

University of Groningen

Effecten van het trainings- en begeleidingstraject 'Streef

Doolaard, S.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2013

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Doolaard, S. (2013). *Effecten van het trainings- en begeleidingstraject 'Streef: Gebruik maken van opbrengsten'*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

G I O N

Gronings Instituut voor Onderzoek van Onderwijs
Grote Rozenstraat 3 9712 TG Groningen



rijksuniversiteit
 groningen / gion

Effecten van het trainings- en begeleidingstraject 'Streef: Gebruik maken van opbrengsten'

Dr. Simone Doolaard

Effecten van het trainings- en begeleidingstraject 'Streef: Gebruik maken van opbrengsten'

Dr. Simone Doolaard

Effecten van het trainings- en begeleidingstraject 'Streef: Gebruik maken van opbrengsten'

Rapportage BOPO-programmalijn Onderwijskwaliteit PO
Deelproject 4 'Professionaliseren en samenwerken' (413-09-064)

Dr. Simone Doolaard
Groningen: GION/RUG
Februari 2013

ISBN 978-90-6690-598-6

2013. GION, Gronings Instituut voor Onderzoek van het Onderwijs

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook zonder voorafgaande schriftelijke toestemming van de directeur van het instituut.

No part of this book may be reproduced in any form, by print, foto print, microfilm or any other means without written permission of the director of the instituut.

Inhoud

Managementsamenvatting.....	1
1 Inleiding.....	5
1.1 Opbrengstgericht werken in het landelijk onderwijsbeleid.....	7
1.2 Theoretische achtergrond.....	10
2 Methode.....	13
2.1 Werving en respons.....	13
2.2 De interventie.....	15
2.2.1 Effectieve interventies.....	15
2.2.2 Opzet en inhoud van de interventie.....	17
2.3 Dataverzameling bij leerkrachten en leerlingen.....	22
2.3.1 Leerkrachtvragenlijst.....	22
2.3.2 Leerlingtoetsen en –achtergrondgegevens.....	22
2.3.3 Analyse van de leerkrachtvragenlijst en de leerlingresultaten.....	23
2.3.4 Beoordeling van de analyse-rapportages.....	25
3 Resultaten.....	27
3.1 Uitvoering en implementatie van de interventie.....	27
3.2 Effecten op leerkrachtattitude, –gedrag en analysevaardigheden.....	27
3.2.1 Effect op leerkrachtattitude- en gedrag.....	28
3.2.2 Effect op analysevaardigheden.....	34
3.3 Effecten op leerlingresultaten.....	36
3.4 Verklaring: leerkrachtattitude- en gedrag als mediator.....	40
4 Samenvatting en conclusies.....	45
Literatuur.....	49
Bijlage.....	51

Managementsamenvatting

Om te onderzoeken in welke mate en hoe leerkrachten de gegevens die zij kunnen ontleen aan het Cito-leerlingvolgsysteem gebruiken en om te onderzoeken wat de effecten zijn op de vorderingen van leerlingen als leerkrachten als team geschoold worden in een beter gebruik daarvan is in het schooljaar 2010-2011 op 40 scholen een interventie, genaamd *Gebruik maken van opbrengsten* uitgevoerd. De interventie betrof een trainings- en begeleidingstraject uitgevoerd door de Rijksuniversiteit Groningen en twee onderwijsadviesorganisaties. Het traject is een van de deelprojecten binnen Streef, een samenhangend programma van 3 interventies op verschillende terreinen van het onderwijsbeleid rond kwaliteit in het PO: referentieniveaus, opbrengstgericht werken en excellentie.

Vooraf en na afloop van *Gebruik maken van opbrengsten* is bij leerkrachten met behulp van een digitale vragenlijst gemeten hoe zij over opbrengstgericht werken en daaraan gerelateerde begrippen denken en hoe ze daarbij handelen. Als onderdeel van *Gebruik maken van opbrengsten* zijn driemaal analyserapporten van leerkrachten verzameld aan de hand waarvan onderzoek is gedaan naar hoe hun analysevaardigheden zich ontwikkelden. En ten slotte zijn voor, tijdens en na het traject toetsresultaten van leerlingen verzameld.

Met behulp van deze gegevens is eerst onderzocht wat het directe effect van *Gebruik maken van opbrengsten* is, namelijk een mogelijke feitelijke verbetering van vaardigheden, daarnaast zijn de meest indirecte effecten onderzocht op de attitude en handelen van leerkrachten in de klas. Als laatste in deze keten zijn de vorderingen van leerlingen geanalyseerd.

De analyse-vaardigheden van leerkrachten zijn aan het eind van *Gebruik maken van opbrengsten* duidelijk hoger dan aan het begin van het traject. Dit resultaat komt overeen met de conclusie van Visscher en Ehren (2011) die aangeven dat een beter gebruik van een leerlingvolgsysteem aan te leren is. Uit de analyse van de vragenlijsten blijkt dat de groep leerkrachten die aan *Gebruik maken van opbrengsten* al iets meer dan de leerkrachten in de controlegroep ervaring had met het analyseren van toetsresultaten, maar dat desondanks zij op de variabelen die het meest expliciet aan de orde zijn geweest meer vooruitgegaan zijn dan de controlegroep. Dit resulteert in 2012 in een situatie waarbij de leerkrachten die aan *Gebruik maken van opbrengsten* hebben deelgenomen een uitgebreidere analyse maken van toetsresultaten, daarop in een breder kader reflecteren en meer gebruik maken van groepsplannen waarin vastgelegd is op welke manier zij het onderwijs in de komende periode gaan differentiëren. Wat betreft de keuzes die daarbij gemaakt worden lijken er geen verschillen en geen vooruitgang op te treden.

De analyses van de vorderingen van leerlingen laten tot nu toe weliswaar een effect zien in de positieve richting maar dit effect is niet significant zodat we feitelijk moeten spreken van het ontbreken van een effect op leerlingniveau. Hoewel er dus in het algemeen een positieve tendens te zien is, is het effect, binnen de gehanteerde grenzen, (net) niet groot genoeg om

significant te zijn. De retentie-effecten zijn zien er minder positief uit. Samenvattende kunnen we stellen dat het licht positieve – zij het niet significante – effect wat optrad in het schooljaar waarin de leerkrachten deelnamen aan de interventie niet vastgehouden wordt in het schooljaar daarna.

Hierdoor is ook de laatste stap waarbij wordt onderzocht of het effect op leerlingniveau verklaard kan worden door een verbetering in attitude of gedrag van leerkrachten niet aan de orde.

Leeropbrengsten stijgen wanneer leerkrachten werken volgens de evaluatieve cyclus. Dat blijkt ook uit de clusteranalyse die op basis van de leerkrachtvragenlijst van Streef gedaan is. Dat leerkrachten zich ontwikkelen door deelname van Streef blijkt ook uit dit onderzoek. Wat kunnen dan redenen zijn dat de leerlingresultaten zich weliswaar in de gewenste richting ontwikkelen maar niet overtuigend zijn en bovendien in het jaar nadat de leerkrachten aan Streef deelnamen weer terugvielen?

Een van de mogelijke verklaringen is af te leiden uit de clusteranalyse waarin 4 typen teams waren onderscheiden: opbrengstgerichte en niet-opbrengstgerichte teams, teams die ‘individualistisch’ genoemd kunnen worden (waarin leerkrachten wel veel analyseren, maar er geen sprake is van overleg en samenwerking binnen het team) en teams die ‘behoudend’ genoemd worden en die ‘gemiddeld’ analyseren, samenwerken, maar weinig gericht zijn op vernieuwingen en verbeteringen in hun handelen. Deze laatste genoemde behalen resultaten die vergelijkbaar zijn met de opbrengstgerichte teams. De ‘individualistische’ teams presteren minder goed. Dit kan duiden op het feit dat veel analyseren maar daar niet in teamverband iets mee doen, een eventueel positief effect van het maken van een goede analyse teniet doet. Mogelijk zijn de leerkrachten tijdens Streef nog te veel bezig met het verbeteren van hun eigen vaardigheden in het analyseren en komt het teamoverleg daarover nog niet goed uit de verf. Als er geen ‘teamdruk’ op ligt, vervliegt het effect van beter analyseren mogelijk weer na het jaar van actieve participatie of voeren leerkrachten de analyse niet meer uit als ze daartoe niet een expliciete opdracht gekregen hebben.

Een andere mogelijke verklaring wordt ook door Streef bevestigd: namelijk dat de stap van analyse naar groepplan en de uitvoering daarvan nog onvoldoende gemaakt wordt. De beide clusters van teams die de hoogste leerlingresultaten behalen scoren allebei hoog op differentiatie. Blijkbaar zijn zij in staat om die stap wel te maken en uit te voeren in de klas. Streef-leerkrachten gaan wel vooruit op analysevaardigheden maar mogelijk niet voldoende op de kennis en vaardigheden die nodig zijn om die vertaalslag te maken en uit te voeren.

Samenvattend concentreren de aanbevelingen voor beleid en de onderwijspraktijk gericht op het verder verbeteren van opbrengstgericht werken in de scholen zich dus op het stimuleren van en aandacht besteden aan de fase van het cyclus van opbrengstgericht werken na het

analyseren van de data: het met elkaar bespreken van de uitkomsten, het gezamenlijk bepalen van 'nieuwe' doelen en het uitstippelen van een weg daarnaar toe. Daarvoor is een gedegen kennis van de leerstof nodig en kennis van op onderzoek gebaseerde didactische aanpassingen.

Vaardigheden die nodig zijn om goed te kunnen differentiëren vragen om ervaring en kunnen op de opleidingen wel aangereikt worden, maar onderzoek naar de ontwikkeling van leerkrachten toont aan dat zij zich in het begin van hun loopbaan eerst moeten concentreren op een aantal meer basale aspecten van het werk. Benodigde kennis van de leerstof, de didactiek en ook het analyseren van gegevens behoren wel al in de opleiding aangeleerd te worden. Met name op dat laatste gebied is een verbetering van de opleiding mogelijk. Waar en hoe gevorderde leerkrachten een 'proeve van bekwaamheid' zouden moeten afleggen is een punt van aandacht voor het beleid.

1 Inleiding

In ‘Scholen voor Morgen’, de kwaliteitsagenda voor het primair onderwijs uit 2007 en vooral in 2011 in het actieplan ‘Basis voor Presteren’ neemt opbrengstgericht werken een belangrijke rol in. Hoewel er geen eenduidige definitie is van opbrengstgericht werken wordt in het algemeen een manier van werken bedoeld waarin heldere doelen gesteld zijn en hoge verwachtingen, waarin vorderingen regelmatig in kaart worden gebracht en geanalyseerd en waarin vervolgens het onderwijs aangepast wordt aan de behoeften van leerlingen. Het betreft een cyclisch proces dat schoolbreed, door het hele team, wordt doorlopen en dat gericht is op het realiseren van zo hoog mogelijke prestaties door alle leerlingen. Opbrengstgericht werken is dus geen doel maar een middel, een manier van werken, volgens sommigen ook een attitude van leerkrachten en schoolleiders. Verschillende onderzoeken (bijvoorbeeld Ledoux, Blok & Boogaard, 2009) hebben aangetoond dat individuele leerkrachten lang niet allemaal voldoende vaardig zijn in het analyseren van toetsresultaten en daarvan dus ook geen gebruik maken bij het vormgeven van hun onderwijs, ook op veel scholen ontbreekt het aan een systematiek om gezamenlijk, in teamverband, resultaten te analyseren en te bespreken.

In het actieplan, waarin de ambities van de landelijke overheid zijn geformuleerd, zijn zowel ‘einddoelen’ geformuleerd op het niveau van de leerling die onder andere bereikt zouden moeten worden doordat scholen meer opbrengstgericht werken, als ‘tussendoelen’ zoals een toename van het percentage scholen dat opbrengstgericht werkt.

Deze ambities staan niet op zichzelf maar zijn verweven met een aantal andere doelstellingen voor het primair onderwijs gericht op het laten excelleren van leerlingen, het bieden van maatwerk aan excellente leerlingen, het beperken van het aantal taal- en rekenzwakke scholen, het definiëren van bekwaamheidseisen voor schoolleiders en het verhogen van de effectieve leertijd. Om de beleidsambities te realiseren is een aantal maatregelen genomen zoals het formuleren van referentieniveaus, het informeren over en het stimuleren van opbrengstgericht werken, het afsluiten van het Bestuursakkoord PO, het verplicht stellen van een leerlingvolgsysteem, het financieren van reken- en taalpilots en meerjarige trajecten op het gebied van onderwijs aan excellente leerlingen.

Om inzicht te krijgen in de implementatie en werking of doeltreffendheid van deze beleidsmaatregelen, dus in de manier waarop en de mate waarin er veranderingen gerealiseerd worden in leerkrachtattitude en –gedrag en in leerlingprestaties heeft de BOPO de onderzoekslijn ‘Kwaliteit PO’ geprogrammeerd met daarin deelonderzoeken rondom referentieniveaus, opbrengstgericht werken en excellentie. De onderzoekslijn wordt uitgevoerd door het GION, Rijksuniversiteit Groningen en het Cito, in samenwerking met de

onderwijsbegeleidingsdiensten Marant en Cedin. Om de onderzoeksvragen van de verschillende deelonderzoeken in samenhang met elkaar te kunnen beantwoorden is een schoolverbeterings- en onderzoeksproject opgezet onder de naam ‘Streef’, waarin op de 3 genoemde thema’s interventies zijn uitgevoerd waarin leerkrachten zijn getraind en begeleid. Deelnemende scholen namen aan een van de interventies deel en participeerde allemaal in het dataverzamelingstraject dat 2,5 jaar duurde en waarin leerkrachten en leerlingen zijn bevraagd en getoetst. Bij alle deelnemende scholen is tweemaal een vragenlijst voor leerkrachten afgenomen, waaraan alle leerkrachten deelnamen. Vijfmaal zijn toetsgegevens bij leerlingen verzameld, eenmaal voorafgaand aan deelname (juni 2010) en 2 keer tweemaal in de schooljaren 2010-2011 en 2011-2012. Van de leerlingen zijn ook verscheidene achtergrondgegevens verzameld.

Naast de genoemde 3 projecten participeerden er scholen in Streef die deelnamen aan *Gebruik maken van opbrengsten* met de middenbouwgroepen 4 en 5 en aan ‘Strategisch leren lezen’ met groep 6 t/m 8. Beide projecten vinden hun oorsprong in TIER, Topinstituut voor evidence based onderwijs onderzoek, waarin de RUG participeert. Voor de scholen maakten het geen verschil aan welk deelproject zij deelnamen, alle scholen draaiden op dezelfde manier mee in het onderzoek.

In deze rapportage wordt verslag gedaan van het deelonderzoek rondom de effecten van scholing en begeleiding op het gebied van opbrengstgericht werken waarin de volgende onderzoeksvragen centraal staan: .

1. In welke mate en hoe gebruiken de leerkrachten de gegevens die ze kunnen ontleen aan het Cito-leerlingvolgsysteem?
2. Wat zijn de effecten op de vorderingen van leerlingen als leerkrachten als team geschoold worden in een beter gebruik?

In 4 regionale schooloverstijgende bijeenkomsten en 5 consultaties op de eigen school werden leerkrachten geschoold en begeleid bij het optimaal gebruik maken van informatie uit leerlingtoetsen. Welke informatie over leerlingen en over de hele groep kun je als leerkracht halen uit zo’n toets en hoe ga je daarmee vervolgens om in je lessen? De bovenschoolse bijeenkomsten werden verzorgd door een RUG-medewerker en hadden meer het karakter van een training, de bijeenkomsten op de eigen school werden verzorgd door een medewerker van een van de onderwijsbegeleidingsdiensten en had meer een begeleidend karakter. Aan dit deelonderzoek namen 40 scholen deel, met hun onderbouw- of bovenbouwleerkrachten, uit de oorspronkelijke groep van 92 Streef-scholen die geworven zijn in de eerste helft van 2010. De ingeschatte tijdsbesteding voor leerkrachten was 40 uur.

In het vervolg van dit inleidende hoofdstuk wordt in 1.1 verder ingegaan op de beleidscontext rondom opbrengstgericht werken (1.1) en de theoretische achtergrond van het begrip (1.2). In

1.2 wordt het begrip opbrengstgericht werken theoretisch ingebed en wordt een overzicht gegeven van eerder onderzoek naar (elementen van) opbrengstgericht werken. Deze uitkomsten van eerder onderzoek vormen de inhoudelijke achtergrond van dit traject.

Doel van dit deelonderzoek was om te onderzoeken op welke manier en in welke mate het handelen van leerkrachten en de resultaten van leerlingen verbeterd kunnen worden door leerkrachten in teamverband te scholen en te begeleiden en samenwerking tussen leerkrachten te stimuleren op het gebied van opbrengstgericht werken. In hoofdstuk 2 wordt daarom, naast het beschrijven van de onderzoeksgroep, de onderzoeksopzet en de effectmetingen bij leerlingen en leerkrachten in gegaan op de interventie en de manier waarop in de interventie aansluit bij literatuur over onderwijsverbetering en effectieve interventies. Hoofdstuk 3 bevat de resultaten van het traject en hoofdstuk 4 de conclusie en discussie.

