rijksuniversiteit groningen / gion, groningse instituut voor onderzoek van onderwijs

REVIEW STUDIE 'ONDERWIJSKWALITEIT PO'

Jaap Scheerens & Simone Doolaard

REVIEW STUDIE 'ONDERWIJSKWALITEIT PO'

Een beschrijving en beoordeling van het onderzoek dat in de periode 2009- 2012 is uitgevoerd naar aspecten van de kwaliteit van het basisonderwijs, in opdracht van de BOPO

413-11-002

Jaap Scheerens & Simone Doolaard

18 maart 2013

Managementsamenvatting 1
1 Inleiding 5
2 De beleidscontext van het uitgevoerde onderzoeksprogramma 5
3 Het programma gezien tegen de achtergrond van relevante wetenschappelijke kennis 8
4 De voornaamste bevindingen van het onderzoeksprogramma 13
5 Beleidsimplicaties en suggesties voor verder onderzoek 16
Referenties 19
Bijlage 1 Presentatie en verslag van de discussie op 5 maart 2013 21
Bijlage 2 Case studies over onderwijssystemen die hun prestaties verbeterd hebben, Comprehensive School Reform en 'School aan zet' 25

Managementsamenvatting

In deze integrerende studie worden de hoofdlijnen van de uitkomsten van onderzoek uit de programmalijn 'Onderwijskwaliteit PO' uit het BOPO-programma IV (2009-2012), genaamd Streef, weergegeven en verklaard in het licht van het gevoerde beleid, de uitkomsten van ander (inter)nationaal onderzoek en theoretische kennis. Dit resulteert in een aantal implicaties voor beleid en onderzoek.

Beleidsachtergrond

Traditioneel wordt de organisatie van het onderwijs in Nederland gekenmerkt door 'soevereiniteit in eigen kring'. De basisgedachte was dat autonome scholen het beste in staat zouden zijn om in te spelen op maatschappelijke veranderingen. Het beleid zette in op het liberaliseren van processen, terwijl er tegelijkertijd meer aandacht kwam voor toetsing op resultaten en verantwoording. Vanaf het begin van de jaren tachtig kwam er meer ruimte voor de evaluatiefunctie in het onderwijs. De beleidsfilosofie kreeg hiermee kenmerken van New Public Management, waarbij verantwoordingsverplichtingen als een counter-balance van geliberaliseerde processen functioneren. Hoewel er dus een sterke evaluatiefunctie werd gerealiseerd, werd deze volgens de Parlementaire Commissie Onderwijsvernieuwingen (2008), ook wel Commissie Dijsselbloem genoemd, niet gebruikt om het beleid daadwerkelijk te evalueren en functioneerden de verschillende evaluatievoorzieningen los van elkaar. De belangrijkste voorstellen van de commissie zijn verwerkt in de Kwaliteitsagenda en daama het Actieplan Basis voor Presteren door het invoeren van leerstandaarden in het basisonderwijs en een meer op onderzoekgegevens gebaseerd onderwijsbeleid. In het beleid is geen afstand genomen van het principe om scholen meer autonomie te geven. In het middenveld zijn nieuwe krachtige structuren ontwikkeld, in de vorm van de PO en VO Raad. Deze voeren een eigen innovatiebeleid – net als de al bestaande onderwijsondersteuningsinstellingen - dat gebaseerd is op ontwikkeling vanuit de basis, vanuit het onderwijsveld in plaats van vanuit het beleid.

Wanneer men de situatie gedurende de afgelopen vijf jaar beziet dan is er een hernieuwing van de rationalisering van beleid in de vorm van evidencebased werken zichtbaar en zijn er tegelijkertijd zowel impulsen tot centralisatie als impulsen tot verdergaande decentralisatie en autonomie van onderwijsinstellingen. De beleidsinstrumenten en sturingsmechanismen zijn daarmee ook in twee 'hoofdgroepen' in te delen: de inzet van de infrastructuur van de onderwijsondersteuning gericht op het verbeteren van de eigen innovatiekracht van de scholen die zelf van onderaf innovaties moeten ontwikkelen en de inzet van de infrastructuur van doelen formuleren, toetsen en evalueren om te 'controleren' en te 'sturen/bijsturen'. De vraag is of dergelijke innovaties niet te groot en veelomvattend zijn om voor zo'n groot deel overgelaten te worden aan de innovatiekracht van de individuele school.

Empirische en theoretische achtergrond

Omdat Nederland, sinds de jaren negentig, ruim heeft deelgenomen aan internationaal vergelijkend onderzoek naar de prestaties van leerlingen in het primair onderwijs zijn er op dit punt veel gegevens beschikbaar die vergeleken kunnen worden met bevindingen uit nationaal onderzoek zoals PPON en onderzoek door de onderwijsinspectie. De scores van de Nederlandse leerlingen liggen in internationaal onderzoek bijna altijd boven het gemiddelde. Op de mondiale ranglijst staat Nederland vaak wat lager vanwege de hoge scores in Aziatische landen. Wat betreft ontwikkelingen in de tijd gaat het te ver om te stellen dat er sprake is van een duidelijke achteruitgang in leerprestaties in het basisonderwijs in Nederland, maar er is in ieder geval geen sprake is van vooruitgang. Opmerkelijk is dat Nederlandse

leerlingen in het hogere score segment relatief meer achteruitgang laten zien. De aanwijzigen voor een bescheiden teruggang in leerprestaties zoals die volgens sommigen naar voren komen in de internationale onderzoeken, worden nauwelijks bevestigd door de meer gedetailleerde nationale peilingonderzoeken. Het in algemene zin redelijk positieve beeld van het Nederlandse onderwijs wordt ook bevestigd in de rapportages van de onderwijsinspectie. Het percentage zeer zwakke basisscholen is gedaald tot beneden de 1%. Een zeer hoog percentage scholen haalt de normen op elementaire condities van goed onderwijs. Deze percentages zijn duidelijk lager op indicatoren van adaptief onderwijs en specifieke begeleiding en zorg. Iets minder dan de helft van de basisscholen voldeed op 1 september 2011 aan alle facetten van kwaliteitszorg en opbrengstgericht werken.

Opbrengstgericht werken is een belangrijk speerpunt in het huidige onderwijsbeleid. In onderwijseffectiviteitsonderzoek is van de verschillende kerncomponenten van opbrengstgericht werken als opbrengst- of prestatiegerichtheid, focus op basisvaardigheden, evaluatie- en monitoring, feedback, gestructureerde planning van didactische situaties, differentiatie en taakgerichte samenwerking tussen leerkrachten vaak een positieve associatie met leerprestaties vastgesteld (zie bijvoorbeeld Scheerens & Bosker, 1997). Hetzelfde geldt voor meer ondersteunende condities, zoals de optimalisering van onderwijstijd, onderwijskundig leiderschap en professionele ontwikkeling van leerkrachten, die eveneens onderstreept worden in de beleidsplannen. Opbrengstgericht werken lijkt dus voldoende evidencebased te zijn. Het synthetiseren van de kennis over onderwijseffectiviteit is echter een complexe aangelegenheid. In enkele recente 'state of the art reviews' (Reynolds e.a. 2012, Muijs e.a., 2012 en Hopkins e.a., 2012) kon weliswaar opnieuw een grote consensus over de in aanmerking komende factoren op school- en klasniveau worden geconstateerd, maar over de relatieve impact van de verschillende factoren is veel minder consensus evenmin als over de manier waarop verschillende componenten op elkaar inwerken (Scheerens, 2012). In de onderwijspraktijk komen de componenten vaak in onderlinge samenhang voor, als min of meer samenhangende conglomeraten zoals ook het concept van opbrengstgericht werken. In de resultaten van meta-analyses op het terrein van onderwijseffectiviteit blijkt relatief de meeste empirische support voor prestatiegerichtheid en een goede leerstofkeuze (opportunity to learn); de resultaten voor evaluatie zijn minder eenduidig, hoewel in veel onderzoeken flinke effecten worden gevonden voor de combinatie van evaluatie en instrumentele (in tegenstelling tot puur beoordelende) feedback.

Opzet en resultaten van Streef

Aan Streef heeft een groep van 120 scholen deelgenomen, die allen - in principe – met een bouw aan een van de deelprojecten deelnam en met het hele team en alle leerlingen aan de dataverzameling. Op die manier was elke school vertegenwoordigd in de 'experimentele' groep en in de controlegroep. De oorspronkelijke opzet om bouwen van scholen random toe te wijzen aan een deelproject kon niet gehandhaafd blijven omdat niet voldoende scholen aan Streef wilden deelnemen onder deze conditie. Voor het toeschrijven van eventuele effecten aan de interventie is deze keuze 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 naar waarschijnlijkheid veroorzaakt worden door de

interventie is – gebruik makend van informatie over de scholen voorafgaand aan deelname – nagegaan in hoeverre er sprake is van vahditeitsbedreigers.
De resultaten van Streef wijzen uit dat het trainen en begeleiden van leerkrachten effecten heeft op hun vaardigheid in het analyseren van toetsresultaten, ze bekijken de resultaten in een breder kader en betrekken daar ook hun eigen handelen bij. Ook maken zij meer gebruik van groepsplannen. Er is weliswaar een positieve trend zichtbaar in de prestaties van leerlingen, maar er wordt geen significant verschil gevonden. Een jaar na de uitvoering van het deelproject is de positieve trend niet meer terug te vinden. Het ontbreken van significante, blijvende resultaten bij leerlingen kan te maken hebben met het onderbelichten van de *inhoud* van de groepsplannen (welke beslissingen nemen leerkrachten naar aanleiding van de analyse, kiezen ze de 'juiste remedie') en het feit dat er geen/te weinig borging van de wijze van analyseren heeft plaatsgevonden.
Van het vastleggen van referentieniveaus om bij leerkrachten een duidelijk beeld te hebben waar ze met leerlingen naar toe moeten werken is vooralsnog weinig effect te verwachten; hoewel meer leerkrachten in 2012 bekend zijn met de niveaus is hun verwachting van het effect gedaald en blijken zij niet goed te weten waar de niveaus eigenlijk liggen. Desgevraagd komen zij op zeer uiteenlopende niveaus uit, met name voor de betere leerhngen. Het onderwijs aan de betere leerlingen laat sowieso te wensen over. Hoewel de erkenning dat deze leerlingen de laatste decennia onvoldoende zijn bediend breed wordt gedragen blijkt het voor leerkrachten moeilijk om aan deze groep in de praktijk voldoende tijd en aandacht te besteden. Deels is dit te wijten aan het gebrek aan materiaal waarvan duidelijk is waar het toe leidt en of deze effecten ook behaald kunnen worden, maar voor een belangrijker deel aan de houding ten opzichte van en de vaardigheid in differentiëren. Inspectieonderzoek bevestigt de complexiteit van deze vaardigheden maar ook de urgentie: meer opbrengstgericht werken maakt verschillen tussen leerlingen nog duidelijker en daarmee ook de noodzaak om hier rekening mee te houden. Ook de kennis van en het inzicht in leerlijnen en methoden maakt dat leerkrachten moeite hebben in het schrappen van onderdelen, het uitbreiden van de leerstof en het stimuleren van complexe vaardigheden bij (betere) leerlingen.

