score van 2 of 1 punten toegekend, en bij geen aanpassingen 0. Maar omdat in het interview gevraagd is naar voorbeelden van onderwijsaanpassingen, kon niet met zekerheid worden vastgesteld of de genoemde voorbeelden een betrouwbare maat voor de kwaliteit van de aanpassingen waren. Toen bleek dat dit item negatief samenhang met de rest, en vanwege het arbitraire karakter van dit item, is besloten het te verwijderen van het formulier. Ook na verwijdering van dit item bleek component 3 niet een betrouwbare schaal ($\alpha=0,37$ op basis van 2 items), en moesten de twee items van de derde component dus als afzonderlijke variabelen worden meegenomen bij verdere analyses. Voor de analyses die nodig zijn voor het beantwoorden van de onderzoeksvragen 2 en 3 bleven 24 items over: een schaal van 20 items voor component 1, 2 afzonderlijke items voor component 2 en 2 afzonderlijke items voor component 3. Deze versie van het formulier is als bijlage bij dit verslag gevoegd (bijlage 2).

Een volgende stap in de ontwikkeling van het scoreformulier was het bepalen of de totaalscores normaal verdeeld waren, om te kunnen vaststellen of een T-toets en regressie-analyse kon worden uitgevoerd. Dit bleek niet zo te zijn: de scores hadden een positieve scheefheid. Uit de normaliteitstest van Kolmogorov-Smirnov bleek dat de verdeling significant afweek van een normale verdeling, zowel voor de schaal van 20 items (0,129; $df=78$; $p=0,003$), als voor de twee items van doelen (0,381; $df=78$; 0,000) als ook voor het totaal van de 24 items (0,140; $df=78$; $p=0,001$).

Aanpassing proefformulier

Uiteraard kon nog op dat moment worden overgegaan op non-parametrische testen, maar zeker voor de eerste component was het vreemd dat de scores niet normaal verdeeld waren. De vraag rees, of deze scheefheid het gevolg zou kunnen zijn van het omschalen van de oorspronkelijk interviewschaal (nooit, incidenteel, 1x per jaar, 2x per jaar) tot de scores op het formulier (0,1,2). Daarom is besloten voor component 1 de scores opnieuw te bepalen, maar nu door gebruik te maken van de zuivere scores op de interview vragen. Hiervoor werd de volgende codering gebruikt: 0=nooit, 1=incidenteel, 2=1x per jaar, 3=2x per jaar. Voor elke analyse is een gemiddelde score berekend door de onafgeronde scores van groep 1 tot en met groep 8 bij elkaar op te tellen en te delen door 8. Voor het bespreken gold opnieuw dat een score van 0 bij uitvoeren, direct zorgde voor een score van 0 bij bespreken. De onafgeronde scores voor de items 1.1 tot en met 1.20 zijn bij elkaar opgeteld en gedeeld door 20. Dit resulteerde dus in een onafgeronde, gemiddelde score per school voor de eerste component. Factor analyse gaf wederom geen onderscheid tussen de verschillende niveaus: naast een grote, algemene factor die 40% van de variantie verklaarde (van de 20 items van de eerste component), werden nog vier kleinere factoren gevonden, maar ook nu kon geen zinvolle verklaring hiervoor worden gegeven. De interne consistentie van de schaal bleek hoog ($\alpha=0,92$ op basis van 20 items). Uit de normaliteitstest van Kolmogorov-Smirnov bleek, dat de verdeling van deze onafgeronde en meer zuivere scores niet significant afweek van een normale verdeling (0,089; $df=82$; $p=0,169$). De scheefheid was nog steeds aanwezig, maar veel lager (van ,571 naar ,173).

Omdat de verdeling van deze zuivere scores niet significant afweek van de normale verdeling, is besloten deze manier van scores te gebruiken voor de verdere analyses. Voor de overige componenten is gekozen voor correlatietechnieken en toetsen die geen normaliteit vereisen.

Verschil tussen Controle en Experimentele Groepen in Mate van OGW

Om te beoordelen of de controle en experimentele groepen significant van elkaar verschillen in de mate van opbrengstgericht werken, werd voor de schaal "Uitvoeren en Bespreken van analyses" een T-test uitgevoerd. Het gemiddelde van de controlegroep lag iets hoger dan dat van de experimentele groep voor wat betreft de gemiddelde score over de items 1.1 tot en met 1.20 (tabel 3). Dit verschil is niet significant ($t=0,422$; $df=80$; n.s.) bij tweezijdige toetsing.

Eveneens is gekeken of de groepen verschillen als een totaalscore over alle 24 items wordt vergeleken. Wederom was de gemiddelde score in de controle groep hoger dan die in de experimentele groep. Ook dit verschil is niet significant ($t= 0,792$; $df= 75$; n.s.) (tabel 3).

Om vervolgens te beoordelen of de scores op de afzonderlijke items van de overige twee componenten van OGW (doelen stellen en aanpassen onderwijs) verschilden tussen de groepen, werd gebruik gemaakt van non-parametrische toetsing. Hieruit bleek het volgende: De controlegroep scoort hoger op item 2.1 (doelen op schoolniveau). Dit verschil is significant ($Z= -4,576$; $p=0,000$). De controlegroep scoort eveneens hoger op item 2.2 (doelen op groepsniveau), maar dit verschil is niet

significant. De controlegroep scoort ook hoger op item 3.1 (besluiten tot aanpassen onderwijs). Dit verschil is significant ($Z = -4,162$; $p = 0,000$). Tenslotte scoort de controlegroep ook iets hoger op item 3.2 (het bewaken van de uitvoering). Dit verschil is echter niet significant.

Tabel 3

Verskil tussen de controle en de experimentele groep voor de schaal "Uitvoeren en bespreken van analyses" (items 1.1 tot en met 1.20) en voor de totaalscore over alle 24 items

Schaal	Controle groep			Exp. Groep			Verskil	Toets	Df	P
	N	M	SD	N	M	SD				
Items 1.1-1.20	40	1,19	0,69	42	1,12	0,81	0,07	T=0,422	80	n.s.
Item 1.1-3.2	37	27,84	14,25	40	25,00	16,97	2,84	T=0,792	75	n.s.

Noot. De kans (P) geldt voor tweezijdige toetsing. Bij de eerste T-toets werden de gemiddelde scores over de eerste 20 items vergeleken; bij de tweede T-toets werd het gemiddelde vergeleken tussen de totaalscores over alle 24 items.

Correlaties tussen Mate van OGW en Toetsscores

Voor het onderzoeken van de samenhang tussen de mate van OGW en leerlingprestaties is gebruikt gemaakt van Pearson voor wat betreft de 20-items schaal. Voor de losse items is gebruik gemaakt van Spearman's Rho. Voorafgaand is onderzocht of de toetsscores binnen de groepen kleiner zijn dan tussen de groepen, omdat anders verdere analyse zinloos zou zijn. Hiertoe is een ANOVA uitgevoerd.

Er zijn significante verschillen tussen de groepen in gemiddelde Z-scores ($F = 2,813$; $p = 0,00$). Er waren echter teveel groepen voor een post hoc analyse of voor controle van de homogeniteit van varianties. De verschillen binnen groepen waren in elk geval kleiner dan tussen de groepen. Om die reden is besloten dat het zinvol was om correlaties te berekenen.

Allereerst is met Pearson Correlation de samenhang berekend tussen de gemiddelde score op de schaal "Uitvoeren en Bespreken van analyses" en de gemiddelde Z-score van een school. Er bleek een lichte positieve samenhang ($r = 0,210$; $p = 0,46$) bij eenzijdige toetsing. Ook voor de Z-scores per leerjaar is de correlatie berekend. Tabel 4 geeft een overzicht van de correlaties tussen de schaal Uitvoeren en Bespreken en de Z-scores per leerjaar.

Tabel 4

Correlatie tussen de score van scholen op de schaal "Uitvoeren en Bespreken van analyses" en de gemiddelde Z-scores per school

Leerjaar	Correlatie	Significantieniveau	N
Gemiddelde over alle leerjaren*	0,21	$p < 0,05$	66
1	0,01	n.s.	62
2	0,26	$p < 0,05$	59
3	0,24	n.s.	65
4	0,15	n.s.	62
5	0,10	n.s.	64
6	0,13	n.s.	60
7	0,01	n.s.	49

*Een gemiddelde Z-score per school werd berekend als van minimaal vier leerjaren de Z-score bekend was

De correlatietabel laat zien dat er zowel voor het totaal als voor de groepen afzonderlijk een positief verband blijkt te zijn tussen OGW en scores. Dit verband is echter klein, en alleen voor leerjaar 2 en voor de totale gemiddelde Z-score per school significant.

Ook is de correlatie berekend tussen de totaalscore op alle 24 items en leerling prestaties van de scholen. Er bleek wederom een kleiner, maar significante positieve correlatie ($r=0,228$; $p=0,037$) voor wat betreft de gemiddelde Z-score per school. Ook voor de gemiddelde Z-scores per school per leerjaar werd voor alle leerjaren een kleine positieve samenhang gemeten. De correlatie was alleen significant voor groep 2 en groep 3 (tabel 5).

Tabel 5

Pearson's correlatie tussen de OGW-score van scholen en de gemiddelde Z-scores per school, op basis van de totaalscore voor alle 24 items

Leerjaar	Correlatie	Significantieniveau	N
Gemiddelde over alle leerjaren*	0,23	$p<0,05$	62
1	0,064	n.s.	58
2	0,274	$p<0,05$	56
3	0,235	$p<0,05$	61
4	0,150	n.s.	58
5	0,086	n.s.	60
6	0,097	n.s.	57
7	0,069	n.s.	46

*Een gemiddelde Z-score per school werd berekend als van minimaal vier leerjaren de Z-score bekend was

Tot slot zijn nog correlaties berekend voor de aparte items. Omdat dit scores betreft op ordinaal niveau, is gebruik gemaakt van Spearman's Rho. Omdat er geen zekerheid was over de kwaliteit van de items is tweezijdig getoetst. Uit de tabel 6 is af te lezen dat er geen duidelijke samenhang bestaat tussen de scores van scholen op de items en de Z-scores van de rekentoetsen.

Tabel 6

Spearman's Rho: de correlatie tussen items 2.1 t/m 3.2 en Z-scores totaal en per leerjaar

Item		Alle leerjaren	1	2	3	4	5	6	7
2.1	R	-,018	-,142	-,180	,033	,124	,099	,296*	-,230
	Sig	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	,019	n.s.
	N	68	63	59	67	64	66	62	51
2.2	R	-,012	,010	-,107	-,063	-,001	,026	,134	-,158
	Sig	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
	N	68	63	59	67	64	66	62	51
3.1	R	,129	,000	-,091	,132	,346**	,032	,189	-,111
	Sig.	n.s.	n.s.	n.s.	n.s.	,006	n.s.	n.s.	n.s.
	N	66	61	57	65	62	64	61	50
3.2	R	,087	,007	,056	,098	-,057	,127	,180	-,070
	Sig.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
	N	66	61	58	65	62	64	60	49

* $p< .05$

** $p< .01$

Voor regressie analyse moet er sprake zijn van een sterk lineair verband ($r > 0,80$) (Baarda, De Goede & Van Dijkum, 2003). Aangezien hiervan geen sprake is, wordt geen regressie-analyse uitgevoerd.