1.1 Opbrengstgericht werken in het landelijk onderwijsbeleid

In de kwaliteitsagenda ‘Scholen voor morgen’ komt de term ‘opbrengstgericht werken’ nog niet expliciet voor. Het verbeteren van taal- en rekenprestaties van kinderen staat op de eerste plaats. In de tekst wordt ingegaan op wat er nodig is om dat te bereiken: ‘Om de taal- en rekenprestaties te verbeteren is het nodig dat leraren en schoolleiders de aanwezige informatie over leerprestaties in de school benutten en leren van en met collega’s en experts, zowel in als buiten de school, over de meest effectieve onderwijsmethoden’ (p. 5). Verwijzend naar eerder onderzoek – onder andere van de Inspectie van het Onderwijs - wordt erop gewezen dat op basisscholen wel al een schat aan informatie aanwezig is maar dat dat nog onvoldoende gebruikt wordt. Men zou gestimuleerd moeten worden om vaardigheden te ontwikkelen om de uitkomsten te analyseren, te begrijpen en te vertalen naar concrete voorstellen voor verbetering. Deze vaardigheden worden dan nog niet ondergebracht onder de term ‘opbrengstgericht werken’ maar onder ‘kwaliteitszorg’. De doelstelling luidt dan ook: ‘In 2011 voldoet minimaal 80 procent van de scholen aan de standaarden van kwaliteitszorg zoals omschreven in het toezichtskader van de inspectie. Dat betekent dat de informatie over onderwijskwaliteit die daaruit voortvloeit, beter wordt benut door leraren. Dit houdt in het systematisch gebruiken van toetsen, de toetsdata interpreteren en gebruiken voor het concrete onderwijsaanbod aan alle leerlingen en als maatstaf: waar sta ik ten opzichte van mijzelf (vorig jaar, parallelklas, etc.), andere, vergelijkbare scholen en de landelijke referentieniveaus’ (p.11).

In de jaren daarna raakt de term ‘opbrengstgericht werken’ meer ingeburgerd. In de eerste voortgangsrapportage van 7 juli 2008 wordt het analyseren en evalueren van toetsen en het op basis daarvan aanpassen van het onderwijs ‘een belangrijk onderdeel’ van kwaliteitszorg genoemd. De ambitie die daarbij aansluit luidt dat in 2011 minimaal 70 procent van de scholen aan de indicatoren van goede kwaliteitszorg zal voldoen. In een voetnoot wordt gemeld dat overleg met de inspectie heeft geleid tot precisering en bijstelling van de eerder geformuleerde ambitie. Van de indicatoren van het bredere aspect kwaliteitszorg gaat het

hierbij om twee specifieke indicatoren: het vergroten van de opbrengstgerichtheid en het evalueren van het onderwijsproces en de opbrengsten. In 2007 voldeed 39% van alle scholen aan beide indicatoren.

In de tweede voortgangsrapportage van 29 juni 2009 worden twee onderdelen genoemd die extra aandacht behoeven. Een daarvan is het percentage scholen dat aan beide specifieke kwaliteitszorgindicatoren voldoet; in het schooljaar 2007/2008 was dat weliswaar gestegen naar 52% maar daarmee lag het nog ver van de beoogde 70% af. Daarom is de inzet op dit terrein in de twee schooljaren daarna geïntensiveerd. Dit betekent dat er nog meer scholen betrokken worden in de lerende netwerken op het terrein van taal en rekenen (de taal- en rekenpilots), er komen financiële middelen voor verbetertrajecten op 50 nieuwe scholen, er komt meer aandacht voor het delen van effectief gebleken praktijkvoorbeelden van opbrengstgericht werken op een aantal regionale bijeenkomsten en meer experts om de scholen te ondersteunen. Ook zullen vanaf augustus 2010 de referentieniveaus wettelijk vastgesteld zijn en kunnen dienen als instrument om opbrengstgericht werken te realiseren.

Op 28 juni 2010 verschijnt de derde voortgangsrapportage. Opbrengstgericht werken wordt daarin 'de rode draad' door de kwaliteitsagenda genoemd. Ten aanzien van de doelstelling op het terrein van kwaliteitszorg c.q. opbrengstgericht werken is een lichte verbetering te zien. Het percentage scholen dat voldoende scoort op de indicatoren voor kwaliteitszorg is 54% (Onderwijsverslag 2008/2009). Op elk van de 5 individuele indicatoren is een stijgende trend zichtbaar. De inspectie concludeert dat de voorwaarden aanwezig zijn maar dat het evalueren van de resultaten en het daarop aanpassen van het onderwijs een knelpunt blijft. Eventuele effecten van de geïntensiveerde inzet waarover in de 2^e voortgangsrapportage wordt gesproken zijn in deze cijfers nog niet zichtbaar omdat het cijfers uit het schooljaar daarvoor betreft. In de opiniepeiling van Researchnet in 2010 en in een meta-analyse van het IVA uit 2010 zeggen respondenten wel vaker activiteiten te ondernemen die onder opbrengstgericht werken vallen.

Overigens wordt hier gesproken over kwaliteitszorg als belangrijk onderdeel van opbrengstgericht werken, waar opbrengstgericht werken in de eerste voortgangsrapportage in 2008 een belangrijk onderdeel van kwaliteitszorg was.

Hoewel er dus signalen zijn dat er in het onderwijs meer opbrengstgericht gewerkt wordt en leerkrachten en schoolleiders er niet negatief tegenover staan is er ook wel weerstand die de verdere uitbreiding van opbrengstgericht werken wel zou kunnen tegenwerken. Die weerstand concentreert zich op drie terreinen:

- Opbrengstgericht werken wordt geassocieerd met de basisvaardigheden. Aan taal en rekenen wordt zeer veel aandacht besteed, waardoor andere vakgebieden en andere vaardigheden, zoals sociaal-emotionele en burgerschapsvaardigheden te weinig aandacht krijgen.

- Opbrengstgericht werken wordt geassocieerd met (veel) toetsen. Met name voor jonge kinderen wordt dit niet altijd een geschikte werkwijze gevonden om oog te houden op de ontwikkeling van kinderen. Toetsen voor deze kinderen zouden te weinig zeggen omdat de ontwikkeling niet lineair is. Meer in het algemeen ageren betrokkenen tegen de monopoliepositie van het Cito, de instelling die het leeuwendeel van de toetsen levert en in de ogen van sommige betrokkenen daardoor te bepalend is in wat er getoetst en dus geleerd wordt.
- Opbrengstgericht werken wordt geassocieerd met het controleren van scholen. Scholen moeten de toetsgegevens aanleveren aan hun bestuur, de inspectie en mogelijk nog andere instanties en daarmee hun onderwijs verantwoorden. Het betreft hier toetsgegevens die ook gebruikt worden om het onderwijs te verbeteren. Dit dubbele perspectief wordt vaak als problematisch gezien.

Op 23 mei 2011 presenteren de nieuwe minister en staatssecretaris 3 actieplannen – een voor het primair onderwijs, een voor het voortgezet onderwijs en een voor leraren – waarin zij het onderwijsbeleid voor de komende regeerperiode uiteenzetten. Het actieplan voor het primair onderwijs, getiteld ‘Basis voor presteren’, is een voortzetting van het beleid zoals dat in de voorgaande regeerperiode is gevoerd met een nadruk op basisvaardigheden en op opbrengstgericht werken. Als streefdoel is geformuleerd dat in het primair onderwijs in 2015 60% van de scholen opbrengstgericht werkt en in 2018 90%, volgens het oordeel van de inspectie waarbij er gerekend wordt met een nulmeting in 2010 waarbij 30% van de scholen opbrengstgericht werkte. In de begeleidende brief wordt aangegeven dat opbrengstgericht werken geen doel is maar een effectief middel om de prestaties naar een hoger plan te brengen.

In het voorjaar van 2012 blijkt uit een peiling van Regioplan (Dekker, Krooneman, Brekelmans & Groenwoud, 2012) naar de manieren waarop scholen inspelen op de in de referentieniveaus beschreven eisen voor taal en rekenen opnieuw een versterking van het opbrengstgerichte werken. Volgens eigen zeggen van de respondenten is er meer aandacht voor taal- en rekenzwakke leerlingen en wordt er meer in niveaugroepen gewerkt. Daarnaast is er draagvlak voor de referentieniveaus, het vaststellen van een minimumniveau en extra aandacht voor de basisvaardigheden.

Het laatste onderwijsverslag van de inspectie (2012a) is minder positief: sinds het vorige onderwijsverslag heeft er geen verbetering plaats gevonden in het percentage scholen dat voldoet aan de indicatoren voor opbrengstgericht werken. Het percentage scholen dat systematisch de resultaten van leerlingen evalueert is iets hoger geworden, maar in algemene zin is er geen verbetering zichtbaar.

In 2011 en 2012 verschijnen tenslotte ook de eindevaluaties van de taal- en rekenverbetertrajecten en de Op Maat-trajecten waarin opbrengstgericht werken een

belangrijke rol speelden (Inspectie van het onderwijs, 2012b; van Schilt-Mol & van Vijfeijken, 2011; PO-Raad, 2012). Uit deze rapportages blijkt dat de resultaten van reken- en begrijpend lezen wel, maar van technisch lezen niet vooruit zijn gegaan. Wat betreft de onderwijsprocessen signaleren van Schilt-Mol en van Vijfeijken een cultuuromslag; er wordt meer belang gehecht aan het zelf uitvoeren van data-analyses en er is een dialoog op gang gekomen over de relatie tussen leerling-resultaten en leerkracht-handelen. Leerkrachten zijn steeds meer bereid kritisch te kijken naar en te reflecteren op eigen handelen en op het handelen van collega's.

Uit dit chronologische overzicht is duidelijk op te maken dat er in de looptijd van de kwaliteitsagenda 'Scholen voor morgen' tot en met de start van het actieplan 'Basis voor presteren' een verschuiving is opgetreden in de definitie van opbrengstgericht werken en de manier waarop dit gemeten wordt. Niet alleen gebruiken verschillende organisaties en instellingen verschillende definities en operationalisaties, ook binnen dezelfde organisatie – zoals bijvoorbeeld de inspectie – zien we verschuivingen. Dat maakt het moeilijk om een duidelijk antwoord te krijgen op de vraag in welke mate er in het basisonderwijs opbrengstgericht gewerkt wordt en of er sprake is van vooruitgang. Duidelijk is wel dat opbrengstgericht werken gezien wordt als middel om de ambities op het gebied van leerlingprestaties te verbeteren. Dat het complexe begrip opbrengstgericht werken bestaat uit school- en klaskenmerken die in principe samenhangen met goede leerlingprestaties komt in de volgende paragraaf aan de orde.

1.2 Theoretische achtergrond

Hoewel de term *opbrengstgericht werken* pas kort wordt gebruikt is het achterliggende principe, de werking, van het concept wordt met name binnen het onderwijseffectiviteitsonderzoek al vele jaren onderzocht (zie bijvoorbeeld Scheerens, Luyten & Steen 2007; Visscher & Ehren, 2011). Gerelateerde termen die daar gebruikt worden zijn dan bijvoorbeeld evaluerend vermogen, prestatiegerichtheid, hoge verwachtingen. Deze termen verwijzen naar de 'smalle' betekenis van opbrengstgericht werken waaronder gerekend worden: het regelmatig toetsen van leerlingen, het analyseren van de resultaten en het trekken van conclusies ten aanzien van het onderwijs aan (individuele) leerlingen. Ook Ledoux, e.a. (2009) hanteren een smalle betekenis van opbrengstgericht werken, door hen meetgestuurd onderwijs genoemd, waaronder zij verstaan het vastleggen van doelen en standaarden, het verzamelen van informatie, het registreren van vorderingen, het interpreteren van vorderingen en het nemen van beslissingen.

Met alleen deze smalle inkleuringen van opbrengstgericht werken verbetert echter niet vanzelfsprekend de kwaliteit van het onderwijs en de prestaties van leerlingen. De smalle betekenis beslaat maar een deel van de cyclus van opbrengstgericht werken waarin ook het uitvoeren van (gedifferentieerd) onderwijs en het evalueren daarvan een rol spelen. Zo

uitgedrukt is opbrengstgericht werken in te passen in de bekende plan-do-check-act-cirkel, waarin dus niet alleen geëvalueerd en gepland wordt maar ook uitgevoerd. De analysevaardigheden van leerkrachten worden het meest genoemd als hiaat in deze cyclus en vanwege het ontbreken daarvan worden ook vraagtekens geplaatst bij de mate waarin de afstemming van het onderwijs op verschillen tussen leerlingen effectief is (Inspectie van het Onderwijs, 2010).

Zoals aangegeven bestaat er voor de afzonderlijke ‘bestanddelen’ van opbrengstgericht werken voldoende bewijs van het feit dat zij samenhangen met leerlingprestaties. In een recente meta-analyse zijn de resultaten door Scheerens, e.a (2007) samengevat (zie Tabel 2.1).

Tabel 2.1 Effecten van school- en instructiefactoren (bron: Scheerens, e.a., 2007)

Schoolfactoren	Gemiddeld effect
Resultaatgerichtheid	0,14
Consensus en samenhang/samenwerking	0,02
Onderwijskundig leiderschap	0,05
Kwaliteit v/h curriculum/gelegenheid tot leren	0,15
Ordelijk klimaat	0,13
Monitoring	0,06
Betrokkenheid van ouders	0,09
Effectieve leertijd	0,15
Huiswerk	0,07
Differentiatie	0,02
Instructiefactoren	
Curriculum	0,08
Leraargestuurd klassenmanagement	0,10
Onderwijsstrategie (directe instructie)	0,09
Onderwijsstrategie (constructivistisch)	0,14
Klimaat/ondersteuning	0,12
Feedback/controle/evaluatie/test	0,07

De grootste effecten zijn de zien bij resultaatgerichtheid, kwaliteit van het curriculum (waarmee bedoeld wordt dat er leerstof wordt aangeboden die past bij waar de leerling behoefte aan heeft), een ordelijk klimaat op school en in de klas, veel effectieve leertijd en een constructivistische aanpak die cognitief georiënteerd is en de nadruk legt op metacognitie en zelfregulerend leren. Monitoren op schoolniveau en het afnemen van toetsen, werk controleren en feedback geven op groepsniveau laten weliswaar positieve effecten zien, maar die zijn wel wisselend en in vergelijking met andere factoren klein. In andere meta-analyses, bijvoorbeeld die van Hattie uit 2009 wordt een veel groter effect gevonden. Onderzoek dat specifiek deze factor onderzoekt laat zien dat er van evaluatie wel aanzienlijke effecten te verwachten zijn, maar alleen als het geven van effectieve, formatieve feedback aan leerlingen daar expliciet onderdeel van uitmaakt.

Opvallend is het geringe effect van differentiatie in de meta-analyse, een bevinding die ook door andere meta-analyses wordt bevestigd. Differentieren neemt in de cyclus van opbrengstgericht werken wel een belangrijke rol in; analyse van leerlingresultaten zal immers (bijna) altijd leiden tot de conclusie dat leerlingen verschillen en diensgevolge andere behoeften hebben. Blok (2004) geeft daarvoor in zijn review van adaptief onderwijs meerdere redenen waaronder de diverse operationalisatie en doelstellingen die leerkrachten hebben van adaptief onderwijs of differentiatie. Sommige leerkrachten zetten differentiatie impliciet in ervoor te zorgen dat de groep bij elkaar blijft (en de ene leerling dus meer zal groeien dan de andere) en anderen hebben juist tot doel iedereen zich op zijn of haar eigen niveau te laten ontwikkelen waardoor prestaties uiteengaan lopen. Deze verscheidenheid aan doelstellingen kan conflictueren met de gebruikelijke onderzoeksmethoden in het onderwijseffectiviteitsonderzoek waarbij de school waar gemiddeld de meeste leerwinst geboekt wordt het meest effectief is.

De effecten van de afzonderlijke bestanddelen zegt nog weinig over het effect van het samengestelde concept opbrengstgericht werken. Eerder onderzoek waarin onderzoekers factoren geclusterd hebben, bijvoorbeeld in varianten van kwaliteitszorg of van adaptief onderwijs laat zien dat er geen sprake is van een optelsom van effecten, maar dat de elementen waaruit een concept is opgebouwd elkaar ook kunnen versterken of juist afremmen.

In dit onderzoek wordt daarom in de interventie aandacht besteed aan de hele cyclus van opbrengstgericht werken, zodat het effect van het samengestelde concept onderzocht kan worden, ook de positie van de verschillende aspecten binnen het concept en hun wisselwerking zal onderzocht worden.

2 Methode

Zoals al eerder aangegeven wordt voor het beantwoorden van de onderzoeksvragen een trainings- en begeleidingstraject met leerkrachten doorlopen op het gebied van opbrengstgericht werken. Hieraan hebben ruim 250 leerkrachten meegedaan verspreid over 40 scholen. In 2.1 wordt eerst ingegaan op de werving en de respons van Streef als geheel en specifiek voor dit deelproject. In 2.2 wordt uitvoerig ingegaan op de interventie en de mate waarin die voldoet aan kenmerken van effectieve interventies. In 2.3 wordt de dataverzameling toegelicht en wordt aangegeven hoe de analyses zijn uitgevoerd.

2.1 Werving en respons

Omdat Streef is uitgevoerd door de RUG en het Cito, in samenwerking met de onderwijsbegeleidingsdiensten Marant – in de regio Arnhem – en Cedin – in de regio Noord Nederland, zijn de scholenbestanden van deze laatste 2 organisaties gebruikt voor het aanschrijven van de scholen. Daarbij zijn vooraf de volgende scholen uitgesloten:

- Zeer kleine scholen
- Zeer zwakke scholen volgens de Onderwijsinspectie
- Scholen die participeerden in lopend onderzoek van het GION

Dit resulteerde in een bestand van ongeveer 500 scholen en bijbehorende besturen die benaderd zijn met een brief en een folder. De wervingsperiode heeft gelopen van januari tot en met juni 2010 en heeft uiteindelijk geresulteerd in een groep van 94 deelnemende scholen. In 2010 zijn daar nog 8 scholen bijgekomen uit de regio Nijmegen omdat er aan deelproject 1 in 2010-2011 te weinig scholen hadden deelgenomen en dit project in de tweede helft van het kalenderjaar 2011 nogmaals is uitgevoerd. Ook voor een van de aan Streef gelieerde TIER-projecten zijn in dat jaar nog 18 scholen geworven die deelgenomen hebben aan het onderzoek.

