

Tussenrapportage

pilots leerwinst/toegevoegde waarde

in het voortgezet onderwijs

Jaap Roeleveld

Kohnstamm Instituut UvA



Het pilotproject wordt uitgevoerd in opdracht van
de Inspectie van het Onderwijs.

Ondersteuning van de scholen: Oberon.

Begeleidend onderzoek: Kohnstamm Instituut.

Inhoudsopgave

1.	Inleiding.....	1
2.	Opzet van de pilots	3
3.	Start van het project: verwachtingen van scholen	4
4.	Het begeleidende onderzoek.....	6
4.1	Inleiding.....	6
4.2	Verzamelde gegevens op de scholen.....	9
4.3	Beschrijvende analyses per school: de schoolrapporten.....	16
4.4	Bespreking schoolrapporten met de scholen	22
5.	Prestatieniveau en leerwinst vergeleken.....	24
6	Toegevoegde waarde.....	30
6.1	Inleiding.....	30
6.2	Beginniveau en leerwinst naar afkomst uit APCG-gebied	32
6.3	Beginniveau en leerwinst naar voordeel van de twijfel.....	36
6.4	Beginniveau en leerwinst naar leerwegondersteuning	40
6.5	Beginniveau en leerwinst naar geslacht	44
6.6	Verschillen tussen scholen en cohorten	47
7	Vooruitblik.....	50
	Referenties.....	51

1. Inleiding

De huidige minister van Onderwijs, Cultuur en Wetenschappen (OCW) zet beleid voort dat gestart is onder het vorige kabinet, onder de noemer Kwaliteitsagenda Primair en Voortgezet Onderwijs. Het kerndoel van dat beleid is streven naar verhoging van prestaties in de basisvaardigheden taal en rekenen. De beleidsinstrumenten die daarvoor al in de vorige kabinetsperiode zijn ontwikkeld zijn onder meer de wettelijke invoering van referentieniveaus en diverse acties om te stimuleren dat scholen meer opbrengstgericht gaan werken. Opbrengstgericht werken is, in de definitie van de Inspectie van het Onderwijs, “systematisch en doelgericht werken aan het maximaliseren van de prestaties van de leerlingen” (Inspectie van het Onderwijs, 2010). De internationale term is *datadriven teaching*, hetgeen aangeeft dat het gaat om onderwijs dat mede wordt gestuurd door analyse van gegevens die scholen hebben over leerresultaten van leerlingen (Ledoux e.a., 2009).

Het actuele beleid is verwoord in de actieplannen ‘Basis voor Presteren’ (primair onderwijs) ‘Beter Presteren’ (voortgezet onderwijs) en ‘Leraar 2020’ (Ministerie OCW, 2011). Verbetering van leerprestaties, ambitieuze doelen en benutten van talent zijn hierin kernbegrippen. De actieplannen bevatten concrete doelen op het gebied van verhogen van de eindopbrengsten basisonderwijs, terugdringen van het aantal zwakke scholen, invoering van een predicaat ‘excellente school’ en toename van het aantal scholen dat op een opbrengstgerichte manier werkt, volgens het oordeel van de inspectie.

Nieuwe elementen in het beleid zijn een centrale eindtoets voor alle leerlingen in het primair onderwijs, tussentijdse toetsen Nederlands, Engels en wiskunde/rekenen in het voortgezet onderwijs, verplichting voor scholen om gebruik te maken van een leerling- en onderwijsvolgsysteem en de ontwikkeling van een nieuwe maat voor het beoordelen van de leeropbrengsten van een school via het meten van de *toegevoegde waarde* van de school. Die ontwikkeling moet plaatsvinden in een serie pilots die in het schooljaar 2011-2012 van start zijn gegaan in het primair onderwijs en een jaar later in het voortgezet onderwijs. Het doel is te komen tot een manier van bepalen van toegevoegde waarde die enerzijds functioneel is voor scholen en hen kan helpen bij zelfevaluatie en opbrengstgericht werken, anderzijds bruikbaar is voor de Inspectie van het Onderwijs bij het beoordelen van de kwaliteit van de leeropbrengsten van scholen, op een manier die zoveel mogelijk objectief en eerlijk is.

In het projectplan voor de pilots in het voortgezet onderwijs (Inspectie van het Onderwijs, 2012) wordt kort ingegaan op de termen leerwinst en toegevoegde waarde:

“De termen ‘toegevoegde waarde’ en ‘leerwinst’ worden vaak naast en door elkaar gebruikt. Ook in het Regeerakkoord en de onderliggende Actieplannen is dat het geval. Feitelijk zijn het echter twee verschillende begrippen.

Onder leerwinst verstaan we de toename van vaardigheden, kennis en/of competenties van individuele leerlingen of groepen van leerlingen, gedurende (een bepaald deel van) de leerweg. De leerwinst wordt altijd bepaald door minimaal twee meetmomenten. Het verschil tussen die metingen maakt de

ontwikkeling van de leerling (of de groep leerlingen) zichtbaar. De metingen moeten eenzelfde schaal omvatten en het moet gaan om genormeerde toetsen.

Onder toegevoegde waarde verstaan we de bijdrage van de school aan de ontwikkeling (de leerwinst) van leerlingen. Het gaat hierbij dus om dat deel dat met enige zekerheid aan de school kan worden toegeschreven.

Leerwinst kan nauwkeurig in kaart worden gebracht door te kijken naar twee of meer in de tijd liggende meetmomenten en daar het verschil tussen te meten. Uit literatuurstudie blijkt dat leerwinstmetingen in het Nederlandse onderwijs het best gebaseerd kunnen worden op individuele metingen van meer dan twee verspreid in de tijd liggende meetmomenten.