Implicaties
De houding van de scholen ten opzichte van de huidige beleidsspeerpunten is relatief positief (in vergelijking met het verleden). Toch is er verbetering mogehjk en heeft het onderzoeksprogramma Streef niet uitgewezen hoe de situatie is bij scholen die zich niet zelf hebben geselecteerd voor onderzoek en ontwikkeling op dit terrein. Ook kan Streef vanwege het loslaten van de randomisering niet met zekerheid bewijzen dat de interventies de effecten (bij leerkrachten) veroorzaken, dit kan hoogstens aannemelijk gemaakt worden. Overigens verdient het wel aanbeveling om te overwegen op welke manier de deelname bereidheid van scholen om aan onderzoek mee te doen gestimuleerd kan worden als men waarde hecht aan evidencebased kennis. Echte experimenten leveren daartoe – hoewel daarover wel discussie is – uiteindelijk (nadat op basis van eerder/ander onderzoek een onderwijskundige theorie geconcretiseerd is in een praktische en overdraagbare interventie) het enige bewijs dat de theorie werkt.
Kennisname van (internationaal) onderzoek en onderwijseffectiviteit onderzoek toont aan dat er evidentie is voor het belang van de huidige speerpunten met betrekking tot opbrengstgericht werken en de aandacht voor goede leerlingen is groot. Voor opbrengstgericht werken is er ook evidentie wat betreft de mogelijk te behalen effecten, evidentie voor excellentie betreft vooral de noodzaak om iets te doen, niet zozeer 'hoe' dat dan zou moeten gebeuren (de inhoud).
Hoewel keuzes in beleidsinstrumenten (m.n. communiceren, stimuleren en subsidiëren) wel te begrijpen zijn in het licht van de historie is het wel de vraag of er echt structurele verbeteringen te verwachten zijn.

Er is een groot 'geloof' in een *bottom-up* benadering, waarbij wordt aangenomen dat de beste resultaten worden bereikt als scholen werken volgens een idee dat zij zelf hebben ontwikkeld. Internationale ervaringen, bijvoorbeeld met *Comprehensive School Reform* programma's in de Verenigde Staten laten zien dat er een alternatief is, dat waarschijnlijk meer doeltreffend en doelmatig is.
Het niveau van de leerkrachten is onvoldoende wat betreft kennis en wat betreft differentiatievaardigheden. De vraag is in hoeverre differentiatie in alle opzichten een haalbaar ideaal is. Zou dit niet eerder ook het werk moeten zijn van experts/onderzoekers/ontwikkelaars? In eerste plaats zou bij leerkrachten in ieder geval de houding aanwezig moeten zijn om te werken met iets dat bewezen effectief is. Ook de vaardigheden in bijvoorbeeld het managen van de klas moeten bij elke leerkracht voldoende zijn om differentiatie in de klas mogehjk te maken, om voldoende *time-on-task* te creëren, die ook nog kwalitatief goed wordt besteed. Voldoende kennis om goede leerstofkeuzes te maken zou eventueel bij ontwikkelaars kunnen liggen (koppeling met methoden en 'recepten' voor oplossing van hiaten, maar toch vraagt dit ook een bepaalde houding van leerkrachten, namelijk om handleidingen te lezen en volgen. Essentieel is ook de houding dat leren leuk is, dat goede prestaties niet 'overdreven' zijn. Tenslotte is het van belang dat de ingezette vernieuwingen beklijven en dat scholen hier zelf sturing geven. Een punt van aandacht is ook het aanbod van verzorgingsinstellingen, waarvan de kwaliteit niet gegarandeerd is. Er is te weinig systematisch, cyclisch werken, te veel ad hoe. Een gunstige conditie is dat het toetssysteem als ruggengraat van opbrengstgericht werken functioneert.

Suggesties voor vervolgonderzoek liggen op het terrein van R&D ten aanzien van ontwikkeling van leermateriaal, haalbaarheid van een alternatief voor de dominante bottom up innovatie en implementatie strategie en onderzoek naar de effectiviteit van de verzorging en ondersteuning bij opbrengstgericht werken en experimenten met *Comprehensive School Reform* benaderingen.

1 Inleiding

Binnen de programmalijn 'Onderwijskwaliteit PO' uit het BOPO-programma IV (2009-2012) hebben het GION/RUG en Cito, in samenwerking met de onderwijsadviesorganisatie Cedin en Marant een samenhangend onderzoeks- en schoolverbeteringsproject genaamd Streef opgezet en uitgevoerd gericht op het verkrijgen van meer kennis over en inzicht in de implementatie en doeltreffendheid van het beleid. Binnen het project zijn deelprojecten uitgevoerd waarbij leerkrachten geschoold en begeleid zijn op het gebied van referentieniveaus, opbrengstgericht werken en differentiatie en excellentie. Effecten zijn vastgesteld op de attitude en het handelen van leerkrachten en de prestaties van leerlingen. In deze integrerende studie worden de hoofdlijnen van de uitkomsten van Streef weergegeven en verklaard in het licht van het gevoerde beleid, de uitkomsten van ander (inter)nationaal onderzoek en theoretische kennis. Vervolgens worden implicaties voor het beleid en vervolgonderzoek besproken.
In paragraaf 2 en 3 wordt achtereenvolgens een kort overzicht gegeven van het gevoerde onderwijsbeleid en van onderzoek op relevante thema's en de theoretische achtergrond. In paragraaf 4 worden de uitkomsten van Streef beschreven en verklaard. In paragraaf 5 komen vervolgens de implicaties daarvan aan de orde.
Op dinsdag 5 maart 2013 zijn de resultaten van deze integrerende studie besproken op de BOPO-dag 2013, ter afsluiting van het programma BOPO IV. Het doel van deze dag was informatie-uitwisseling en discussie tussen onderzoekers enerzijds en BOPO, beleidsmedewerkers van OCW en andere gebruikers anderzijds. In Bijlage 1 is de presentatie opgenomen die op die dag door de onderzoekers is gehouden en het verslag van de daaropvolgende discussie.

2 De beleidscontext van het uitgevoerde onderzoeksprogramma

Traditioneel wordt de organisatie van het onderwijs in Nederland gekenmerkt door 'soevereiniteit in eigen kring', later voorzien van een nieuw 'new public management jasje' zoals verwoord in het akkoord van het zogenoemde Schevenings Beraad Bestuurlijke Vernieuwing (1994), met een korte onderbreking in de jaren 70 waarin een meer constructief beleid de overhand had. Jaren voor het constructieve onderwijsbeleid van Minister van Kemenade verscheen er al een notitie van een van de onderwijskoepels die de titel droeg: 'Bouwen aan de relatief autonome school' (Nederlandse Katholieke Schoolraad, 1979). Tijdens centrum rechtse kabinetten in de jaren tachtig, met onderwijsministers als Pais en Deetman, waren differentiatie, meer vrijheid van scholen en verder gaande autonomie de slagwoorden. Achteraf gezien moet de constructieve onderwijspolitiek slechts als een kortstondige tegenbeweging binnen een doorlopende traditie van schoolautonomie worden gezien. Wel is een zekere golfbeweging te signaleren in de periode van 1970 tot heden. Pijlers van de constructieve onderwijspolitiek waren een centrale aansturing van onderwijsinnovatie en rationalisering in de zin van de ambitie om het beleid wetenschappelijk te onderbouwen. De ervaringen met de innovatie-experimenten, zoals het Middenschool Experiment, wezen uit dat hier weinig van terecht kwam, zo bleek het zelfs niet mogelijk de proefprogramma's fatsoenlijk te evalueren. In de jaren tachtig en negentig was, zoals gezegd autonomie weer volop troef, en waren er weinig ambities om te komen tot een meer wetenschappelijk onderbouwd onderwijsbeleid. De Parlementaire Commissie Onderwijsvernieuwingen (ofwel de Commissie Dijsselbloem) legde wat dit betreft de vinger op de zere plek. Het Studiehuis werd ontmaskerd als niet berustend op solide wetenschappelijke inzichten en van een empirische effect evaluatie was opnieuw geen sprake. In de periode na Dijsselbloem trad een nieuwe fase in, waarbij deels werd teruggegrepen op het

rationaliteitideaal dat ook aanwezig was in de constructieve onderwijspolitiek, onder de noemer van evidencebased beleid. Nieuw beleid zou gebaseerd moeten zijn op wetenschappelijke evidentie, het liefst verkregen op basis van 'echte' experimenten, dat wil zeggen, onderzoeksdesigns met aselecte toewijzing aan behandelingscondities. Aan decentrahsatie en autonomie van scholen werd niet getornd, integendeel, de autonomie van scholen werd zo mogelijk nog uitgebreid en sinds jaren is Nederland, blijkend uit OESO gegevens, het land met de grootste schoolautonomie ter wereld. Toch werden er onder de regimes van de onderwijsministers Plasterk en Van Bijsterveld tegelijkertijd impulsen gegeven die meer centrale sturing inhielden. Enerzijds hadden deze vooral op de evaluatie functie betrekking, invoering van referentieniveaus, discussie over leerstandaarden en het voornemen tot verplichtstelling van een eindtoets basisonderwijs. Anderzijds werden duidelijke beleidsprioriteiten gesteld in zogenoemde kwaliteitsagenda's en actieplannen. Hierin wordt een accent gelegd op het verbeteren van prestaties in basisvakken als rekenen en taal, met tevens een toenemende aandacht voor prestaties in het hogere segment van de verdeling, onder de noemer van de bevordering van excellent presteren. Veel van de prioriteiten uit de kwaliteitsagenda's komen aan bod in het concept van 'opbrengstgericht werken'. Tevens werden nieuwe instrumenten ingezet, zoals het beleid met betrekking tot doorlopende leerlijnen, het onderzoeksprogramma Onderwijs Bewijs en de oprichting van een speciaal instituut (netwerk) ter bevordering van evidencebased onderwijs en beleid (TIER). In Tabel 1, ontleend aan Scheerens en Ehren (2009), wordt de zojuist beschreven situatie samengevat en nader geïllustreerd.

Tabel 1. Beleidsinstrumenten en rationalisering van centrale en decentrale impulsen van het onderwijsbeleid (ontleend aan Scheerens & Ehren, 2009)

	Centrale impulsen	Decentrale impulsen
Beleidsinstrumenten	Concretere eindtermen PO Leerstandaarden Maatschappelijke stage Kwahteitsagenda, doorlopende leerlijnen Opbrengstgericht werken	Arbeidsvoorwaarden Lumpsum financiering Bottom up onderwijsinnovatie gestimuleerd door VO en PO raad
Rationalisering	Directie Kennis TIER (Netwerk ten behoeve van evidencebased beleid) PROO (Programma Raad Onderwijskundig Onderzoek) Onderwijsinspectie Cohort Studies Internationaal Vergelijkend Assessment onderzoek (TIMSS, PISA) Examens	Praktijkgericht onderzoek geïnitieerd door VO en PO Raad R&D van Landelijke Pedagogische Centra

In het beleid gericht op innovatie en kwaliteitsverbetering wordt enerzijds gekapitaliseerd op de autonomie van scholen, maar is er anderzijds sprake van de zojuist genoemde meer centralistische aspecten. De spanning tussen deze twee impulsen komt bijvoorbeeld aan de orde, door in de kwaliteitsagenda's en innovatie filosofie enerzijds te kapitaliseren op *bottom up* verandering, geleid door autonome scholen en netwerken van scholen, en anderzijds *evidencebased* werken, liefst op basis van wetenschappelijk verantwoorde beleidsexperimenten, te propageren. Dergelijke experimenten vragen een schoolexteme regie, interventies die een zekere uniformiteit hebben, en gestandaardiseerde gegevensverzameling. Zaken die moeilijk te rijmen zijn met ontwikkeling van onderop.