Conclusie en Discussie

Aan het begin van dit verslag werden de volgende onderzoeksvragen geformuleerd:

1. Hoe ziet een beredeneerde maat voor Opbrengstgericht werken eruit?
2. Is er voor aanvang van de interventie Focus verschil in de mate van opbrengstgericht werken tussen experimentele en controle scholen?
3. Wat zeggen de verzamelde gegevens bij experimentele en controle scholen van Focus I over de relatie tussen de mate van opbrengstgericht werken van een school en hun leerlingprestaties?

De resultaten leiden tot de volgende conclusies met betrekking tot deze onderzoeksvragen:

Ad 1. Een simpele maat voor OGW, zoals van tevoren bedoeld was te ontwikkelen, is op basis van de data uit de gestructureerde interviews niet realiseerbaar. Het interview leent zich niet voor het volledig vatten van de mate van OGW in een totaalscore per school. Wel kan op basis van de informatie over het uitvoeren en bespreken van analyses een veel gedetailleerdere indruk worden verkregen dan op de huidige maat van de inspectie omtrent dit aspect van OGW. De vier losse items die zijn gevormd over doelen stellen het aanpassen van het onderwijs, meten slechts een beperkt deel van deze componenten van OGW.

Ad 2. Met het formulier zoals bijgevoegd als bijlage 2 is een aantal analyses uitgevoerd. Ten eerste is onderzocht of er verschil is in opbrengstgerichtheid tussen controle en experimentele scholen. Hoewel de maat dus te wensen overlaat, en slechts een matig valide beeld geeft van opbrengstgericht werken, zijn de afzonderlijke delen gebruikt om te onderzoeken of er verschillen zijn. De gemiddelde scores van de controle groep liggen telkens iets hoger dan die van de experimentele groep. Over de eerste schaal van 20 items (uitvoeren en bespreken van analyses) verschillen de controle en experimentele groep licht, maar dit verschil is niet significant. Ditzelfde geldt voor het totaal van 24 items. Voor wat betreft de losse items is er alleen een significant verschil gevonden voor de items 2.1 (het stellen van doelen op schoolniveau) en 3.1 (hoe vaak maatregelen tot het aanpassen van onderwijs betrekking hebben op alle leerlingen).

Ad 3. Ook is met de afzonderlijke delen van het scoreformulier onderzocht of er een samenhang is tussen (de delen van) OGW en leerling prestaties. Er lijkt een positieve samenhang te zijn tussen de totaalscore van zowel de 20 als de 24 items en de leerling prestaties. Deze correlaties zijn echter slechts in beperkte mate significant. Er is geen regressie analyse gedaan, omdat de samenhang daar niet groot genoeg voor was. De grootste significante samenhang die gevonden werd is 0,346. Dit was de correlatie tussen item 3.1 en de Z-score van leerjaar 6. Voor de andere leerjaren werd met betrekking tot dit item geen significante samenhang gevonden, en voor de groepen 2 en 7 zelfs een negatieve samenhang. De correlatie tussen Z-scores en alle 24 items was iets hoger dan tussen Z-scores en de eerste 20 items, maar dit kan verklaard worden door het grotere aantal items.

Op basis van de antwoorden is een viertal conclusies mogelijk:

- het formulier meet de mate van opbrengstgerichtheid niet valide/betrouwbaar.
- de Z-scores geven een vertekend beeld van de prestatie van een school.
- opbrengstgericht werken hangt niet positief samen met prestaties.
- een derde variabele beïnvloedt de uitkomst van de analyses.

Voor elke conclusie zijn argumenten. Het formulier laat wel degelijk te wensen over. Het open karakter van de vragen met betrekking tot doelen en het aanpassen van het onderwijs veroorzaken, dat het toekennen van een score bemoeilijkt wordt. Voor wat betreft de doelen zou duidelijk beoordeeld moeten worden of het om haalbare, uitdagende doelen gaat. Uit dit onderzoek is niet overtuigend gebleken dat het uitdrukken van doelen in vaardigheidsscores tot hogere prestaties leidt. Ook wat betreft de onderwijsaanpassingen is het lastig een valide en betrouwbare meetmethode te ontwikkelen. Daarbij is niet transparant, wanneer een onderwijsaanpassing goed is. Een combinatie van de maat die de Inspectie voor het rekenen/wiskunde onderzoek gebruikte en de eerste 20 items van dit scoreformulier zou wellicht een meer valide meetinstrument zijn.

Helaas gaven de toetsscores nogal wat problemen. Er waren niet voor elk leerjaar scores voor veel leerlingen. Soms hadden scholen alleen toetsscores voor groep 1 en 2, of geen score voor een

bepaald leerjaar. Bovendien waren de afnamedata van de toetsen niet overal gelijk, en waren er soms leerlingen die meerdere toetsscores hadden (bijvoorbeeld een score op E6 en E7), of was een toets twee keer afgenomen. Dubbele scores zijn zoveel mogelijk handmatig verwijderd, maar aangezien het een bestand van ruim 10.000 leerlingen betrof, zullen hier ongetwijfeld scores gemist zijn. Ook het omschalen van oude toetsen naar nieuwe scores gaf wat problemen, waardoor 350 leerling scores uit het bestand verwijderd moesten worden. Ook waren er soms minder dan 10 scores per leerjaar, wat misschien veroorzaakt wordt doordat niet alle leerlingen in de klas de toets hadden gemaakt. Het is dus de vraag hoe betrouwbaar de scores weergeven wat het prestatieniveau van een school is.

Een derde mogelijke conclusie is, dat opbrengstgericht werken niet samenhangt met hogere leerling prestaties. Mogelijk worden de prestaties van leerlingen met name bepaald door de kwaliteit van de lessen van de leerkracht en wordt die nog slechts beperkt beïnvloed door OGW zoals dat in dit onderzoek is gemeten.

Tenslotte is het heel goed mogelijk dat een derde variabele de uitkomst van de analyses heeft beïnvloed. Zo is er geen rekening gehouden met de gewichten van leerlingen, omdat deze niet voor alle scholen bekend waren. Ook is geen rekening gehouden met het geslacht van leerlingen, de SES of de intelligentie. Er waren ook geen gegevens bekend over de achtergrond van leerkrachten: hun leeftijd, hun intelligentie, hun opleidingsniveau, hun ervaring. Nu is er door het toepassen van matching de gelijkheid tussen de groepen bevorderd op een aantal van deze factoren, maar niet op allemaal.

Noot. Bijlagen kunnen worden opgevraagd (m.abbes@student.utwente.nl).

Literatuurlijst

- Abbes, M. (2012). *Literatuurstudie naar de relatie tussen opbrengstgericht werken en leerling-prestaties* (bachelor thesis). Enschede: Universiteit Twente.
- Abbott, M., Atwater, J., Lee, Y., & Edwards, L. (2011). A Data-Driven Preschool PD Model for Literacy and Oral Language Instruction. *NHSA Dialog*, 14(4), 229-245.
- Baarda, D. B., De Goede, M. P. M. & Van Dijkum, C. J. (2003). Basisboek statistiek met SPSS. Groningen/Houten: Wolters-Noordhoff bv.
- Bursuck, W. D., Smith, T., Munk, D., Damer, M., Mehlig, L., & Perry, J. (2004). Evaluating the impact of a prevention-based model of reading on children who are at risk. *Remedial and Special Education*, 25(5), 303-313. doi: 10.1177/07419325040250050401.
- Carlson, D., Borman, G. D., & Robinson, M. (2011). A Multistate District-Level Cluster Randomized Trial of the Impact of Data-Driven Reform on Reading and Mathematics Achievement. *Educational Evaluation and Policy Analysis*, 33(3), 378-398. doi:10.3102/0162373711412765
- 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 (2012). *De Staat van het Onderwijs*. Onderwijsverslag 2010/2011. Utrecht: Inspectie van het Onderwijs.
- Jones, E. D., & Krouse, J. P. (1986, april). The Effectiveness of data-based instruction by student teachers in classrooms for students with mild learning handicaps. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. <http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED281836>
- Locke, E. A. & Latham, G. (2002). Building a practically useful theory of goal setting and task motivation. *The American psychologist*, 57(9), 705-17.
- Lai, M. K., McNaughton, S., Amituanai-Tolosa, M., Turner, R., & Hsiao, S. (2009). Sustained Acceleration of Achievement in Reading Comprehension: The New Zealand Experience. *Reading Research Quarterly*, 44(1), 30-56.

- Marsh, J. A., McCombs, J. S., & Martorell, F. (2010). How Instructional Coaches Support Data-Driven Decision Making Policy Implementation and Effects in Florida Middle Schools. *Educational Policy*, 24(6), 872-907. doi: 10.1177/0895904809341467
- McCaffrey, D. F., Hamilton, L. S., & Rand Education, S. M. C. A. (2007). *Value-Added Assessment in Practice: Lessons from the Pennsylvania Value-Added Assessment System Pilot Project*. [Technical Report. TR-506-CC]. Santa Monica (CA): RAND Corporation.
- 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(2), 203-228. <http://dx.doi.org/10.1080/09243453.2011.652126>
- Meijer, J., Ledoux, G., Elshof, D.P. (2011). Gebruikersvriendelijke leerlingvolgsystemen in het primair onderwijs. Amsterdam: Kohnstamm Instituut.
- OCW (2011). *Opbrengsten, maak er werk van!* Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap.
- Sexton, J. (2010). *Leveling the Playing Field: Increasing Student Achievement through Data-Driven Ability Grouping and Instructional Practices* [Online Submission].
- Slavin, R. E., Cheung, A., Holmes, G., Madden, N. A., & Chamberlain, A. (2011). *Effects of a data-driven district reform model*. <http://www.cddre.org/images/Effects%20of%20a%20Data%20Driven%20District%20Reform%20Model%20January%202011.pdf>
- Staman, L., Visscher, A. J. & Luyten, H. (2012). De effecten van het eerste jaar opbrengstgericht werken met Focus. Paper gepresenteerd tijdens de Onderwijs Research Dagen 2012. Enschede: Universiteit Twente.
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, 42(8), 795-819. doi: 10.1002/pits.20113.
- Thoemmes, F., & Kim, E. S. (2011). A Systematic Review of Propensity Score Methods in the Social Sciences. *Multivariate Behavioral Research*, 46, 90-118.
- Visscher, A. J. & Ehren, M. (2011). De eenvoud en complexiteit van opbrengstgericht werken. Enschede: Universiteit Twente. <http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2011/07/13/de-eenvoud-en-complexiteit-van-opbrengstgericht-werken.html>
- Visscher, A., Peters, M. & Staman, L. (2010). *Het Focus-project: opbrengstgericht werken op basis van prestatie-feedback*. Enschede: Universiteit Twente.

Literatuurstudie naar de Relatie tussen Opbrengstgericht Werken en Leerlingprestaties

Marieke Abbes
Universiteit Twente Enschede
6 augustus 2012

Samenvatting

Als scholen Opbrengstgericht werken, dan wordt hiervan een positief effect verwacht op de prestaties van leerlingen. In deze literatuurstudie is onderzocht, hoe overtuigend voorgaand onderzoek dit heeft aangetoond. Er werden 12 studies gevonden die de relatie onderzoeken tussen

Opbrengstgericht werken en taal- en/of rekenprestaties van leerlingen in het basisonderwijs. Omdat het grotendeels buitenlands onderzoek betrof, is gebruik gemaakt van de zoektermen data-based teaching en data-driven decision making.