Toegevoegde waarde is veel lastiger in kaart te brengen en tot een maat te maken: er moet correctie plaatsvinden op een aantal achtergrond- en contextkenmerken (waarbij er nooit honderd procent zekerheid is dat alle correctiefactoren 'gepakt' zijn) en er moet vervolgens een schatting gemaakt worden van wat aan de school kan worden toegeschreven.

In de pilots gaat het zowel om leerwinst als om de toegevoegde waarde van een school. Op een open wijze zal nader verkend worden hoe de leerwinst van leerlingen in het VO beter recht kan worden gedaan, en hoe dit mogelijk kan leiden tot indicaties van de toegevoegde waarde van een school.”

De pilots Leerwinst en toegevoegde waarde zijn in het voortgezet onderwijs in 2012 van start gegaan. In dit rapport wordt verslag gedaan van de voortgang in en de opbrengsten van het eerste schooljaar 2012/2013.

Vooraf merken we nog op dat de uitgangssituatie voor de pilots in het voortgezet onderwijs complex is in vergelijking met die in het basisonderwijs. In het voortgezet onderwijs is veel minder dan in het basisonderwijs sprake van het systematisch volgen van de voortgang van leerlingen in een leerlingvolgsysteem (Oomens, Van Aarsen & Hulsen, 2012; Meijer, Ledoux, & Elshof 2011)¹. Bovendien is er geen sprake van een vergelijkbaar curriculum, maar bieden scholen onderwijs op verschillende niveaus, van vmbo basisberoepsgericht tot aan vwo en gymnasium. Daardoor is het moeilijker opbrengsten van scholen onderling te vergelijken.

Het doel van de pilots is daarmee vooral het verkennen van mogelijkheden die scholen voor voortgezet onderwijs hebben om de ontwikkeling van hun leerlingen systematisch te volgen en de uitkomsten daarvan te gebruiken in hun eigen kwaliteitsbeleid en in te zetten bij de ontwikkeling naar meer opbrengstgericht werken. Daarbij hoort dan ook een verkenning van de mogelijkheden en beperkingen van het gebruik van leerwinst en toegevoegde waarde bij het beoordelen van de opbrengsten van scholen, afdelingen van scholen en klassen.

¹ Overigens worden ook in het basisonderwijs de mogelijkheden van leerlingvolgsystemen lang niet altijd optimaal gebruikt (Meijer, Ledoux & Elshof, 2011).

2. Opzet van de pilots

Het project heeft tot doel gedurende twee schooljaren informatie te verzamelen over de mogelijkheden om tot een bruikbare, valide maat voor leerwinst te komen, evenals informatie over de mogelijkheden voor een gestandaardiseerde methode voor het bepalen van de toegevoegde waarde van een school voor voortgezet onderwijs. Om tot die maten en methoden te komen, wordt aangesloten bij de huidige schoolpraktijk, zoals de 27 pilotafdelingen die laten zien. De werkwijze om leerwinst en toegevoegde waarde te bepalen moet (statistisch) valide en betrouwbaar zijn, maar ook recht doen aan de inspanningen van de school en rekening houden met de kenmerken van de schoolpopulatie.

Belangrijk kenmerk is dat er wordt uitgegaan van de huidige schoolpraktijk en dat er gebruik gemaakt gaat worden van gegevens zoals die op dit moment bij de scholen aanwezig zijn. In het voortgezet onderwijs staat het gebruik van leerlingvolgsystemen, waarmee de vooruitgang van leerlingen systematisch kan worden bijgehouden, nog in de kinderschoenen. En waar dergelijke systemen wel worden gehanteerd worden zij zelden ingezet voor opbrengstgericht werken (Oomens, Aarsen & Hulsen, 2012). Overigens leert de ervaring in het basisonderwijs dat, ook wanneer er wel op grote schaal gewerkt wordt met leerlingvolgsystemen, er vaak gebrek aan deskundigheid op school is wat betreft het benutten van de mogelijkheden die deze systemen bieden (zie bijvoorbeeld Meijer, Ledoux & Elshof, 2011). Een van de opbrengsten van de pilots zou dan ook moeten zijn de scholen handvatten te geven om de ontwikkeling van hun eigen leerlingen systematisch te volgen. Voor schoolmanagers en schoolbestuurders moeten daarmee instrumenten ter beschikking komen om het eigen onderwijs te beoordelen en te verbeteren.

In de richting van de Inspectie dienen de pilots informatie op te leveren over hoe leerwinst en toegevoegde waarde van afdelingen en scholen voor voortgezet onderwijs kunnen worden bepaald en mogelijk een rol kunnen spelen bij het beoordelen van de opbrengsten van scholen. Daarbij gaat het enerzijds om een betrouwbare en valide manier van beoordelen. Maar anderzijds dient zo'n beoordeling ook transparant en begrijpelijk voor de scholen en besturen te zijn. En de te benodigde gegevens moeten zonder te veel administratieve last voor de scholen algemeen beschikbaar zijn.

De pilots zijn opgezet door een projectgroep onder leiding van de Inspectie van het Onderwijs en de directie Voortgezet Onderwijs van OCW. Deze projectgroep wordt ondersteund door Oberon, dat de opdracht heeft om de scholen die deelnemen aan de pilot te ondersteunen en om de projectorganisatie te verzorgen. Het wetenschappelijk onderzoek bij de pilot wordt uitgevoerd door het Kohnstamm Instituut.