Scholen konden bij aanmelding aangeven wat het project van hun voorkeur was en met welke bouw zij mee wilden doen. Oorspronkelijk was in het onderzoeksvoorstel opgenomen dat scholen random met een boven- (groep 5 t/m 8) of onderbouw (groep 1 t/m 4) aan een van de projecten zou worden toegewezen. Dit bleek in de praktijk niet uitvoerbaar. Elke school had een sterke voorkeur voor een van de deelprojecten en wilde niet deelnemen als zij pas later te horen zouden krijgen aan welk deelproject ze mee konden doen. De keuze voor een bepaald deelproject paste bij de huidige ontwikkeling van de school en deelname aan een ander deelproject niet. Alle scholen zijn vervolgens ingedeeld bij de deelproject van hun keuze met de bouw van hun keuze. Deze quasi-experimentele opzet heeft een belangrijk nadeel ten opzichte van de oorspronkelijke experimentele opzet; het toeschrijven van eventuele effecten

aan de interventie is problematischer. Immers, de motivatie om aan het deelproject mee te doen, en het feit dat het project inhoudelijk aansluit bij de schoolontwikkeling kunnen een belangrijke bijdrage hebben gehad aan eventuele effecten. Vasthouden aan de voorgestelde experimentele opzet zou echter betekenen dat het benodigde aantal scholen niet behaald zou worden en dat scholen mee moesten doen aan een project waarvoor zij misschien minder gemotiveerd zouden zijn. De kans op tussentijdse uitval is dan ook aanmerkelijk groter. Om toch een uitspraak te kunnen doen over of eventuele effecten veroorzaakt worden door de interventie is – gebruik makend van informatie over de scholen voorafgaand aan deelname – nagegaan in hoeverre er sprake is van validiteitsbedreigers.

Doordat elke school (in principe) deelneemt met een bouw aan een deelproject, is elke school tegelijkertijd experimentele en controleschool. Doet een school bijvoorbeeld met de bovenbouw mee aan deelproject 1 dan maakt de onderbouw van deze school deel uit van de controlegroep voor de scholen die met de onderbouw meedoen aan deelproject 2. Deze opzet heeft als voordeel dat de controlescholen waarschijnlijk wat betreft motivatie, attitude ten opzichte van het onderwijsbeleid en betrokkenheid bij het project en het onderzoek niet heel veel afwijken van de experimentele groep. De school doet immers zelf ook mee aan Streef. Tegelijkertijd is er ook een nadeel; er is altijd kans op ‘weglekken’ van de interventie. Onderbouwleerkrachten kunnen zaken oppikken van hun bovenbouwcollega’s waardoor ze niet helemaal ‘onbehandeld’ zijn. Dit wordt deels opgeheven omdat een bepaalde bouw van een school die aan deelproject A meedoet als controlegroep geldt voor andere scholen die aan deelproject B meedoen. Feit is echter wel dat de deelprojecten onderling niet volledig verschillen, allemaal hebben ze immers als doel de kwaliteit van het onderwijs te verbeteren. In de afsluitende vragenlijst aan de scholen is daarom ook een aantal vragen gesteld over de mate waarin het project bijvoorbeeld in teamvergaderingen is besproken.

Van de oorspronkelijke groep van 94 Streef-scholen hebben zich 40 scholen aangemeld voor *Gebruik maken van opbrengsten* (17 scholen in de regio Friesland/Flevoland, 16 in de regio Groningen/Drenthe en 7 in de regio Arnhem). Van een deel van de deelnemende scholen (11) hebben uiteindelijk alle teamleden meegedaan, dus niet alleen de onder- of alleen de bovenbouwleerkrachten. Dit resulteert in 21 ‘onderbouwteams’ en 28 ‘bovenbouwteams’. De scholen die alleen met de onderbouw- of allen met de bovenbouwleerkrachten hebben deelgenomen verschillen op schoolgrootte en leerlingpopulatie (uitgedrukt in het percentage leerlingen met een gewicht van 1,2) niet van de andere scholen die aan Streef meedoen. De groep scholen die met de onder- en bovenbouw meedoen is gemiddeld genomen kleiner dan de andere scholen; de deelnemende scholen hebben een leerlingpopulatie van gemiddeld bijna 100 leerlingen, de andere scholen van rond de 200 leerlingen. Deelname met het hele team is vanzelfsprekend voor een (zeer) kleine school ook voor de hand liggend.

Of de scholen ook inhoudelijk, op de voor Streef relevante thema's verschillen in de voormeting komt aan de orde in paragraaf 3.2.1, waar ook de effecten op attitude en gedrag van leerkrachten wordt beschreven. Deze vraag is in het bijzonder relevant omdat anders dan vooraf de bedoeling was de scholen niet random zijn ingedeeld maar zich zelf voor een bepaald project hebben opgegeven.

2.2 De interventie

Om een verbetering van de onderwijskwaliteit, meer specifiek een toename van het opbrengstgericht werken, en uiteindelijk een verbetering van leerlingresultaten te bewerkstelligen hebben de deelnemende scholen/leerkrachten gedurende 1 schooljaar (2010-2011) een trainings- en begeleidingstraject gevolgd. In 2.2.1. wordt eerst aangegeven op welke empirische evidentie op het gebied van effectieve interventies het trainings- en begeleidingstraject is gebaseerd. Daarna wordt in 2.2.2. een overzicht gegeven van de opzet en de inhoud van de interventie.

2.2.1 *Effectieve interventies*

Onderzoek naar effectieve interventies is met name uitgevoerd in de school improvement-traditie, interventies zijn immers altijd gericht op het verbeteren van het onderwijs. Een interventie is maar één onderdeel van een groter geheel aan activiteiten gericht op onderwijsverbetering. In het geval van opbrengstgericht werken is deelname aan de interventie maar een van de manieren waarop zij te maken krijgen met opbrengstgericht werken. De landelijke overheid probeert bijvoorbeeld ook via het Projectbureau Kwaliteit – en later School aan zet - de scholen te bereiken, de landelijke media en de vakbladen besteden aandacht aan opbrengstgericht werken en de inspectie heeft de mate van opbrengstgerichtheid opgenomen in het toezichtkader. Deze aspecten kunnen de effectiviteit van de interventie beïnvloeden, maar dat ligt buiten het bereik van de interventie. Over het algemeen wordt aan een dergelijke ketenbenadering waarin betrokkenen op verschillende niveaus meewerken een positieve uitwerking toegeschreven (zie bijvoorbeeld Fullan, 2007; Elmore, 2003). Bij de huidige interventie is het dan dus aannemelijk dat zowel op de controlescholen als de interventiescholen meer opbrengstgericht gewerkt wordt, de vooruitgang zou op de interventiescholen groter moeten zijn dan op de controlescholen.

Wat betreft de effectiviteit van de interventie kan onderscheid gemaakt worden in het voortraject en de interventie zelf. Het voortraject mondt uit in de 'beslissing' van een school of een leerkracht om aan de interventie deel te nemen. Idealiter kiest een potentiële deelnemer er zelf voor om deel te nemen omdat hij/zij een probleem ervaart waarvan hij/zij vindt dat het opgelost moet worden. Daarbij heeft hij/zij het idee daar zelf iets aan te kunnen doen en het idee dat de interventie daar iets zal bijdragen. De werving voor het trainings- en

begeleidingstraject *Gebruik maken van opbrengsten* is op deze manier uitgevoerd (hoewel dat in de oorspronkelijke experimentele opzet niet het geval was). Scholen kozen vrijwillig voor Streef en binnen Streef hebben zij zelf de keuze voor dit deelproject gemaakt en voor de bouw waarmee ze deelnamen. In enkele gevallen zijn scholen door hun bestuur (zwaar) gestimuleerd om mee te doen. Degene die de beslissing nam om deel te nemen was in de meeste gevallen de directie. Op sommige scholen is deelname met het team besproken, op andere scholen is het team ingelicht over deelname. Bij deze scholen kwam het voor dat leerkrachten niet vrijwillig deelnamen.

Wat betreft de reden om deel te nemen op schoolniveau gaven veel schoolleiders aan dat zij van mening waren dat de vaardigheden van hun leerkrachten op gebied van het analyseren van toetsen en het vervolgens gedifferentieerd lesgeven verbeterd kon worden evenals de leerlingresultaten (in de breedte of op bepaalde deelgebieden). Ook veel leerkrachten gaven aan dit noodzakelijk te vinden en hoewel men over het algemeen wel van mening was daar zelf een rol in te spelen, stuitte het daadwerkelijk tijd besteden aan het analyseren van toetsen en het aanpassen van het eigen gedrag in de groep bij enkele leerkrachten wel op weerstand. Meestal lag het argument dan in de benodigde tijd, een moeilijke klas of een onderwijsopvatting waarin toetsen van (jonge) leerlingen niet past.

Wat betreft de opzet en de inhoud van de interventie levert onderzoek (bijvoorbeeld Verloop, 2003) op verschillende aspecten aanwijzingen voor belangrijke kenmerken:

- Aansluiting bij eigen ervaring en praktijk
- Aandacht voor weerstand, moeilijkheden
- Combineren en verbinden van kennis, vaardigheden en attitude
- Gebruik van activerende werkvormen
- Tijd en ruimte voor 'uitproberen'
- Begeleid uitproberen in de eigen praktijk en nabespreking
- Uitwisseling met collega's
- Inbedding in team/school, sturing door de directie
- Aandacht voor borging
- De duur van de interventie
- Deskundigheid van de trainer en begeleider

Aan het eind van de volgende paragraaf, waarin de opzet en de inhoud van de interventie wordt beschreven wordt op deze aspecten teruggekomen en wordt duidelijk gemaakt in welke mate de interventie voldoet aan de kenmerken van een effectieve interventie.

2.2.2 Opzet en inhoud van de interventie

Op basis van een chronologische beschrijving van de bijeenkomsten wordt hieronder de opzet en de inhoud van de interventie beschreven. Van de bovenschoolse bijeenkomsten zijn er per regio steeds 2 georganiseerd; een voor onderbouwleerkrachten en een voor bovenbouwleerkrachten.

Omdat Streef een samenwerking betrof tussen de Rijksuniversiteit Groningen en het Cito is het Cito-LOVS in het traject leidend geweest. Bovendien biedt dit systeem voor het analyseren van toetsresultaten meer mogelijkheden dan de andere systemen die van oorsprong schooladministratiesystemen zijn. In de praktijk bleek het traject ook wel bruikbaar voor scholen die een ander systeem gebruiken, maar enkele opdrachten zouden daarop beter aangepast kunnen worden.

1^e bijeenkomst (bovenschools, eind september)

De eerste bijeenkomst is een grotendeels plenaire bijeenkomst waarin een toelichting wordt gegeven op het begrip opbrengstgericht werken en op het traject *Gebruik maken van opbrengsten*. Hierin wordt gebruik gemaakt van een filmpje van de CD 'Kijken naar opbrengsten' van de PO-Raad waarop mensen uit het onderwijsveld vertellen over hoe ze het opbrengstgericht werken in hun school handen en voeten geven en wat de effecten daarvan zijn. Daarnaast is er aandacht voor het landelijk beleid ten aanzien van opbrengst gericht werken en de positie van de inspectie daarin. Er worden resultaten gepresenteerd van onderzoek onder alle leerkrachten van de Streef-scholen uit het schooljaar 2010-2011 waarin hen een aantal vragen is gesteld die te maken hebben met opbrengstgericht werken. Benadrukt wordt tevens dat opbrengstgericht werken, anders dan veel andere onderwijsvernieuwingen, een basis heeft in wetenschappelijk onderzoek en theorievorming over 'wat werkt in het onderwijs'. In dit onderzoek blijkt al jaren dat nadruk op basisvaardigheden, het hebben van hoge verwachtingen en systematisch evalueren factoren zijn die samenhangen met goede leerlingresultaten.

In het tweede deel van de bijeenkomst wordt een begin gemaakt met het presenteren van de analysemogelijkheden van het computerprogramma van het Cito-LOVS. Voor beide onderdelen is een powerpoint presentatie beschikbaar. Op de CD staat tevens een document met een overzicht van alle beschikbare rapportage- en analysemogelijkheden. Bij bijeenkomst 3 en 5 zijn aanvullend overzichten opgenomen van de in die bijeenkomst behandelde analyses. Als huiswerkopdracht voor de 2^e bijeenkomst op de eigen school maken de leerkrachten een analyse van de resultaten van de leerlingen uit hun eigen groep op basis van de toetsen aan het eind van het voorgaande schooljaar. Ze maken hiervoor gebruik van het stappenplan (zie CD) dat in deze 1^e bijeenkomst is geïntroduceerd. Dit stappenplan is door de deelnemende onderwijsadviesorganisaties omgezet in een analyseformulier waarin de leerkrachten stapsgewijs de benodigde gegevens in kunnen vullen. Er is een leeg exemplaar en twee ingevulde exemplaren bijgevoegd, een voor de onderbouw en een voor de bovenbouw (de

onderbouw-versie kent ook nog een bijlage met een aantal rapportages uit het Cito-LOVS). Er wordt de scholen geadviseerd als team een vakgebied te kiezen waarvoor men de analyses helemaal uitvoert. Op deze manier kan voorkomen worden dat de opdracht te omvangrijk wordt en kan bewerkstelligd worden dat de opdracht goed aansluit bij de actuele schoolontwikkeling. Is men bijvoorbeeld bezig met het rekenonderwijs, dan draagt het uitvoeren van de analyses voor het vakgebied rekenen bij aan die ontwikkeling en wordt het huiswerk minder als een belasting ervaren. Nadeel is dat er in het traject in de bovenschoolse bijeenkomsten bijna geen aandacht is voor vakinhouden. In de bijeenkomsten op schoolniveau is die gelegenheid er wel, maar zal het afgestemd moeten worden met elke individuele school. Bij de bovenschoolse bijeenkomsten kan de trainer wel materiaal meenemen, bijvoorbeeld vanuit de taal- en rekenpilots om leerkrachten attent te maken op het bestaan ervan en om hen op die manier te stimuleren hiervan gebruik te maken.

Opdracht 3 (eveneens een huiswerkopdracht) is een oproep om vragen, opmerkingen t.a.v. het Cito-LOVS-computerprogramma, de toetsen, opbrengstgericht werken e.d. door het geven zodat daar in volgende bijeenkomsten aandacht aan besteed kan worden.

2^e bijeenkomst (eigen school)

Tijdens deze bijeenkomst is de huiswerkopdracht op de eigen school besproken. Op de CD staat een voorbeeld agenda en een document waarin een aantal suggesties zijn opgenomen voor de aanpak en inhoud van deze bijeenkomst. In de praktijk bleek de vaardigheid van de leerkrachten (zelfs binnen de scholen) ver uiteen te lopen. De bijeenkomsten op de eigen school zullen dus niet op elke school dezelfde invulling hebben. Op sommige scholen zullen de leerkrachten zeer basale vragen hebben met betrekking tot het computersysteem (op sommige scholen zal het zelfs noodzakelijk zijn om de analyse ter plekke met behulp van het computerprogramma van het Cito-LOVS te maken), terwijl bij andere scholen veel meer de vertaling naar een groepsplan (versie van Cedin toegevoegd op de CD) en de praktijk in de groep centraal zal staan. Essentie van deze 2^e bijeenkomst is om de basis analysemogelijkheden van het Cito-LOVS te kunnen gebruiken.

3^e bijeenkomst (bovenschools, november)

Deze bijeenkomst start met een plenair gedeelte waarin op basis van een uitleg over de inhoud en structuur van de Cito-toetsen aandacht is voor meer geavanceerde analysemogelijkheden, zoals het alternatieve profiel, foutenanalyse en categorieënanalyse, en voor overwegingen t.a.v. de keuze en afname van de toetsen. De uitleg over de inhoud en de structuur van de Cito-toetsen begint met opdracht 4 waarin leerkrachten zelf een aantal toetsitems op volgorde moeten leggen van moeilijkheid (document met toetsitems staat op CD, onder bijeenkomst 3), vervolgens wordt gepresenteerd hoe de moeilijkheid van de items door leerlingen wordt ervaren en welke plaats de items in de toetsen hebben. Op deze manier worden leerkrachten zich ervan bewust dat toetsen ‘overlappen’, dat er een vaardigheidsschaal onder ligt en dat het

soms dus nuttiger is om een leerling een makkelijkere of een moeilijker toets voor te leggen. Bovendien wordt benadrukt dat je om goed te analyseren waar leerlingen op achterblijven inzicht moet hebben in de inhoud van de toets en de inhoud van de leerstof. Vervolgens gaan de leerkrachten zelf aan de slag met het computerprogramma. Zij kunnen daarbij kiezen voor welke groep (2, 4, 5 of 7) zij opdracht 5 gaan maken waarin achtereenvolgens de alternatieve profielen, foutenanalyse en categorieënanalyse aan de orde komen. Voor de snelle leerkrachten is er nog een extra opdracht waarin subgroepen aangemaakt worden en waarin een verantwoorde toetskeuze gemaakt moet worden voor een spellingstoets op basis van resultaten van een eerdere toets en een overzicht van de inhoud van de spellingstoetsen. Voor het maken van opdracht 5 staan op de CD verschillende aanvullende documenten: antwoorden van toetsen voor de verschillende groepen die ingevoerd moeten worden om de analyses te kunnen uitvoeren en het overzicht van spellingcategorieën voor de extra opdracht. De huiswerkopdracht voor bijeenkomst 4 is om opnieuw een analyse te maken van toetsresultaten, nu gebruik makend van de afname medio schooljaar.

4^e bijeenkomst (eigen school)

Deze bijeenkomst kent eenzelfde inhoud als bijeenkomst 2, alleen is de toetsafname nu actueel. Essentieel in de bespreking is nu de vertaling van de analyse naar de praktijk in de groep. Ook nu zal het niveau erg uiteenlopen. Sommige scholen zullen al werken met groepsplannen, daarvoor de analyses gebruiken en doelstellingen formuleren voor de volgende toetsafname, op sommige scholen zal het ‘technische’ aspect van het maken van een analyse nog steeds de grootste nadruk hebben. Met name op deze laatste scholen is het nuttig om tijdens de bespreking toegang te hebben tot het computerprogramma, een beamer/whiteboard en de eigen data.