Het recent gestarte programma 'School aan Zet', biedt mogelijk een nieuwe variant in de dynamiek tussen centrale impulsen, de sturing van de kwaliteitsverbetering op schoolniveau en de autonomie van scholen. De school aan zet benadering is enerzijds gericht op Opbrengstgericht werken en anderzijds op HRM/de school als lerende organisatie. Op de website wordt aangegeven dat experts scholen bezoeken om de scholen te assisteren bij zaken waar men al mee bezig is. De rol van de experts is die van kritisch meedenker ('critical friend'). Het programma biedt mogelijkheden tot professionele ontwikkeling, master classes, en biedt handreikingen wat betreft instrumenten en hulpmiddelen; tevens is het de bedoeling dat goede praktijken uitgewisseld worden.
De benadering bij School aan Zet heeft meer gemeen met de vigerende schoolbegeleidingspraktijk in Nederland dan met de aanpak in Comprehensive School Reform Programs. Het belangrijkste verschil is dat er geen sprake is van externe, wetenschappelijk beproefde programma's. Uit het overzicht van de experts ontstaat de indruk dat deze vooral knowhow hebben op schoolorganisatie niveau, terwijl het vakdidactische element minder vertegenwoordigd lijkt te zijn. Niet duidelijk is of de voorziene monitoring en evaluatie zodanig ontworpen is dat het succes van het overheidsbeleid toetsbaar is. Evenmin of er uitgewerkte plannen zijn, bijvoorbeeld wat betreft de manier waarop opbrengstgericht werken en het functioneren van de school als lerende organisatie op een geïntegreerde manier vorm moeten krijgen. Zie ook Bijlage 2.

Een iets andere insteek om de structurele voorzieningen die deel uit maken van de beleidscontext te beschrijven is te kijken naar enerzijds de ondersteuningsstructuur voor het onderwijs, die een specifieke functie heeft bij onderwijsinnovatie en anderzijds naar de voorzieningen op het terrein van evaluatie, monitoring en assessment.

Wat betreft onderwijsondersteuning heeft de verzorgingsstructuur een specifieke plaats. Traditioneel werkt deze structuur in dienst van de scholen en wordt niet aangestuurd door hogere bestuurlijke niveaus. Dit betekent dat deze verzorgingsstructuur niet specifiek inzetbaar is voor het aanpakken van school of netwerk overstijgende kennisvragen die aan de orde zijn bij het realiseren van idealen als opbrengstgericht werken.

In een onderwijsstelsel dat een gedecentraliseerd karakter heeft, met veel autonomie van scholen en leerkrachten, moet evaluatie niet alleen als een controle mechanisme worden gezien, maar ook als een sturend en bijsturend gegeven. Dergelijke evaluatie gerelateerde sturing vindt in principe op meerdere niveaus plaats: op systeem niveau, op schoolniveau en op klasniveau. Feedback kan worden gebruikt door een hogere bestuurslaag om beleidsbeslissingen te onderbouwen, maar ook op 'horizontaal' niveau kan feedback worden gebruikt voor het bijstellen van plannen en handelingen. In het eerste geval is er sprake van externe evaluatie in het kader van 'accountability', in het tweede geval van interne evaluatie, als onderdeel van verbetering en leren. Zeker wanneer beleid expliciet gericht is op verhoging van kwaliteit, ligt het voor de hand dat peiling en evaluatie van kwaliteit een centrale rol heeft in dit beleid. Het is zelfs mogelijk om evaluatie als motor van kwaliteitsverbetering te zien. Bij alle autonomie heeft het Nederlandse onderwijs toch enkele belangrijke meer centralistische kenmerken, namelijk met de centrale examens in het voortgezet onderwijs en de (bijna) nationale toepassing van de Cito-eindtoets in het basisonderwijs. Volgens economen als Bishof en Woessmann (Woessmann e.a., 2009) hang het hebben van centrale examens positief samen met het prestatieniveau van landen op internationale assessment toetsen, waarbij men ertoe neigt deze relatie als causaal te interpreteren.

In het kader van een overzichtsstudie ten behoeve van de OESO schetsen Scheerens (2012) een beeld van de evaluatie voorzieningen in Nederland, gedifferentieerd naar systeem evaluatie, school evaluatie, leerkracht evaluatie en leerling evaluatie. Alles bij elkaar blijkt er in Nederland sprake van een breed scala aan procedures en instrumenten voor systeem, school en leerling evaluatie, terwijl leerkracht evaluatie nauwelijks ontwikkeld is. Samenvattend kan worden gesteld dat de kwaliteitsverbeteringsfunctie van evaluatie, examinering en assessment op de volgende punten tot uitdrukking komt:

- het formuleren van toetsbare standaarden die als richtinggevend voor het onderwijs functioneren;
- het stimuleren van prestatie motivatie op het niveau van individuele leerlingen (om te slagen voor een examen);
- het stimuleren van opbrengstgerichtheid bij scholen en eventueel ook leerkrachten;
- het stimuleren van leerprocessen op meerder niveaus, op basis van feedback en doelgerichte verbeteringsactiviteiten

De gesignaleerde tweespalt in het op kwaliteitsverbetering gerichte beleid, met enerzijds veel nadruk op autonomie en anderzijds zekere impulsen tot centrale sturing wordt versterkt door de invloed van de ondersteuningsstructuur, die sterk geënt is op ontwikkeling van onderop, en door de voorzieningen op het terrein van evaluatie, assessment en monitoring die meer aangrijping bieden voor centrale aansturing.
Aangenomen wordt dat de thema's die in de programmalijn 'Kwaliteit' van de BOPO zijn onderzocht, evenals de manier waarop het onderzoek kon worden uitgevoerd, beïnvloed zijn door de hierboven beschreven kenmerken van de beleidscontext.

3 Het programma gezien tegen de achtergrond van relevante wetenschappelijke kennis

De onderzoekslijn van de BOPO die in deze studie centraal staat heeft betrekking op het bevorderen van de kwaliteit van het basisonderwijs. In dit onderdeel wordt dit thema in een bredere context geplaatst door te kijken naar wat er uit ander, nationaal en internationaal onderzoek bekend is over de kwahteit van het Nederlandse basisonderwijs en naar onderzoeksgegevens waarin kwaliteitsbevorderende factoren centraal staan. Hierbij wordt ingegaan op de vraag in hoeverre de internationale kennis verdieping en verklaring biedt voor de Nederlandse context. De voornaamste kennisbronnen die hierbij worden gebruikt zijn enerzijds evaluatie en assessment onderzoek en anderzijds onderzoek naar onderwijseffectiviteit.

Evaluatie- en assessment onderzoek
De scores van de Nederlandse leerlingen liggen in internationaal onderzoek bijna altijd ruim boven het gemiddelde. Op de, mondiale ranglijst staat Nederland vaak wat beneden de absolute top vanwege de hoge scores in landen als Japan, Singapore, Korea en Taiwan. Hoewel er bij recente uitvoeringen van de internationale onderzoeken sprake is van een lichte achteruitgang bij sommige vakken, kan dit niet als een duidelijke trend worden gezien (Luyten, 2011).De aanwijzingen voor een teruggang in leerprestaties zoals die voor een deel naar voren komen in de internationale onderzoeken, worden nauwelijks bevestigd door de meer gedetailleerde nationale peilingonderzoeken (PPON) die sinds 1987 worden uitgevoerd door Cito. Alleen voor het domein rekenen/wiskunde zijn duidelijke ontwikkelingen waar te nemen. De prestaties op basale onderwerpen zijn duidehjk verbeterd, maar daar staat een even sterke negatieve trend tegenover voor bewerkingsopgaven. Bij de interpretatie van deze gegevens zou men eventueel rekening

kunnen houden met het gegeven dat de omvang van het speciaal basisonderwijs in de loop der tijd is afgenomen, terwijl het totale leerhngenbestand is toegenomen. Het in algemene zin redelijk positieve beeld van het Nederlandse onderwijs zoals dat uit internationale onderzoeken en de peilingonderzoeken van Cito naar voren komt wordt ook bevestigd in de onderwijsverslagen van de inspectie (2009, 2012). De percentages zeer zwakke scholen zijn in op de peildatum van 1 september 2011 gedaald tot beneden de 1%, zowel in het primair als het secundair onderwijs. Verder blijkt dat een zeer hoog percentage de normen haalt op opbrengstindicatoren en op elementaire condities van goed onderwijs. Deze percentages zijn duidelijk lager op indicatoren die te maken hebben met adaptief onderwijs en specifieke begeleiding en zorg. Iets minder dan de helft van de basisscholen voldeed op 1 september 2011aan alle facetten van kwaliteitszorg en opbrengstgericht werken. Voor een ruimere analyse van de kwaliteit van het Nederlandse basis- en voortgezet onderwijs zij verwezen naar Luyten (2011).
Opmerkehjk is overigens dat Nederlandse leerlingen in het hogere score segment op sommige internationale toetsen, door de tijd gezien achteruitgang laten zien, dit bleek speciaal bij lezen in het basisonderwijs. Vermeer en Van der Steeg (2011, p 16-19) hebben de verandering in de scores op internationale toetsen berekend voor elk percentiel van de scoreverdeling. Op die manier kan worden nagegaan of de verandering consistent is op de verschillende scoreniveaus of juist verschilt per scoreniveau. Het algemene beeld is dat Nederlandse leerlingen aan de onderkant van de verdeling hun internationale positie behouden, maar dat er in het hoger scorende segment van de Nederlandse leerlingen sprake is van achteruitgang. Deze trend doet zich zowel in het voortgezet onderwijs voor als in het primair onderwijs. De grootste achteruitgang in het hogere segment wordt aangetroffen in het vak lezen binnen het primair onderwijs (vergelijking PIRLS 2001 en 2006). In PIRLS 2006 zit de onderste 25% van de Nederlandse scoreverdeling in de internationale top 5. Maar bij leesvaardigheid presteert de bovenste 25% niet meer in de top 10. Hierbij presteren dus ook de modale leerlingen minder dan hun internationale collega's.
De meest recente resultaten op internationaal vergelijkend assessment onderzoek; gegevens op het terrein van leesvaardigheid in het basisonderwijs, zoals onderzocht in PIRLS (2011) en op natuuronderwijs en rekenen wiskunde (TIMSS, 2011), laten een stabiel hoog niveau zien voor lezen en natuuronderwijs, en een significante achteruitgang voor rekenen en wiskunde, in vergelijking met 1995. Ook op deze meest recente toetsen wordt het beeld van zeer sterk scoren in het lagere segment, maar duidelijk minder sterk in het hogere score segment bevestigd.