3. Start van het project: verwachtingen van scholen

Het project is gestart met een gezamenlijke bijeenkomst in Utrecht. Daarna is met alle scholen een intakegesprek gehouden. Van de gesprekken zijn, per school, verslagen gemaakt. Voor deze tussenrapportage heeft Oberon de gesprekken als volgt kort samengevat.

Na een gezamenlijke startbijeenkomst op 2 november 2012 is door de begeleiders van Oberon op alle scholen een intakegesprek gehouden. Het doel van deze gesprekken was om meer zicht te krijgen op de gegevens die scholen in pilot wilden gebruiken voor leerwinst en toegevoegde waarde en op de wijze waarop en de mate waarin scholen opbrengstgericht werken. Ook werd aan de scholen gevraagd welke behoeften zij hadden aan verdere ondersteuning rondom leerwinst en toegevoegde waarde.

Uit de gesprekken bleek dat de scholen gebruik maken van diverse toetsen en schooladministratiesystemen. Ook verschilden de scholen in de wijze waarop zij opbrengstgericht werken en in de ervaring die ze daarmee hebben.

De verwachtingen van scholen over de pilot waren wisselend. Een deel van de scholen wil er met behulp van de pilot voor zorgen dat de Inspectie van het onderwijs ook de achtergrondkenmerken van hun leerlingen meeneemt in de beoordeling. Zij gaven aan een lastige leerlingpopulatie te hebben waar de Inspectiebeoordeling voor hun gevoel te weinig rekening mee houdt. Andere scholen hopen met behulp van de pilot meer handvatten te krijgen om het onderwijs te verbeteren.

Sommige scholen gaven aan niet alleen te willen kijken naar de cognitieve ontwikkeling van hun leerlingen, maar ook op zoek te zijn naar manieren om de niet-cognitieve ontwikkeling meetbaar te maken. Bij een paar scholen leidt deze wens tot verwarring over het begrip toegevoegde waarde. Deze scholen verstaan onder toegevoegde waarde de zaken die de school leerlingen leert naast cognitieve dingen. In de pilot gaat het bij toegevoegde waarde echter om de bijdrage van de school aan de leerwinst van leerlingen, ongeacht of deze leerwinst betrekking heeft op de cognitieve ontwikkeling of op andere aspecten.

Tot slot is in de intakegesprekken aan de scholen gevraagd aan welke ondersteuning zij behoefte hebben. Sommige scholen hadden vooral behoefte aan het bespreken van de schoolrapportage van de eerste pilotmeting. Anderen hadden meer behoefte aan een workshop waarin dieper wordt ingegaan op wat je als school kunt met gegevens uit diverse toetsen of rondom opbrengstgericht werken in bredere zin. Er waren ook scholen die meer wilden weten over de reden van afstroom van leerlingen in de onderbouw of juist van het succes van hun leerlingen in het vervolgonderwijs.

De bevindingen van de intakegesprekken zijn gebruikt om:

- de scholen in te delen in de verschillende onderzoeksscenario's die in de pilot verder zijn uitgewerkt;

- de dataverzameling af te stemmen op de door scholen gebruikte systemen;
- het ondersteuningsaanbod zoveel mogelijk af te stemmen op de wensen van de scholen (ook gebaseerd op de bevindingen uit de besprekingen van de eerste schoolrapportages).

De in deze samenvatting genoemde onderzoeksscenario's hadden betrekking op de verschillende instrumenten, die scholen in de pilot wilden gaan gebruiken om de ontwikkeling van hun leerlingen in beeld te brengen (zie voor een overzicht paragraaf 4.2 in dit rapport) .

Zoals gemeld was er bij een aantal scholen de behoefte om ook de niet-cognitieve ontwikkeling van leerlingen in het project te betrekken. Omdat daar in de praktijk allerlei verschillende instrumenten voor worden aangeboden en gebruikt is besloten om in het kader van de pilot een tweetal instrumenten aan de scholen aan te bieden, die gebruikt worden in het landelijke COOL5-18 cohortonderzoek. Het betreft vragenlijsten op het domein van motivatie en van burgerschapscompetenties. Op deze manier kan ervaring worden opgedaan met de mogelijkheden om ook zulke niet-cognitieve competenties te betrekken bij de beoordeling van de ontwikkeling van leerlingen en de opbrengsten van scholen. In de toekomst zal mogelijk kunnen worden aangesloten bij de beoordeling van sociale opbrengsten van het onderwijs, zoals die door de Inspectie van het onderwijs wordt ontwikkeld.

4. Het begeleidende onderzoek

4.1 Inleiding

Leerwinst en toegevoegde waarde

In het project Pilots leerwinst en toegevoegde waarde in het voortgezet onderwijs wordt gezocht naar bruikbare indicatoren voor de kwaliteit van een school, of van een afdeling in een school². De aandacht gaat met name uit naar indicatoren die betrekking hebben op leeropbrengsten bij leerlingen. We gebruiken hier de term leeropbrengsten in ruime zin, ongeacht of deze op het cognitieve vlak liggen (zoals leeropbrengsten binnen de afzonderlijke schoolvakken) of op het sociaal-emotionele domein (zoals motivatie, zelfvertrouwen of burgerschapscompetenties).

We onderscheiden in dit project twee typen indicatoren: indicatoren op basis van leerwinst, dan wel indicatoren op basis van toegevoegde waarde³.