5^e bijeenkomst (bovenschools, maart)

In deze bijeenkomst gaan de leerkrachten eerst in gemengde groepen aan de slag met opdracht 7. M.b.v. deze opdracht vindt er een uitwisseling plaats van ‘good practices’ op het gebied van opbrengst gericht werken, analyseren en in het bijzonder differentiatie in de groep. Het goed analyseren van leerlingresultaten zal (bijna) altijd leiden tot de conclusie dat er gedifferentieerd moet worden. Differentiëren vraagt veel van leerkrachten: er moeten keuzes gemaakt worden voor een nadruk op convergent of op divergent differentiëren (laat ik de leerlingen steeds verder uit elkaar lopen of houd ik ze zoveel mogelijk bij elkaar), keuze voor werkvormen en leerstofaanbod passend bij de doelstelling en goed klassenmanagement. Leerkrachten worden vooraf gestimuleerd om elkaar kritisch te bevragen, alleen dan kan deze werkvorm iedereen iets opleveren dat hij/zij mee naar ‘huis’ neemt.

Een van de vragen binnen opdracht 7 betreft het bekijken en becommentariëren van ingevulde analyseformulieren. De bijgevoegde formulieren verschillen heel erg in kwaliteit (zowel in mate waarin ze compleet zijn als diepgang van analyse en keuze van vervolgstappen) en staan

op de CD. De bedoeling is dat de leerkrachten een lijstje maken met kenmerken waaraan een goede analyse moet voldoen. In dit geval zijn het dus analyses van anonieme andere leerkrachten, die geen profijt hebben van de feedback. Een alternatief is om leerkrachten (van verschillende scholen) elkaar te laten bevragen op de analyse zodat ze feedback van elkaar krijgen en gedwongen worden om over hun eigen analyse na te denken.

In het tweede deel van de bijeenkomst worden de laatste analysemogelijkheden (van de module zelfevaluatie) toegelicht en voorgedaan.

6^e en 7^e bijeenkomst (eigen school)

Aansluitend bij de inhoud van bijeenkomst 5 voert de cursusleider op elke school een aantal observaties in de groep uit. De nadruk ligt hier op differentiatie. De observaties worden individueel/in het team nabesproken. Observatieformulieren zijn opgenomen op de CD.

8^e bijeenkomst (bovenschools, mei/juni)

In de eerste helft van de bijeenkomst wordt met behulp van een aantal vragen/stellingen als terugblik op het traject nog eenmaal de cyclus van opbrengstgericht werken doorlopen. Elke leerkracht krijgt een stemkastje waarmee hij/zij aan kan geven hoe hij/zij denkt en handelt. De uitkomsten van de ‘stemming’ worden besproken en vergeleken met de situatie voordat we met Streef begonnen. Er is ruimte voor discussie en het geeft leerkrachten inzicht in hoe ‘ver’ zij zelf zijn en hoe ‘ver’ hun school is in vergelijking met andere leerkrachten en scholen en in vergelijking met een jaar geleden.

Daarna wordt aandacht besteed aan een aantal discussiepunten/vragen/opmerkingen die elke keer weer opkomen als het over dit thema gaat. Dit wordt gedaan aan de hand van een aantal stellingen. Op basis van de meningen, die we opnieuw met behulp van de stemkastjes meten, worden 1 of 2 stellingen geselecteerd die uitgebreider aan de orde komen. Dat zijn stellingen waar de meningen over verdeeld zijn. Over deze stellingen wordt een debat (zoals in het Lagerhuis) georganiseerd. Leerkrachten bereiden zich in groepjes voor op het debat door hun argumenten voor of tegen met elkaar te bespreken, vooraf te bedenken welke tegenargumenten de andere ‘partij’ zal gebruiken en een strategie voor het debat te bepalen (2 documenten met tips voor het voeren van een debat zijn op de CD opgenomen). Tot slot wordt het debat gevoerd door +/- 3 voor- en 3 tegenstanders. Na het debat kan het ‘publiek’ nog reageren of kan opnieuw gestemd worden.

Het debat is geen doel maar een werkvorm om de deelnemers te stimuleren na te denken over waarom ze bepaalde keuzes maken. Het levert dus niet directe concrete handreikingen op, maar is meer gericht op het stimuleren van reflectie.

9^e bijeenkomst (eigen school)

Streef sluit af met een bijeenkomst op de eigen school waarin nu de toetsresultaten van het eind van het schooljaar worden besproken: zijn de resultaten behaald die de leerkrachten zich tot doel hadden gesteld? Tevens is er aandacht voor de borging van deze manier van werken: hoe kun je er als school voor zorgen dat deze manier in de volgende schooljaren nog wordt verbeterd en/of volgehouden wordt?

Effectieve kenmerken van *Gebruik maken van opbrengsten*

In paragraaf 2.2.1 is een aantal kenmerken genoemd van effectieve interventies, waarmee bij de ontwikkeling van *Gebruik maken van opbrengsten* rekening is gehouden.

Om effectief te kunnen zijn moet er door de deelnemers voldoende tijd aan de interventie besteed worden. De totale begeleidende tijd is minimaal 20 uur ($4 * 2,5$ uur + $5 * 2$ uur), voor het maken van de opdrachten is ook minimaal 20 uur nodig. Gezien de reacties van de deelnemers is die tijd door het overgrote deel ook daadwerkelijk aan de interventie besteed.

Door de afwisseling van bovenschoolse en schoolgebonden bijeenkomsten, die respectievelijk meer op training en meer op begeleiding waren gericht, is gezorgd voor een afwisseling van kennis en vaardigheden en voor een afwisseling van algemene en schoolspecifieke invulling. Attitude-aspecten, inclusief weerstanden, zijn zowel bovenschools als schoolgebonden besproken in discussieopdrachten waarbij zowel gewerkt is in groepjes van leerkrachten van verschillende scholen en verschillende jaargroepen als in teamverband.

Voor de uitvoering is samenwerking aangegaan met 2 onderwijsadviesorganisaties, waarbij overeengekomen is dat zij personeel zou inzetten dat inhoudelijk kennis had van de verschillende aspecten van opbrengstgericht werken en praktische vaardigheden bezat om leerkracht individueel en teams daarbij te begeleiden. Deze begeleiders zijn vooraf getraind en bovendien bij elke bovenschoolse bijeenkomst aanwezig geweest om er zorg voor te dragen dat zij goed op de hoogte waren van wat daar aan de orde is geweest.

De deelnemers hebben zowel wat betreft het analyseren als wat betreft het gedifferentieerd werken in de groep gelegenheid gehad om te oefenen en daarop feedback gekregen. Analyse-rapporten zijn besproken evenals de observatie in de groep.

Samenvattend kan gesteld worden dat *Gebruik maken van opbrengsten* in opzet en inhoud voldoet aan de meeste kenmerken van een effectieve interventie. Minder sterke punten zijn de mate waarin er aandacht is besteed aan de inbedding in het team, de sturing door de directie en de borging in de toekomst. De inbedding in het team is vanwege het onderzoek niet in de interventie aangezet. Het onderzoeksdesign is immers zo opgezet dat de niet-deelnemende bouw deel uitmaakt van de controlegroep voor andere scholen. Wanneer de verworvenheden van de interventie meteen breed gedeeld zouden worden, verliest het design zijn vergelijkingsmogelijkheden.

Een ander aspect waarop de doelstelling van de interventie soms in strijd was met de eisen die vanwege het onderzoeken van de interventie zijn gesteld is enerzijds de noodzaak om zoveel mogelijk aan te sluiten bij de individuele ontwikkeling van scholen en leerkrachten en anderzijds de noodzaak om een zo uniform mogelijke interventie uit te voeren. Is de interventie op elke school anders dan is het heel moeilijk om succes of falen ergens aan toe te schrijven. Ook het feit dat de begeleiding op de scholen door 5 verschillende begeleiders is uitgevoerd (waarbij er ook nog een wisseling heeft plaatsgevonden) zorgt er al voor dat er variatie in de interventie is geweest. Zoals gezegd is aansluiten bij de schoolsituatie ook noodzakelijk om iets te kunnen bereiken op een school. Zoals uit hoofdstuk 4 zal blijken was de beginsituatie van de scholen zeer verschillend en waren dus ook de punten die op de bijeenkomsten op de scholen zijn besproken, hoewel vaak qua onderwerp gelijk, heel anders qua niveau. Waar de ene leerkracht, als het gaat om het analyseren van de vooruitgang van een leerling, nog moet leren hoe je de lijn moet interpreteren, vraagt de andere leerkracht zich af of de afgebeelde vooruitgang normaal is of afwijkend.

2.3 Dataverzameling bij leerkrachten en leerlingen

2.3.1 *Leerkrachtvragenlijst*

Om de beginsituatie en de veranderingen in leerkrachtgedrag en attitude in kaart te brengen is in het voorjaar van 2010, voorafgaand aan de verschillende deelprojecten binnen Streef en in het voorjaar van 2012, na afloop van de deelprojecten een digitale vragenlijst bij alle leerkrachten op de deelnemende scholen afgenomen. De selectie en de ontwikkeling van de vragen en de resultaten op de voormeting in 2010 zijn beschreven in Deunk, Doolaard & Hofman (2011). De vragenlijst bevat vragen op alle terreinen die in de deelprojecten aan de orde komen: referentieniveaus, professionalisering op het gebied van opbrengstgericht werken en differentiatie en excellentie. In de bijlage is een overzicht opgenomen van de variabele die voor dit deelproject van belang zijn. Daarin is ook kort aangegeven wat de inhoud van de schalen en indexen is.

2.3.2 *Leerlingtoetsen en –achtergrondgegevens*

Bij de leerlingen van alle deelnemende scholen worden de volgende gegevens uit het Cito-leerlingvolgsysteem verzameld:

- Groep 1 & 2: Ordenen/Rekenen voor Kleuters, Taal voor Kleuters
- Groep 3, 4 & 5: Technisch lezen (Leestempo), Begrijpend lezen (groep 4 en 5), Spelling, Rekenen-wiskunde
- Groep 6, 7 en 8: Spelling, Begrijpend lezen, Rekenen-wiskunde en in groep 8 de Cito-eindtoets

Van leerlingen die doubleren worden de gegevens uit het Cito-leerlingvolgsysteem gewoon verzameld. De scores van doubleurs en onvertraagden kunnen immers op dezelfde schaal worden afgebeeld.

De scholen nemen de toetsen in eigen beheer volgens de handleiding af, voeren de gegevens in in het digitale Cito-LOVS en verzenden deze gegevens naar het Cito. Scholen die dit systeem niet gebruiken sturen een Excel-bestand of een papieren uitdraai op.

Van de leerlingen worden daarnaast de volgende achtergrondgegevens verzameld: geslacht, etniciteit (geboorteland van de ouders), leerlinggewicht en leeftijd.

Omdat veel scholen de achtergrondgegevens (behalve leeftijd en geslacht) vaak niet in het Cito-LOVS hebben opgenomen zijn in het voorjaar van 2012 deze gegevens via de post/email bij de scholen verzameld, tegelijkertijd met het afnemen van de leerkrachtvragenlijst.

2.3.3 Analyse van de leerkrachtvragenlijst en de leerlingresultaten

Voor het vaststellen van de effecten van deelname aan *Gebruik maken van opbrengsten* op de attitude en het gedrag van leerkrachten en op de resultaten van leerlingen wordt er uit de totale deelnemersgroep van Streef een experimentele en een controlegroep geformeerd. De experimentele groep is voor wat betreft de analyse van de leerkrachtresultaten vanzelfsprekend de groep leerkrachten die aan deelproject 2 heeft deelgenomen. De controlegroep bestaat uit leerkrachten die aan geen enkel project hebben meegedaan of aan een project dat inhoudelijk geen gelijkenis vertoont met *Gebruik maken van opbrengsten*. De controlegroep zou immers zo ‘onbehandeld’ mogelijk moeten zijn. Dat betekent dat er ook leerkrachten zijn die helemaal buiten de analyses vallen omdat zij noch tot de experimentele groep gerekend worden, noch tot de controlegroep.

Voor wat betreft de resultaten van de onderbouw en de bovenbouwleerlingen is een gelijksoortige procedure gevolgd. In tabel 2.1 is weergegeven hoeveel scholen en welke scholen leerlingen leveren voor de analyse van het effect op onderbouw- en op bovenbouwleerlingen.

De meer inventariserende vragen over attitude en handelen van leerkrachten te beantwoorden zijn eenvoudige frequentiestaten, gemiddelden e.d. berekend, waar relevant uitgesplitst naar kenmerken van de scholen zoals deelname aan een bepaald deelproject. Om veranderingen tussen 2010 en 2012 te onderzoeken zijn gepaarde t-toetsen worden uitgevoerd. Om verschillen tussen ‘experimentele’ en controlescholen in veranderingen tussen 2010 en 2012 te onderzoeken is getoetst of de verschilcores voor beide groepen verschillen.

Voor het onderzoeken van de relatie tussen leerkracht- en schoolkenmerken en leerlingvorderingen zijn clusters scholen gevormd die verschillen in hun manier van (opbrengstgericht) werken en is onderzoek welk van deze clusters samenhangt met de beste leerlingprestaties.

Tabel 2.1 Selectie van ‘experimentele’ en controlescholen

	Scholen	Overlap	Controlegroep	
A	Deelproject 1 (referentieniveaus Cito) 2010 alleen onderbouw	2	1 school (252) heeft met de bovenbouw met project 2 (F) meegedaan	
B	Deelproject 1 (referentieniveaus Cito) 2010 alleen bovenbouw	3	1 school (220) heeft ook met project 3 (J) meegedaan	
C	Deelproject 1 (referentieniveaus Cito) 2010 onder- en bovenbouw	8	-	
D	Deelproject 1 (referentieniveaus Cito) 2011	8	-	
E	Deelproject 2 alleen onderbouw	10	1 school (244) heeft ook met project 4 (K) meegedaan	OB met B, K + F + J
F	Deelproject 2 alleen bovenbouw	17	1 school (252) heeft met de onderbouw met project 1 (A) meegedaan	BB met A + E + J
G	Deelproject 2 onder- en bovenbouw	11	1 school (231) heeft ook met project 3 (J) meegedaan	OB met B,K + F + J BB met A + E + J
	Deelproject 2 OB	21		OB met B, K (3+18-4)→N=17 + F (17-school 252)→N=33 + J (9-school 231)→N= 41
	Deelproject 2 BB	28		BB met A (2-school 252)→N=1 + E (10-school 244)→N=10 + J (9-school 231)→N= 18
H	Deelproject 2 MB 2010 (TIER)	6	2 scholen (259, 263) hebben ook aan project 4 (K) meegedaan	
I	Deelproject 2 MB 2011 (TIER)	18	1 school (277) heeft ook aan project 4 (K) meegedaan	
J	Deelproject 3 (Excellentie)	9	Zie B en G	
K	Deelproject 4 (Leren lezen) (TIER)	18	4 scholen (244, 259, 263, 277) hebben ook aan andere projecten meegedaan (E, H, I)	
	Totaal aantal deelnemende scholen	110-7 ‘dubbele’ scholen =103		
L	Geen deelname (wel aangemeld maar nooit begonnen of heel snel afgevallen)	17 (14%)		
	Totaal aantal aangemelde scholen	120		

Voor het nagaan van de effecten van – met name - deelproject 2 is per leerstofdomein, per cohort een multilevelanalyse uitgevoerd met de voortoets als covariaat. Elke analyse levert een effectgrootte van deelname van de bouw op. Deze effectgroottes zijn vervolgens samengevoegd in een meta-analyse per leerstofdomein. Dit is zowel uitgevoerd voor de directe effecten (in het jaar van deelname) als de retentie-effecten (een jaar na deelname).

2.3.4 Beoordeling van de analyse-rapportages

In 2.2 is beschreven dat de deelnemende leerkrachten 3 keer in de looptijd van de interventie is gevraagd een analyse-rapportage te maken en in te leveren, eenmaal aan het begin van de interventie, eenmaal na de afname van de Cito-LOVS-toetsen medio 2010-2011 en eenmaal na de afname van de Cito-LOVS-toetsen aan het eind van het schooljaar. Geen enkele leerkracht heeft daadwerkelijk driemaal een analyse ingeleverd en slechts enkele leerkrachten hebben dat tweemaal gedaan zodat het niet mogelijk was om individuele ‘groei’ te analyseren. Wel was het mogelijk om de analyses die aan het begin van de interventie zijn ingeleverd te vergelijken met analyses die tijdens het tweede of het derde inlevermoment zijn ingeleverd.

Tabel 2.2 Aantal ingeleverde analyses van de leerkrachten

	Meetmoment 1 September 2010	Meetmoment 2 Februari 2011	Meetmoment 3 Juli 2011	Totaal
Leerkrachten die een analyse hebben ingeleverd	57	30	3	92
Leerkrachten die twee meetmomenten een analyse hebben ingeleverd	3	2	1	3
Leerkrachten die drie meetmomenten een analyse hebben ingeleverd	0	0	0	0

Deze analyses zijn beoordeeld met behulp van een beoordelingsformulier. Bij het ontwikkelen van het beoordelingsformulier is aangesloten bij het stappenplan dat in de interventie wordt aangeboden en kent drie categorieën (groepsniveau, leerlingniveau, oorzaken en gevolgen) en een totaalscore. Binnen de categorieën kan voor de verschillende beoordelingspunten 0 ‘niet beantwoord’, 1 ‘beantwoord’ of – indien relevant - 2 punten ‘beantwoord en geanalyseerd’ worden behaald. Per categorie wordt het cijfer omgezet naar een cijfer tussen de 0 en 10, waarna de totaalscore het gemiddelde van de 3 categorieën is. Alle analyses van het eerste inlevermoment zijn door twee beoordelaars beoordeeld waarbij een Cohen’s Kappa werd behaald van 0,63 op de totaalscore. Op grond van dit resultaat zijn de volgende analyses door een beoordelaar beoordeeld.