Op grond van de resultaten lijkt er een positieve samenhang te bestaan tussen opbrengstgericht werken en leerlingprestaties, maar overtuigend bewijs voor een positief effect is niet gevonden. In dit verslag is een overzicht gegeven van de conclusies uit de 12 gevonden studies. Vervolgens is beschreven hoe de kwaliteit van de onderzoeksopzetten en de betreffende interventies van invloed kunnen zijn geweest op deze conclusies.

1^e begeleider: Dr. A.J. Visscher

2^e begeleider: L. Staman MSc

“Onderzoek toont aan dat leerresultaten verbeteren als scholen daadwerkelijk opbrengstgericht gaan werken” (Inspectie van het onderwijs, 2010).

De Onderwijsinspectie trekt deze conclusie na een recent onderzoek waarbij gekeken is naar de samenhang tussen de opbrengstgerichtheid van een school en de leerlingprestaties van dezelfde school. Het geloof in het effect van opbrengstgericht werken op leerprestaties is de reden voor de onderwijsinspectie om in het toezicht op scholen te betrekken in welke mate scholen opbrengstgericht werken. Deze maatregel sluit aan op het beleid van het ministerie van Onderwijs, Cultuur en Wetenschappen dat erop gericht is basisvaardigheden in het primair onderwijs te verbeteren. Al in de Kwaliteitsagenda Primair Onderwijs van 2007 (OCW) is opbrengstgericht werken (OGW) een speerpunt. OGW wordt door de inspectie omschreven als het systematisch en doelgericht werken aan het maximaliseren van de prestaties van leerlingen (OCW, 2011). Met systematisch wordt dan bedoeld dat de OGW cyclus, die bestaat uit meten, doelen stellen en maatregelen nemen, structureel wordt ingezet in het onderwijsproces (op vaste meetmomenten, bijvoorbeeld elk half jaar).

Dat er van opbrengstgericht werken grote resultaten worden verwacht, blijkt uit het bestuursakkoord van 17 januari 2012 (Ministerie van Onderwijs, Cultuur en Wetenschap en PO-raad, 2012). Daarin worden ambitieuze doelen gesteld. Ten eerste dat in 2015 het aantal zwakke scholen maximaal 35 is (waar dat in 2010 nog 57 was) en ten tweede dat in 2015 het aantal taalzwakke en rekenzwakke scholen ten opzichte van 2010 gehalveerd is (van 238 naar maximaal 119). Het doel voor 2015 is dat 60 % van de basisscholen opbrengstgericht werkt (waar dat 30 % in 2010 was), dat alle scholen meetbare doelen formuleren (70% in 2010). Ook worden verhoogde prestaties van scholen verlangt op voor opbrengstgerichte vroegschoolse educatie alsmede het hebben van een uitdagend aanbod voor cognitief talentvolle leerlingen.

Opbrengstgericht werken kan worden uitgesplitst in drie componenten: het in kaart brengen van de beginsituatie, het stellen van haalbare, realistische doelen en het kiezen van de juiste strategieën om die doelen te bereiken (Visscher & Ehren, 2011). Het in kaart brengen van de beginsituatie kan verbeterd worden door een beter gebruik van het Leerlingvolgsysteem (LVS). Het bepalen van doelen zou in lijn met de theorie van Goal Setting (Locke & Latham, 2002) kunnen worden gedaan: uitdagende SMART doelen stellen. Bij de derde component, het kiezen van de juiste strategie, is het van belang dat leraren niet alleen kennis hebben van verschillende instructiemethoden en mogelijke onderwijsmaatregelen, maar ook dat zij strategieën kiezen die aangetoond effectief zijn (Slavin, Cheung, Holmes, Madden & Chamberlain, 2011). In de literatuur wordt dit veelal aangeduid met de term evidence-based.

Ook in andere landen is er een sterke tendens om te investeren in het beter benutten van informatie voor het nemen van instructiebeslissingen. Het No Child Left Behind beleid in de VS is erop gericht prestaties van zwakke scholen te verbeteren, en daarom zijn vele maatregelen getroffen en projecten opgezet. Deze projecten, zoals professionaliseringstrainingen of coachingstrajecten zijn bedoeld om scholen en leerkrachten te leren hun onderwijsactiviteiten beter af te stemmen op meetgegevens, en instructiemethoden te kiezen die aansluiten bij actuele inzichten over effectiviteit. Er is een sterk geloof dat data—in het bijzonder toetsresultaten van leerlingen—belangrijke informatiebronnen zijn die het nemen van beslissingen zouden moeten ondersteunen (Marsh, 2010).

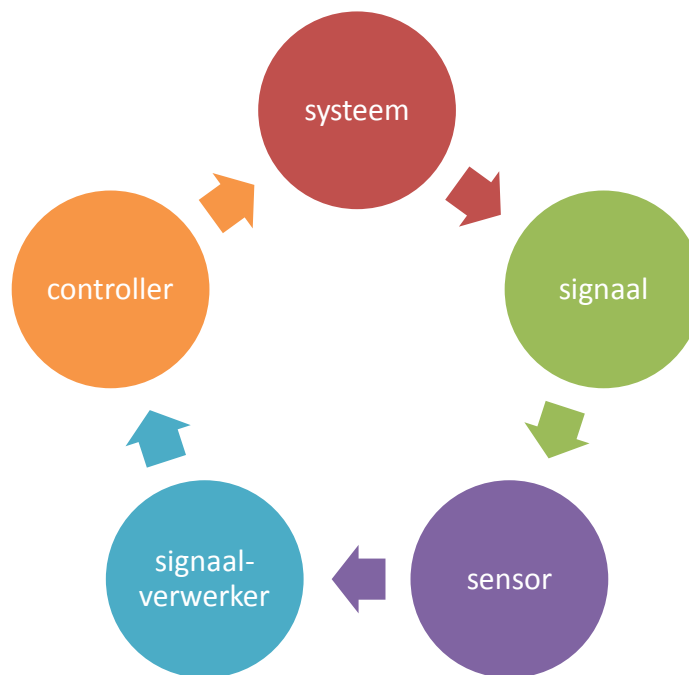
De vraag rijst, of beter gebruik van data en het afstemmen van onderwijsactiviteiten op die data, ook daadwerkelijk leidt tot betere leerlingprestaties. De theoretische basis, prestatiefeedback in combinatie met goal setting, is sterk. Maar betekent dit ook dat deze trend van dataproductie en -consumptie, met alle interventies en projecten die daartoe zijn ondernomen, zorgen voor hogere opbrengsten? Om antwoord te vinden op deze vraag, is een literatuurstudie uitgevoerd naar de relatie tussen opbrengstgericht werken en leerlingprestaties. In de volgende alinea wordt eerst ingegaan op het theoretische principe waarop opbrengstgericht werken is gebaseerd: feedback. Feedback is informatie over iemands prestaties of begrip (Hattie & Timperley, 2007). De kern van opbrengstgericht werken is het benutten van informatie over prestaties van leerlingen voor het nemen van de juiste instructiebeslissingen. Daarom wordt de vorm van feedback die bij opbrengstgericht werken wordt ingezet ook wel *prestatiefeedback* genoemd.

Conceptueel Kader

Feedback

De oorsprong van het concept feedback

Wat is feedback? De term feedback vindt zijn oorsprong in de cybernetica (de wetenschap van de besturing van biologische en mechanische systemen). De term werd later ook populair binnen de onderwijskunde. Halverson (2010) constateert dat het onderwijsproces sindsdien mede bestudeerd werd vanuit het systeemdenken. Feedback is informatie die wordt teruggekoppeld, om processen binnen een systeem te reguleren. In de meest eenvoudige vorm bestaat feedback uit vier onderdelen: signalen, sensoren, signaalverwerkers en controllers (Richardson, 1991; von Bertalanffy, 1969). Signalen zijn dan de verstuurde informatie, die wordt opgevangen door sensoren. De signaalverwerkers stellen de betekenis van het signaal vast (interpretatie en waardering) en controllers beslissen vervolgens welke actie er moet worden ondernomen (Halverson). In figuur 1 wordt dit proces grafisch weergegeven.



Figuur 1. De feedback cyclus in de meest eenvoudige vorm

Formatieve feedback in het onderwijs

Bij opbrengstgericht werken kan deze cyclus ook gebruikt worden om de werking van prestatie-feedback uit te leggen. Het proces is dan het onderwijsleerproces. In feite functioneert het systeem optimaal wanneer het maximale leerprestaties voor elke leerling oplevert. Feedback verschaft informatie over waar het niet zo goed gaat, en daar wordt dan bijgesteld met als doel dat het daarna beter gaat. Men noemt dit ook wel formatieve feedback. Het systeemdenken voor het onderwijs is een manier van denken die kan bijdragen aan het vormen van een lerende organisatie (Senge, 1992). Feedback staat daarbij centraal, met daarop volgend gezamenlijke reflectie op de resultaten. Dit wordt gezien als een manier om los te komen van bestaande misvattingen of tekortkomingen (CPS, 2011).

In de literatuur over feedback en opbrengstgericht werken wordt veel gesproken over de cyclus Plan-do-check-act, een evaluatie-cyclus waar de OGW cyclus van is afgeleid. Voor dit literatuuronderzoek wordt gebruik gemaakt van de eenvoudige feedback cyclus van figuur 1, omdat daar erg mooi de verschillende onderdelen van het proces duidelijk worden. En het beeld van een systeemproces helpt eraan te herinneren hoe belangrijk het is dat alle onderdelen goed werken.

Feedback is een van de meest invloedrijke factoren op leerprestaties (Hattie, 2009). Formatieve feedback kan betrekking hebben op de leerling, die vervolgens leeractiviteiten moet

aanpassen, of op de leraar, die onderwijsactiviteiten moet aanpassen. Effectieve feedback geeft antwoord op drie vragen: “Where am I going? How am I going? Where to next?” (Hattie en Timperley, 2007; Chappuis & Stiggins, 2002). Of, anders gezegd, wat is het leerdoel? Waar staat de leerling ten opzichte van dit leerdoel? Wat is de beste stap om de leerling richting het doel te krijgen? In de ideale situatie (volgens Hattie en Timperley) zoeken zowel de leerling als de docent het antwoord op deze drie vragen.

De invloed van feedback is echter afhankelijk van de kwaliteit ervan (Hattie en Timperley, 2007; Shute, 2008). Je moet van feedback vooral kunnen leren hoe je het beter kunt doen. Er is veel onderzoek gedaan naar feedback van *leraren aan leerlingen*. Deze vorm van feedback moet aan een aantal voorwaarden voldoen, wil ze effect hebben op leerprestaties. Zo moet effectieve feedback informatie bevatten over de taak, of over bruikbare leerstrategieën. Ook moet feedback gerelateerd zijn aan leerdoelen. Feedback moet duidelijk, doelgericht en betekenisvol zijn. Daarnaast moet het aansluiten bij voorkennis, en logische verbanden zichtbaar maken voor de leerling. Dat de feedback moet aansluiten bij voorkennis, betekent ook dat feedback niet zinvol is wanneer de leerstof echt niet begrepen wordt. Dan is aanvullende instructie zinvoller dan feedback. Ook zorgt effectieve feedback ervoor dat leerlingen actief informatie gaan verwerken. Feedback moet daarom goed te begrijpen zijn voor de leerling, en geen bedreiging vormen voor iemands persoon. Feedback werkt vooral wanneer er een cultuur is van het willen leren van fouten.