Indicatoren op basis van *leerwinst* zijn afgeleid van de leerwinst die de leerlingen van de school (of afdeling) in een nader aan te geven periode hebben geboekt. Idealiter is het voor de bepaling van de leerwinst nodig dat leerlingen een begintoets en een eindtoets afleggen. Aangenomen dat de toetsscores op één schaal kunnen worden gebracht, komt het verschil tussen de beginmeting en de eindmeting in dat geval overeen met de leerwinst.

Als de betreffende toetsen landelijk genormeerd zijn kan voor het beoordelen van de behaalde leerwinst (gemiddeld per klas of per school) een vergelijking gemaakt worden met de leerwinst die landelijk behaald wordt. Op die manier wordt de leerwinst dus vergeleken met landelijke gemiddelden: er wordt vergeleken met 'doorsnee-leerlingen'.

Maar scholen trekken een zeer verschillend leerlingenpubliek. Met een beginmeting weten we iets over het niveau waarmee de leerlingen de school binnen komen. Maar daarnaast spelen er ook andere factoren mee, die bevorderend of belemmerend kunnen zijn voor de (cognitieve en sociale) ontwikkeling van de leerlingen. Dat kan een individuele problematiek betreffen, maar ook de sociale achtergrond van leerlingen speelt mee. Bij sommige leerlingen is het moeilijker vooruitgang te boeken dan bij andere leerlingen. Hoe beter dat in kaart kan worden gebracht, des te beter kunnen we ook de bijdrage van de school zelf aan de ontwikkeling van leerlingen inschatten.

Daarmee komen we bij de indicatoren op basis van *toegevoegde waarde*. Deze worden geoperationaliseerd als dat deel van de leerwinst dat als de unieke bijdrage van de school kan worden beschouwd. In dit type indicatoren wordt een correctie toegepast voor verschillen tussen leerlingen

² In het working paper 'Indicatoren voor schoolkwaliteit op basis van onderwijsopbrengsten' gaan we uitgebreider in op verschillende modellen om de opbrengsten van scholen te beoordelen.

³ In de internationale literatuur wordt overigens leerwinst (d.w.z. rekening houden met het beginniveau van leerlingen) vaak al aangeduid met '*value added*', de Engelse term voor toegevoegde waarde.

waar scholen geen grip op hebben, zoals intelligentie (of leerpotentieel) of meer of minder gunstige omstandigheden thuis. Via statistische modellen, waarin scholen met andere scholen worden vergeleken, worden toetscores of examencijfers van leerlingen ontleed in afzonderlijke componenten. Alleen de component die geen statistische samenhang vertoont met aanvangsverschillen tussen leerlingen (zoals intelligentie, omstandigheden thuis, voorkennis) wordt beschouwd als een schatting van de toegevoegde waarde. Met andere woorden, het onverklaarde deel in de kennisgroei van leerlingen wordt beschouwd als de unieke bijdrage van de school.

Er is een belangrijk verschil tussen de beide typen indicatoren. Leerwinstindicatoren hebben een absoluut karakter: schattingen voor de leerwinst voor één school zijn onafhankelijk van schattingen van de leerwinst voor andere scholen. Beoordeling van de behaald leerwinst in termen van voldoende of onvoldoende gebeurt door vergelijking met landelijke gemiddelden, afkomstig uit de normeringsonderzoeken van toetsontwikkelaars.

Indicatoren op basis van toegevoegde waarde hebben daarentegen een relatief karakter: schattingen van de toegevoegde waarde voor één school zijn afhankelijk van de vorderingen die leerlingen op andere scholen boeken. Pas door opbrengsten van scholen met verschillende leerlingpopulaties met elkaar te vergelijken kan bepaald worden of een school relatief veel of juist weinig vooruitgang boekt met de leerlingen.

Omdat scholen over het algemeen alleen over eigen vorderingengegevens beschikken – en niet over die van andere scholen – zijn indicatoren op basis van leerwinst voor hen veel makkelijker te bepalen dan indicatoren op basis van de toegevoegde waarde. Daar komt nog bij dat voor het schatten van het laatste type indicatoren statistische expertise onontbeerlijk is.

Onderzoeksvragen

Voor het onderzoek bij de pilot zijn de volgende vragen geformuleerd:

1. Welke mogelijkheden/instrumenten zijn er om de leerwinst van leerlingen in kaart te brengen en te meten?
2. Welke modellen zijn mogelijk om de toegevoegde waarde van een afdeling te kunnen bepalen?
3. Op welke wijze kunnen de bevindingen uit de pilotscholen gebruikt worden om zich als school te verantwoorden over hun opbrengsten?
4. Op welke wijze kunnen de bevindingen uit de pilotscholen ingepast worden in de (herziene) opbrengstensystematiek van de Inspectie van het Onderwijs?

In deze tussenrapportage gaan we allereerst in op de eerste onderzoeksvraag. Bij deze vraag is van belang welke instrumenten, zoals toetsen, de pilotscholen gebruiken. Er moet nagegaan worden of deze instrumenten geschikt zijn om ook de ontwikkeling (leerwinst) van leerlingen te volgen. Een ander belangrijk onderdeel is ook de vraag of de door de school geregistreerde gegevens volledig zijn en of het om meerdere meetmomenten gaat. In paragraaf 4 wordt nader ingegaan op de kwaliteit van de

verzamelde gegevens. In paragraaf 5 wordt een vergelijking gemaakt tussen het prestatieniveau van leerlingen op de scholen in vergelijking met de leerwinst die zij geboekt hebben.