3 Resultaten

3.1 Uitvoering en implementatie van de interventie

De interventie *Gebruik maken van opbrengsten* is uitgevoerd in de periode september 2010 tot juli 2011 in drie regio's. Voor de regio Groningen zijn de bovenschoolse bijeenkomsten gehouden in Eelde, voor de regio Friesland in Heerenveen en voor de regio Arnhem/Nijmegen in Arnhem. De bovenbouwbijsamenkomst in Friesland/Flevoland was het grootst met ruim 70 deelnemers, de onderbouwbijsamenkomst in Arnhem het kleinst met 20 deelnemers. Niet iedereen was elke bijsamenkomst aanwezig, op de bijsamenkomsten op de eigen school waren wel alle deelnemende teamleden aanwezig, bij de bovenschoolse bijsamenkomsten – waarvoor sommige deelnemers ruim een uur moesten rijden – was elke keer 60 tot 80% van de deelnemers aanwezig.

De bijsamenkomsten op de scholen, die geleid werden door een medewerker van de onderwijsadviesorganisatie, werden steeds bezocht door bijna alle leerkrachten van de bouw die aan het project deelnam. Vanzelfsprekend is hiervoor een praktische reden maar er zijn ook inhoudelijke redenen gegeven voor de hogere deelname aan de schoolbijsamenkomsten. Op de schoolbijsamenkomsten stonden altijd de eigen leerlingen centraal; leerkrachten moesten voorafgaand aan de bijsamenkomst bijvoorbeeld zelf een analyse maken die vervolgens besproken werd, de bevindingen van de klassenobservaties op de eigen school werden besproken of de mogelijkheden om in de eigen klas te differentiëren. Ook een belangrijke reden waarom de bijsamenkomsten op de eigen school hoger gewaardeerd en beter bezocht werden was de mogelijkheid om aan te sluiten bij de kennis en ervaring die leerkrachten al hadden in het analyseren van leerlingresultaten. Die kennis en ervaring liep zeer sterk uiteen; sommige leerkrachten hadden zelf nog nooit een analyse gemaakt of zelf kennis gemaakt met het Cito-LOVS, waar in de interventie mee gewerkt werd. Door middel van gedifferentieerde 'hands-on' opdrachten is wel geprobeerd ook in de bovenschoolse bijsamenkomsten gedifferentieerd te werken, en de 'voorlopers' te gebruiken om leerkrachten die minder vaardig waren, meer bezwaren hadden en negatiever stonden ten opzichte van opbrengstgericht werken, een stap in de goede richting te krijgen, maar in de bijsamenkomsten op de eigen school lukt dat vanzelfsprekend beter dan bovenschools. De ongelijke beginsituatie zorgde er ook voor dat de ervaren taakbelasting voor de ongeoefende leerkrachten aanzienlijk hoger leek te liggen dan voor de leerkrachten die de huiswerkopdrachten, zoals het maken van een analyse, al uitvoerden in het kader van de reguliere cyclus op schoolniveau.

3.2 Effecten op leerkrachtattitude, –gedrag en analysevaardigheden

Om te bepalen welk effect deelname aan *Gebruik maken van opbrengsten* heeft gehad op leerkrachten is alle leerkrachten van alle scholen die aan Streef hebben deelgenomen

tweemaal een digitale vragenlijst voorgelegd; een keer voor de start van Streef en een keer na afloop van alle deelprojecten. Door de antwoorden van de leerkrachten die aan *Gebruik maken van opbrengsten* te vergelijken met leerkrachten die aan andere en/of aan geen enkel deelproject hebben deelgenomen kan het effect van deelname aan dit specifieke deelproject worden bepaald. In 3.2.1 worden de resultaten hiervan beschreven. Naast deze effectbepaling is als afstudeeronderzoek door een studente van de master Onderwijskunde een analyse gemaakt van de analyses die door de leerkrachten zijn gemaakt en ingeleverd als huiswerkopdracht gedurende het project. Hiervan wordt verslag gedaan in 3.2.2.

3.2.1 Effect op leerkrachtattitude- en gedrag

In 2010 vulden 793 respondenten (79%) de vragenlijst voor de voormeting in, in 2011 vulden nog eens 79 later geworven respondenten de voormeting in. In 2012 vulden 634 respondenten (55%) de nameting in, 447 respondenten vulden de lijst tweemaal in en in het totaal hebben 1059 respondenten een- of tweemaal de vragenlijst ingevuld. Bij de voormeting in 2011 en bij de nameting is onderscheid gemaakt tussen respondenten die directeur of intern begeleider zijn en respondenten die groepsleerkracht zijn of directeur/intern begeleider en structureel minimaal 4 dagdelen voor de groep staan. Deze laatste groep wordt hierna aangeduid als leerkracht. In deze rapportage zijn alleen de respondenten die de leerkrachtversie in 2010 en/of 2012 hebben ingevuld (N=884).

Van de respondenten die in 2012 hebben deelgenomen en van de respondenten die de vragenlijst tweemaal hebben ingevuld heeft ongeveer de helft actief deelgenomen aan een van de deelprojecten van Streef, de andere helft heeft niet actief deelgenomen maar werkt op een school waarvan andere leerkrachten wel hebben deelgenomen. Om het effect van deelname aan *Gebruik maken van opbrengsten* na te gaan zijn de leerkrachten opgedeeld in 3 groepen: een 'experimentele' groep met actieve deelnemers waarvan verwacht kan worden dat deelname hun attitude en gedrag heeft kunnen veranderen, een controlegroep van leerkrachten die niet aan dit deelproject hebben deelgenomen en niet actief aan een van de andere deelprojecten en een groep die buiten de analyse wordt gehouden omdat zij wel actief meegedaan hebben aan een deelproject dat inhoudelijk overeenkomsten vertoont met *Gebruik maken van opbrengsten*. Bij deze groep zou eenzelfde verandering in attitude en gedrag kunnen optreden die niet toe te schrijven is aan *Gebruik maken van opbrengsten*. Het betreft hier bijvoorbeeld de leerkrachten die hebben deelgenomen aan het deelproject over referentieniveaus.

In Tabel 3.1 is een overzicht gegeven van de aantallen leerkrachten in de verschillende groepen.

Tabel 3.1 Aantal leerkrachten in ‘experimentele’ en controlegroep

	Alleen voormeting	Alleen nameting	Voor- en nameting	Totaal
‘Experimentele groep’	0	15	72	87
Controlegroep	176	62	165	403
Verwijderd	239	47	108	394
Totaal	415	124	345	884

In de vragenlijst zijn vragen opgenomen die betrekking hebben op alle onderwerpen die in Streef aan de orde komen: referentieniveaus, opbrengstgericht werken en excellentie. In Deunk e.a (2011) is een verantwoording gegeven van de achtergrond, inhoud, betrouwbaarheid en validiteit van de vragenlijst en de resultaten op de voormeting in 2010 over alle leerkrachten en scholen heen.

In Tabel 3.2 (zie de bijlage voor een korte toelichting op de inhoud van de schalen en indexen) is weergegeven wat de resultaten van de experimentele en controlegroep zijn in 2010 en 2012 en hoe groot de groei, dan wel daling is, tussen 2010 en 2012.

Tabel 3.2 Uitkomsten op de schalen en indexen

	2010		2012		2012-2010		Verskil tussen exp en contr in verschilscore significant?
	Exp	Contr	Exp	Contr	Exp	Contr	
Samenwerken	2,56	2,50	2,40	2,46	-0,02	-0,05	Nee
Beroepsattitude ^{3,4}	4,20	4,19	4,39	4,40	0,22	0,18	Nee
Metacognitie	3,84	3,98	3,83	3,97	-0,02	0,00	Nee
Prestatie-informatie ^{3,4}	3,99	3,96	4,16	4,09	0,14	0,11	Nee
Explicitering doelen ²	2,17	2,01	2,47	2,16	0,29	0,07	Nee
Attitude opbrengstgericht werken ^{2,3}	2,71	2,54	2,96	2,71	0,31	0,05	Nee
Breed kader ^{2,3}	2,24	2,23	2,86	2,47	0,67	0,25	Nee
Reflectie ^{2,3}	2,41	2,46	2,68	2,55	0,22	0,04	Ja
Basisvaardigheden ³	3,51	3,47	3,61	3,58	0,20	0,04	Nee
Rapport ^{1,2,3}	3,41	2,99	3,79	3,38	0,33	0,22	Nee
Dif_niveau	1,59	1,56	1,82	1,68	0,25	0,15	Nee
Dif_actie	1,84	2,02	1,94	1,95	0,19	0,03	Nee
Dif_attitude	1,86	1,87	1,86	1,95	0,06	0,01	Nee

1=in 2010 is er een verschil tussen de exp en de controlegroep

2=in 2012 is er een verschil tussen de exp en de controlegroep

3=het verschil tussen 2010 en 2012 is significant voor de experimentele groep

4=het verschil tussen 2010 en 2012 is significant voor de controlegroep

Daarbij is ook steeds aangegeven of de experimentele groep en de controlegroep verschilden in 2010, of ze verschilden in 2012 of de experimentele groep vooruit of achteruit is gegaan tussen 2010 en 2012, of de controlegroep vooruit of achteruit is gegaan tussen 2010 en 2012

en ten slotte of het verschil tussen 2010 en 2012 anders was voor de experimentele groep dan voor de controlegroep, met andere woorden, of een groei of daling in de experimentele groep wel of niet ook bij de controlegroep zichtbaar is en dus niet toegeschreven kan worden aan deelname aan *Gebruik maken van opbrengsten*.

Omdat de waarden in de kolom 2010 en 2012 steeds betrekking hebben op alle leerkrachten die in dat jaar de vragenlijst hebben ingevuld en de waarden in de kolom 'verschil' alleen op de leerkrachten die de vragenlijst in 2010 en 2012 hebben ingevuld is de verschillscore niet exact het verschil tussen de waarde in de kolom 2010 en de waarde in de kolom 2012.

Op slechts een van deze schalen en indexen is te zien dat de deelnemers aan *Gebruik maken van opbrengsten* een significant andere ontwikkeling doormaken tussen 2010 en 2012 dan de leerkrachten die niet deelgenomen hebben. Het betreft de schaal reflectie. Deze schaal wordt in Tabel 3.4 verder uitgewerkt. Andere opvallende variabelen zijn:

- Attitude opbrengstgericht werken en Breed kader, waarop de experimentele groep niet verschilt van de controlegroep in 2010 maar wel in 2012. De verandering (=voortgang) binnen de experimentele groep tussen 2010 en 2012 is significant maar niet groter of kleiner dan in de controlegroep. Hoewel van deze variabelen dus niet te zeggen is dat de deelnemers meer voortgaan dan de niet-deelnemers, is er bij de deelnemers wel sprake van een voortgang die we bij de controlegroep niet zien en is het eindresultaat dat de leerkrachten in de experimentele groep in 2012 positiever staan ten opzichte van opbrengstgericht werken dan de controle groep en dat ze meer dan de leerkrachten in de controlegroep de resultaten van hun leerlingen in een breder kader bekijken door ze bijvoorbeeld te vergelijken met het landelijk gemiddelde, met eerdere resultaten van dezelfde leerlingen of met voorgaande leerjaren.
- Rapport, waarvoor geldt dat de experimentele groep wel een voortgang doormaakt tussen 2010 en 2012 en in 2012 hoger uitkomen dan de controlegroep. Voor deze variabele geldt echter ook nog dat de experimentele groep in 2010 ook al hoger scoorde dan de controlegroep. De leerkrachten in de experimentele groep gebruikten dus al meerdere analyse- en rapportagemogelijkheden uit het leerlingvolgsysteem in 2010, maar zijn daar tussen 2010 en 2012 nog uitgebreider gebruik van gaan maken.
- Beroepsattitude en Prestatie-informatie, waarbij zowel de experimentele groep als de controlegroep voortgaan tussen 2010 en 2012. Voortgang op deze variabelen is er dus bij alle leerkrachten. Dat betekent dat er waarschijnlijk een andere oorzaak aan ten gronde ligt dan deelname aan *Gebruik maken van opbrengsten*. Mogelijk is hier sprake van een algehele tendens in het onderwijsveld. Een andere mogelijkheid is dat de voortgang samenhangt met opgave voor en deelname aan Streef op schoolniveau. Ook de leerkrachten die in de controlegroep zitten zijn immers verbonden aan een school die aan een van de Streefprojecten heeft meegedaan, weliswaar heeft dan niet de leerkracht zelf maar een of meerdere collega's deelgenomen waarbij niet bekend is of en op welke

manier de ‘opbrengsten’ van deelname in het team besproken zijn en op die manier ook collega’s hebben besmet.

- Tenslotte zien we bij de mate waarin de doelen geëxpliciteerd worden dat de experimentele groep in 2012 hoger scoort, dat wil zeggen vaker doelen bespreekt in het team en vastlegt, dan de controlegroep en dat het belang dat leerkrachten hechten aan de basisvaardigheden in de experimentele groep stijgt tussen 2010 en 2011.

Op de variabelen die iets te maken hebben met differentiatie, dus met het handelen in de groep op basis van de resultaten en de analyse van toetsen verschilt de experimentele groep niet van de controlegroep, noch in 2010, noch in 2012. Ook is er op deze variabelen geen ontwikkeling te zien tussen 2010 en 2012.

Twee variabelen worden in Tabel 3.3 en 3.4. nog nader uitgewerkt. Het betreft de variabelen die inhoudelijk het dichtst liggen bij de inhoud van *Gebruik maken van opbrengsten*. Het betreft de variabelen Reflectie en Breed kader. In Breed kader wordt heel concreet gevraagd op welke manier de leerkracht de toetsresultaten bekijkt: zijn er speciale leerlingen waarvan de resultaten nader bestudeerd worden, worden er vergelijkingen gemaakt met het landelijk gemiddelde of met vorige jaren en worden de resultaten bezien in het licht van de ontwikkeling van de individuele leerling. De antwoordmogelijkheden waren nooit (1), soms (2) en meestal (3). Uit Tabel 4.3 blijkt dat bijna alle leerkrachten meestal de toetsresultaten van zwakke leerlingen nader bestuderen en de resultaten bezien in het licht van de ontwikkeling van de individuele leerling. Dat deden ze al in 2010 en ook in 2012 geldt dit voor alle leerkrachten ongeacht hun deelname aan *Gebruik maken van opbrengsten*. Het minst vergelijken de leerkrachten de prestaties van leerlingen met de prestaties in hetzelfde leerjaar in vorige jaren. Toen dit tijdens *Gebruik maken van opbrengsten* ter sprake kwam gaven ze aan dat niet zo nuttig te vinden omdat het een andere groep leerlingen was die misschien wel gewoon wat beter of wat slechter was, waardoor het niet te achterhalen is waarom prestaties nu iets beter of iets slechter zijn dan in voorgaande jaren. Het maken van een dergelijke trendanalyse vonden ze meer iets voor een intern begeleider of directie.

Het item waarop de experimentele groep meer vooruitgaat dan de controlegroep is het vergelijken van resultaten van verschillende vakken. Zowel op leerlingniveau als op schoolniveau zijn hier binnen het leerlingvolgsysteem specifieke analysemogelijkheden voor, die bij leerkrachten nog relatief onbekend waren.

Tabel 3.3 Breed kader

	2010		2012		2012-2010		Verschil tussen exp en contr in verschillscore significant?
	Exp	Contr	Exp	Contr	Exp	Contr	
Ik kijk preciezer naar de prestaties van de zwakke leerlingen.	2,87	2,95	2,95	2,92	0,08	-0,04	Nee
Ik kijk preciezer naar de prestaties van de excellente leerlingen. ⁴	2,37	2,51	2,63	2,56	0,14	0,19	Nee
Ik vergelijk verschillende vakken.	2,19	2,33	2,54	2,32	0,38	-0,04	Ja
Ik kijk hoe mijn leerlingen het doen in vergelijking met de leerlingen die vorig jaar in mijn groep zaten. ²	2,00	1,72	2,28	1,79	0,27	0,15	Nee
Ik kijk hoe mijn leerlingen het doen in vergelijking met het landelijk gemiddelde. ^{2,4}	2,38	2,10	2,52	2,26	0,17	0,30	Nee
Ik kijk terug naar eerdere toetsgegevens van de leerlingen	2,83	2,75	2,86	2,93	0,08	0,09	Nee

1=in 2010 is er een verschil tussen de exp en de controlegroep

2=in 2012 is er een verschil tussen de exp en de controlegroep

3=het verschil tussen 2010 en 2012 is significant voor de experimentele groep

4=het verschil tussen 2010 en 2012 is significant voor de controlegroep

In de schaal Reflectie gaat het niet om het vergelijken van de ‘cijfers’ maar om het interpreteren en het zoeken naar oorzaken en zo nodig oplossingen. Bijvoorbeeld door het terugkijken naar hoe er instructie gegeven is over een bepaald onderwerp, naar de dekking van de toets, naar welke stof er precies aan bod is geweest. Ook hier loopt de schaal van nooit (1) naar meestal (3). Op deze aspecten verschillen de experimentele en de controlegroep niet van elkaar en is er ook geen sprake van voor- of achteruitgang op bepaalde items terwijl in tabel 4.2 wel bleek dat over de hele schaal genomen de experimentele groep een grotere groei had doorgemaakt dan de controlegroep. Blijkbaar wordt de vooruitgang bepaald door een algehele vooruitgang op alle items waardoor leerkrachten in de experimentele groep meer/vaker zijn gaan reflecteren op hoe bepaalde toetsresultaten tot stand gekomen zijn en wat daar mogelijke oorzaken voor zijn.