Zoals genoemd kan feedback niet alleen betrekking hebben op de leerling, maar ook op de leraar. Hattie (2009) concludeert na jaren studie naar feedback:

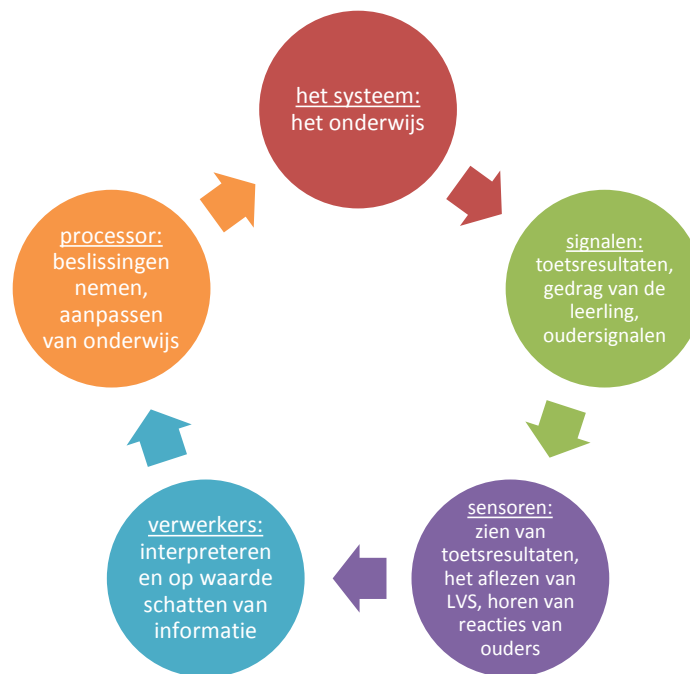
*“The mistake I was making was seeing feedback as something **teachers provided to students**... feedback was most powerful when it is from the **student to the teacher**.”* (p. 173)

Hij bedoelt daarmee dat leraren open moeten staan voor feedback van leerlingen, wat zij wel en niet begrijpen, waar zij fouten maken, wanneer ze verkeerde concepties hebben, of wanneer de leerstof niet overkomt. Dan zullen lesgeven en leren daadwerkelijk op elkaar aansluiten, en zal feedback een krachtig instrument zijn in het onderwijsleerproces. Deze vorm van feedback, informatie die de leerling verschaft over de stand van zaken in zijn leren en presteren, vormt de kern van opbrengstgericht werken.

OGW en daarmee ook deze literatuurstudie richt zich primair op de activiteiten van de leerkracht of van de school, en niet op de beslissingen die de leerling zelf neemt op basis van feedback. Het gaat hier dus om het aanpassen van het onderwijs op basis van feedback. Dit kan op leerkrachtniveau, door het aanpassen van onderwijsactiviteiten, maar feedback kan ook op schoolniveau werken, waarbij informatie gebruikt wordt voor beleidsbeslissingen. Op leerkrachtniveau of op schoolniveau is alleen het aanwezig zijn van informatie nog geen effectieve feedback.

De verschillende onderdelen van de feedback cyclus in het onderwijs

Welke schakels in de feedback cyclus van invloed zijn, is het best weer te geven met het eenvoudige model van figuur 1, maar nu toegespitst op het onderwijs (figuur 2).



Figuur 2. De feedback cyclus toegespitst op het onderwijssysteem

Hoe de verschillende onderdelen van de feedback cyclus zoals weergegeven in figuur 2 werken binnen het onderderwijssysteem, wordt in de volgende alinea's uitgelegd.

Het systeem: het onderwijs

Het lastige van het onderwijssysteem is dat het een complex systeem is. Je weet eigenlijk nooit wanneer het proces optimaal is. Wanneer worden maximale leerprestaties behaald? Het systeem heeft geen standaardreactie op een genomen beslissing. Het is niet altijd van tevoren te voorspellen wat het effect is van een onderwijsaanpassing. Daarom is het belangrijk dat het een continue proces is van doen en meten, een evaluatieve cyclus. De Inspectie verwoordt dit met de term systematisch.

Een andere moeilijkheid is dat er niet altijd overeenstemming is over het doel van het systeem. Waar werkt het naartoe? Er is geen eenduidige definitie voor goed onderwijs, maar over het algemeen wordt de kwaliteit van een school bepaald op basis van de opbrengsten: wat heeft de school toegevoegd aan het startniveau van de leerling. Daarbij ligt het accent vaak op cognitieve vaardigheden zoals taal en rekenen, maar ook sociaal-emotionele opbrengsten horen bij de opbrengsten van een onderwijssysteem en zelfs de tevredenheid van leerlingen, ouders, en het vervolgonderwijs (Oberon, 2008).

Tenslotte is een onderwijssysteem geen machine: het bestaat niet uit te programmeren onderdelen, maar uit mensen met een eigen wil, die eigen keuzes maken en waarbij gedrag niet altijd te voorspellen is.

Het signaal: de toetsresultaten, gedrag van de leerling, signalen van ouders

Informatie wordt zichtbaar door signalen die leerlingen afgeven in de vorm van toetsresultaten, waarneembaar gedrag of gesprekken met ouders. Het is echter onmogelijk om alle signalen van een leerling te meten, aangezien er ook processen zijn die niet waarneembaar zijn zoals de ontwikkeling van de hersenen. Het spreekt voor zich dat het belangrijk is dat toetsen valide zijn, en dat wat ze meten goed moet aansluiten bij de doelen van het onderwijs.

De sensor: de opvanger van het signaal, de ogen/oren van de leerkracht, leerling, of schoolleider, al dan niet via een Leerlingvolgsysteem.

Een klas vol leerlingen en slechts één leerkracht, met twee ogen en twee oren. Dat niet alle signalen worden opgevangen zal niemand verbazen. Uit onderzoek blijkt dat leerkrachten vaak moeite hebben om data te verwerken, te analyseren en te interpreteren (Earl & Katz, 2006). Tegenwoordig maken

scholen steeds meer gebruik van een leerlingvolgsysteem, wat resulteert in nog meer signalen om te verwerken. Scholen verzuipen soms in de grote hoeveelheid informatie.

De signaalverwerker: de leerkracht, de leerling of schoolleider die het signaal moet interpreteren en op waarde moet schatten

De invoering van het leerlingvolgsysteem heeft er wel voor gezorgd dat de informatie overzichtelijk wordt gemaakt. Daarmee wordt de informatie uit toetsen of observaties overzichtelijk gepresenteerd, vaak ook nog beeldend. Een dergelijk systeem biedt ook mogelijkheden om het op waarde schatten van toetsresultaten te ondersteunen. Zo is het mogelijk resultaten te vergelijken met andere leerlingen, met voorgaande jaren, met andere scholen, en kan de vaardigheidsgroei per leerling worden gemeten. De interpretatie van die gegevens is nog een volgende stap. Het interpreteren van informatie brengt onzekerheid met zich mee (Bennett, 2011). Hoe interpreteert een leerkracht een fout antwoord op een vraag? Is het een slordigheidsfoutje, het niet begrepen hebben van de stof, het missen van benodigde vaardigheden, kan een leerling zich niet concentreren? Formatieve feedback is het gebruiken van informatie om onderwijs- en leerprocessen te verbeteren (Halverson, 2010). Maar hoe dit uiteindelijk gebeurt, is afhankelijk van de interpretatie van data door een persoon (of meerdere personen), en de consequenties die men daaraan verbindt. Om te interpreteren moeten de scores gecombineerd worden met de overige informatie over de leerling (gedrag, vriendjes, thuissituatie), maar ook met gegevens over geboden onderwijs, en alle aanpassingen binnen het onderwijs (tijd, instructiemethoden, pedagogische technieken, materialen, begeleiding, groeperingsvorm, e.a.) die de prestaties zouden kunnen verhogen. Het is belangrijk dat scholen kennis hebben van wat werkt. Helaas is er binnen de onderwijskunde geen eensluidend oordeel over wat werkt. Over de te nemen beslissingen (welke onderwijsactiviteiten het meest geschikt zijn om het leerdoel te bereiken) zullen de meningen dan ook verschillen.

De controller: de leerkracht, de leerling of de schoolleider neemt een beslissing om al dan niet leer- of onderwijsactiviteiten aan te passen.

De laatste stap in de cyclus is het aanpassen van onderwijsactiviteiten (door de leerkracht), of van maatregelen op schoolniveau (door de schoolleider, al dan niet samen met het team). In de praktijk blijkt dat leerkrachten en scholen deze laatste stap lang niet altijd maken. Van het meten en interpreteren van informatie alleen wordt het onderwijs niet beter. Waar het uiteindelijk om gaat is het bijstellen van onderwijsactiviteiten, zoals het verbeteren van de instructie of het verhogen van de effectieve leertijd (Ledoux, Blok & Bogaard, 2009). Een belangrijk aspect van deze laatste stap is het overtuigd stellen van ambitieuze, uitdagende doelen. Uitdagende doelen zorgen ervoor dat feedback wordt gezocht en gebruikt (Hattie, 2009).

Als alle schakels van de cyclus goed werken, pas dan werkt de feedback optimaal, en zouden de leerprestaties het hoogst moeten zijn. Maar hoe weet je of de cyclus goed werkt? Onderzoek naar de kwaliteit van de afzonderlijke schakels is al moeilijk, laat staan onderzoek naar het geheel.

Data-Driven Decision Making

Het onderzoek binnen de onderwijskunde naar formatieve feedback richtte zich in eerste instantie op het klasniveau. Black en William (1998) hebben uitgebreid onderzoek gedaan naar de rol van feedback, waarbij ze op het niveau van de leerkracht en de leerling keken. Later werd feedback op schoolniveau ook onderzocht. Door de recente vraag vanuit de overheid en de maatschappij om meer transparant beleid, en om verantwoording voor resultaten en processen, heeft het onderwerp de laatste jaren meer aandacht gekregen. Dientengevolge zijn er tal van professionaliserings-interventies van de grond gekomen om scholen te leren hun beslissingen te baseren op meetbare informatie. In de Verenigde Staten werd in 2001 de No Child Left Behind act aangenomen, en in Engeland in 2002 het soortgelijke initiatief Every Child Matters. Dit was het begin van een reeks interventies gericht op de verschillende stappen uit de feedbackcyclus. De achtergrondgedachte van deze onderwijservormingen is dat wanneer scholen over de benodigde gegevens beschikken, opleiders dan in de positie zouden zijn om sterke en zwakke aspecten van het onderwijs te diagnosticeren. Het aanpassen van het onderwijs op basis van deze informatie zou leiden tot beter onderwijs (Earl & Katz, 2006). Het wijd verspreiden van leerlingvolgsystemen op scholen zal hier ongetwijfeld mee samenhangen. Veel trainingen die volgen op deze beleidsvoornemens richtten zich vooral op het betrouwbaar en valide

meten van leerlingprestaties, op het kiezen van onderwijsactiviteiten die berusten op recent onderzoek, en op het verantwoorden van beslissingen. In de literatuur zijn artikelen hierover te vinden door de zoektermen data-driven (of data-based) decision making (of teaching) te gebruiken. Marsh beschrijft data-driven decision making (DDDM) als volgt:

“DDDM in education refers to teachers, principals, and administrators systematically collecting and analyzing various types of data, including input, process, outcome, and satisfaction data, to guide a range of decisions to help improve the success of students and schools.” (Marsh, Pane & Hamilton, 2006, pag1.)