De beschrijvende analyses, die aan de scholen zijn teruggedoorgerapporteerd, leveren ook een bijdrage aan de beantwoording van de derde onderzoeksvraag. In die analyses wordt niet alleen de gemiddelde vooruitgang van de leerlingen vergeleken met landelijke normen, maar wordt ook nagegaan of de school een 'moeilijker' of 'makkelijker' leerlingpopulatie heeft dan landelijk. Scholen kunnen zo hun eigen 'verhaal achter de cijfers' onderbouwen.

Bij de tweede onderzoeksvraag gaat het om de vraag hoe bij het beoordelen van de ontwikkeling van leerlingen op de school rekening kan worden gehouden met verschillen in de leerlingpopulatie van de school. In paragraaf 5 wordt geïllustreerd welke verschillen in onderwijsresultaten er bestaan tussen leerlingen met verschillende (achtergrond)kenmerken. Ook wordt nader onderzoek gedaan naar de omvang van de verschillen tussen scholen, wanneer gecorrigeerd is voor verschillen in hun leerlingpubliek.

Onderzoeksvraag vier zal pas later in het pilot project aan de orde kunnen komen.

4.2 Verzamelde gegevens op de scholen

Zoals eerder opgemerkt wordt in de pilots uitgegaan van gegevens, zoals die op de scholen zelf aanwezig zijn. Het idee hierachter is dat scholen zelf, met hun eigen gegevens, leerwinst en toegevoegde waarde moeten kunnen gaan bepalen en beoordelen.

In het voortgezet onderwijs worden onderwijsresultaten, cijfers en toetsscores vooral gebruikt om te determineren: waar is de leerling op zijn/haar plek (op- en afstroom), kan de leerling overgaan of moet hij/zij een jaar over doen? Maar informatie over onderwijsresultaten zou beter benut kunnen worden door meer systematisch de ontwikkeling van leerlingen te volgen. Hoe staat de leerling er voor bij instroom op de school; hoe ontwikkelt hij/zij zich vervolgens gedurende de schooljaren? Hoe gaan de onderwijsresultaten vooruit (*leerwinst*)?

Omdat scholen onderling verschillen in het soort leerlingen dat instroomt in de school is het ook van belang zicht te krijgen op kenmerken van die instroom. De instromende leerlingen vormen het “materiaal” waarmee de scholen moeten gaan werken (Bosker, 2012). Bij de ene leerlingen kan immers ‘makkelijker’ vooruitgang en leerwinst geboekt worden dan bij de andere leerling.

In het project zijn verschillende soorten gegevens van de scholen verkregen:

- kenmerken van leerlingen uit de schooladministraties
- scores op een aantal verschillende toetsen, die gebruikt kunnen worden om de cognitieve ontwikkeling van leerlingen in beeld te brengen
- eindexamencijfers
- gegevens uit de vragenlijsten Motivatie en Burgerschapscompetenties

Niet alle scholen hebben met al deze gegevens meegedaan met de pilot. Op basis van de wensen en mogelijkheden van de scholen zijn er enkele onderzoeksscenario's onderscheiden:

- Scenario 1: toetsgegevens uit het Cito Volgsysteem voortgezet onderwijs.
- Scenario 2: toetsgegevens van andere landelijk genormeerde en methode-onafhankelijke toetsen:
 - Diataal;
 - TOA;
 - ROUTE VO (234VO);
 - Deviant-toetsen.
- Scenario 3: gegevens die betrekking hebben op andere resultaten dan de hiervoor genoemde toetsen:
 - motivatie, self-efficacy ('school zelfvertrouwen') en taakoriëntatie;
 - burgerschapscompetenties;
 - eindexamengegevens, indien mogelijk in relatie tot toetsgegevens onder 1 of 2.

Afhankelijk van de deelname van de scholen aan één of meer scenario's konden in het onderzoek verschillende soorten analyses worden uitgevoerd en naar de scholen worden teruggerapporteerd. Hierna gaan we nader in op de verschillende soorten verzamelde gegevens.

Gegevens uit de schooladministraties

Van de scholen zijn gegevens verkregen van de leerlingen, die in schooljaar 2012/2013 op de school zaten. Algemene gegevens zijn leerlingnamen, geboortedata, geslacht en het (interne) stamnummer van de leerling.

Daarnaast zijn gegevens verkregen over het leerjaar en het type onderwijs (vmbo-basis, vmbo-kader, vmbo-gl/tl, havo, vwo; en combinaties daarvan in brugklassen). Zoals gezegd zijn dit gegevens vanuit het huidige schooljaar. Als een leerling nu, bijvoorbeeld, in leerjaar 3 in een vmbo-tl klas zit weten we niet precies op welk niveau die leerling ook gestart is in het voortgezet onderwijs. Dat kan in een brugklas geweest zijn met vmbo-tl, maar er kan ook sprake zijn geweest van op- of afstroom. In de analyses en de schoolrapporten zijn de leerlingen ingedeeld naar het huidige type onderwijs.

In de ideale situatie wordt voor de analyse van leerwinsten en van toegevoegde waarde uitgegaan van een cohortbenadering: alle leerlingen, die in een bepaald leerjaar instromen in de school (een cohort) doen mee met de toetsen en worden door de jaren heen gevolgd. Dan kan ook op- en afstroom en uitval (of verhuizing) van leerlingen verdisconteerd worden. In deze pilot moeten we echter uitgaan van de 'foto' van schooljaar 2012/2013.