Deze, deels internationale, gegevens over de kwahteit van het basis- en voortgezet onderwijs in Nederland zijn informatief omdat ze de gerichtheid op verdere prestatieverbetering, in het vigerende onderwijsbeleid, in perspectief plaatsen. Kort gezegd komt dit erop neer dat de kwaliteit van het onderwijs in Nederland al over een redelijk lange periode, tien a vijftien jaar, goed is, maar dat handhaving van deze positie misschien in het geding komt gezien lichte daling op enkele internationale projecten. Uiteraard wordt het op prestatieverbetering gerichte beleid zeer expliciet gemotiveerd wanneer men streeft naar het scoren binnen de internationale top vijf van de rangordes op internationale toetsen. Meer recente analyses die wijzen op een minder riante positie van Nederland in het hoogste segment van de scoreverdehng kunnen worden gezien als ondersteuning voor het streven om meer aandacht te besteden aan 'excellentie'.
Hoewel de gepresenteerde, deels internationale, gegevens over de kwaliteit van de opbrengsten van het Nederlandse basis- en voortgezet onderwijs dus een goede basis bieden voor een meting en beoordeling van kwaliteit, zeggen ze op zichzelf nog niets over een verklaring van het kwaliteitsniveau. Het blijkt

bijzonder moeilijk om op basis van de internationale assessment onderzoeken, zoals TIMSS en PISA te weten te komen hoe het komt dat het ene land zoveel beter presteert dan het andere. De soms bijna anekdotische verklaringen van de uitstekende prestaties van Finland illustreren dit. Over de goede prestaties van Nederland bestaan zeker ook hypothesen, (vergelijk bijvoorbeeld Doebert e.a., 2004), waarbij onder meer examens en goed ontwikkelde interne evaluatie op scholen zijn genoemd. Echter, analyses die de gegevens uit de achtergrond vragenlijsten uit PISA in verband hebben gebracht met de leerprestaties, bieden weinig houvast (Luyten, e.a. 2005; Scheerens, 2013). De schoolvariabele die het nog relatief het beste doet in PISA is een ordelijk schoolklimaat.
Om meer inzicht te krijgen in de invloed van door het beleid en de school te beïnvloeden condities op leerprestaties is daarom een andere kennisbasis, namelijk die van onderwijseffectiviteit onderzoek de aangewezen bron. Het op kwaliteit gerichte overheidsbeleid in het basisonderwijs, zoals dat is geformuleerd in de Kwaliteitsagenda en het Actieplan, evenals het concept van opbrengstgericht werken, kapitaliseren op door de school beheersbare factoren, die ook in onderwijseffectiviteit onderzoek veelvuldig aan de orde zijn gekomen. Een reflectie op de kennisbasis die op dit onderzoek gebaseerd is, biedt daarmee (post hoe) een globaal beeld van de empirische onderbouwing van dit beleid, maar zo mogelijk eveneens nadere indicaties over de verwachtingen die men van sommige beleidsmaatregelen mag hebben.

Onderzoek naar onderwijseffectiviteit
Een poging om een samenvatting te geven van de stand van zaken van de resultaten van onderzoek naar onderwijseffectiviteit is een onderneming die, gezien de omvang van het materiaal en het feit dat het hier om een 'bewegend doel' gaat, hooguit een partieel resultaat kan opleveren. De snelste manier om een beeld te krijgen is zich te baseren op meta-analyses. Daarbij valt vervolgens op dat er, voor vrijwel alle factoren, grote verschillen zijn in de uitkomsten tussen meta-analyses.

In Tabel 2 zijn resultaten van meta-analyses van Hattie, (2009), Scheerens e.a., (2007), Creemers en Kyriakides (2008) en Seidel en Shavelson, (2007), evenals uitkomsten van TIMSS, naast elkaar gezet.
De coëfficiënten in Tabel 2 zijn te interpreteren als correlaties. Volgens geldende normen moeten correlaties beneden de .20 als klein worden opgevat, .50 als gemiddeld en .80 als hoog. Over de onderwijskundige betekenis van de 'kleine' effecten is echter verdere discussie mogelijk, door deze bijvoorbeeld te vergelijken met de vooruitgang gebaseerd op een schooljaar (Scheerens e.a., 2007)
De analyse van Hattie (2009) is gebaseerd op een synthese van niet minder dan 800 meta-analyses. Wat opvalt, is dat de effectgroottes die hij rapporteert voor de meeste variabelen aanmerkelijk hoger liggen dan in de drie andere meta-analyses, waarvan resultaten vermeld zijn in Tabel 2. Een mogelijke verklaring die Hattie hiervoor aangeeft, in een vergelijking van sommige van zijn resultaten met die van Seidel en Shavelson, is dat in dit onderzoek rekening is gehouden met achtergrondvariabelen van de leerlingen. Dit suggereert dat dit in het onderzoek waarover hij rapporteert niet gebeurd is (Hattie, 2009, p 202). Een problematisch aspect van het werk van Hattie is dat dit soort methodologische punten niet te verifiëren zijn op basis van zijn rapportage. Het is daarom, zonder replicatie onderzoek, niet mogelijk te verklaren waarom er zulke grote verschillen zijn in de uitkomsten van meta-analyses. Over de uitkomsten vermeld in Tabel 2 valt op te merken dat Hattie's resultaten, zeker voor de instructievariabelen, naar algemene conventies, tenderen naar 'medium' en die van de andere auteurs klein, tot verwaarloosbaar. Variabelen waarover een zekere mate van overeenkomst bestaat tussen de meta-analyses in Tabel 2 zijn een ordelijk klimaat en het onderwijzen van leerstrategieën.

Tabel 2. Resultaten van recent meta-analyses

Schoolfactoren	Scheerens e.a., 2007	Hattie, 2009	Creemers & Kyriakides, 2008	> 50% significantie in internationaal vergelijkend onderzoek
Consensus & cohesie	.02	-	.16	X
Ordelijk klimaat	.13	.17	.12	X
Monitoren & evalueren	.06	.31	.18	X
Curriculum/gelegenheid tot leren	.15	-	.15	X
Huiswerk	.07	.15	-	
Effectieve leertijd	.15	.19	-	
Betrokkenheid ouders	.09	.25	-	
Prestatiegerichtheid	.14	-	-	
Schoolleiderschap	.05	.18	.07	
Differentiatie	.02	.06	-	
Leerkracht-/klasfactoren	Idem	Idem	Seidel & Shavelson, 2007	
Leertijd	.08	.17	.03	X
Klassenmanagement	.10	.26	.00	
Gestructureerd onderwijs	.09	.30	.02	X
Aandacht voor strategieën	.22	.35	.22	
Constructivistisch onderwijs	.14	.37		
Klasklimaat	.12	.27		
Feedback & monitoren	.07	.33	.01	

Coefficients are correlations; as Hattie presents effect sizes in terms of d, these were divided by two to arrive at comparable estimates. Column 5 shows a cross when a specific variable addressed in secondary analyses based on TIMSS 1995, TIMSS 1999 and TIMSS 2003 was found significant for 50% or more of the number of times it was used; based on TIMSS 1995, TIMSS 1999 and TIMSS 2003. Sources Van der Kley, 2009: Scheerens, e.a. 2007. Seidel & Shavelson, 2007

Wat de uitkomsten betreft bestaat een zekere consensus over het positieve effect van 'curriculum/gelegenheid tot leren', dat wil zeggen dat er een goede aansluiting is tussen de leerstof die onderwezen en de leerstof die getoetst wordt en tevens over het effect van het leren van leerstrategieën. Factoren die in het vigerende op kwaliteit gerichte beleid een rol hebben zijn verder prestatie gerichtheid, effectieve leertijd en vooral ook evaluatie, feedback en monitoring. Tenslotte spelen ook samenwerking en differentiatie een rol. Voor prestatiegerichtheid en tijd worden kleine maar relatief niet onbelangrijke effecten gevonden; de uitkomsten voor evaluatie en feedback zijn wisselend, voor samenwerking en differentiatie zijn de effecten zeer klein.

Enkele recente meta-analyses (Lomos, Hofman & Bosker, 2011; Scheerens, 2012; Hendriks e.a., 2012; Scheerens, 2013) bieden eveneens relevante informatie over variabelen die in het kwaliteitsbeleid en opbrengstgericht werken een belangrijke plaats hebben.

Het onderzoek van Lomos, e.a. (2011) had betrekking op de effecten van professionele leergemeenschappen. In een meta-analyse, uitgevoerd op 5 onderzoeken, vonden de onderzoekers een gemiddelde effect-grootte van d=.25; te vergelijken met een correlatie van .13. Er werden vijf kernfacetten van professionele leergemeenschappen onderscheiden: reflecterende dialoog, 'ontprivatisering' van de onderwijspraktijk, feedback die betrekking heeft op instructie, gedeelde doelen en een collectieve gerichtheid op het leren van leerlingen. In de 5 onderzoeken werden de hierboven genoemde kernfacetten van professionele leergemeenschappen redelijk teruggevonden. Wel liepen de operationaliseringen enigszins uiteen. In elk van de vijf operationaliseringen was het facet samenwerken aanwezig, terwijl in vier van de vijf operationaliseringen tevens sprake was van enigerlei oriëntatie op

leerprestaties (interpretatie J.S.). Zoals door de auteurs wordt opgemerkt zijn deze twee facetten ook afzonderlijk onderzocht in meta-analyses, waarbij bleek dat het effect van de samenwerkingscomponent beduidend lager is dan voor prestatie-gerichtheid (Scheerens & Bokser, 1997, vonden een effectgrootte (r) van respectievelijk .03 en .14). In de eerder besproken meta-analyse van Scheerens e.a. (2007) was de effectgrootte voor samenwerking r=.02 en voor prestatiegerichtheid r=.14.
In het onderzoek van Scheerens (2012), werd de gemiddelde effectgrootte van schoolleiderschap vastgesteld, op basis van 14 onderzoeken, waarin de effecten via zogenoemde indirecte effectmodellen waren vastgesteld. De gemiddelde effectgrootte die werd gevonden was .06.
In het (nog lopende) onderzoek van Hendriks, Scheerens en Steen, naar de effecten van evaluatie en assessment op school en klasniveau, werden de volgende effectgroottes gevonden: .08 voor het effect van evaluatie op schoolniveau, gebaseerd op 10 onderzoeken, .07 voor het effect van evaluatie op klasniveau (18 onderzoeken) en .03 voor het effect van toetsen (13 onderzoeken). Uit dit onderzoek kwam tevens naar voren dat relatief grotere effecten werden gevonden, naarmate de evaluatie gepaard ging met feedback. Dit beeld wordt versterkt in resultaten van andere meta-analyses, onder meer die van Hattie (2009) en Kluger en DeNisi, 1996, waarin aanmerkelijk hogere effectgroottes werden gevonden (correlaties in de orde van .25 tot .50).
In een recent afgeronde review studie over de effecten van instructie tijd op school, huiswerk en onderwijs buiten de officiële lesuren, rapporteren Scheerens (2013) gemiddelde effectgroottes die in eerdere meta-analyses over onderzoek naar deze factoren zijn gevonden. Deze zijn respectievelijk .18, .15 en .09 (uitgedrukt als correlaties). In de eigen meta-analyses werden aanmerkelijk kleinere effecten gevonden, namelijk van .05 voor instructietijd op school en eveneens .05 voor huiswerk (voor buitenschools leren waren er onvoldoende bruikbare onderzoeken voor het uitvoeren van een kwantitatieve meta-analyse).