Slavin, Cheung, Holmes, Madden en Chamberlain (2011) spreken over *Data Driven Reform*. De essentie van deze op data gerichte onderwijsvernieuwing is het leren verzamelen van actuele informatie over leerlingprestaties, het vervolgens analyseren van de oorzaken van cijfers, en het op basis hiervan nemen van die onderwijsmaatregelen waarvan een positief effect wordt verwacht op leerlingprestaties.

In Nederland wordt de term meetgestuurd onderwijs gebruikt, maar ook de termen Opbrengstgericht Onderwijs of *Opbrengstgericht Werken*. Deze laatste term wordt door de Inspectie van het Onderwijs gehanteerd, en zal ook gebruikt worden bij het formuleren van de onderzoeksvraag voor deze literatuurstudie.

Onderzoeksvraag

Naar aanleiding van de enorme aandacht voor meetgestuurd onderwijs of opbrengstgericht werken, en de veronderstelde positieve effecten ervan op de leerprestaties, is de volgende onderzoeksvraag opgesteld voor deze literatuurstudie:

Hoe overtuigend is aangetoond dat opbrengstgericht werken een positief effect heeft op de prestaties van basisschoolleerlingen?

Methode

Het onderwerp opbrengstgericht werken sluit nauw aan bij meetgestuurd onderwijs, data-driven/data-based decision making, formative assessment en assessment for learning. Ook de concepten formative feedback en prestatiefeedback beslaan hetzelfde onderwerp. Om niet te verdrinken in de enorme hoeveelheid literatuur, is gekozen om in het literatuuronderzoek alleen artikelen mee te nemen die de termen data-driven en data-based gebruiken (in onderwijskundig onderzoek), en in de Nederlandse literatuur de term Opbrengstgericht. Hier is voor gekozen, omdat deze artikelen altijd gaan over feedback die gebruikt wordt voor het aanpassen van het onderwijs, en bijvoorbeeld niet over feedback die gebruikt wordt voor het aanpassen van leeractiviteiten door leerlingen. Bovendien zijn dit de meest recente termen voor dit onderwerp. Er werd aangenomen dat in actuele artikelen die met deze zoektermen werden gevonden, de belangrijkste conclusies van eerdere onderzoeken waren verwerkt. Vervolgens zijn deze zoektermen gecombineerd (door het gebruik van AND) met de term “achievement”. Aanvankelijk werd ook de term “outcomes” toegevoegd, maar dit gaf veel ruis. De resultaten zijn vervolgens gefilterd, waarbij gezocht is naar artikelen die uitspraken doen over een causaal of correlatieel verband tussen OGW of meetgestuurd onderwijs en prestaties van leerlingen.

Voor het zoeken is gebruik gemaakt van verschillende databases: Web of Science, Eric, Scopus en Picarta. In tabel 1 is te zien welke zoektermen en filters gebruikt zijn per database, en hoeveel artikelen dit heeft opgeleverd. De artikelen zijn beoordeeld op bruikbaarheid aan de hand van de volgende criteria:

- Alleen artikelen waarin onderzoek wordt beschreven naar de relatie tussen opbrengstgericht werken en leerlingprestaties (en niet om meningen van leerkrachten, de relatie met dropout rates of beschrijvingen van het implementatieproces). De voorkeur gaat uit naar experimenteel onderzoek, maar ook correlatieel onderzoek wordt meegenomen.
- De steekproef/onderzoeksgroep is niet heel specifiek (bijv. kinderen met ADHD of blinde kinderen).

Tabel 1

Aantal gevonden artikelen per gebruikte database, zoektermen en filters

Database	Zoektermen	Aantal artikelen
Web of Science	(Achievement OR outcomes) AND (Data-based OR data-driven)	1975, veel ruis door outcomes
	Zonder outcomes	151
	Filter: education special, education educational research, education scientific disciplines	62
	Abstracts gelezen en beoordeeld aan de hand van criteria	5
	Volledige artikelen gelezen en beoordeeld aan de hand van criteria	3 (Bursuck, Carlson, Marsh)
Eric	(Achievement OR outcomes) AND (Data-based OR data-driven)	578
	Filter: Alleen research reports	136
	Abstracts gelezen en beoordeeld aan de hand van criteria	16
	Artikelen gelezen	4 (Abbott, Jones, McCaffrey, Sexton)
Picarta	Opbrengstgericht OF meetgestuurd	41
	Titels gescreend, bij twijfel artikelen opgezocht en beoordeeld aan de hand van criteria	1 (Inspectie van het onderwijs)
Scopus	(Achievement) AND (Data-based OR data-driven)	60
	Filter: social sciences	
	Titels, abstracts en zo nodig hele artikel gelezen en beoordeeld aan de hand van criteria	1 (Stecker)

Op basis van de zoektermen, het filter en de beoordeling aan de hand van de criteria zijn negen artikelen geselecteerd voor deze literatuurstudie. Daarnaast zijn nog drie effectstudies ontdekt tijdens het doorlezen van artikelen over OGW en via een tip van de begeleider van dit onderzoek. De in totaal twaalf artikelen zijn in de referentielijst aangegeven met een *.

Resultaten

Inleiding

De gevonden studies hebben betrekking op lees- en/of rekenprestaties op basisscholen. De resultaten van de onderzoeken naar het effect van OGW op rekenen en lezen verschillen. Over het algemeen geven quasi-experimentele studies positievere resultaten dan experimentele studies. 6 van de 12 studies doen uitspraken over rekenprestaties, 11 van de 12 studies (tevens) over leesprestaties. Van alle effectstudies wordt over het algemeen de meeste waarde gehecht aan gerandomiseerde experimentele studies. Een controlegroep met voor- en natesten zorgen ervoor dat andere verklaringen voor gevonden verbanden ontdekt kunnen worden, en het randomiseren is de beste manier om bias

(bijvoorbeeld door selectie) te voorkomen (Slavin, 2008). Bij een zuiver experiment, met gerandomiseerde groepen, een pre-test en een post-test is vergelijkbaarheid tussen controle groepen en experimentele groepen het grootst, en kunnen eventuele effecten toegeschreven worden aan de interventie. Dit Aan de uitkomst van deze experimentele studies wordt daarom de meeste waarde gehecht. Vijf van de gevonden artikelen hebben betrekking op experimentele studies. Een overzicht van alle gevonden studies en hun kenmerken is gegeven in tabel 2. De opzet van deze tabel is aangepast overgenomen van Slavin, Lake en CDDRE (2007).

Tabel 2

Studies opgenomen in literatuurstudie en hun kenmerken

Artikel	Opzet	Duur	N	Groep	Kenmerken steekproef	Bewijs van gelijkheid groepen	Soort interventie	Gebruikte (voor-/) natest	Resultaat onderzoek
Experimentele studies									
Carlson, Borman en Robinson (2011)	Experiment (rekenen en lezen)	1 jaar	500 scholen, 59 districten	Grade 3-8	Scholen met lage scores (uitgekozen door district leiders)	Random assignment van districten, covariates per staat	CDDRE, invoeren benchmark tests en consulting	State administred achievement test	Pos+sig voor rekenen, pos maar niet sig voor lezen
Marsh, McCombs en Martorell (2010)	Experiment (lezen, rekenen)	1 jaar	86 scholen, ruim 70.000 leerlingen	Grade 6-8	Florida, zowel grote als kleine districten	random assignment per district; gecorrigeerd voor school en leerling kenmerken	Florida's reading coach program	FCAT (Florida's Comprehensive Assessment Test)	Pos +sig maar klein effect, van coaches die assessment data reviewen
Slavin, Cheung, Holmes, Madden, & Chamberlain (2011)	Beide: Experiment en quasi-experiment (triangulatie) (lezen, rekenen)	Experiment 1 jaar; quasi-experiment 4 jaar.	59 districten	Grade 5 en grade 8	Zie Carlson	Random assignment van districten; demografische kenmerken en eerdere prestaties onderzocht	CDDRE, invoeren benchmark tests en consulting	Standaard schoolniveau scores	Positieve effecten voor lezen en rekenen, maar vooral significante effect vanaf het derde jaar
Jones & Krouse (1986)	Experiment (lezen, rekenen)	8 weken	23 (13/10) leraren	Grade 3-6	Kinderen met mild learning handicaps	Leraren zijn random verdeeld over groepen	Minstens 1x per week supervisor, evalueren instructie, data gebruik. In 1 ^e wk 3x 2 uur seminar. (ook DI)	-Passage reading test van Fuchs -arithmetic computation test - reading comprehension	Pos+sig voor lezen, geen sign. effect voor rekenen.
Stecker (2005)	Review van 3 experimenten uit jaren 70-90	Minstens 7 weken (lezen, rekenen)	Verschillende, maar vooral kleine steekproeven	Kindergarten – grade 8	Kinderen met licht tot matige leerbepervingen	Alleen experimenten	CBM 2x per week data collectie, gebruik data voor aanpassen instructie	Verschillende, voor lezen vooral woordjes oplezen binnen tijd	Alleen effect als ook instructie wordt aangepast

Artikel	Opzet	Duur	N	Groep	Kenmerken steekproef	Bewijs van gelijkheid groepen	Soort interventie	Gebruikte (voor-/) natest	Resultaat onderzoek
Hoogstaande Quasi-experimentele studies									
McCaffrey & Hamilton (2007).	Quasi-experiment	1 jaar (lezen + rekenen)	Cohort 1: 62 districten; cohort 2: 32 districten	5 en 8	Pennsylvania	controlegroep, districten gematcht op basis van o.a. test scores op schoolniveau, gebruik van propensity scores	PVAAS, districten voorzien van informatie van VAA systeem	PSSA test scores	Geen effect (rekenen en lezen)
Correlationeel onderzoek									
Inspectie van het onderwijs	Quasi-experiment (rekenen)	N.v.t.	166 scholen	Trend groep 8 (6-4) en groep 6 (4)	Landelijke representatieve steekproef	n.v.t.	Geen	n.v.t.	Betere prestaties rekenen als scholen opbrengstgerichter zijn
Quasi-experimentele studies									
Bursuck, Smith, Munk, Damer, Mehlig & Perry (2004)	Quasi experiment	2 jaar (lezen)	3 scholen, totaal 93 II exp./ 85 controle	Kindergarten, grade 1, grade 2	High poverty, Minority diverse, veel II do not reach standard scores reading	Controleschool toegevoegd eind jaar 2, gekozen op basis van testgegevens begin project, demografie en vergelijkbare prestaties	Project PRIDE multitiered model (3 tiers), professionele ontwikkeling, DDDM	DIBELS	Gemiddelden significant hoger voor lezen; minder at risk kinderen
Abbott, Atwater, Lee & Edwards (2011)	Quasi-experiment	2 jaar (lezen)	27 leerkrachten, 9 klassen	Preschool, k1-k3	90% op of onder armoedegrens	Niet. Geen controlegroep	ERF project. 2 tiers. Professionele ontwikkeling en coaching. 104 uur per jaar per leerkracht.	TOPEL natest	Verhoging scores in lezen (iedereen haalt na 2 jaar gemiddelde standaard score)
Lai, McNaughton, Amituanai-Tolao, Turner, & Hsiao (2009)	Quasi-experiment	3 jaar (lezen)	7 scholen	3-8	New Zealand	Niet. Geen controlegroep. Wel drie meetpunten.	Learning Schools Model: 2 delen, effectieve instructie en datagebruik	PAT (progressive ach. test), STAR (Supplementary Tests of Achievement in Reading)	1 jaar meer voortgang dan verwacht (in totaal over 3 jaar)