Tabel 3.4 Reflectie

	2010		2012		2012-2010		Verschil tussen exp en contr in verschillscore significant?
	Exp	Contr	Exp	Contr	Exp	Contr	
Ik kijk terug naar de lesstof die aan bod is geweest	2,83	2,75	2,93	2,86	0,21	0,10	Nee
Ik kijk terug naar het werk van de leerlingen uit het afgelopen blok	2,38	2,42	2,60	2,55	0,32	0,10	Nee
Ik kijk terug naar het gedrag en/of welbevinden van de leerlingen in de afgelopen periode	2,11	2,27	2,49	2,44	0,08	-0,19	Nee
Ik kijk terug naar eerdere toetsgegevens van de leerlingen	2,60	2,63	2,74	2,56	0,08	0,09	Nee
Ik kijk of de inhoud van de toets gedekt wordt door de methode	2,41	2,52	2,68	2,54	0,23	-0,06	Nee
Ik kijk terug naar hoe ik instructie heb gegeven	2,21	2,41	2,66	2,45	0,40	-0,02	Ja
Ik kijk terug naar de indeling van mijn groepsplan	2,31	2,27	2,68	2,52	0,22	0,38	Nee

Een van de punten waar tijdens de reflectie naar teruggekeken kan worden is het huidige groepsplan waarin voor een bepaalde periode is vastgelegd welke leerstof, op welke manier aan welke leerlingen (in niveaugroepen) wordt aangeboden, wie alleen aan de basisinstructie mee doet en wie in aanmerking komt voor verlengde instructie. Veel scholen zijn de laatste jaren gaan werken met groepsplannen om het overzicht te behouden bij het werken in verschillende niveaus, gecombineerd met leerlingen met een individuele leerlijn. Na een toets kan het nodig zijn om het groepsplan te herzien, bijvoorbeeld omdat leerlingen beter in een andere niveaugroep passen of omdat bepaalde leerlingen een onderdeel nog niet beheersen en daarin nog herhaling nodig hebben.

In de nameting is veel expliciet gevraagd naar het gebruik van groepsplannen, per leerjaar per vakgebied (zie Tabel 3.5). Daaruit komt een duidelijk beeld: voor rekenen en technisch lezen wordt door ongeveer 3 kwart van de respondenten een groepsplan gebruikt, waarbij de leerkrachten van groep 1 enigszins achter blijven. Voor begrijpend lezen wordt een groepsplan minder vaak gebruikt, ongeveer de helft van de leerkrachten werkt daar met een groepsplan. Opvallend is dat de leerkrachten die aan deelproject 2 hebben meegedaan structureel meer gebruik maakt van groepsplannen, in enkele groepen kan het verschil voor bepaalde vakgebieden wel oplopen tot ruim 20% van de respondenten. Vanwege het feit dat deze gegevens alleen bekend zijn uit de nameting is niet met zekerheid te zeggen dat dit het gevolg is van deelname aan Streef, mogelijk is ook dat de leerkrachten van deze scholen vanwege het gebruik van groepsplannen behoefte hadden aan scholing op het gebied van het gebruik maken van toetsresultaten om zo hun groepsplannen beter te kunnen onderbouwen en evalueren.

Tabel 3.5 Gebruik van groepsplannen

Groep	N	Rekenen, % ja		Taal, % ja		Technisch lezen, % ja		Begrijpend lezen, % ja	
		Con	Exp	Con	Exp	Con	Exp	Con	Exp
1	73	58	69	54	69				
2	71	75	93	69	100				
3	53	70	70			74	90		
4	53	73	88			81	88	33	50
5	55	73	83			62	89	41	59
6	61	68	82			61	82	40	77
7	66	71	86					48	68
8	57	69	81					44	70

Samenvattend kan geconcludeerd worden dat uit de vergelijking van de stand van zaken in 2010 en 2012 blijkt dat de groep leerkrachten die aan *Gebruik maken van opbrengsten* al iets meer dan de leerkrachten in de controlegroep ervaring had met het analyseren van toetsresultaten, maar dat desondanks zij op de variabelen die het meest expliciet aan de orde zijn geweest meer vooruitgegaan zijn dan de controlegroep. Dit resulteert in 2012 in een situatie waarbij de leerkrachten die aan *Gebruik maken van opbrengsten* hebben deelgenomen een uitgebreidere analyse maken van toetsresultaten, daarop in een breder kader reflecteren en meer gebruik maken van groepsplannen waarin vastgelegd is op welke manier zij het onderwijs in de komende periode gaan differentiëren. Wat betreft de keuzes die daarbij gemaakt worden lijken er geen verschillen en geen vooruitgang op te treden.

In de volgende paragraaf komt aan de orde of de zelfrapportage van leerkrachten met betrekking tot hun analysevaardigheden ook ondersteund wordt door de beoordeling van de ingeleverde analyserapporten. In 3.3 worden daarna de leerlingresultaten van de experimentele scholen en de controlescholen vergeleken om vast te kunnen stellen of de deelname van leerkrachten aan *Gebruik maken van opbrengsten* ook gevolgen heeft voor de resultaten van leerlingen en zo ja of dat dan verklaard kan worden door de vooruitgang in de variabelen zoals gemeten in de leerkrachtvragenlijst, in het bijzonder de variabelen waarop een vooruitgang te zien is.

3.2.2 Effect op analysevaardigheden

In aanvulling op de zelfrapportage van leerkrachten is hun analysevaardigheid gemeten door hun analyserapportages te beoordelen op 3 categorieën: de analyse op groepsniveau, op leerlingniveau en de analyse van oorzaken en oplossingen. Op elke categorie kan, zoals aangegeven in hoofdstuk 3 een cijfer tussen de 1 en 10 behaald worden. Op basis daarvan wordt een totaal cijfer voor de analyse bepaald. Er wordt een vergelijking gemaakt tussen analyses die aan het begin van het project zijn gemaakt, de voormeting, en analyses die na een half jaar of na een jaar zijn gemaakt.

De analyses van het eerste meetmoment lopen zeer uiteen. Een enkele analyse scoort een ruime voldoende maar slechts 4 van de 57 ingeleverde analyses (7%) heeft hoger dan een 7 gehaald. Het gemiddelde van het eerste analysemoment is 3,9. Bij sommige leerkrachten ontbrak het aan kennis van het leerlingvolgsysteem, anderen hadden bepaalde onderdelen overgeslagen; door 25% van de leerkrachten was er helemaal niet gereflecteerd op oorzaken en oplossingen. Enkele leerkrachten hadden slechts een uitdraai van een rapportage uit het leerlingvolgsysteem ingeleverd. Voor deze eerste analyse (in oktober) is de leerkrachten gevraagd gebruik te maken van de toetsresultaten aan het eind van het vorig schooljaar. De leerlingen zaten toen in een andere groep en waarschijnlijk ook bij een andere leerkracht. Een analyse over deze gegevens maken spreekt misschien minder aan dan op het tweede moment (februari) wanneer de leerkrachten al een halfjaar hard aan het werk zijn met hun groep leerlinge en dan een analyse moeten maken over de toetsen die zij zelf hebben afgenomen. Desalniettemin kan op basis van deze uitkomst wel geconcludeerd worden dat ten tijde van het eerste meetmoment veel leerkrachten niet over de analysevaardigheden beschikken om een goede analyse te maken.

De analyses van het tweede meetmoment waren een stuk constanter. Al deze analyses zijn aan de hand van het analyseformat ingevuld. Deze 35 analyses bevatten een analyse op groepsniveau en een analyse van de oorzaken en de oplossingen. Nog slechts 2 van de 35 analyses mist de categorie leerlingniveau (6%). Het gemiddelde van de analyses ligt een stuk hoger, 7,7 en de standaarddeviatie is kleiner geworden.

Tabel 3.6 Gemiddelden en standaarddeviaties van beide meetmomenten

Categorie	Moment 1	Moment 2	Vershil	t-waarde
Groepsniveau	5.2 (2,8)	8,0 (1,9)	2,8	-5.80
Leerling niveau	1.9 (2,9)	6,6 (2,2)	4,7	-8.71
Oorzaken en oplossingen	3.9 (2,6)	8,1 (1,9)	4,3	-9.01
Totale beoordeling	3.9 (2,3)	7,7 (1,5)	3,8	-9.64

Bekijken we de analyses per beoordeelde categorie dan valt op dat de groei binnen de categorie groepsniveau het kleinst is, maar wel significant. In deze categorie valt het op dat veel leerkrachten aangeven dat zij niet kunnen aflezen in het leerlingvolgsysteem hoe het geanalyseerde resultaat zich verhoudt tot het landelijk gemiddelde of op vergelijkbare scholen. Bij veel analyses is deze vraag niet beantwoord. Ook in het tweede meetmoment werd deze vraag vaak niet beantwoord. Verder viel het op dat veel leerkrachten geen foutenanalyse maken of hadden gemaakt van de te analyseren toets. Het belang van het maken van een foutenanalyse bij genormeerde toetsen wordt door verschillende auteurs een zeer waardevolle handeling bij het analyseren genoemd (Mandinach, Honey & Light, 2006; Onderwijsinspectie, 2010). Een foutenanalyse kan helpen bij het goed afstemmen van het onderwijsleerproces.

De groei is het grootst bij de categorie leerlingniveau. In deze categorie was het, net zoals bij het groepsniveau, opvallend dat veel leerkrachten niet beschikten over een foutenanalyse waardoor er niet genoemd kan worden hoe de leerling scoort op specifieke onderdelen. Ook viel het in deze categorie op dat er veel gescoord werd in beantwoord of niet beantwoord maar dat er niet vaak dieper op in werd gegaan waardoor de beoordeling ‘beantwoord en geanalyseerd’ kon worden aangekruist. De standaarddeviatie blijft hoger dan dat die bij de andere categorieën is. De scores binnen deze categorie blijven dus vrij ver uit elkaar liggen.

Opvallend binnen de categorie oorzaken en oplossingen is dat er een verschuiving plaats heeft gevonden van de kolom “beantwoord” op het beoordelingsformulier naar de kolom “beantwoord en geanalyseerd”. Deze categorie heeft daarom ook een grote groei in gemiddelde gemaakt bij het tweede meetmoment ten opzichte van het eerste meetmoment. Er werd duidelijk meer verteld en verder gekeken naar oorzaken voor de resultaten en wat dat voor consequenties zou kunnen hebben voor de inrichting van het onderwijs.

Vanwege het feit dat maar heel weinig leerkrachten meerdere keren een analyse hebben ingeleverd was het niet mogelijk om na te gaan of leerkrachten zichzelf significant verbeterd hebben na begeleiding binnen het project. Slechts 3 leerkrachten hebben op twee momenten een analyse ingeleverd. Deze 3 leerkrachten zijn fors vooruit gegaan maar vanwege een te kleine groep kunnen we hieruit geen conclusies trekken.

Samengevat bevestigt de ‘objectieve’ beoordeling van de analyserapportages dus de zelfrapportage van leerkrachten in de vragenlijst wat betreft de vooruitgang die geboekt is. Uit de analyse van de leerkrachtvragenlijst bleek ook dat de leerkrachten die aan *Gebruik maken van opbrengsten* mee gingen doen in 2010 al wel meer ervaring had dan de controlegroep. Dit is ogenschouw genomen is het wel opmerkelijk dat er op de voormeting zo laag wordt gescoord. Mogelijk hebben de leerkrachten hier in de voormeting een iets te rooskleurig beeld gegeven van hun capaciteiten.

Hoewel de analysevaardigheden dus aanmerkelijk gegroeid lijken is er nog wel ruimte voor verbetering met name ten aanzien van het analyseren van fouten om zo in meer detail te weten te komen waar hiaten bij leerlingen liggen. Daarbij moet wel opgemerkt worden dat de meeste analyses dateren van halverwege *Gebruik maken van opbrengsten*. Aan het eind van het traject is de leerkrachten nogmaals de opdracht gegeven een analyse uit te voeren, nu van de toetsafname aan het eind van het schooljaar maar vanwege tijdsdruk in die tijd van het schooljaar hebben niet veel leerkrachten die analyse ook daadwerkelijk ingeleverd.

3.3 Effecten op leerlingresultaten

Het uiteindelijk beoogde effect van de interventie *Gebruik maken van opbrengsten* was het verbeteren van de resultaten van leerlingen. In de vorige paragraaf hebben we gezien dat de leerkrachten die aan de interventie hebben meegedaan op een aantal aspecten meer

voortuitgaan dan leerkrachten die niet aan de interventie hebben meegedaan. Dat is een voorwaarde om ook een effect te kunnen verwachten op de prestaties van leerlingen. Wat betreft de verbetering van leerlingprestaties wordt gekeken naar de directe effecten in het jaar van de interventie en naar retentie-effecten een jaar later. In paragraaf 3.4 wordt tenslotte onderzocht of de mate opbrengstgericht werken ook daadwerkelijk de verschillen tussen scholen in de prestaties van leerlingen kan verklaren.

Om het effect op leerlingprestaties te onderzoeken zijn voor alle cohorten de leerlingen van de experimentele scholen vergeleken met die van de controlescholen volgens het overzicht in hoofdstuk 2. Welke groepen en toetsen er precies zijn meegenomen in weergegeven in Tabel 3.7 voor de directe effecten en in Tabel 3.8 voor de retentie-effecten. Voor de retentie-effecten zijn minder groepen en leerstofdomeinen gebruikt omdat er bij de andere groepen/leerstofdomeinen te weinig scholen overbleven om een goede analyse te maken.

Tabel 3.7 Leerstofdomeinen en groepen waarop directe effecten zijn gebaseerd

Leerstofdomein	Pretest in schooljaar 2009-2010	Posttest in schooljaar 2010-2011	N (experimentele scholen)	N (controle scholen)
Taal kleuters	E1 TvK'04	E2 TvK'11	4	18
	E1 TvK'11	E2 TvK'11	8	10
Technisch lezen	E2 TvK'04	E3 Leestempo	3	8
	E2 TvK'11	E3 Leestempo	-	-
	E3 Leestempo	E4 Leestempo	2	10
	E4 Leestempo	E5 Leestempo	10	6
	E5 Leestempo	E6 Leestempo	9	2
Begrijpend lezen	E2 TvK'04	E3 Begrijpend lezen	6	22
	E2 TvK'11	E3 Begrijpend lezen	5	4
	E3 Begrijpend lezen	E4 Begrijpend lezen	11	21
	E4 Begrijpend lezen	M6 Begrijpend lezen	17	13
	M5 Begrijpend lezen	M7 Begrijpend lezen	18	13
	M6 Begrijpend lezen	M8 Begrijpend lezen	11	10
Spelling	E2 TvK'04	E3 Spelling	5	25
	E2 TvK'11	E3 Spelling	5	7
	E3 Spelling	E4 Spelling	13	30
	E4 Spelling	E5 Spelling	20	14
	E5 Spelling	E6 Spelling	18	14
	E6 Spelling	E7 Spelling	16	9
Rekenen	E1 Ordenen	E2 Ordenen	3	10
	E1 Ordenen	E2 RvK	9	13
	E2 Ordenen	E3 Rek/wis	13	33
	E3 Rek/wis	E4 Rek/wis	13	33
	E4 Rek/wis	E5 Rek/wis	20	15
	E5 Rek/wis	E6 Rek/wis	19	15
	E6 Rek/wis	E7 Rek/wis	19	13

Tabel 3.8 Leerstofdomeinen en groepen waarop retentie-effecten zijn gebaseerd

Leerstofdomein	Pretest in schooljaar 2009-2010	Posttest in schooljaar 2011-2012	N (experimentele scholen)	N (controle scholen)
Spelling	E3 Spelling	E5 Spelling	7	20
	E4 Spelling	E6 Spelling	13	7
	E5 Spelling	E7 Spelling	8	7
Rekenen	E1 Ordenen	E3 Rek/wis	8	15
	E2 Ordenen	E4 Rek/wis	8	16
	E3 Rek/wis	E5 Rek/wis	8	16
	E4 Rek/wis	E6 Rek/wis	14	7
	E5 Rek/wis	E7 Rek/wis	12	8

Om het effect van *Gebruik maken van opbrengsten* te bepalen zijn eerst per cohort het effect van de interventie bepaald met behulp van een multilevel analyse. Daarna zijn de uitkomsten van die analyses in een meta-analyse ingebracht om het gezamenlijke effect van de interventie te berekenen. De resultaten van de afzonderlijke multilevel analyses en van de meta-analyse zijn voor de directe effecten in Tabel 3.9 weergegeven en voor de retentie-effecten in Tabel 3.10. Per cohort is het effect weergegeven, de standaardfout en de variantie in het effect. Daarnaast het 90% betrouwbaarheidsinterval, de Z-score en de bijbehorende p-waarde. De cohorten waar het effect significant afwijkt van 0 zijn gearceerd.

Voor alle leerstofdomeinen wijkt het effect als alle cohorten samen worden genomen niet significant af van 0 als er gebruik wordt gemaakt van een betrouwbaarheidsinterval van 90%. Bij elk samengevat leerstofdomein ligt de ondergrens van het interval immers onder 0 en de bovengrens boven 0 en omvat het interval dus de waarde 0, wat aanduidt dat er geen significant effect optreedt.

Van de 25 geschatte effectgroottes zijn er – zij het niet allemaal significant - 18 positief en 7, negatief, slechts eenmaal laat een bovenbouwgroep een (niet significant) negatief effect zien en bij begrijpend lezen zijn de effecten allemaal positief. Bij alle leerstofdomeinen (behalve taal in de kleutergroep) is te zien dat het 90%-betrouwbaarheidsinterval zich voor het grootste deel aan de positieve kant van de 0 bevindt (technisch lezen: -0,02 tot 0,40; spelling: -0,08 tot 0,17; begrijpend lezen: -0,02 tot 0,22; rekenen: -0,03 tot 0,17).

Hoewel er dus in het algemeen een positieve tendens te zien is, is het effect niet groot genoeg om significant te zijn.