De resultaten van meta-analyses op het terrein van onderwijseffectiviteit, zoals hierboven gepresenteerd laten zien dat de meeste empirische support wordt gevonden voor prestatiegerichtheid en een goede leerstofkeuze (curriculum/gelegenheid tot leren); de resultaten voor evaluatie zijn minder eenduidig, hoewel in veel onderzoeken flinke effecten worden gevonden voor de combinatie van evaluatie en instrumentele (in tegenstelling tot puur beoordelende) feedback. Samenwerking blijkt vooral effect te sorteren wanneer er sprake is van prestatiegerichte, taakgerichte samenwerking. Over het algemeen blijkt instructietijd wel een betekenisvol effect te hebben, maar in de praktijk is dit effect moeilijk te scheiden van het goed gebruik van deze tijd, met name door een goede leerstofkeuze (curriculum/gelegenheid tot leren) en goede instructie. De minste empirische ondersteuning wordt gevonden voor schoolleiderschap en differentiatie.
Wanneer de resultaten van onderwijseffectiviteit onderzoek worden weergegeven op basis van correlaties (of andere effectmaten) tussen afzonderlijke variabelen en leerprestaties, wordt in feite een sterk gestileerd beeld van de werkelijkheid gegeven. In de onderwijspraktijk komend deze variabelen meer in onderlinge samenhang voor, als min of meer samenhangende conglomeraten. Het aardige is dat in nogal wat operationele definities van op het oog simpele concepten als tijd, leiderschap, evaluatie en professionele leergemeenschappen deze complexiteit 'binnensluipt' doordat de concepten in feite meerdimensionaal gedefinieerd zijn. Bij evaluatie, assessment en monitoring lijkt de effectgrootte afhankelijk van de 'toevoeging' van feedback, en eventueel zelfs het preluderen op gerichte actie, volgende op die feedback. Bij leertijd wordt door sommige onderzoekers het begrip 'academische leertijd' gebruikt, dit veronderstelt dat de tijd besteed wordt met het aanbieden van leerstof met het juiste

moeilijkheidsniveau. In feite gaat het daarbij dus om een conglomeraat van tijd, aangeboden leerinhouden en differentiatie. In de literatuur over schoolleiderschap worden eveneens vaak multidimensionale constructen gebruikt, zoals geïntegreerd schoolleiderschap en verdeeld (distributief) schoolleiderschap. Er zijn enige aanwijzingen dat een smallere invulling van onderwijskundig leiderschap wellicht het meest effectief is (bijvoorbeeld Robinson e.a., 2008; Hattie, 2009), maar de gebrekkige kwaliteit van het meeste onderzoek op dit terrein maakt zo'n conclusie hachelijk. In het onderzoek naar professionele leergemeenschappen blijken samenwerking en prestatiegerichtheid hoofdcomponenten, met daarbij de kanttekeningen dat samenwerking op zich meestal maar geringe effecten laat zien, terwijl die voor prestatiegerichtheid wat groter zijn.

De bovenstaande constateringen die verwijzen naar meer dimensionale conglomeraten van effectiviteitbevorderende factoren sluiten goed aan bij de analyse van opbrengstgericht werken. Het concept van opbrengstgericht werken is eveneens een conglomeraat van verschillende effectiviteitbevorderende condities. Deze worden bijvoorbeeld helder beschreven in Visscher en Ehren (2011). Wanneer we bij opbrengstgericht werken uitgaan van een retro-actieve cyclus, wordt er begonnen met toetsing van leervorderingen, en het interpreteren daarvan, onder meer in termen van leerstof die individuele leerhngen minder of meer beheersen. Het bieden van handreikingen aan docenten om dit soort interpretaties te doen, kan worden gezien als feedback. Het gebruik maken van deze diagnostische informatie om concrete doelen en targets te stellen voor individuele leerlingen of voor leergroepen heeft te maken met een focus op leren en leerprestaties. Voor zover dit binnen scholen in samenwerking tussen leerkrachten, gestimuleerd door de schoolleiding gebeurt, zijn we op het terrein van instructie gecentreerde professionele leergemeenschappen en onderwijskundig leiderschap. De didactische analyse die nodig is om (eventueel gezamenlijk) te zoeken naar de beste benaderingen en methoden om de gestelde concrete doelen te bereiken, appelleert aan vakdidactische kennis, '*pedagogicalcontentknowledge*' inclusief kennis over instructiestrategieën. Op basis van de individuele toetsgegevens is in principe ook individualiseren, differentiërend en adaptief onderwijs mogelijk, waarbij differentiatie in tempo en beschikbaar te stellen leertijd een van de opties is.

Bij opbrengstgericht werken komt dus een hele serie effectiviteitbevorderende kenmerken aan de orde. De meest prominente zijn: prestatiegerichtheid als focus, het gebruiken van evaluatie en feedback, doelgericht werken en didactische analyse. Samenwerking en onderwijskundig leiderschap kunnen worden gezien als ondersteunende condities.

4 De voornaamste bevindingen van het onderzoeksprogramma

Het GION/RUG en Cito hebben een samenhangend onderzoeksprogramma opgezet, in samenwerking met twee onderwijsadviesorganisaties waarin drie interventiestudies zijn uitgevoerd waarvan de effecten vastgesteld zijn op basis van gegevens van twee dataverzamelingsprojecten, een bij leerlingen en een bij leerkrachten en schoolleiders. Waarbij de laatst genoemde ook een aantal op zichzelf staande onderzoeksvragen heeft. De drie interventiestudies betreffen trainings- en begeleidingstrajecten voor teams van leerkrachten op het terrein van de drie beleidsspeerpunten; het werken met referentieniveaus, opbrengstgericht werken en excellentie. Werving en communicatie naar de scholen heeft plaatsgevonden onder de naam 'Streef', een schoolverbeterings- en onderzoekproject.

Voor het samenhangende onderzoeksprogramma zijn de volgende hoofdvragen geformuleerd:

1. Hoe kijken leerkrachten aan tegen de referentieniveaus? Kunnen ze ermee werken?
2. Welke professionaliseringsactiviteiten ondernemen (teams van) leerkrachten in het kader van een kwaliteitsverhoging? En wat zijn de effecten op de gecijferd- en geletterdheid van de leerlingen?
3. Hoe gaan leerkrachten om met de gevraagde extra nadruk op de basisvaardigheden taal/lezen en rekenen? Wat zijn de effecten hiervan op de gecijferd- en geletterdheid van de leerlingen? En wat zijn de eventuele neveneffecten voor andere domeinen?
4. Welke activiteiten ondernemen scholen om excellentie (niet alleen cognitief opgevat) bij leerlingen te bevorderen, en wat zijn de effecten hiervan op de ontwikkeling bij de betreffende leerlingen?

Aan Streef heeft een groep van 120 scholen deelgenomen, die allen met in principe één bouw aan één van de deelprojecten deelnam en met het hele team en alle leerlingen aan de dataverzameling. Door dit gekruiste design was elke school vertegenwoordigd in de 'experimentele' groep en in de controlegroep. De oorspronkelijke opzet om bouwen van scholen aselect toe te wijzen aan een deelproject kon niet gehandhaafd blijven omdat niet voldoende scholen aan Streef wilden deelnemen onder deze conditie. Voor het toeschrijven van eventuele effecten aan de interventie is deze keuze problematischer, immers, de motivatie om aan het deelproject mee te doen, en het feit dat het project inhoudelijk aansluit bij de schoolontwikkeling kunnen een bijdrage hebben gehad aan eventuele effecten. Het loslaten van de randomisering leidt er dus toe dat hoogstens aannemelijk gemaakt kan worden dat de interventies de effecten (bij leerkrachten en leerlingen) veroorzaken. Alleen echte experimenten kunnen uiteindelijk (nadat op basis van eerder/ander/quasi-experimenteel onderzoek een onderwijskundige theorie geconcretiseerd is in een praktische en overdraagbare interventie) het bewijs leveren dat de theorie werkt (Onderwijsraad, 2006; Bosker, 2008a, 2008b; Gravemeijer & Kirschner, 2007, 2008).
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. Door het loslaten van de aselecte toewijzing is een situatie gecreëerd waarin scholen 'extra' gestimuleerd werden om de beleidsspeerpunten te implementeren zodat inzicht verkregen kon worden in de vraag hoe leerkrachten daar in een optimale situatie mee omgaan en van profiteren. Optimaal omdat ze zelf gekozen hebben voor deelname en er kosteloos bij begeleid werden. Heeft een dergelijke interventie geen effect dan is de kans dat stimulerend overheidsbeleid – op afstand – wel effect heeft klein.

In het deelproject over het werken met referentieniveau is met leerkrachten zelf ervaring opgedaan in het formuleren van niveaus en zijn prestaties van hun eigen leerlingen vergeleken met de vastgestelde en de zelf opgestelde niveaus. In het deelproject over opbrengstgericht werken zijn leerkrachten getraind en begeleid in het analyseren van toetsresultaten van hun eigen leerlingen en het vertalen daarvan naar een plan voor het onderwijs na de toets. In het deelproject Differentiatie en excellentie zijn leerkrachten van 9 scholen een schooljaar lang gevolgd in en geadviseerd over het werken met excellente leerlingen.

De resultaten van Streef wijzen uit dat het trainen en begeleiden van leerkrachten effecten heeft op hun vaardigheid in het analyseren van toetsresultaten, ze bekijken de resultaten in een breder kader en betrekken daar ook hun eigen handelen bij. Dit blijkt zowel uit de zelfrapportage in de vragenlijst als uit de objectieve beoordeling van hun analyserapporten. Ook maken zij meer gebruik van groepsplannen. Er is weliswaar een positieve trend zichtbaar in de prestaties van leerlingen, maar er wordt geen significant verschil gevonden. Een jaar na de uitvoering van het deelproject is de positieve trend niet meer terug te

vinden. Het ontbreken van significante, blijvende resultaten bij leerlingen kan te maken hebben met het onderbelichten van de *inhoud* van de groepsplannen (welke beslissingen nemen leerkrachten naar aanleiding van de analyse, kiezen ze de 'juiste remedie') en het feit dat er geen/te weinig borging van de wijze van analyseren heeft plaatsgevonden.
Van het vastleggen van referentieniveaus om bij leerkrachten een duidelijk beeld te hebben waar ze met leerlingen naar toe moeten werken is vooralsnog weinig effect te verwachten. Hoewel meer leerkrachten in 2012 – en alle schoolleiders en intern begeleiders - bekend zijn met de niveaus is hun verwachting van het effect gedaald, praat minder dan een derde van de leerkrachten erover op school en blijken zij niet goed te weten waar de niveaus eigenhjk liggen.
Waar de respondenten bij het definieren van het minimumniveau nog redelijk in de buurt zitten van het niveau dat landelijk door 90% van de leerlingen op dit moment wordt bereikt, is dat voor het fundamenteel en het streefniveau bepaald niet het geval. Gemiddeld genomen ligt de lat daar veel hoger dan het landelijk bereikte niveau. Daarmee leggen de leerkrachten hun standaard duidelijk hoger vast dan de wettelijke referentieniveaus 1F en 1S en lopen zij zeer uiteen, met name voor de betere leerlingen.
Het onderwijs aan de betere leerlingen laat sowieso veel te wensen over. Hoewel de erkenning dat deze leerlingen de laatste decennia onvoldoende zijn bediend breed wordt gedragen, blijkt het voor leerkrachten moeilijk om aan deze groep in de praktijk voldoende tijd en aandacht te besteden. Deels is dit te wijten aan het gebrek aan materiaal, waarvan duidelijk is waar het toe leidt en of deze effecten ook behaald kunnen worden, maar voor een belangrijker deel aan de houding ten opzichte van en de vaardigheid in differentieren. Inspectieonderzoek bevestigt de complexiteit van deze vaardigheden maar ook de urgentie: meer opbrengstgericht werken maakt verschillen tussen leerhngen nog duidelijker en daarmee ook de noodzaak om hier rekening mee te houden. Ook de kennis van en het inzicht in leerlijnen en methoden maakt dat leerkrachten moeite hebben in het schrappen van onderdelen, het uitbreiden van de leerstof en het stimuleren van complexe vaardigheden bij (betere) leerlingen. Soms is er echter ook sprake van 'onjuiste' aannames die verhinderen dat een leerkracht de goede leerling voldoende uitdaging biedt, bijvoorbeeld: 'heel goede leerlingen moeten zelfstandig aan de slag kunnen met verrijkingsmateriaal' of 'als goede leerlingen hun reguliere werk afhebben moet je ze belonen in plaats van hen iets moeilijks aanbieden' of 'zelf mogen kiezen wat je gaat doen zorgt ervoor dat ze gemotiveerd blijven'.
De ervaringen van leerlingen die in Streef systematisch met verrijkingsmateriaal werken zijn positief en minder positieve ervaringen komen deels voort uit het feit dat leerlingen niet vanaf het begin van hun schoolloopbaan voldoende uitgedaagd zijn en pas in hogere leerjaren geconfronteerd worden met hogere eisen. Vanzelfsprekend vinden ook zij niet al het werk leuk en zijn er ook leerlingen waarbij het minder goed werkt. Net als bij leerlingen die op een gemiddeld niveau functioneren, of daaronder, zijn er onder hen die cognitief op een hoog niveau functioneren leerlingen die meer zorg van de leerkracht vragen bijvoorbeeld vanwege hun gedrag. Leerkrachten lijken dat van cognitief hoog presterende leerlingen minder goed te accepteren: 'die moeten toch snappen hoe je je moet gedragen'. Of de leerlingen feitelijk meer vooruitgaan dan op basis van hun eerdere resultaten is te verwachten en zich sociaal-emotioneel goed ontwikkelen is niet te beantwoorden met de beschikbare data. De onderzoeksgroep is heel klein, waarbij de grens van wie er wel en niet onder excellente leerlingen valt van school tot school verschilde. Bovendien hebben ze geen eenduidige interventie ondergaan waaraan een eventueel effect aan toe te schrijven is.
Differentiatie is een essentieel onderdeel van de cyclus van opbrengstgericht werken voor alle leerhngen. Uit de vragenlijst voor leerkrachten blijkt dat er nog verbeteringen nodig zijn op gebied van het optimaal