Artikel	Opzet	Duur	N	Groep	Kenmerken steekproef	Bewijs van gelijkheid groepen	Soort interventie	Gebruikte (voor-/) natest	Resultaat onderzoek
Quasi-experimentele studies (vervolg)									
McNaughton, Lai en Hsiao (2012)	Quasi-experiment	3 jaar (lezen)	Verschillende cohorten, 671 leerlingen	4-8	Like-groep=lage scores reading comprehension assessment, unlike groep= minder lage niveaus	Niet. Geen controlegroep	Zie boven	STAR reading comprehension, repeated measures om acceleratie te laten zien	Significant verschil met verwachte scores (3 tot 4 maanden winst per jaar)
Sexton, 2010	Quasi-experiment	10 weken (lezen)	1 school, 10 leerlingen	Grade 5	Lage scores benchmark tests	Niet. Geen controlegroep. Ook onrealistische setting.	Ability grouping, gedetailleerde item analyse en afstemmen instructie op vaardigheid en voorkeuren leerlingen	Benchmark test, niet verder omschreven	Positief effect leesprestaties

Noot. De studies zijn ingedeeld in drie groepen: Experimentele studies, hoogstaande quasi-experimentele studies en (overige) quasi-experimentele studies. Een onderzoek werd ingedeeld bij experimentele studies als er sprake was van een controle en een experimentele groep, een pre-test en een post-test, en wanneer de respondenten random aan groepen waren toegewezen. Een onderzoek is ingedeeld bij hoogstaande quasi-experimentele studies wanneer er sprake was van een pretest en een posttest, een controle en een experimentele groep, en als daarnaast de respondenten aan de groepen werden toegewezen op basis van propensity score matching. Onderzoeken die niet aan deze voorwaarden voldeden, werden ingedeeld bij de quasi-experimentele studies.

De belangrijkste resultaten en conclusies uit deze literatuurstudie worden in de volgende paragrafen gerapporteerd: eerst voor rekenen, en vervolgens voor lezen.

Resultaten voor Rekenen

De eerste experimentele studie naar het verband tussen Data-Driven Decision Making (DDDM) en rekenprestaties is dat van Carlson, Borman en Robinson (2011). Zij hebben een experimenteel onderzoek opgezet, waaraan meer dan 500 scholen uit 59 districten meededen. In dit grootschalige onderzoek werd het effect gemeten van een 1-jarige interventie data-driven reform. Dit reform initiatief is geïmplementeerd door het John Hopkins Center for Data-Driven Reform in Education (CDDRE). Het eerste jaar van deze 3 jaar durende interventie bestond uit het invoeren van benchmark tests (standaard tests) na elk kwartaal en intensieve consultatie gericht op het leren omgaan en interpreteren van data. In jaar 2 en 3 richtte het project zich ook nog op het zoeken naar en toepassen van evidence-based instructiemethoden, maar dit onderzoek betrof alleen het eerste jaar van de interventie. De uitkomst is positief en significant voor rekenen. De resultaten zijn gecorrigeerd voor district en school effecten., doordat er gebruik is gemaakt van een multilevel-analyse. Het verschil tussen de beide groepen is 0,06 SD op leerling-niveau.

Ook Marsh, McCombs en Martorell (2010) beschrijven een experimentele studie naar het effect van DDDM op leerling prestaties. Het onderzoek betrof 86 scholen uit 8 districten. De interventie bestond uit het inzetten van een leescoach, iets wat op grote schaal gebeurt sinds de invoering van een leesstimuleringsinitiatief door de staat, genaamd: Just Read, Florida!. Het leescoach-programma volgt geen specifiek model, maar stimuleert en leidt op verschillende manieren de inzet van leescoaches op scholen. Het algemene doel is leerprestaties van leerlingen verhogen door het coachen van leerkrachten in het inzetten van effectieve, op onderzoek gebaseerde instructie. Voor het testen is gebruik gemaakt van scores op een standaard test: de Florida's Comprehensive Assessment Test Sunshine State Standards (FCAT-SSS). Hoewel de coaches oorspronkelijk werden ingezet om leesprestaties te verbeteren, werd in dit onderzoek ook onderzocht of het coachen effect had op rekenprestaties. Het rekendeel van deze test bevatte namelijk tekstrijke onderdelen. De onderzoekers vonden een significant positief effect voor rekenen, hoewel het effect klein was, en ook kleiner dan dat voor lezen. De conclusie is gebaseerd op regressie analyses, waarbij gecontroleerd wordt voor leerlingkenmerken en schoolkenmerken, en eerdere prestaties van leerlingen. Er zijn met behulp van een kwalitatief onderdeel van het onderzoek verschillende coachactiviteiten onderscheiden. Het effect is gevonden voor dat onderdeel van het coachen waarbij de coach assessment data analyseerde en besprak. Een verhoging van 1 SD van de hoeveelheid tijd die een coach aan dit onderdeel besteedt, zou een verhoging van 0,05 SD voor leerlingprestaties opleveren.

Slavin, Cheung, Holmes, Madden & Chamberlain (2011) onderzochten het effect van dezelfde interventie als die van Carlson et al. (59 districten). Omdat de controledistricten een jaar later dan de experimentele districten begonnen met de invoering dan de experimentele districten, werden de data van het eerste jaar gebruikt voor het experimentele deel van dit onderzoek. Wel zijn de data van de overige 3 jaren gebruikt voor een aanvullend deel van het onderzoek, een quasi-experiment waarbij de onderzoekers groepen vormen op basis van matching. Ze geven een inschatting van het effect in latere jaren van het project. Er werd in het experimentele deel voor groep 5 geen significant effect gevonden in het eerste jaar, maar wel in groep 8 (0.17). Met behulp van de informatie uit de quasi-experimentele deel concluderen zij dat vooral effect gevonden wordt na 3 of 4 jaar interventie, hoewel de effectgroottes nogal variëren (van -0.15 tot 0.49). Tijdens latere jaren van de interventie werd vooral ook getraind in het toepassen van effectieve instructiemethoden.

Jones en Krouse (1986) vonden geen significant effect voor rekenen in hun onderzoek. De experimentele groep werd hier getraind in het identificeren van belangrijke leerproblemen op grond van data, het ontwerpen en implementeren van effectieve instructie-interventies, en het maken van valide beoordelingen van prestatieverhoging over een kortdurende periode. De interventie bestond uit drie 2 uur durende seminars in de eerste week van het onderzoek, gevolgd door 7 weken waarin minimaal 1x per week een supervisor de leerkracht bezocht. De looptijd van dit onderzoek was erg kort, namelijk slechts 8 weken. Het uitblijvende effect zou hiermee verband kunnen houden. Slavin (2008) adviseert om onderzoeken die korter duren dan 12 weken niet mee te nemen in een literatuurstudie.

Ook Stecker (2005) vindt geen effect van DDDM op rekenprestaties. In deze studie worden experimentele studies van minimaal 7 weken gereviewed. Het gaat om interventies vergelijkbaar met DDDM, onder de naam Curriculum Based Measurements (CBM). De experimenten zijn uitgevoerd tussen 1970 en 1990. CBM is een term die voor 1990 gehanteerd werd, en waarmee databenutting van prestatiemetingen bedoeld werd, ten behoeve van het afstemmen van het onderwijs. Het had dan voornamelijk betrekking op het speciaal onderwijs, voor de afstemming op individuele leerlingen. Stecker vindt drie onderzoeken naar effecten van CBM op leerling prestaties. De conclusie van dit onderzoek luidt: alleen het verschaffen van informatie over prestaties is niet voldoende. DDDM heeft pas effect als ook het onderwijs wordt aangepast op basis van de informatie. Stecker noemt daarbij het belang van het stellen van ambitieuze, realistische doelen. Dit sluit aan bij de componenten van opbrengstgericht werken, zoals gesteld door Visscher en Ehren (2011).

Naast deze vijf artikelen waarin het experimentele studies betreft, is voor rekenen ook een quasi-experimentele studie gevonden. Het betreft een kwalitatief hoogstaand onderzoek (McCaffrey, Hamilton & Rand Education, 2007). Hoewel hierbij geen sprake was van randomisatie, maar van matching, ligt deze studie qua kwaliteit toch dicht bij de experimentele studies dan bij de overige gevonden quasi-experimentele studies. Gerandomiseerde experimenten vormen het hardste bewijs voor gevonden effecten, maar grote, goed gecontroleerde gematchte studies bevatten ook belangrijke informatie (Slavin, 2008). In de studie van McCaffrey et al. werden de controle en experimentele groep gematcht op basis van demografische gegevens en eerdere prestaties. Bovendien was hier sprake van een grootschalig onderzoek: het ging om 93 districten, daaruit ruim 400 scholen. De leerlingprestaties werden vergeleken op basis van standaard toetsen (PSSA=Pennsylvania System of School Assessment), en niet op toetsen die sterk gerelateerd waren aan de interventie die alleen de experimentele groep onderging. Hierdoor werd voorkomen dat effecten te groot zouden worden ingeschat (Slavin & Madden, 2008). Een gevonden effect uit deze studie zou dus ook een overtuigend argument zijn, waaraan veel waarde kan worden gehecht.

De interventie die in dit onderzoek beschreven werd, is het beschikbaar stellen van het Pennsylvania Value Added Assessment System (PVAAS), een systeem van statistische technieken om testresultaten te beoordelen. Aan scholen werd door de districten informatie van leerlingsscores op standaard toetsen meegedeeld. In deze studie is echter geen effect gevonden. In het artikel is met behulp van een kwalitatief deel van het onderzoek onderzocht waar dit aan zou kunnen liggen. De conclusie luidt dat de mogelijkheden van het Value Added Assessment systeem nog niet volledig benut werden. Uit vragenlijsten afgenomen op de experimentele scholen bleek zelfs dat veel scholen, hoewel de mogelijkheden er zijn, helemaal geen gebruik maakten van het systeem. Dit sluit aan bij de conclusie van Visscher en Ehren (2011) dat scholen vaak hun leerlingvolgsysteem onderbenutten. Een andere verklaring voor het uitblijven van effect kan zijn dat deze interventie geen aandacht besteedt aan het leren analyseren en interpreteren van data en het aanpassen van instructie. De interventie bestaat hier namelijk alleen uit het toegankelijk maken van data voor scholen. Uit literatuur blijkt dat dit niet voldoende is, en dus ook uit dit onderzoek.