Verder is aan de scholen gevraagd om een aantal nadere kenmerken van de leerlingen. Met deze gegevens willen we een indicatie krijgen van de uitgangssituatie, waarmee leerlingen in de school instromen. Idealiter willen we beschikken over een schatting van het cognitieve niveau en eerdere schoolprestaties van leerlingen. En daarnaast over indicaties rond een meer of minder stimulerend thuisklimaat en eventuele bijzondere omstandigheden.

Voor het aanvangsniveau van de leerlingen is gevraagd naar de score op de Cito Eindtoets basisonderwijs of de NIO en het advies voor voortgezet onderwijs dat de leerling heeft gekregen. Gebleken is dat dit niet op alle scholen systematisch geregistreerd is. Het advies, dat voor alle leerlingen verstrekt wordt, is soms maar deels opgenomen in de administratie. En zowel de Cito-eindtoets als de NIO worden niet op alle basisscholen afgenomen. En ook wanneer deze wel zijn afgenomen wordt de score niet altijd systematisch geregistreerd in de schooladministratie van de school voor voortgezet onderwijs.

Een andere indicatie voor het aanvangsniveau is het al dan niet vertraagd zijn in de eerdere loopbaan door het basisonderwijs. Deze is voor het eerste leerjaar te bepalen uit de geboortedatum van de leerlingen, die voor bijna alle leerlingen is opgenomen in de administratiebestanden. Bij de hogere leerjaren is niet meer scherp te bepalen of vertraging al bij instroom bestond of dat die is opgelopen op de pilotschool zelf.

Wat betreft indicaties van het thuisclimaat zouden we het liefst beschikken over het opleidingsniveau en de etnische herkomst van de ouders. Uit tal van onderzoeken (zie bijvoorbeeld Onderwijsraad, 2013) blijken deze kenmerken samen te hangen met de onderwijsresultaten van de kinderen. Vanwege de gewichtenregeling wordt in het basisonderwijs doorgaans geregistreerd of de ouders lage opleidingen hebben, hoewel het onderscheid tussen middelbare en hogere opleidingen ook daar vaak ontbreekt. In het voortgezet onderwijs is de opleiding van de ouders echter onbekend. Als grove benadering van de thuissituatie kan daar alleen gewerkt worden met het al dan niet wonen in een zogenaamd APCG: een armoedeprobleemcumulatiegebied (SCP/CBS, 2007). Vanuit de postcodes van leerlingen in de schooladministraties kan deze indicator voor vrijwel alle leerlingen bepaald worden; incidenteel ontbrak zo'n postcode in de aangeleverde gegevens.

Of er sprake is van bijzondere omstandigheden bij de leerlingen, wordt zeer wisselend geregistreerd door de scholen. Onduidelijk is ook of dit systematisch is gebeurd. In het algemeen is wel bekend of er sprake is van leerwegondersteuning (lwoo).

Wanneer scholen de ontwikkeling van leerlingen systematisch willen gaan bijhouden en ook willen nagaan hoe verschillende groepen leerlingen zich wat dat betreft ontwikkelen, dan zou een eerste stap moeten zijn om goed te registreren wat er bekend is over de 'uitgangssituatie' van de startende leerlingen: wat is bekend over hun eerdere loopbaan (advies, cito- en of nio, vertraging), wat voor speciale omstandigheden zijn er.

Toetsscores

In de pilot is gebruik gemaakt van een aantal niet methoden gebonden toetsen, die betrekking hebben op de domeinen Nederlandse taal, Rekenen/wiskunde en Engels⁴. Om de vooruitgang van leerlingen op deze domeinen scherp in beeld te krijgen zijn er prestatie metingen van leerlingen nodig op tenminste twee tijdstippen. Leerwinst wordt pas echt meetbaar als de toetsscores op die twee meetmomenten ook op eenzelfde schaal zijn uit te drukken.

Van de bij de scholen gebruikte toetsen voldoen aan deze eisen:

- de Cito toetsen uit het volgsysteem voortgezet onderwijs (VVO, voorheen VAS);
- de DiaTaal toetsen;
- de toetsen van Route VO (voorheen Toets 234VO).

De twee andere toetsen, die scholen in de pilot gebruiken, TOA en Deviant, worden gebruikt om de beheersing van de referentieniveaus voor taal en rekenen mee te bepalen. Er worden geen vaardigheidsscores, die op eenzelfde schaal kunnen worden geplaatst, mee bepaald.

⁴ Deze toetsen worden uitgebreider besproken in het working paper "Methode-onafhankelijke toetsen in het voortgezet onderwijs".

Verder moeten de toetsen ook valide en betrouwbaar zijn en voorzien zijn van landelijke normen. Een van de eisen die we in dit verband aan de toetsen kunnen stellen is dat ze als voldoende of goed zijn beoordeeld door de COTAN (Commissie Testaangelegenheden Nederland van het Nederlands Instituut van Psychologen). Van de in de pilot gebruikte toetsen zijn alleen de Cito-toetsen en de Route VO toetsen door de COTAN beoordeeld en op de meeste criteria als voldoende of goed beoordeeld. Bij deze toetsen bestaan ook landelijke normtabellen. DiaTaal is(nog) niet beoordeeld door de COTAN, maar wel wetenschappelijk onderbouwd (Hacquebord e.a., 2005). DiaTaal geeft normtabellen die jaarlijks worden aangepast. Van TOA en Deviant is geen landelijk normeringsonderzoek bekend.