gebruik maken van verschillende prestatieniveaus, het aanbieden van verschillende verwerkingsactiviteiten na een toets en de houding tegenover de wenselijkheid en haalbaarheid van differentiatie. Leerkrachten scoren in 2010 en 2012 laag tot gemiddeld op de verschillende differentiatieonderdelen en zouden hier dus meer ondersteuning bij kunnen gebruiken. Ook zouden zij meer kunnen samenwerken. De meeste samenwerking betreft het uitwisselen van lesmateriaal of het praten over de ontwikkeling van nieuw lesmateriaal met collega's. Lesobservaties zijn een belangrijke bron van informatie bij professionele ontwikkeling, maar slechts een-derde van de leerkrachten observeert wel eens bij een collega in de les.

Tenslotte blijkt er op grond van de leerkrachtvragenlijst sprake te zijn van 4 typen teams. In *opbrengstgerichte* teams wordt veel samen gewerkt, gereflecteerd en gedifferentieerd, teamleden hechten veel waarde aan rekenen en taal. *Individualistische* teams hechten wel veel waarde aan de basisvaardigheden en voeren uitgebreide rapportages uit, maar werken weinig samen en reflecteren en differentiëren weinig. *Behoudende* teams hechten weinig waarde aan innovatie en professionalisering, zijn gemiddeld in mate van samenwerking en aandacht voor prestaties en differentiëren veel. *Niet-opbrengstgerichte* teams tenslotte hechten weinig belang aan basisvaardigheden en opbrengsten. De attitudes en het handelen van de leerkrachten in de schoolteams zijn van invloed op de leerwinst van de leerlingen. Leerlingen van opbrengstgerichte en behoudende leerkrachtenteams boeken meer vooruitgang dan leerlingen van individualistische en niet opbrengstgerichte teams. De opbrengstgerichte en behoudende teams hebben gemeen dat zij relatief vaak differentiëren in de klas. Behoudende teams staan echter minder open voor onderwijsvernieuwingen en zien minder noodzaak zichzelf te bhjven ontwikkelen op professioneel vlak. Het valt op dat het maken van uitgebreide toets-rapportages, zoals de individuahstische teams doen, niet genoeg is om extra leerwinst te boeken. Leerkrachten moeten ook naar de informatie uit de rapportages handelen en moeten met elkaar als team werken aan en reflecteren op hun onderwijs. Dit geeft het belang aan van ondersteuning bij het schoolbreed implementeren van opbrengstgericht werken, een van de uitgangspunten van de Streefprojecten.

Samengevat kan op basis van de analyses gëstëld worden dat beter toetsgebruik en een positievere houding tegenover aspecten van opbrengstgericht werken een belangrijke basis vormen voor kwaliteitsverbetering in het onderwijs en hiermee een eerste stap zijn in de richting van prestatieverhoging van leerhngen. Daarmee zijn scholen er echter nog niet: voor het vergroten van de leerwinst van leerlingen lijken een voortrekkende rol van de schoolleider en intem begeleider en vooral een opbrengstgericht, samenwerkend en differentiërend team essentieel om de ingezette ontwikkeling verder uit te breiden en te borgen.

5 Beleidsimplicaties en suggesties voor verder onderzoek

Interpretatie van de uitkomsten in het licht van beleidscontext en kennisbasis

Bij de analyse van het onderzoek in de programmalijn onderwijskwaliteit zijn opmerkingen te maken over het onderzoeksproces, c.q. de condities waaronder dit heeft plaatsgevonden en over de inhoudelijke uitkomsten van het onderzoek.

De procesgang van het onderzoek maakte (opnieuw) duidelijk dat het niet eenvoudig is om in alle opzichten de medewerking van scholen aan onderwijsonderzoek te verkrijgen. Wat dit betreft zijn de ervaringen met de aan de orde zijnde programmalijn zeker niet slecht, in het bijzonder wanneer het gaat om de response van scholen. Toch bleek het niet mogelijk de voorgenomen *randomized field trial* te realiseren. In dit opzicht reflecteert de procesgang de spanning die er bestaat tussen school autonomie, en

'bottom up' ontwikkeling aan de ene kant, en wetenschappelijke eisen die gelden voor *evidencebased* werken. In het programma in kwestie zijn acceptabele oplossingen gevonden door met een quasi experimentele alternatieve opzet te werken. Op basis hiervan werd het mogelijk om de centrale onderzoeksvragen te beantwoorden, zij het niet met de methodologische kwaliteitsgaranties die inherent zijn aan een echt experimentele opzet.

Confrontatie met resultaten uit ander, vooral internationaal vergelijkend, onderzoek kan worden gezien als een ondersteuning van het onderwijsbeleid dat is ingezet met de kwaliteitsagenda, de actieplannen beter presteren en het propageren van opbrengstgericht werken. Hoewel de kwaliteit van het Nederlandse basisonderwijs, naar internationale maatstaven, goed is, is het verstandig om de vinger aan de pols te houden en het ambitieniveau hoog te blijven stellen. Recent onderzoek biedt ondersteuning voor de aandacht die er in het beleid wordt gegeven aan de prestaties in het hogere segment van de score verdeling, omdat Nederland het in dit segment, ook in het basisonderwijs wat minder goed doet. Het aan de orde zijnde onderzoeksprogramma reflecteerde op dit aspect van 'excellentie'.

Er is een goede aansluiting tussen de principes van opbrengstgericht werken en de kennisbasis op het terrein van onderwijseffectiviteit. Een deel van de factoren die centraal staan in deze kennisbasis, zoals effectieve leertijd, evaluatie en feedback, differentiatie en samenwerking tussen leerkrachten zijn tevens te zien als pijlers van opbrengstgericht werken. Deze aansluiting kan worden gezien als een onderbouwing van het onderwijsbeleid op dit gebied. De wetenschappelijke literatuur biedt ondersteuning voor de verwachting dat de factoren op schoolniveau, die benadrukt worden, inderdaad tot kwaliteitsverbetering zullen leiden.

Tevens biedt deze aansluiting een basis om de inhoudelijke resultaten van het onderzoek te duiden en in een breder perspectief te plaatsen.

Het algemene beeld dat uit het onderzoek naar voren komt is dat scholen, leerkrachten en leerlingen niet onwelwillend staan tegenover opbrengstgericht werken en differentiatie. Tegelijkertijd blijkt dat men moeite heeft met de rationeel technische inslag van de aanpak, wat bijvoorbeeld blijkt uit de enigszins moeizame omgang met de referentieniveaus. Zoals ook uit ander vergelijkbaar onderzoek (Visscher & Ehren, 2011) naar voren komt, kost het de nodige moeite om scholen goed te leren werken met toetsgegevens, en nog meer om de resultaten van toetsgegevens te gebruiken voor didactische analyse en de daarop volgende stap van gedifferentieerde leerstofkeuze en onderwijsbijstelling. Bij deze laatste stappen werd een lacune geconstateerd wat betreft onderwijsleermateriaal, in het bijzonder in het kader van de speciale aandacht voor hoogpresterende leerhngen. De moeite die scholen ook volgens dit onderzoek hebben met differentiatie is bekend uit de onderzoeksliteratuur, en verklaart waarom gedifferentieerd onderwijs (in theorie een geweldig idee) in de praktijk niet duidelijk geassocieerd is met leerprestaties.

Beleidsaanbevelingen

Het onderzoek wekt de indruk dat er in de onderwijspraktijk geen grote problemen zijn wat betreft de legitimiteit en de acceptabiliteit van het beleid. Wat de haalbaarheid betreft is er een wat genuanceerder beeld, er worden vorderingen gemaakt en die hjken ook effect te hebben, tevens worden knelpunten gesignaleerd, vooral op het terrein van het werken met referentieniveaus en doelstellingen, geschikt onderwijsleermateriaal, didactische analyse en differentiatie. De voorzichtig positieve resultaten zijn echter fragiel en vragen om continuïteit van de ingezette beleidslijn.

Wat de proceskant van de beleidsimplementatie betreft is gewezen op de spanning tussen *evidencebased* werken enerzijds en zeer grote schoolautonomie en een daar bij aansluitende *bottom up* innovatie en

implementatie strategie anderzijds. Voor een deel moet deze context als een gegeven worden opgevat en dit zou kunnen betekenen dat men ambities over *evidencebased* werken naar beneden zou moeten bijstellen, en tevens dat men in feite genoegen neemt met een implementatie strategie die weinig efficiënt is. Overwogen zou echter kunnen worden om meer handreikingen te bieden aan de scholen op het terrein van leermateriaal en onderwijsmethoden. Dit impliceert dat scholen niet alles zelf hoeven te ontwikkelen en dat in mindere mate het wiel steeds opnieuw uitgevonden behoeft te worden. Hoewel in een geheel andere nationale context gesitueerd, bieden de resultaten van zogenoemde *Comprehensive School Reforms* in de Verenigde Staten, mogelijk een alternatief, waarbij er een grotere rol is voor sterker gestructureerde externe inbreng, van beproefde methoden.
De aanbeveling die in de samenvatting van de resultaten van het onderzoek is gedaan, namelijk om in de opleiding van docenten te zorgen voor een betere toerusting tot opbrengstgericht werken, wordt onderschreven. Tevens zou het aanbeveling verdienen om scherper toe te zien op de kwaliteit van de ondersteuning die de scholen inkopen (zie suggesties voor verder onderzoek).
Wat de inhoud van opbrengstgericht werken en bevordering van excellentie betreft heeft het onderzoek de vinger gelegd op deelgebieden die meer aandacht behoeven. Dit betreft hoofdaspecten als het werken met referentieniveaus en doelstellingen, didactische analyse en interpretatie van toetsgegevens en differentiatie. De hierboven genoemde aanbevelingen, namelijk continuïteit van het beleid, maatregelen in de sfeer van initiële opleidingen, professionele ontwikkeling en externe handreikingen op het gebied van leermateriaal en systematische werkwijzen, zouden tot verbetering op deze punten kunnen leiden.