In het laatste artikel dat gevonden is voor rekenen is geen sprake van een interventie, maar het betreft een correlatieel onderzoek. Het is het enige gevonden onderzoek dat is uitgevoerd in Nederland, en het werd uitgevoerd in opdracht van de Inspectie van het onderwijs (2010). Voor de studie werd gebruik gemaakt van een representatieve landelijke steekproef van 166 scholen. Er is onderzocht of scholen die opbrengstgericht werken, betere rekenprestaties halen dan scholen die minder opbrengstgericht werken. Dit is gebeurd door de opbrengstgerichtheid in kaart te brengen met behulp van een uitgebreid waarderingskader, waarna onderzocht werd of opbrengstgerichte scholen vaker voldoende rekentoetsresultaten en positieve trends van rekenresultaten behalen. De conclusie is positief: *“Op opbrengstgerichte scholen presteren leerlingen beter op rekentoetsen. Bovendien is op opbrengstgerichte scholen de trend die zich in de leerresultaten manifesteert vaker gunstiger dan op niet opbrengstgerichte scholen”* blz. 34 (Inspectie van het onderwijs, 2010). Op welke manier de analyses werden uitgevoerd of welke testen gebruikt zijn, wordt niet nauwkeurig beschreven.

Resultaten voor Lezen

De vijf genoemde experimentele studies en het gematchte quasi-experiment van McCaffrey en Hamilton meten niet alleen of DDDM effect heeft op rekenprestaties, maar ook of DDDM effect heeft op leesprestaties. De inhoud en uitvoering van de onderzoeken staat in de vorige alinea beschreven

(Resultaten voor Rekenen). Daarom volgen hier alleen de resultaten van deze onderzoeken voor wat betreft het effect van DDDM op leesprestaties.

Carlson et al. (2011) vonden net als voor rekenen een positief effect op leesprestaties, maar voor lezen was dit effect niet significant. Slavin et al. (2011), die het effect van dezelfde interventie onderzochten, vonden na 1 jaar interventie wel een significant effect voor groep 8 (0.26), maar niet voor groep 5. In combinatie met het quasi-experimentele deel van hun studie concludeerden zij dat de effecten pas vanaf het derde jaar na aanvang duidelijk te zien waren. Marsh et al. (2010) vonden een klein maar significant positief effect. Het effect was wel iets groter dan dat voor rekenen.

Jones & Krouse (1986) vonden geen effect voor rekenen, maar wel voor lezen, terwijl het een kortdurend onderzoek betrof. Stecker (2005) vond net als voor rekenen, ook voor lezen geen effect. Hier geldt zoals bij rekenen al genoemd, dat de verklaring gevonden werd in het feit dat instructie vaak niet werd aangepast aan de conclusies die volgen op het verzamelen en interpreteren van data. McCaffrey, Hamilton & Rand Education (2007) tenslotte vonden net als voor rekenen geen effect voor lezen. De conclusie luidde wederom dat het beschikbaar stellen van data niet genoeg is om leerling prestaties te verhogen.

Naast de genoemde experimentele studies en de studie van McCaffrey et al. zijn er voor lezen ook een aantal quasi-experimentele studies gevonden. De resultaten van deze studies hebben minder gewicht, omdat ze om verschillende redenen meer risico lopen op bias en over- of onderschatting van gevonden effecten. Ook is het bij deze onderzoeken minder zeker dat de gevonden samenhangen ook een causaal verband hebben. In de hierna beschreven quasi-experimenten werden overal positieve effecten op leesprestaties gevonden.

Bursuck, Smith, Munk, Damer, Mehlig & Perry (2004) beschrijven een quasi-experimenteel onderzoek in het kader van het project PRIDE (Preventing and Remediating Reading Problems Through Early Identification and Direct teaching of Early Literacy Skills). Dit is een vierjarig project gericht op het verhogen van leesprestaties, dat bestaat uit twee componenten: DDDM en professionele ontwikkeling van leerkrachten. Leerkrachten sorteren leerlingen op basis van assessments in verschillende niveaugroepen, "Tiers" genoemd. Tier 1 was klassikale instructie. Leerkrachten werden getraind in het gebruiken van effectieve strategieën en krijgen pedagogische tips. In Tier 2 kregen leerlingen die meer moeite hadden met lezen extra oefening in kleine groepjes. In Tier 3 zaten kinderen die achterbleven op de tweemaandelijks assessments. Hier bestond de instructie uit dagelijkse intensieve instructie in nog kleinere groepjes. De professionele ontwikkeling van leerkrachten bestond uit workshops, jaarlijkse trainingen en coaching.

In het derde jaar werd een controlegroep toegevoegd, die gelijk was qua leerlingkenmerken en prestatieniveau, gebaseerd op data van voor aanvang van het project. De leerlingprestaties voor het meten van effect van de interventie bestonden uit toetscores in het derde jaar, dus na twee jaar interventie. Er werden significante verschillen gevonden tussen de experimentele groep (102 leerlingen) en de controle groep (73 leerlingen). Deze verschillen werden duidelijker toen een overzicht werd gegeven van het percentage dat na het 3^e jaar de benchmark, een algemene standaard voor lezen, haalde.

Abbott, Atwater, Lee & Edwards (2011) hebben een quasi-experimentele studie naar het effect van DDDM op leesprestaties beschreven. De interventie van deze studie was een project binnen ERF (Early Reading First), een nationaal investeringsproject gericht op het professionaliseren van leerkrachten op het gebied van lesonderwijs. Het model dat in dit onderzoek wordt beschreven verbindt professionele ontwikkeling met DDDM door training en coaching. Het is een driejarig project, waarin leerlingen op basis van assessment geplaatst worden in 2 "tiers". Tier 1, net als in project PRIDE, is de klassikale instructie groep, en tier 2 is de extra instructie groep voor laagpresteerders. Het onderzoek betreft 9 klassen op een school. De 27 leerkrachten ontvingen gemiddeld 104 uur per jaar aan training/coaching. Na twee jaar worden effecten gemeten op leesprestaties van leerlingen. Hoewel hier geen controle groep is, wordt wel een verhoging geconstateerd op de standaard test Topel, een test waarbij meerdere aspecten van het lezen gemeten worden.

Lai, McNaughton, Amituanai-Tolosa, Turner, & Hsiao (2009) en McNaughton, Lai en Hsiao (2012) beschrijven het effect van een project uitgevoerd in Nieuw Zeeland, genaamd Learning Schools Model. Het is een driejarig project, dat zich zowel richt op DDDM als op het geven van effectieve instructie. Het originele onderzoek is beschreven door Lai et al., en betreft de driejarige interventie op

7 scholen. McNaughton et al. repliceren het onderzoek van Lai et al. met twee andere groepen: één groep lijkt sterk op de groep zoals in het onderzoek van Lai et al. (de “like” groep), de andere groep wijkt juist sterk af (de “unlike” groep). Er zijn geen controlescholen, maar er wordt wel gebruik gemaakt van gestandaardiseerde tests. Er wordt een baseline vastgesteld, waarna de verwachte groei van leerlingprestaties vergeleken wordt met de werkelijke. In beide onderzoeken, en voor alle drie de groepen, wordt een positief resultaat gevonden. Per jaar van het project is er gemiddeld 3 à 4 maand leerwinst ten opzichte van het verwachte leerresultaat, wat voor de duur van het driejarige project leidt tot de leerwinst van een heel jaar.

Het laatste artikel in deze literatuurstudie betreft een onderzoek van kleine omvang (Sexton, 2010). De schrijver noemt de gebruikte methode “action research”, maar zijn opzet bevat een aantal elementen die het bijna onmogelijk maken uitspraken te doen die generaliseerbaar zijn. De interventie die wordt beschreven wordt uitgevoerd door de onderzoeker zelf, die leerkracht is. Hij selecteert tien leerlingen uit zijn klas, die hij wekelijks test. Hij geeft die leerlingen extra instructie in leesvaardigheid, en kiest de invulling van de instructie op basis van gedetailleerde item-analyse en op basis van voorkeuren van leerlingen. Er is geen controle groep, er vindt geen statische bewerking van de gegevens plaats, en er worden slechts resultaten beschreven over een periode van 10 weken. In het artikel wordt wel geconcludeerd dat er een positief effect is, dat gemeten wordt door een verhoging van scores op een gestandaardiseerde (benchmark) test. Met dit onderzoek wordt misschien aangetoond dat intensieve aangepaste instructie effect heeft op prestaties, maar de vraag is of de interventie realistisch is. Wellicht heeft elke vorm van instructie van deze intensiteit effect op leesprestaties.

Conclusie en Discussie

Aan het begin van dit verslag werd deze onderzoeksvraag geformuleerd:

Hoe duidelijk is aangetoond dat opbrengstgericht werken een positief effect heeft op prestaties van basisschoolleerlingen?

De twaalf artikelen die gevonden zijn voor het beantwoorden van deze vraag geven een redelijk beeld van hoe duidelijk het effect van OGW op leerling prestaties eigenlijk is aangetoond. De resultaten voor rekenen lijken sterk op die voor lezen.

Het positieve effect van opbrengstgericht werken op rekenprestaties dat gevonden is door de inspectie is interessant, maar vraagt om verder onderzoek. Voor het beoordelen of dit verband causaal is, en men dus mag spreken van een effect, is experimenteel onderzoek nodig. De voorkeur gaat uit naar zuivere experimenten. Voor rekenen zijn er vijf artikelen gevonden die zulke experimenten beschrijven. Er worden positieve effecten gevonden in de experimentele studies van Carlson, Borman en Robinson (2011), Slavin et al. (2011) en Marsh, McCombs en Martorell (2010). Bij deze drie onderzoeken betreft de interventie training en/of coaching in databenutting, data-analyse en data-interpretatie. Slavin et al. vindt grotere effecten voor latere jaren in de interventie, waarin ook het onderwijs daadwerkelijk wordt aangepast. Dit sluit aan bij de resultaten van Stecker (2005), die geen positief effect vindt, omdat scholen hun onderwijs niet aanpassen. Ook Jones en Krouse (1986) vinden geen positief effect, maar hier duurde de interventie slechts 8 weken. In het quasi-experimenteel, maar goed opgezette onderzoek van McCaffrey, Hamilton en Rand Education (2007) wordt ook geen positief effect gevonden. Hier is eveneens geen sprake van juiste interpretatie van data en de aanpassing van het onderwijs: het data-analyse systeem wordt vaak niet eens gebruikt. DDDM/OGW lijkt een positief effect te hebben op leesprestaties, als niet alleen data geanalyseerd wordt, maar deze ook juist wordt geïnterpreteerd en het onderwijs erop wordt aangepast. Het aanleren van effectieve instructiemethoden vergroot het effect van een interventie.