Voor de meeste scholen zijn, na toestemming van de scholen hiervoor, de toetsscores centraal verkregen bij de toetsaanbieders (Cito, Diataal, TOA). Het per leerling koppelen van deze scores aan gegevens uit de schooladministraties heeft de nodige problemen opgeleverd. De verkregen gegevens bestaan uit een groot aantal losse bestanden (per afnamejaar en per toets). Bij Diataal en TOA is daarbij veelal ook het (unieke) stamnummer opgenomen, waaronder de leerlingen ook bekend zijn in het administratiebestand. Maar bij Cito zijn de leerlingen alleen identificeerbaar aan de hand van hun naam, geboortedatum en geslacht. Vanwege spelfouten, verwisseling van voor- en achternaam, ontbrekende geboortedata, enzovoorts, is de koppeling daar niet volledig geslaagd. Voor statistische doeleinden hoeft dit niet zo'n probleem te zijn, maar wel als men de ontwikkeling van een individuele leerling wil kunnen volgen.

Wanneer de scholen zelf de ontwikkeling van de toetsscores van leerlingen systematisch willen gaan volgen zou het aan te bevelen zijn dat scholen de uitkomsten van de verschillende toetsen ook zelf in hun leerlingvolgsystemen gaan registreren. Of dat gebruik wordt gemaakt van, bijvoorbeeld, de mogelijkheid om in het Cito portal ook het stamnummer van de leerlingen op te nemen. Maar het zou nog beter zijn als er een automatische koppeling tot stand zou kunnen komen tussen het volgsysteem en de toetsregistratiesystemen.

Tot slot merken we nog op dat al de gebruikte toetsen algemene vaardigheden meten. In de curricula van de onderzochte scholen komen natuurlijk ook allerlei andere vaardigheden aan de orde, bijvoorbeeld bij de beroepsgerichte vakken in het vmbo tot aan de klassieke vorming en talen op het gymnasium. Voor zulke vakken zijn over het algemeen echter geen landelijk genormeerde instrumenten voorhanden, die het mogelijk maken om opbrengsten van scholen (in termen van leerwinst en toegevoegde waarde) met elkaar te vergelijken.

Eindexamencijfers

Er zijn enkele pilotscholen die ook hun opbrengsten in termen van eindexamengegevens nader willen analyseren, bij voorkeur in verband gebracht met (toets)gegevens van leerlingen bij instroom in de school. Het aantrekkelijke daarvan is dat de schoolloopbaan als geheel vanaf het begin (start leerjaar 1) tot en met het eindexamen onderzocht wordt. Tegelijk brengt zo'n langere periode van tenminste vier leerjaren ook een aantal problemen met zich mee: er is tussentijds sprake van zittenblijven, op- en

afstroom en in- en uitstroom op de school. De bij de scholen verkregen eindexamencijfers bieden niet goed zicht op dat hele tussenliggende proces.

Vragenlijsten Motivatie

Deze vragenlijst wordt gebruikt in het landelijke cohortonderzoek COOL⁵⁻¹⁸ (Driessen e.a., 2012; Zijsling e.a., 2012; zie ook www.cool5-18.nl.) Uit de vragenlijst kunnen scores op de hieronder omschreven schalen worden bepaald. De schalen gaan over doelen voor leren. Zo werken sommige leerlingen graag hard, omdat ze het beter willen doen dan andere leerlingen (Prestatie), anderen doen hun best op school, omdat ze graag complimentjes en beloningen ontvangen (Extrinsiek).

Bekwaming (mastery) betreft de mate waarin leerlingen gericht zijn op het ontwikkelen van hun eigen vaardigheden en het verbeteren daarvan. Items zijn bijvoorbeeld: 'Ik zie graag dat mijn schoolwerk steeds beter wordt' en 'Ik doe goed mijn best op school omdat ik mijn schoolwerk interessant vind'.

Prestatie (performance) gaat over je best doen op school om beter te presteren in vergelijking met andere leerlingen en het sociaal gezien verkrijgen van aanzien. Items zijn bijvoorbeeld: 'Ik ben alleen maar tevreden als ik een van de besten van de klas ben' en 'Ik doe mijn best op school zodat ik de leider van een groepje word'

Sociaal (social). Leerlingen die hier hoog op scoren, waarderen het sociale aspect van leren: ze voelen zich betrokken bij andere kinderen en waarderen de omgang met andere kinderen. Voorbeelditems zijn 'Ik help graag andere kinderen goede cijfers te halen op school' en 'Ik werk op school liever samen met andere kinderen dan alleen'.

Extrinsiek (extrinsic) gaat over harder willen werken voor school als je complimentjes en beloningen ontvangt. Items zijn bijvoorbeeld: 'Ik krijg graag complimentjes voor goed schoolwerk' en 'Als je op school 'bonuspunten' zou kunnen krijgen, zou ik beter mijn best doen'.

Met **taakmotivatie** wordt bedoeld de mate waarin leerlingen gericht zijn op het ontwikkelen van hun vaardigheden, op leren en het begrijpen van schoolwerk. Leerlingen zijn daarbij gericht op hun eigen prestaties en hoe die zich ontwikkelen en niet op de vergelijking met andere leerlingen. Items zijn bijvoorbeeld: 'Ik ben tevreden als ik op school iets heb geleerd dat ik begrijp' en 'Als ik op school iets niet meteen snap, ga ik er juist extra mijn best voor doen'.

Met **cognitief zelfvertrouwen** (self-efficacy) wordt bedoeld de mate waarin leerlingen vertrouwen in zichzelf hebben dat zij kunnen slagen op school. Hier zijn items bijvoorbeeld: 'Ik weet zeker dat dit jaar alles op school me wel zal lukken' en 'Ik kan bijna alles op school, als ik het maar blijf proberen'.