Verder onderzoek
De volgende thema's voor verder onderzoek vloeien voort uit de resultaten en aanbevelingen van deze studie:
- Meer evidencebased materiaal ontwikkelen, door middel van gestructureerde R&D, dat aansluit bij diagnostische patronen die voortkomen uit de toetsresultaten. Koppeling met PROO onderzoek op het thema excellentie.
- Beleidsonderzoek naar de haalbaarheid van het werken met sterker extern ontwikkeld didactisch materiaal, en de haalbaarheid van echte experimenten.
- Onderzoek naar de effectiviteit van de ondersteuning en onderwijsverzorging bij opbrengstgericht werken.
- Experimenten geïnspireerd op de principes van de Amerikaanse *Comprehensive School Reform* programma's.
- Onderzoek naar de bijdrage van factoren op schoolniveau voor opbrengstgericht werken, leiderschap, samenwerking, en kwaliteit van ondersteuning.

Referenties

Borman, G.D., Hewes, G.M., Overman, L.T. & Brown, S. (2003). Comprehensive school reform and achievement: a meta-analysis. *Review of Educational Research. 73*, 125 -230.

Bosker, R.J. (2008a)Tripliek: een onnodig gecompliceerde voorstelling van zaken. *Pedagogische Studien, 85*, 305-308.

Bosker, R.J. (2008b). Naar meer evidence-based onderwijs! *Pedagogische Studien, 85*, 305-308.

Creemers, B. &Kyriakides, L. (2008). *The Dynamics Of Educational Effectiveness: A Contribution To Policy, Practice And Theory In Contemporary Schools.* New York: Routledge.

Gravemeijer, K.P.E. &Kirschner, P.A. (2007). Naar meer evidence-based onderwijs? *PedagogischeStudien, 84*, 195-197.

Gravemeijer, K.P.E. &Kirschner, P.A. (2008).Dupliek: eentesimpelevoorstelling van zaken. *Pedagogische Studien, 85*, 195-197.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York: Routledge.

Hendriks, M., Scheerens, J. & Steen, R. (2012). A meta-analysis of school and classroom evaluation. Manuscript. Enschede: University of Twente.

Hopkins, D., Stringfield, S., Harries, A., Stoll, L. & Mackay, T., (2012). School and System Improvement: A State of the Art Review. *School Effectivenessand School Improvement,* (submitted).

Inspectie van het Onderwijs (2009). De *Staat van het Onderwijs, Onderwijsverslag 2008/2009.* Utrecht: Inspectie van het Onderwijs.

Inspectie van het Onderwijs (2012). De *Staat van het Onderwijs, Onderwijsverslag 2010/2011.* Utrecht: Inspectie van het Onderwijs.

Kluger, A.N. & A. DeNisi (1996).The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254.

Lomos, C., Hofland, R. &Bosker, R. (2011). Professional communities and learning achievement- A meta-analysis. *School Effectiveness and School Improvement* (22)121-148.

Luyten, H. (2011). The quality of Education in the Netherlands, as expressed by Attainment and Achievement Indicators. In: Scheerens, J., Luyten H. & Van Ravens, J. *Perspectives on Educational Quality. Illustrative Outcomes on Primary and Secondary Education in the Netherlands.*London, New York: Springer.

Mc Kinsey & Company (2010), How the world's most improved school systems keep getting better. London: Mc Kinsey & Company.

Muijs, D., Creemers, B., Kyriakides, L., Van der Werf, G., Timperley, H. & Earl, L. (2012). Teaching Effectiveness. A State of the Art Review. *School Effectiveness and School Improvement,* (submitted).

Nederlandse Katholieke Schoolraad (1979). *Bouwen aan de relatief autonome school.* NKSR.

OECD (2010). *Strong performance and successful reformers.Lessons from PISA for the United States.* Paris: OECD Publishing.

Onderwijsraad (2006). *Naar meer evidence-based onderwijs.* Den Haag: Onderwijsraad.

Parlementaire Commissie Onderwijsvernieuwingen (2008). *Tijd voor Onderwijs.*Den Haag: Ministerie van OCW.

Reynolds, D., Sammons, P., De Fraine, B., Townsend, T., Van Damme, J., Teddlie, C. &Stringfield, S. (2012). Educational Effectiveness Research (EER): A state of the Art Review. *School Effectiveness and School Improvement*, (submitted).

Robinson, V.M.J., Lloyd, C. & Rowe, K.J. (2008). The impact of leadership on student outcomes. An analysis of the differential effect of leadership type. *Educational Administrative Quarterly,* 44 (5), 635-674.

Seidel, T. &Shavelson, R.J. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Review of Educational Research 77*(4), 454-499.

Scheerens, J. &Bosker, R. (1997). *The foundations of educational effectiveness.* Oxford: Pergamon.

Scheerens, J., Luyten, H., Steen, R. & Y. Luyten-de Thouars (2007). *Review and meta-analyses of school and teaching effectiveness.* Enschede: University of Twente, Department of Educational Organization and Management.

Scheerens, J. & Ehren, M.C.M. (2009). Voortmodderen met het onderwijs: over sturing en zelforganisatie in het onderwijs, vanuit een onderwijskundige optiek. *NederlandsTijdschriftvoorOnderwijsrecht, 2*1(1), 5-25.

Scheerens, J. (2012). School, Teaching and System Effectiveness, some Comments with three State of the Art Reviews. Submitted.

Scheerens, J. (2013). (ed.) *Productive Time in Education. A review of the effectiveness of teaching time at school, homework and extended time outside school hours.* Enschede: University of Twente.

Schevenings Beraad Bestuurlijke Vernieuwing (1994). *Gezamenlijke richtinggevende uitspraken met betrekking tot de gewenste ontwikkelingen in de bestuurlijke verhoudingen in primair en voortgezet onderwijs.* Den Haag: SDU.

Vermeer, N. & Van der Steeg, M., (2011). Onderwijsprestaties Nederland in Internationaal Perspectief. CPB Achtergronddocument bij CPB Policy Brief 05, 2011. Den Haag: CPB.

Visscher, A.J.&Ehren, M. (2011). *De eenvoud en complexiteit van opbrengstgericht werken.* [Enschede]: [Universiteit Twente, Vakgroep Onderwijsorganisatie en -management].

Woessmann, L., Luedemann, E., Schuetz, G. & West, M.R. (2009). *School Accountability, Autonomy and Choice around the World.* Cheltenham, UK/ Northampton, MA, USA: Edward Elgar.

Bijlage 1 Presentatie en verslag van de discussie op 5 maart 2013

Het streven naar kwaliteit in scholen voor primair onderwijs

Resultaten en reflectie

BOPO Programmalijn: Kwaliteit PO

Jaap Scheerens & Simone Doolaard

5 maart 2013

Achtergrond en opzet

- Scholen voor morgen & Basis voor presteren
- Referentieniveaus, Opbrengstgericht werken & Excellentie
- Samenhangend onderzoeks/schoolverbeteringsprogramma: Streef (RUG/GION, Cito ism Cedin, Marant)
- Training en begeleiding van leerkrachten
- Effectonderzoek: leerkrachtvragenlijst & leerlingresultaten
- Werving & 'toewijzing'
- Gekruist design

Resultaten I

- Ontwikkeling bij schoolteams
  - Deelname aan Streef leidt tot meer gebruik van toetsresultaten/reflectie en positievere houding t.o.v. OGW en mogelijkheden prestaties te verbeteren
  - Leerdoelen in teamverband? ½ v/d leerkrachten
  - Differentiatie en (echt) samenwerken blijft achter
  - Opbrengstgerichte en behoudende teams leveren betere resultaten dan individualistische en niet-opbrengstgerichte teams
- Werken met referentieniveaus
  - Leerkrachten variëren, desgevraagd schatten ze 1F en 1S veel hoger in dan wettelijk vastgelegd
  - Deelname leidt tot meer bekendheid, begrip en realistische verwachtingen van leerlingen
  - Gemiddeld zijn verwachtingen getemperd
  - Deelname heeft geen effect op leerlingen

Resultaten II

- Gebruik maken van opbrengsten
  - Analyse-rapportages verbeteren, reflecteren in breder kader, meer groeusplannen
  - In leerlingresultaten een positieve tendens in jaar van deelname, daarna terugval
- Differentiatie en excellentie
  - Alle scholen van mening dat ze atwel 'iets doen', urgentie duidelijk, veel onrekerheid (roep om materiaal), buiten > binnen
  - Feitelijks situatie in de klas ≠ mondelinge/schriftelijke mfo
  - Knelpunten: vooroordelen (moeilijk werk=straf, keuze=fijn, zelfstandigheid), differentieren in klassikale instructie, eigen beheersing lesstof, visie/doelstelling/evaluatie, focus op echt HB, klassenmanagement, versnippering, voorbereiding, interpretatie reacties/gedrag leerlingen

Beleidsconsequenties

- Referentieniveaus in dienst van opbrengstgericht werken
- Ondanks bedenkingen, positieve houding t.o.v. OGW
- OGW leidt automatisch tot differentiatie > differentiëren vraagt complexe vaardigheden die slechts ½ v/d leerkrachten beheersen. Onderzoek naar entwikkefing van leerkrachten (in fases) bevestigt dit. Wat is startbekwaam? En hoe/wanneer is een leerkracht volleerd?
- Aandacht voor de basis: klassenmanagement, kennis van leerlijnen, didactische vaardigheden, evidence based methoden (wat werkt waartoe en voor welke leerlingen?)
- Onderzoek naar (kosten)effectiviteit aanpak excellente leerlingen (beoogde effect?)

Beleidscontext

- De inzet van de infrastructuur van de onderwijsondersteuning gericht op het verbeteren van de eigen innovatiekracht van de scholen die zelf van onderaf innovaties moeten ontwikkelen
- Infrastructuur van doelen formuleren, toetsen en evalueren om te 'controleren' en te 'sturen/bijsturen'.
- Waar wringt deze context?
  - Bij het uitvoeren van randomized field trials
  - Bij het voorzien in kennislacunes bij scholen

Onderzoeksprogramma en andere gegevens over kwaliteit en effectiviteit van het Nederlandse onderwijs

- Beleid gericht op excellentie wordt ondersteund door resultaten uit internationaal onderzoek, waaruit blijkt dat Nederlandse leerlingen in het basisonderwijs in het hogere score segment enigszins achterblijven
- Er is een goede aansluiting tussen de principes van OGW en de kennisbasis op het terrein van onderwijseffectiviteit. Een deel van de factoren die centraal staan in deze kennisbasis, zoals effectieve leertijd, evaluatie en feedback, differentiatie en samenwerking tussen leerkrachten zijn tevens te zien als peilers van opbrengstgericht werken.