Ook voor lezen zijn de resultaten gemengd. Net als voor rekenen, geldt voor lezen dat de experimentele studies, alsmede de quasi-experimentele studie van McCaffrey en Hamilton alleen positieve effecten rapporteren als er minstens sprake is van het leren interpreteren van data. Carlson et al. en Slavin et al., bij wie op het moment van natest nog geen sprake is van het hebben aangeleerd van effectieve instructiemethoden, vonden soms wel, en soms geen significante effecten. Ronduit positieve effecten zijn gevonden in de quasi-experimentele studies. Hoewel bij deze studies sprake kan zijn van bias en overschatting van het effect, zijn ook de interventies behoorlijk intensief. Ze zijn niet alleen

gericht op databenutting, maar ook op het trainen van interpretatie van data en het aanleren van effectieve instructie-methoden. Hoewel de experimenten vaak geen controle groep hebben, zeggen de verhoogde scores op de benchmark testen wel degelijk iets. Als er voor aanvang van de interventie sprake was van een groot percentage leerlingen dat de benchmark niet haalde, en na de interventie een veel groter percentage dat die benchmark haalt, dan heeft de interventie waarschijnlijk in de betreffende setting effect gehad. Toch is dit niet zeker; er kan ook sprake zijn van een Hawthorne effect. Aangezien de resultaten van meerdere onderzoeken in dezelfde richting wijzen, versterkt dit het bewijs voor de stelling dat het professionaliseren van leerkrachten op het gebied van databenutting en –interpretatie, samen met het aanleren van effectieve instructie strategieën, een positief effect heeft op leesprestaties.

Uit het conceptueel kader kwam reeds naar voren dat feedback bestaat uit het waarnemen van informatie, het interpreteren van die informatie en het nemen van maatregelen. OGW is dan ook meer dan alleen het afnemen van toetsen. Juist voor het leren interpreteren van data is training en coaching nodig: wat is de achterliggende reden voor tegenvallende prestaties? Hoe gedetailleerd is de beschikbare informatie en wat zegt die informatie nu precies? Tenslotte is het ook nog niet zo voor de hand liggend dat leerkrachten meteen weten welke onderwijsmaatregelen het meeste effect zullen hebben; het is belangrijk dit aspect ook te betrekken bij professionaliseringsinterventies. Want, zo stellen Marsh et al. (2010): *“Taking action is often more challenging and requires more creativity than does analysis”*.

Discussie

De vraag is, waarom er bij sommige onderzoeken wel, en bij andere onderzoeken geen positieve effecten worden gevonden. Bij de quasi-experimentele studies worden positieve resultaten gerapporteerd, maar effectgroottes worden niet genoemd. Bij de experimentele studies lijken de effecten af te hangen van de kwaliteit en de intensiteit van de interventies. De interventies in de gevonden artikelen verschillen aanzienlijk van elkaar. Dit heeft onder andere te maken met het feit dat opbrengstgericht werken uit meerdere componenten bestaat. De eerste component, het vaststellen van de beginsituatie, gebeurt in de onderzoeken op verschillende manieren. De tweede component, het stellen van doelen, wordt alleen meegenomen bij het Nederlandse onderzoek. De laatste component, het aanpassen van het onderwijs, wordt op verschillende manieren uitgewerkt. Soms worden in een interventie een aantal evidence-based instructiemethoden onderwezen, soms worden leerkrachten geprofessionaliseerd in het zelf zoeken van instructiemethoden die aangetoond effectief zijn.

Toch is een effect niet altijd gerelateerd aan de inhoud van de interventie. Zo worden bij de interventie van CDDRE na 1 jaar al effecten gevonden voor rekenen (door Carlson et al.) en voor zowel lezen als rekenen in groep 8 (door Slavin et al.). Terwijl in dat eerste jaar de interventie alleen gericht was op het leren lezen van data, en nog niet op het aanpassen van het onderwijs. Ook de onderzoeksofzet is van invloed op het vinden van effecten. Zo vonden Carlson et al. bijvoorbeeld aanvankelijk geen effect, maar wel nadat een multilevel analyse werd uitgevoerd waarbij covariaten op school- en districtniveau werden meegenomen.

Vervolgens is het niet alleen van belang dat de gevonden effecten statistisch significant zijn, maar ook de omvang ervan substantieel is. Slechts in enkele onderzoeken worden effectgroottes vermeld, hetgeen vergelijken van effecten bemoeilijkt. De effecten die vermeld worden liggen beneden 0.5 en zijn dus gemiddeld of klein. Toch is een klein effect op school- of leerkrachtniveau al interessant, omdat een groot deel van leerlingprestaties verklaard wordt door leerlingkenmerken zoals intelligentie en SES. Als er slechts een klein effect bestaat, maar er is wel gecontroleerd voor belangrijke covariaten op school- en leerlingniveau, dan is dit toch een teken dat opbrengstgericht werken de moeite waard is om in te investeren.

Is het onderzoek naar de samenhang tussen OGW en leerling prestaties voldoende om een eenduidige conclusie te trekken? Nee. Het aantal onderzoeken naar het effect van opbrengstgericht werken op prestaties van leerlingen dat gevonden is in deze literatuurstudie is gering. Er zijn slechts vijf artikelen gevonden over experimentele studies en één hoogstaande quasi-experimentele (matching) studie. Daarbij is niet altijd gecorrigeerd voor alle mogelijke covariaten. Met meer, en vooral nauwkeuriger opgezet experimenteel onderzoek kan overtuigender worden aangetoond dat opbrengstgericht werken effect heeft op prestaties. Daarbij is het ook interessant te onderzoeken, onder welke voorwaarden dit effect het grootst is.

Literatuurlijst

- *Abbott, M., Atwater, J., Lee, Y., & Edwards, L. (2011). A Data-Driven Preschool PD Model for Literacy and Oral Language Instruction. *NHSA Dialog*, 14(4), 229-245.
- Bennett, R.E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18 (1), 5-25. <http://dx.doi.org/10.1080/0969594X.2010.513678>.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- *Bursuck, W. D., Smith, T., Munk, D., Damer, M., Mehlig, L., & Perry, J. (2004). Evaluating the impact of a prevention-based model of reading on children who are at risk. *Remedial and Special Education*, 25(5), 303-313. doi: 10.1177/07419325040250050401.
- *Carlson, D., Borman, G. D., & Robinson, M. (2011). A Multistate District-Level Cluster Randomized Trial of the Impact of Data-Driven Reform on Reading and Mathematics Achievement. *Educational Evaluation and Policy Analysis*, 33(3), 378-398. doi: 10.3102/0162373711412765
- Chappuis, S. & Stiggins, R. J. (2002). Classroom Assessment for Learning. *Educational Leadership*, 60 (1), 40-43.
- CPS (2010). Hogere leeropbrengsten door versterken van bestuurlijk en onderwijskundig handelen.
- Earl L. & S. Katz (2006). *Leading Schools in a Data-Rich World*. Thousand Oaks (CA): Corwin Press.
- Halverson, R. (2010). School Formative Feedback Systems. *Peabody Journal of Education*, 85(2), 130-146.
- Hattie, J. (2009). *Visible Learning*. Routledge: Abingdon.
- Hattie, J.A.C., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- *Inspectie van het Onderwijs (2010). *Opbrengstgericht werken in het basisonderwijs. Een onderzoek naar opbrengstgericht werken bijrekenen-wiskunde in het basisonderwijs*. Utrecht: Inspectie van het Onderwijs.
- *Jones, E. D., & Krouse, J. P. (1986, april). The Effectiveness of data-based instruction by student teachers in classrooms for students with mild learning handicaps. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. <http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED281836>
- *Lai, M. K., McNaughton, S., Amituanai-Tolosa, M., Turner, R., & Hsiao, S. (2009). Sustained Acceleration of Achievement in Reading Comprehension: The New Zealand Experience. *Reading Research Quarterly*, 44(1), 30-56.
- Ledoux, G., Blok, H. & Boogaard, M. (2009). *Opbrengstgericht werken. Over de waarde van meetgestuurd onderwijs*. Amsterdam: SCO Kohnstamm Instituut.
- Locke, E. A. & Latham, G. (2002). Building a practically useful theory of goal setting and task motivation. *The American psychologist*, 57(9), 705-17.
- *Marsh, J. A., McCombs, J. S., & Martorell, F. (2010). How Instructional Coaches Support Data-Driven Decision Making Policy Implementation and Effects in Florida Middle Schools. *Educational Policy*, 24(6), 872-907. doi: 10.1177/0895904809341467
- Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education*. Santa Monica (CA): Rand Education. http://www.rand.org/pubs/occasional_papers/OP170.html
- *McCaffrey, D. F., Hamilton, L. S., & Rand Education, S. M. C. A. (2007). *Value-Added Assessment in Practice: Lessons from the Pennsylvania Value-Added Assessment System Pilot Project*. [Technical Report. TR-506-CC]. Santa Monica (CA): RAND Corporation.
- *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(2), 203-228. <http://dx.doi.org/10.1080/09243453.2011.652126>
- Ministerie van Onderwijs, Cultuur en Wetenschap en PO-raad (2012). *Bestuursakkoord Primair Onderwijs 2012-2015*. <http://www.poraad.nl/sites/www.poraad.nl/files/bestuursakkoordgetekend.pdf>
- Oberon (2008). *Opbrengstgericht werken door scholen voor primair en voortgezet onderwijs* (onderzoek in opdracht van de Onderwijsraad). Utrecht: Oberon.

- OCW (2007). *Scholen voor morgen. Samen op weg naar duurzame kwaliteit in het primair onderwijs*. (Kwaliteitsagenda Primair Onderwijs). Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap.
- OCW (2011). *Opbrengsten, maak er werk van!* Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap.
- Richardson, G. (1991). *Feedback thought in social science and systems theory*. Philadelphia: University of Pennsylvania Press.
- Senge, P. M., (1992). *De vijfde discipline*. Schiedam: Scriptum Management.
- *Sexton, J. (2010). *Leveling the Playing Field: Increasing Student Achievement through Data-Driven Ability Grouping and Instructional Practices* [Online Submission].
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189. doi: 10.3102/0034654307313795
- Slavin, R. E. (2008). Perspectives on Evidence-Based Research in Education--What Works? Issues in Synthesizing Educational Program Evaluations. *Educational Researcher*, 37(1), 5-14.
- *Slavin, R. E., Cheung, A., Holmes, G., Madden, N. A., & Chamberlain, A. (2011). *Effects of a data-driven district reform model*.
<http://www.cddre.org/images/Effects%20of%20a%20Data%20Driven%20District%20Reform%20Model%20January%202011.pdf>
- Slavin, R. E., Lake, C., & Center for Data-Driven Reform in Education. (2007). *Effective Programs in Elementary Mathematics: A Best-Evidence Synthesis* (Version 1.2). Center for Data-Driven Reform in Education.
- Slavin, R. E., & Madden, N. A. (2008). *Understanding bias due to measures inherent to treatments in systematic reviews in education*. Paper presented at the annual meetings of the Society for Research on Effective Education. Virginia: Christal City.
- *Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, 42(8), 795-819. doi: 10.1002/pits.20113.
- Visscher, A. & Ehren, M. (2011). *De eenvoud en complexiteit van opbrengstgericht werken*. Enschede: Universiteit Twente. Verkregen via: <http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2011/07/13/de-eenvoud-en-complexiteit-van-opbrengstgericht-werken.html>
- Visscher, A., Peters, M. & Staman, L. (2010). *Het Focus-project: opbrengstgericht werken op basis van prestatie-feedback*. Enschede: Universiteit Twente.
- Von Bertalanffy, L. (1969). *General system theory: Foundations, development, applications*. In: Halverson, R. (2010). School Formative Feedback Systems. *Peabody Journal of Education*, 85(2), 130-146.