Tabel 3.9 Resultaten van de multilevel- en meta-analyse (directe effecten)

	Interventie in ...	d	se d	var d	90%	90%	z	P
					benedengrens	bovengrens		
Taal kleuters	groep 2 (TvK'04)	0,09	0,30	0,09	-0,40	0,59	0,31	0,76
	groep 2 (TvK'11)	-0,52	0,32	0,10	-1,05	0,00	-1,65	0,10
	Samenvattend	-0,20	0,22	0,05	-0,55	0,16	-0,90	0,37
Technisch lezen	groep 3 (TvK'04)	-0,15	0,24	0,06	-0,54	0,25	-0,60	0,55
	groep 4	-0,22	0,45	0,20	-0,95	0,51	-0,50	0,62
	groep 5	0,53	0,20	0,04	0,21	0,85	2,70	0,01
	groep 6	0,13	0,29	0,08	-0,35	0,60	0,44	0,66
	Samenvattend	0,19	0,13	0,02	-0,02	0,40	1,51	0,13
Spelling	groep 3 (TvK'04)	0,33	0,33	0,11	-0,21	0,87	1,01	0,31
	groep 3 (TvK'11)	0,72	0,36	0,13	0,12	1,31	1,99	0,05
	groep 4	-0,15	0,16	0,03	-0,42	0,13	-0,88	0,38
	groep 5	0,14	0,14	0,02	-0,09	0,37	1,02	0,31
	groep 6	-0,18	0,16	0,03	-0,44	0,08	-1,12	0,26
	groep 7	0,17	0,22	0,05	-0,19	0,52	0,77	0,44
	Samenvattend	0,04	0,08	0,01	-0,08	0,17	0,55	0,58
Begrijpend lezen	groep 3 (TvK'04)	0,17	0,23	0,05	-0,20	0,54	0,74	0,46
	groep 3 (TvK'11)	0,63	0,28	0,08	0,17	1,08	2,26	0,02
	groep 4	0,05	0,17	0,03	-0,23	0,34	0,30	0,77
	groep 5	0,03	0,14	0,02	-0,21	0,26	0,18	0,86
	groep 6	0,03	0,14	0,02	-0,19	0,26	0,23	0,82
	groep 7	0,11	0,21	0,04	-0,23	0,45	0,53	0,60
	Samenvattend	0,10	0,07	0,01	-0,02	0,22	1,35	0,18
Rekenen	groep 2 (Ordenen)	-0,74	0,32	0,10	-1,26	-0,21	-2,29	0,02
	groep 2 (RvK)	0,01	0,35	0,12	-0,56	0,59	0,04	0,97
	groep 3	0,14	0,19	0,04	-0,18	0,46	0,73	0,46
	groep 4	-0,13	0,18	0,03	-0,43	0,17	-0,71	0,48
	groep 5	0,16	0,09	0,01	0,01	0,31	1,71	0,09
	groep 6	0,03	0,14	0,02	-0,19	0,26	0,24	0,81
	groep 7	0,29	0,25	0,07	-0,13	0,71	1,15	0,25
	Samenvattend	0,07	0,06	0,00	-0,03	0,17	1,11	0,27

Als we naar de afzonderlijke cohorten kijken dan zien we wel bij enkele cohorten een significant positief effect. Deelname van de leerkrachten van de onderbouw leidt bij spelling en bij begrijpend lezen tot een positief effect (als de toets Taal voor Kleuters, 2011 is afgenomen in groep 2) op de prestaties van leerlingen in groep 3. Deelname met de bovenbouwgroepen leidt tot meer leerwinst in groep 5 op technisch lezen en op rekenen. Slechts eenmaal is er sprake van een negatief effect: als zowel bij de voormeting eind groep 1 als bij de nameting eind groep 2 de toets Ordenen wordt gebruikt zien we bij de scholen uit de experimentele scholen minder leerwinst dan bij de controlescholen. Overigens bestaat de groep experimentele scholen hier slechts uit 3 scholen.

Samenvattend betekent dit dat er onder de gehanteerde grenzen (net) geen sprake is van een positief effect op de vorderingen van leerlingen in het schooljaar waarin hun leerkrachten deelnamen aan *Gebruik maken van opbrengsten*.

De retentie-effecten zijn weergegeven in Tabel 3.10. Deze zien er minder positief uit. De effectgrootte is even vaak (niet-significant) positief als negatief, bij alle cohorten en beide leerstofdomeneinen omvat het betrouwbaarheidsinterval de waarde 0 en het resulterende effect wijkt vanzelfsprekend ook niet af van 0.

Samenvattende kunnen we stellen dat het licht positieve – zij het niet significante – effect wat optrad in het schooljaar waarin de leerkrachten deelnamen aan de interventie niet vastgehouden wordt in het schooljaar daarna.

Tabel 3.10 Resultaten van de multilevel- en meta-analyses (retentie-effecten)

Interventie in ...	d	se d	var d	90%		z	P	
				benedengrens	bovengrens			
Spelling	groep 4	-0,17	0,24	0,06	-0,57	0,22	-0,73	0,47
	groep 5	-0,10	0,23	0,05	-0,47	0,28	-0,43	0,66
	groep 6	0,11	0,21	0,04	-0,23	0,45	0,52	0,60
	Samenvattend	-0,04	0,13	0,02	-0,25	0,17	-0,31	0,75
Rekenen	groep 2	0,15	0,22	0,05	-0,21	0,51	0,68	0,50
	groep 3	0,28	0,22	0,05	-0,09	0,65	1,27	0,21
	groep 4	-0,04	0,22	0,05	-0,40	0,32	-0,18	0,85
	groep 5	-0,09	0,20	0,04	-0,41	0,24	-0,44	0,66
	groep 6	0,02	0,20	0,04	-0,32	0,35	0,08	0,94
	Samenvattend	0,05	0,09	0,01	-0,10	0,21	0,57	0,57

3.4 Verklaring: leerkrachtattitude- en gedrag als mediator

In paragraaf 3.2 stond de verandering in de attitude en het handelen van leerkrachten als effect van deelname aan het Streef-deelproject *Gebruik maken van opbrengsten* centraal, in 3.3 de verandering in veranderingen in leerwinst van leerlingen over de periode van 1 en 2 jaar. Hadden er op beide terreinen veranderingen plaatsgevonden ten gevolge van deelname aan de interventie dan was een interessante vervolgvraag geweest of de veranderingen die in attitude en handelen van leerkrachten te zien was de veranderingen in leerwinst konden verklaren. Met andere woorden; welke elementen van opbrengstgericht werken waarop leerkrachten vooruitgang hadden geboekt kunnen de veranderingen in leerwinst verklaren. Er is echter geen significante verandering in leerwinst opgetreden die verband houdt met deelname aan de interventie, dus het beantwoorden van deze vraag zal weinig opleveren. Wel is het interessant om meer inzicht te krijgen in de vraag of en hoe opbrengstgericht werken samenhangt met leerlingprestaties, welke varianten het best werken, welke elementen binnen het complexe begrip opbrengstgericht werken daaraan het meeste bijdragen. In Streef is immers het begrip

opbrengstgericht werken breed opgevat, dus niet alleen het analyseren van toetsgegevens, maar ook het gebruiken en stellen van doelen, individueel en in teamverband, het vertalen van de analyse naar een gedifferentieerd groepsplan en het uitvoeren daarvan. Dit inzicht kan leiden tot een betere duiding van de effecten en het uitblijven daarvan ten gevolge van de interventie *Gebruik maken van opbrengsten*.

Voor het uitvoeren van de analyse¹ worden eerst op basis van de resultaten van de leerkrachten op de voormeting van de leerkrachtvragenlijst (op 9 aan opbrengstgericht werken gerelateerde variabelen, zie Tabel 3.11) vier clusters van opbrengstgericht werken op schoolniveau gedefinieerd.

Tabel 3.11 Samenstelling van de clusters

	Opbrengstgericht	Individualistisch	Behoudend	Niet opbrengstgericht
Samenwerking	+	-	0	0
Beroepsattitude	+	0	--	0
Reflectie	+	-	0	0
Prestatie-informatie	+	+	0	-
Basisvaardigheden	+	+	0	-
Attitude opbrengsten	+	0	0	-
Differentiatie actie	+	-	+	0
Uitgebreide rapportage	0	++	0	0
Praten over referentie-niveaus	0	0	-	0

++ betekent: meer dan 1 sd meer dan het gemiddelde

+ betekent: tussen 1 en 0,5 sd meer dan het gemiddelde

0 betekent: minder dan 0,5 sd meer/miner dan het gemiddelde

- betekent: tussen 1 en 0,5 st minder dan het gemiddelde

-- betekent: meer dan 1 sd minder dan het gemiddelde

Schoolteams die het meest *opbrengstgericht* zijn, scoren bovengemiddeld (0,5 tot 1 standaard deviatie boven het gemiddelde) op vrijwel alle meegenomen kenmerken. Deze teams werken samen, reflecteren, hechten waarde aan basisvaardigheden en opbrengsten en differentiëren ook in hun handelen. Daarnaast zijn er teams die wel belang hechten aan basisvaardigheden en het delen van prestatie-informatie en die uitgebreide rapportages maken, maar die niet samenwerken en reflecteren en die ook niet handelen naar de informatie uit de rapportages gegevens door differentiatie toe te passen. De leerkrachten in deze teams werken *individualistisch* en zijn niet teamgericht. Een derde groep teams heeft een *behoudende* aanpak. Ze zijn niet innovatief: ze praten niet over referentieniveaus en hechten weinig belang aan extra professionalisering. Ze passen echter wel differentiatie toe in hun onderwijs.

¹ De analyse waarvan hier alleen de belangrijkste resultaten worden vermeld, wordt uitvoeriger besproken in het eindverslag van de afname van de leerkrachtvragenlijst van Streef (Deunk & Doolaard, in voorbereiding). Hier worden alleen de resultaten vermeld.

Tenslotte zijn er schoolteams die *niet opbrengstgericht* zijn. Zij zijn gemiddeld in de mate van samenwerking en gebruik van toetsrapportages en differentiatie, maar hechten weinig belang aan basisvaardigheden en opbrengsten.

De leerprestaties (vorderingen na 1 jaar en na 2 jaar, zoals in paragraaf 3.3) van leerlingen van opbrengstgerichte teams zijn vervolgens vergeleken met de drie andere typen teams. Er is per leerjaar en per inhoudsdomein gekeken of de variant van opbrengstgericht werken van het leerkrachtenteam samenhangt met leerwinst van de leerlingen. In Tabel 3.12 en 3.13 is te zien dat de resultaten niet een volledig consistent beeld opleveren. De verwachting is immers dat de drie varianten die afgezet zijn tegen het meest opbrengstgerichte team minder hoge leerwinst bij leerlingen bewerkstelligen. Hoewel we dat niet bij elk leerstofdomein en elk cohort zien komt die verwachting wel grotendeels uit. Bij 61 van de 90 uitgevoerde analyses is een negatief – zij het niet altijd significant – effect zichtbaar, 12 effecten zijn significant. Dat betekent dat een opbrengstgerichte team meer leerwinst boekt dan het type team dat een negatief effect laat zien.

Tabel 3.12 Effect op leerwinst over 1 jaar

	Posttest in ...	Individualistisch	Behoudend	Niet opbrengstgericht
Taal kleuters	groep 2 (TvK'04)	1,40	-0,72	-3,42
Technisch lezen	groep 5	-8,34	-8,49	-6,02
Spelling	groep 3 (TvK'04)	0,17	4,59	1,75
	groep 4	1,94	0,99	-0,42
	groep 5	-0,85	-0,89	0,04
	groep 6	-0,06	-1,21	-0,49
	groep 7	0,77	1,09	0,93
Begrijpend lezen	groep 3 (TvK'04)	-7,11	-3,75	-3,28
	groep 4	1,35	-0,81	1,45
	groep 5	-2,59	-2,55	-2,27
	groep 6	-2,08	-3,99	-4,37
	groep 7	-1,93	-2,36	2,15
Rekenen	groep 2 (Ordenen)	-0,19	13,51	-0,31
	groep 3 (Ordenen)	-8,20	-3,03	-2,97
	groep 4	-2,24	-1,09	-1,43
	groep 5	-0,08	-0,42	-0,85
	groep 6	-3,52	0,25	-1,84
	groep 7	-0,14	2,26	1,26

Vet=significant negatief, *cursief*=significant positief

Drie van de 29 positieve effecten is significant. De significant negatieve effecten zien we vooral bij het cluster 'Individualistisch', dat gekenmerkt worden door weinig samenwerking, reflectie en differentiatie en door een uitgebreid gebruik van analyserapportages. Het cluster dat de karakterisering 'Behoudend' heeft gekregen doet nauwelijks onder voor het

opbrengstgerichte cluster. De leerkrachten in deze teams onderscheiden zich door een lage beroepsattitude; ze zijn minder dan leerkrachten in andere teams bezig met het aanpassen van hun onderwijs, met onderwijsverbeteringen en professionalisering. Ook zijn zij minder goed op de hoogte van de referentieniveaus dan andere leerkrachten.

Tabel 3.13 Effect op leerwinst over 2 jaar

	Posttest in ...	Individualistisch	Behoudend	Niet opbrengstgericht
Technisch lezen	groep 5			
Spelling	groep 3 (TvK'04)	0,13	-0,15	-0,10
	groep 4	-3,91	0,70	-0,18
	groep 5	1,07	0,41	-0,40
	groep 6	-0,93	1,05	0,88
	groep 7	-1,14	1,29	0,74
Begrijpend lezen	groep 4	-9,22	-7,74	-1,31
	groep 5	-0,36	-1,77	0,44
Rekenen	groep 3 (Ordenen)	-1,37	-0,93	3,77
	groep 4	-13,96	-4,85	-4,30
	groep 5	-6,54	-1,91	-1,07
	groep 6	1,11	0,82	-0,52
	groep 7	-3,27	-3,11	-0,65

Vet=significant negatief, *cursief*=significant positief

In hoofdstuk 4 worden de resultaten samengevat en met elkaar in verband gebracht.

4 Samenvatting en conclusies

Om te onderzoeken in welke mate en hoe leerkrachten de gegevens die zij kunnen ontleen aan het Cito-leerlingvolgsysteem gebruiken en om te onderzoeken wat de effecten zijn op de vorderingen van leerlingen als leerkrachten als team geschoold worden in een beter gebruik daarvan is in het schooljaar 2010-2011 op 40 scholen een interventie, genaamd *Gebruik maken van opbrengsten* uitgevoerd. De interventie betrof een trainings- en begeleidingstraject uitgevoerd door de Rijksuniversiteit Groningen en twee onderwijsadviesorganisaties. Het traject is een van de deelprojecten binnen Streef, een samenhangend programma van 3 interventies op verschillende terreinen van het onderwijsbeleid rond kwaliteit in het PO: referentieniveaus, opbrengstgericht werken en excellentie.

Bij de ontwikkeling van *Gebruik maken van opbrengsten* is gebruik gemaakt van onderzoeksresultaten op het gebied van effectief onderwijs en effectieve onderwijsverbetering. Het traject bevat zowel training – bovenschools – als begeleiding op de werkplek waarbij direct gewerkt wordt met gegevens van eigen leerlingen en de vertaling naar de eigen onderwijspraktijk.

Hoewel de beginsituatie van de deelnemers sterk uiteenliep was het in de bijeenkomsten op de eigen school goed mogelijk om aan te sluiten bij de vragen van de individuele scholen. In de bijeenkomsten op bovenschoolsniveau profiteerden met name de scholen die iets konden opsteken van de ‘voorlopers’.

Vooraf en na afloop van *Gebruik maken van opbrengsten* is bij leerkrachten met behulp van een digitale vragenlijst gemeten hoe zij over opbrengstgericht werken en daaraan gerelateerde begrippen denken en hoe ze daarbij handelen. Als onderdeel van *Gebruik maken van opbrengsten* zijn driemaal analyserapporten van leerkrachten verzameld aan de hand waarvan onderzoek is gedaan naar hoe hun analysevaardigheden zich ontwikkelden. En ten slotte zijn voor, tijdens en na het traject toetsresultaten van leerlingen verzameld. Met behulp van deze gegevens is het mogelijk om eerst te onderzoeken wat het directe effect van *Gebruik maken van opbrengsten* is, namelijk een mogelijke feitelijke verbetering van vaardigheden, parallel daaraan kan op basis van de zelfrapportage gemeten worden wat het effect is op hun attitude en gedrag onder andere in de klas. Toegenomen vaardigheden in het analyseren van toetsgegevens zullen immers waarschijnlijk pas effect hebben op vorderingen van leerlingen als de verbeterde analyse ook effect heeft in de groep, met name in differentiatie in de groep. Die differentiatie zal tot uiting moeten komen in het stellen van reële doelen voor alle leerlingen, passend aanbod van lesstof en instructie op verschillende niveaus. Als laatste in deze causale keten zijn dus de vorderingen van leerlingen geanalyseerd.

Als er eerst gekeken wordt naar de eerste stap in de effectevaluatie dan zien we dat de analyse-vaardigheden aan het eind van *Gebruik maken van opbrengsten* duidelijk hoger liggen dan aan het begin van het traject. Een relativering is hier wel noodzakelijk: omdat slechts

enkele leerkrachten twee- of zelfs driemaal een analyse hebben ingeleverd was het niet mogelijk om op leerkrachtniveau een voor- en een nameting met elkaar te vergelijken. Het was slechts mogelijk om de ingeleverde analyses aan het begin ‘gemiddeld’ te vergelijken met de later ingeleverde analyses. Bovendien is er geen sprake van een controlegroep. De vooruitgang is echter zo duidelijk dat er toch gesproken kan worden van een direct effect van *Gebruik maken van opbrengsten*. Deze conclusie bevestigt de conclusie van Visscher en Ehren (2011) die aangeven dat een beter gebruik van een leerlingvolgsysteem aan te leren is. Als vervolgens gekeken wordt naar het meer indirecte effect van *Gebruik maken van opbrengsten* op basis van de leerkrachtvragenlijst, waarbij het wel mogelijk is om de vergelijking tussen de situatie voor en na *Gebruik maken van opbrengsten* op leerkrachtniveau met elkaar te vergelijken dan zien we dat uit de vergelijking van de stand van zaken in 2010 en 2012 blijkt dat de groep leerkrachten die aan *Gebruik maken van opbrengsten* al iets meer dan de leerkrachten in de controlegroep ervaring had met het analyseren van toetsresultaten, maar dat desondanks zij op de variabelen die het meest expliciet aan de orde zijn geweest meer vooruitgegaan zijn dan de controlegroep. Dit resulteert in 2012 in een situatie waarbij de leerkrachten die aan *Gebruik maken van opbrengsten* hebben deelgenomen een uitgebreidere analyse maken van toetsresultaten, daarop in een breder kader reflecteren en meer gebruik maken van groepsplannen waarin vastgelegd is op welke manier zij het onderwijs in de komende periode gaan differentiëren. Wat betreft de keuzes die daarbij gemaakt worden lijken er geen verschillen en geen vooruitgang op te treden. De vooruitgang in analysevaardigheden wordt dus bevestigd door de beoordeling van analyserapportages aan het begin en halverwege/het eind van *Gebruik maken van opbrengsten*.