Stellingen

- Het onderstrepen van zowel de autonomie van scholen in de Kwaliteitsagenda's en Actieplannen als 'evidence based' werken leidt tot interne tegenstrijdigheid
- Uitkomsten van onderwijseffectiviteit onderzoek kunnen worden gezien als ondersteuning van OGW
- Het beschreven onderzoeksprogramma toont aan dat er draagvlak is bij basisscholen voor OGW
- Positieve effecten van OGW blijven achter bij de verwachtingen, vanwege onvoldoende differentiatievaardigheden en kennis van de leerstof
- Het verdient aanbeveling in toekomstige projecten ruimte te bieden om af te wijken van de gangbare 'ontwikkeling van onderop' strategie

Stellingen (vervolg)

- Het ontsluiten en beschikbaar stellen van relevante kennis met betrekking tot OGW op centraal niveau behoefte meer aandacht
- Evaluatie onderzoek naar de doeltreffendheid en de doelmatigheid van het functioneren van de onderwijsverzorging bij de ondersteuning van scholen is gewenst
- Verplichte deelname aan (evaluatie-)onderzoek zou voortaan moeten worden opgenomen als bekostigingsvoorwaarde bij door de overheid gesubsidieerde verbetering- en vernieuwingsprojecten

Samenvatting van de discussie n.a.v. de presentatieop 5 maart 2013

In de eerste reactie vanuit het ministerie van OCW wordt waardering uitgesproken voor het onderzoek en wordt aangegeven dat er op verschillende manieren al vervolg is gegeven aan het beleid dat ten grondslag ligt aan het uitgevoerde onderzoek. Na de kwaliteitsagenda en het actieplan is een bestuursakkoord gesloten en is School aan Zet in het leven geroepen om nog meer scholen 'op maat' te stimuleren om de onderwijskwaliteit onder andere via opbrengstgericht werken te verbeteren. Via deze weg, maar bijvoorbeeld ook door het inzetten van adviseurs en verbeterteams voor (zeer) zwakke scholen wordt geprobeerd om weliswaar initiatief van onderop te belonen, maar ook te voorkomen dat elke school het wiel opnieuw uit gaat vinden en iedereen zijn of haar eigen 'leuke ideetje' gaat uitvoeren.
Net als bij de bespreking van de andere programmalijnen ontstaat een discussie over in hoeverre scholen nu daadwerkelijk autonomie ervaren en autonomie hebben gezien de hoeveel verantwoording die ze moeten afleggen en die de autonomie inperkt. Een tweede gemeenschappehjk discussiepunt, gerelateerd aan het eerste punt, is de wenselijkheid en haalbaarheid van het stimuleren en zelfs opleggen van een bepaalde uniforme, evidencebased aanpak. Enkele onderzoekers dringen hierop aan omdat uit veel onderzoek blijkt dat er een grote verscheidenheid aan invullingen van beleid (deels waarschijnlijk veroorzaakt door het lumpsum toekennen van een budget in plaats van geoormerkt) in de praktijk is, waarvan geen, onduidehjke of zeer tijdelijke effecten meetbaar zijn. Andere onderzoekers en beleidsmakers beschouwen dit als een groot goed en dringen aan op adaptief onderwijsbeleid dat tegemoet komt aan de fase waarin een school zich bevindt en de capaciteit om zelf de kwaliteit van het onderwijs te verbeteren en te borgen.

Een tweede onderwerp betreft de vraag hoe men evidencebased beleid kan voeren. Daarvoor is het in de eerste plaats nodig dat er voldoende kennis is over wat evidencebased is. In het uitgevoerde onderzoek binnen deze programmalijn, maar ook binnen de andere programmalijnen bleek het moeilijk/onmogelijk om een echt experiment op te zetten en/of was het zeer moeilijk om voldoende (representatieve) respondenten te vinden. Er wordt gepleit voor het aangaan van afspraken met scholen, het sluiten van akkoorden waarin deelname aan onderzoek wordt gegarandeerd (zo mogelijk zou dit middels het beschikbaar stellen van tijd/geld gefaciliteerd moeten worden).

Tenslotte is er gediscussieerd over de deskundigheid van de leerkracht. In elk van de programmalijnen is de deskundigheid van de leerkracht aan de orde gekomen, maar in deze programmalijn is die deskundigheid het meest expliciet aan de orde geweest omdat de focus lag op de persoon van de leerkracht. In de Streef-projecten zijn leerkrachten geschoold en begeleid tot in het daadwerkelijke werken in de groep. In de onderzoeken bleken sommige leerkrachten heel goed in staat te zijn om opbrengstgericht te werken en te differentiëren voor zowel zwakke als gemiddelde en sterke leerlingen, terwijl het anderen ontbrak aan kennis, vaardigheden en attitude. Het blijft de vraag waarom de een dit wel kan en de ander niet. Achtergrondkenmerken als leeftijd en ervaring blijken daarbij nauwelijks een rol te spelen.
In aansluiting op stelling 4 is gediscussieerd over de vraag wat te verwachten is van (beginnende) leerkrachten en op welke manier de ontwikkeling van leerkrachten die al aan het werk zijn gemeten, geregistreerd en gestimuleerd zou kunnen worden. Zo werd geopperd dat bepaalde kennis en vaardigheden misschien wel helemaal niet verwacht zouden mogen worden van leerkrachten. Kan bijvoorbeeld van leerkrachten verwacht worden dat zij voor elke leerling in hun klas een verantwoorde

leerlijn op alle vakgebieden moet kunnen bepalen of is dit meer het werk van leerplanontwikkelaars? Ook het goed kunnen differentiëren, in hoeverre kan dit gefaciliteerd worden door bijvoorbeeld de methode?

Tenslotte werden nog enkele algemene opmerkingen gemaakt over de term en het concept opbrengstgericht werken. Vanuit onderwijseffectiviteitsperspectief wordt opbrengstgericht werken gezien als een combinatie van meerdere effectiviteitsbevorderende kenmerken gericht op het behalen van zo hoog mogelijke leerlingprestaties (op de basisvaardigheden). Tijdens de discussie bleek dat dit vanuit bestuurskundig perspectief niet onmiddellijk evident is.

Bijlage 2 Case studies over onderwijssystemen die hun prestaties verbeterd hebben, Comprehensive School Reform en 'School aan zet'

Naar aanleiding van commentaar van de BOPO op het ingediende rapport, wordt in deze bijlage stilgestaan bij enkele relevante buitenlandse rapporten over systeemverandering en schoolverbetering. Het lijkt erop alsof in het BOPO commentaar de internationale studies van Mc Kinsey worden gezien als ondersteuning van de Nederlandse 'bottom up' traditie bij onderwijsverbetering. Dat is maar ten dele het geval. In de studies van Mc Kinsey en ook in het verwante OECD rapport over 'strong performers', nemen prestatiegegevens van hoog niveau, en 'alingment' tussen verschillende beleidsniveau's een belangrijke plaats in, evenals gestandaardiseerde instructiebenaderingen die uitzicht bieden op goede prestaties. Tevens wordt in het BOPO commentaar gesteld, dat het recent ingezette 'School aan zet' overeen zou komen met Comprehensive School Reform programma's, waaraan in het rapport gerefereerd wordt. Dit is echter niet het geval. 'School aan zet', de term zegt het al, is veeleer een voortzetting van de vigerende, laat vele bloemen bloeien, traditie van schoolbegeleiding in Nederland.

Mc Kinsey & Company (2010), How the world's most improved school systems keep getting better.
Stage-gewijze accenten in maatregelen:

- Fase 1: het ondersteunen van basics, accent op primaire proces
- Fase 2: consolideren van een goed basisniveau door middel van 'high quality performance data'
- Fase 3: het ontwikkelen van een sterke professionele status van leerkrachten en schoolleiders
- Fase 4: accent op de school als organisatie, professionele leergemeenschappen en 'peer basedleaming'

OECD (2010).Strong performance and successful reformers.Lessons from PISA for the United States. Paris: OECD Publishing:

- Developing a commitment to education and conviction that all students can achieve high levels
- Ambitious standards aligned with high-stakes gateways and instructional systems (well ahgned testing system)
- Developing more capacity at the point of delivery (high quality teachers and school leaders)
- A work organization in which the teachers can employ their potential: management, accountability and knowledge management (flat organization, away from Tayloristic management, school autonomy)
- Institutionalizing improved instructional practice (diagnostic skills, encyclopedic repertoire, students enthralled, devoted to the improvement of their craft)
- Aligning incentive structures and engaging stakeholders (high stakes examination systems, in collaboration with stakeholders)
- External, professional and parent oriented accountability
- Investing resources where they can make most of a difference (strong teachers aligned to weak students)
- Balancing local responsibihty with a capable centre with authority and capacity to act (state sets clear expectations)
- Importance of work based training (transition school work)

- Coherence and alignment across levels, policies and practices, and sustained input (policy implementation)
- An outwards outlook of the system (responsiveness)

Borman, G.D., Hewes, G.M., Overman, L.T., & Brown, S., (2003) Comprehensive school reform and achievement: a meta-analysis. Review of Educational Research. 73, 125 -230
A Comprehensive School Reform Program:

- Employs proven methods for student learning, teaching and school management that are founded on scientifically based research and effective practices and have been replicated successfully in schools;
- Integrates instruction, assessment, classroom management professional development, parental involvement, and school management;
- Provides high quality and continuous teacher and staff professional development and training;
- Includes measurable goals for student academic achievement and establishes benchmarks for meeting these goals;
- Is supported by teachers, principles, administrators and other staff throughout the school;
- Provides support for teachers, principals, administrators an other school staff by creating shared leadership and a broad base of responsibihty for reform efforts;
- Provides for the meaningful involvement of parents and the local community in planning, implementing , and evaluating school improvement activities;
- Uses high-quality external technical support and assistance from an entity that has experience and expertise in schoolwide reform and improvement, which may include an institution of higher education;
- Includes a plan for the annual evaluation of the implementation of the school reforms and the student results achieved:
- Identifies the available federal, state, local, and private financial and other resources that schools can use to coordinate services that support and sustain the school reform effort; and
- Meets one of the following requirements: Either the program has been found, through scientifically based research, to significantly improve the academic achievement of participating students; or strong evidence has been shown that the program will significantly improve the academic achievement of participating children.

School aan zet

'De School aan Zet experts helpen scholen (po, vo &(v)so) bij het formuleren van doelstellingen, vormgeven van verbeteraanpakken en houden de vinger aan de pols wat betreft de resultaten. School aan Zet verspreidt effectieve aanpakken en er is ruime gelegenheid voor leren van elkaar. Binnen School aan Zet wordt de gehele schoolorganisatie in beweging gebracht: bestuursniveau, schoolleiders, IB-ers, leraren, etc. Het doel is om verandercapaciteit binnen de schoolorganisatie te ontwikkelen. De thema's binnen School aan Zet sluiten aan bij de bestuursakkoorden PO en VO (opbrengstgericht werken, HRM/lerende organisatie, basisvaardigheden/kernvakken taal, lezen rekenen/wiskunde, excellentie, omgaan met verschillen en wetenschap/techniek. Binnen school aan zet worden ook masterclasses, conferenties en netwerken georganiseerd op de kernthema's. Aan het programma School aan Zet is een proces- en effectevaluatie gekoppeld.' (Citaat commentaar BOPO)

De school aan zet benadering is enerzijds gericht op Opbrengstgericht werken en anderzijds op HRM/de school als lerende organisatie. Op de web site wordt aangegeven dat experts scholen bezoeken, om de scholen te assisteren met zaken waar men al mee bezig is. De rol van de experts is die van kritisch meedenker ('critical friend'). Het programma biedt mogelijkheden tot professionele ontwikkeling, master classes, en biedt handreikingen wat betreft instrumenten en hulpmiddelen; tevens is het de bedoehng dat goede praktijken uitgewisseld worden.
De benadering bij School aan Zet heeft meer gemeen met de vigerende schoolbegeleidingspraktijk in Nederland dan met de aanpak in Comprehensive School Reform Programs. Het belangrijkste verschil is dat er geen sprake is van externe, wetenschappelijk beproefde programma's. Uit het overzicht van de experts ontstaat de indruk dat deze vooral knowhow hebben op schoolorganisatie niveau, terwijl het vakdidactische element minder vertegenwoordigd lijkt te zijn. Niet duidelijk is of de voorziene monitoring en evaluatie zodanig ontworpen is dat het succes van het overheidsbeleid toetsbaar is. Evenmin of er uitgewerkte plannen zijn, bijvoorbeeld wat betreft de manier waarop opbrengstgericht werken, en het functioneren van de school als lerende organisatie op een geïntegreerde manier vorm moeten krijgen.