De tweede onderzoeksvraag betreft het effect van het traject op de vorderingen van leerlingen. De analyses hiervan laten tot nu toe weliswaar een effect zien in de positieve richting maar dit effect is niet significant zodat we feitelijk moeten spreken van het ontbreken van een effect op leerlingniveau. Van de 25 geschatte effectgroottes zijn er – zij het niet allemaal significant - 18 positief en 7, negatief, slechts eenmaal laat een bovenbouwgroep een (niet significant) negatief effect zien en bij begrijpend lezen zijn de effecten allemaal positief. Bij alle leerstofdomeinen (behalve taal in de kleutergroep) is te zien dat het 90%-betrouwbaarheidsinterval zich voor het grootste deel aan de positieve kant van de 0 bevindt. Hoewel er dus in het algemeen een positieve tendens te zien is, is het effect, binnen de gehanteerde grenzen, (net) niet groot genoeg om significant te zijn.

De retentie-effecten zijn zien er minder positief uit. De effectgrootte is even vaak (niet-significant) positief als negatief, bij alle cohorten en beide leerstofdomeinen omvat het betrouwbaarheidsinterval de waarde 0 en het resulterende effect wijkt vanzelfsprekend ook niet af van 0. Samenvattende kunnen we stellen dat het licht positieve – zij het niet significante – effect wat optrad in het schooljaar waarin de leerkrachten deelnamen aan de interventie niet vastgehouden wordt in het schooljaar daarna.

Hierdoor is ook de laatste stap waarbij wordt onderzocht of het effect op leerlingniveau verklaard kan worden door een verbetering in attitude of gedrag van leerkrachten niet aan de orde.

Leeropbrengsten stijgen wanneer leerkrachten werken volgens de evaluatieve cyclus (Ledoux et al., 2009; Inspectie van Onderwijs, 2010). Dat blijkt ook uit de clusteranalyse die op basis van de leerkrachtvragenlijst van Streef gedaan is. Dat leerkrachten zich ontwikkelen door deelname van Streef blijkt ook uit dit onderzoek. Wat kunnen dan redenen zijn dat de leerlingresultaten zich weliswaar in de gewenste richting ontwikkelen maar niet overtuigend zijn en bovendien in het jaar nadat de leerkrachten aan Streef deelnamen weer terugvielen?

Een van de mogelijke verklaringen is af te leiden uit de clusteranalyse waarin 4 typen teams waren onderscheiden: opbrengstgerichte en niet-opbrengstgerichte teams, teams die ‘individualistisch’ genoemd kunnen worden (waarin leerkrachten wel veel analyseren, maar er geen sprake is van overleg en samenwerking binnen het team) en teams die ‘behoudend’ genoemd worden en die ‘gemiddeld’ analyseren, samenwerken, maar weinig gericht zijn op vernieuwingen en verbeteringen in hun handelen. Deze laatst genoemde behalen resultaten die vergelijkbaar zijn met de opbrengstgerichte teams. De ‘individualistische’ teams presteren minder goed. Dit kan duiden op het feit dat veel analyseren maar daar niet in teamverband iets mee doen, een eventueel positief effect van het maken van een goede analyse teniet doet. Mogelijk zijn de leerkrachten tijdens Streef nog te veel bezig met het verbeteren van hun eigen vaardigheden in het analyseren en komt het teamoverleg daarover nog niet goed uit de verf. Als er geen ‘teamdruk’ op ligt, vervliegt het effect van beter analyseren mogelijk weer na het jaar van actieve participatie of voeren leerkrachten de analyse niet meer uit als ze daartoe niet een expliciete opdracht gekregen hebben. Dit is in overeenstemming met een aantal informele reacties die we na Streef van deelnemende scholen kregen, namelijk dat ze er veel geleerd hadden, maar een stok achter de deur nodig hadden om op de ingeslagen weg door te gaan.

Een andere mogelijke verklaring wordt ook door Streef bevestigd: namelijk dat de stap van analyse naar groepplan en de uitvoering daarvan nog onvoldoende gemaakt wordt. De beide clusters van teams die de hoogste leerlingresultaten behalen scoren allebei hoog op differentiatie. Blijkbaar zijn zij in staat om die stap wel te maken en uit te voeren in de klas. Streef toont aan dat analysevaardigheden van leerkrachten toenemen na de professionalisering. Dit kan betekenen dat door het verbeteren het interpreteren van toetsgegevens, de beslissingen die gemaakt worden voor het aanpassen van het onderwijs ook verbeteren. De analyse die de leerkracht maakt hoeft echter nog niets te zeggen over de daadwerkelijke verbetering van het doorvoeren van beslissingen in de klas. Wanneer de analyse goed is, zal de leerkracht over genoeg kennis moeten beschikken om gedegen beslissingen te nemen en deze in te voeren in de klas. Wanneer de kennis voor het maken van

onderwijsinhoudelijke beslissingen aanwezig is, is de uitvoering hiervan cruciaal. De uitvoering is bijvoorbeeld ook afhankelijk van goed klassenmanagement. Pas dan kan er verwacht worden dat er ook een stijging van de leeropbrengsten zal optreden. Voor een verbetering van leerlingresultaten zal dus zeker binnen of naast de professionalisering op het gebied van het gebruik maken van opbrengsten ook aandacht moeten zijn voor kennis of effectieve differentiatie en uitvoering van differentiatie, een van de moeilijkere aspecten van het werk van een leerkracht waarvoor een beroep wordt gedaan op complexe vaardigheden. Deze verklaring wordt ondersteund door de peiling van Regioplan (Dekker e.a., 2012) waarin de respondenten uit het onderwijs zelf aangaven dat ze wat betreft professionalisering vooral ondersteuning voor het verbeteren van de didactische vaardigheden missen.

Overigens is het analyseren zelf een vaardigheid die heel goed al binnen de opleiding van leerkrachten aangeleerd kan worden. Toekomstige leerkrachten die nu een opleiding aan een academische pabo volgen zullen in de toekomst waarschijnlijk beter toegerust zijn op dit gebied omdat onderzoeksvaardigheden, inclusief toetstheorie daar standaard aangeboden worden. Dat geldt ook voor de meer theoretische kennis op het gebied van evidence based kennis over effectieve interventies in de klas.

Samenvattend concentreren de aanbevelingen voor beleid en de onderwijspraktijk gericht op het verder verbeteren van opbrengstgericht werken in de scholen zich dus op het stimuleren van en aandacht besteden aan de fase van het cyclus van opbrengstgericht werken na het analyseren van de data: het met elkaar bespreken van de uitkomsten, het gezamenlijk bepalen van ‘nieuwe’ doelen en het uitstippelen van een weg daarnaar toe. Daarvoor is een gedegen kennis van de leerstof nodig en kennis van op onderzoek gebaseerde didactische aanpassingen. In een van de andere deelprojecten van Streef zijn leerkrachten een jaar lang in hun klas gevolgd en is – specifiek voor de goede leerlingen – onderzocht welke beslissingen leerkrachten nemen ten aanzien van leerstofaanbod, didactische werkvormen, differentiatie en organisatie. Ook daaruit bleek dat het regelmatig ontbreekt aan kennis van en inzicht in effectieve manieren om het onderwijs aan te passen aan verschillen tussen leerlingen. In de rapportage waarin de opbrengsten van de verschillende onderdelen van Streef worden geïntegreerd (Doolaard, in voorbereiding) wordt hier verder op ingegaan.

Vaardigheden die nodig zijn om goed te kunnen differentiëren vragen om ervaring en kunnen op de opleidingen wel aangereikt worden, maar onderzoek naar de ontwikkeling van leerkrachten toont aan dat zij zich in het begin van hun loopbaan eerst moeten concentreren op een aantal meer basale aspecten van het werk. Benodigde kennis van de leerstof, de didactiek en ook het analyseren van gegevens behoren wel al in de opleiding aangeleerd te worden. Met name op dat laatste gebied is een verbetering van de opleiding mogelijk. Waar en hoe gevorderde leerkrachten een ‘proeve van bekwaamheid’ zouden moeten afleggen is een punt van aandacht voor het beleid.

Literatuur

- Blok, H. (2004). Adaptief onderwijs: betekenis en effectiviteit. *Pedagogische Studien*, 81(1), 5-27.
- Dekker, B., Krooneman, P.J., Brekelmans, J. & Groenwoud, M. (2012). *Onderwijsinspanningen taal en rekenen in PO, VO en MBO*. Amsterdam: Regioplan.
- Deunk, M. I., Doolaard, S., & Hofman, R. H. (2011). *Attitude en gedrag van basisschoolleerkrachten met betrekking tot het verbeteren van leerlingresultaten. Resultaten van de beginmeting bij leerkrachten*. Groningen: GION.
- Deunk, M.I. & Doolaard, S. (in voorbereiding). *Attitude en handelen van basisschoolleerkrachten met betrekking tot het verbeteren en borgen van leerlingresultaten*. Groningen: GION.
- Doolaard, S. (in voorbereiding). *Het streven naar kwaliteit in scholen voor primair onderwijs. Integraal eindrapport van onderzoeken in de BOPO-programmalijn onderwijskwaliteit PO 2009-2012*. Groningen: GION.
- Eekelen, I.M. van, Vermunt, J.D., & Boshuizen, H.P.A. (2006). Exploring teachers' will to learn. *Teaching and Teacher Education*, 22(4), 408-423.
- Elmore, R. (2003). *Knowing the right thing to do: School improvement and performance-based accountability*. WashingtonDC: NGA Center for Best Practices.
- Fullan, M. (2007). *The new meaning of educational change*. New York/London: Teacher College Press.
- Hattie, J. A. C. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London, New York: Routledge.
- Inspectie van het Onderwijs. (2010). *Opbrengstgericht werken in het basisonderwijs*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs. (2012). *Onderwijsverslag 2010/2011*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het onderwijs. (2012b). *Monitor verbetertrajecten taal en rekenen 2008/2009, 2009/2010 en 2010/2011*. Utrecht: Inspectie van het Onderwijs.
- Ledoux, G., Blok, H., & Boogaard, M. (2009). *Opbrengstgericht werken. over de waarde van meetgestuurd onderwijs*. Amsterdam: SCO-Kohnstamm Instituut.
- Mandinach, E.B., Honey, M. & Light, D. (2006). *A theoretical framework for data-driven decision making*. Paper presented at the AERA: San Francisco.
- Ministerie van Onderwijs (2007). *Kwaliteitsagenda Scholen voor morgen*. Den Haag: MOCW.
- Ministerie van Onderwijs (2011). *Actieplan Basis voor Presteren*. Den Haag: MOCW.
- PO Raad (2012). *Op opbrengsten gericht*. Utrecht: PO Raad
- Scheerens, J., Luyten, H. & Steen, R. (2007). *Review and Meta-analyses of School and Teaching Effectiveness*. Enschede: University of Twente.
- Schilt-Mol, T. van & Vijfeijken, M. van (2011). *Taal- en leesverbetertrajecten in het basisonderwijs. De opbrengsten van 3 jaar werken aan taal en lezen*. Tilburg: IVA.
- Verloop, N. (2003). *Onderwijskunde*. Groningen: Noordhoff Uitgevers.
- Visscher, A., & Ehren, M. (2011). *De eenvoud en complexiteit van Opbrengstgericht werken*. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap.
- Vrieling, S., Hogeling, L. & Brukx, D. (2009). *Opiniepeiling kwaliteitsagenda PO*. Nijmegen: ResearchNed.
- Wahlstrom, K.L. & Louis, K.S. (2008). How teachers experience principal leadership: The roles of professional community, trust, efficacy, and shared responsibility. *Educational Administration Quarterly*, 44(4), 458.

Bijlage

Relevante variabelen uit de leerkrachtvragenlijst

	Omschrijving	Range en betrouwbaarheid
Samenwerken	Schaal van 10 items op basis van Wahlstrom & Louis (2008) waarin gevraagd is hoe vaak leerkrachten observeren, feedback geven, uitwisselen en overleggen.	1=(vrijwel) nooit tot 5=1x per week <i>In 2010: $\alpha=0,70$</i>
Beroepsattitude	Schaal van 5 items op basis van Van Eekelen e.a. (2006) om de flexibiliteit waarmee leerkrachten hun lesgeven aanpakken en de mate waarin zij zich verantwoordelijk voelen voor de voortgang van de leerlingen in kaart te brengen. Voorbeelditem: Ik trek het mijzelf aan als een leerling zwak presteert.	1=helemaal oneens tot 5=helemaal eens <i>In 2010: $\alpha=0,69$</i>
Metacognitie	Schaal van 8 items die aangeeft in hoeverre de leerkracht belang hecht aan het aanleren van metacognitieve vaardigheden bij leerlingen	1=helemaal oneens tot 5=helemaal eens <i>In 2010: $\alpha=0,75$</i>
Explicitering doelen	Samengestelde variabele die aangeeft of de leerkracht weet welk niveau hij/zij aan het eind van het jaar met de leerlingen wil behalen en in hoeverre dat geëxpliciteerd is op papier, in de methode, in overleg met team of intern begeleider.	1= Nee, niveau is niet geëxpliciteerd. 2= Ja, de leerkracht weet het voor zichzelf of het staat in de methode. 3=Ja, de leerkracht heeft het voor zichzelf of samen met de intern begeleider opgeschreven of besproken. 4=Ja, het is met het team of op een andere manier besproken/opgeschreven.
Attitude opbrengstgericht werken	Index van 4 items die meet hoe opbrengstgericht een leerkracht is. Voorbeelditem: Ik streef altijd naar een hoger niveau voor mijn leerlingen, ook als ze al goed presteren.	1=weinig tot 4=veel opbrengstgericht
Breed kader	Index die aangeeft hoe breed het kader is waarin de leerkracht de toetsresultaten bekijkt.	1=Geen analyse 2=Basale analyse: de lk maakt een nadere analyse van de resultaten van de zwakke leerlingen en kijkt naar de eerdere toetsresultaten van de leerlingen. 3=Naast de basale analyse vergelijkt de leerkracht de toetsresultaten soms of meestal met de landelijke gegevens en gebruikt minstens 1 van de andere vergelijkingen 4=Uitgebreide analyse
Reflectie	Schaal van 6 items die aangeeft op welke manier de leerkracht terugkijkt na de afname van een toets. Voorbeelditem: Ik kijk terug naar hoe ik instructie	1=nooit, 2=soms, 3=altijd <i>In 2010: $\alpha=0,70$</i>

	heb gegeven	
Basisvaardigheden	Schaal van 3 items die aangeeft hoe belangrijk de leerkracht de nadruk op basisvaardigheden vindt. Voorbeelditem: Ik vind het belangrijk om in het onderwijs een sterke nadruk op rekenen te leggen.	1=helemaal oneens tot 5=helemaal eens <i>In 2010: $\alpha=0,70$</i>
Rapport	Index die aangeeft van hoeveel en welke rapportagemogelijkheden een leerkracht gebruikt.	1=De lk maakt niet systematisch gebruik van de rapportagemogelijkheden. 2=Systematisch gebruik van een van de rapportagemogelijkheden (leerling of groep). 3=Systematisch gebruik van de mogelijkheden op groeps- en leerlingniveau. 4=De lk gebruikt daarnaast de rapportagemogelijkheden van foute antwoorden en/of leerlingen met een eigen leerlijn.
Prestatie-informatie	Schaal van 4 items die meet in welke mate leerkrachten het belangrijk vinden prestatie-informatie te delen met leerlingen. Voorbeelditem: Ik vind het goed om lln te vertellen wat ik hen de komende tijd ga leren.	1=helemaal oneens tot 5=helemaal eens <i>In 2010: $\alpha=0,69$</i>
Dif_niveau	Samengestelde variabele op basis van het aantal niveaugroepen en het aantal leerlingen op een individuele leerlijn. Door de manier van hercoderen geldt <i>niet</i> hoe meer niveaus, hoe beter; bij heel veel individuele leerlijnen en niveaus in de groep krijgt men juist een wat lagere score, omdat blijkt dat men op die manier niet profiteert van de voordelen voor leerlingen en leerkrachten van een (gedeeltelijk) klassikale aanpak.	1=Geen differentiatie in niveaus 2=Matige differentiatie in niveaus 3=Optimale differentiatie in niveaus
Dif_actie	Index van 7 items die meten hoeveel differentiatie-activiteiten een leerkracht neemt als uit een toets blijkt dat een leerling de stof nog niet of helemaal heeft begrepen. Voorbeelditem: ik bied de leerling verrijkingsstof aan	1=nooit tot 3=meestal
Dif_attitude	Index van 4 items die meten hoe de leerkracht over differentiatie denkt. Voorbeelditem: Ik vind dat de leerkracht zijn of haar aandacht moet verdelen tussen zwakke en excellente leerlingen	1=helemaal oneens tot 5=helemaal eens