De eerste vier schalen zijn een vertaalde versie van de Inventory of School Motivation van Ali & McInerney (2004). Taakmotivatie is gemeten door gebruik te maken van een schaal door Seegers, Van Putten & De Brabander (2002). Bij cognitief zelfvertrouwen is gebruik gemaakt van de schaal 'self-efficacy' die afkomstig is van de 'Patterns of Adaptive Learning Survey (PALS)' (Midgley e.a., 2000). Alle

genoemde schalen laten ook op de gegevens in dit pilot onderzoek een goede betrouwbaarheid zien (Cronbach's alpha van .78 tot .88).

De vragenlijst is door leerlingen van een aantal pilotscholen online ingevuld. Identificatie van de leerlingen en koppeling van deze gegevens aan de administratiebestanden van de scholen moest gebeuren aan de hand van leerlingnamen en geboortedata. Net als bij de toetscores bleek dat niet altijd te lukken, bijvoorbeeld doordat leerlingen geen naam hebben ingevuld.

Vragenlijsten Burgerschapscompetenties

Ook deze vragenlijst wordt gebruikt in het landelijke cohortonderzoek COOL⁵⁻¹⁸ (Driessen e.a., 2012; Zijlsing e.a., 2012; zie ook www.cool5-18.nl.) Het meetinstrument burgerschap (Ten Dam e.a., 2010) is ontwikkeld als antwoord op toenemende aandacht voor burgerschapsvorming in het onderwijs. In het instrument is het begrip competentie uiteengelegd in vier componenten: kennis, reflectie, vaardigheden en attituden. Het begrip burgerschap is geoperationaliseerd aan de hand van vier centrale sociale taken:

- democratisch handelen : het aanvaarden van en bijdragen aan een democratische samenleving;
- maatschappelijke verantwoordelijkheid: medeverantwoordelijkheid nemen voor de leefgemeenschappen waartoe men behoort;
- omgaan met conflicten: betreft (lichte) conflictsituaties of belangentegenstellingen waarbij de jongere zelf 'partij' is;
- omgaan met verschillen: betreft sociale, culturele, religieuze en uiterlijke verschillen.

Bij elk van deze vier centrale taken worden de componenten kennis, reflectie, vaardigheden en attituden met een aantal items gemeten. Het instrument omvat:

- 28 reflectie-items.
Deze gaan over reflecteren, nadenken over. De voorgelegde vraag is: 'hoe vaak denk jij na over...', bijvoorbeeld '...of er naar leerlingen wordt geluisterd op jouw school?'; '.....hoe het komt dat mensen van verschillende afkomst weinig met elkaar omgaan? .
- 15 vaardigheden-items.
Hier moet de leerling aangeven hoe goed hij/zij zichzelf ergens in vindt. De voorgelegde vraag is 'hoe goed ben jij in...', bijvoorbeeld '...indenken hoe een ander zich voelt en daarmee rekening houden?'; '...je taalgebruik aanpassen aan degene met wie je spreekt?'

- 24 attituden-items.
Deze gaan over iets vinden (menen), willen, tot iets bereid zijn. De voorgelegde vraag is 'hoe goed past een uitspraak bij jou?'. Uitspraken zijn bijvoorbeeld: 'Als iemand in de klas het ergens mee oneens is, moet hij/zij de kans krijgen om dat uit te leggen'; 'Ik vind het leuk om iets te weten van verschillende soorten geloof'.
- 27 kennis-items.
Deze gaan over weten, begrijpen, inzicht hebben in wat je het beste kunt doen in een specifieke situatie. Bij de kennis-items kiezen leerlingen het beste antwoordalternatief uit bij een vraag, bijvoorbeeld: 'Alle kinderen hebben recht: a) op zakgeld, b) om te kiezen bij wie je woont, c) op onderwijs'. In dit geval is optie c het goede antwoord.

In de analyses en schoolrapporten worden de vier hoofdschalen Attitude, Vaardigheid, Reflectie en Kennis gebruikt. De score op Vaardigheid is dan bijvoorbeeld de samenvatting van alle vaardigheden met betrekking tot democratisch handelen, maatschappelijke verantwoordelijkheid, omgaan met conflicten én omgaan met verschillen. We hebben de betrouwbaarheid van de vier schalen ook op de gegevens in dit pilot onderzoek nagegaan en er blijkt steeds sprake te zijn van een goede betrouwbaarheid (Cronbach's alpha steeds groter dan .85).

4.3 Beschrijvende analyses per school: de schoolrapporten

In de schoolrapporten zijn naar de scholen beschrijvende gegevens teruggekoppeld. Deze wisselen weliswaar naar het soort gegevens (type toets, motivatie en burgerschap, eindexamencijfers), waarmee de school deelneemt aan de pilot, maar volgen zoveel mogelijk wel dezelfde opzet.

Kenmerken van de leerlingeninstroom

De scholen in de pilot verschillen in hun leerlingpopulatie. Dat hangt uiteraard samen met welke afdeling van de school meedoet in het onderzoek, maar ook scholen die met een zelfde soort afdeling meedoen kunnen onderling nog sterk verschillen, bijvoorbeeld naar het aandeel leerlingen met lwoo, het aandeel leerlingen uit een armoedeprobleemcumulatiegebied, het aandeel leerlingen dat met een hoger of juist een lager advies aan de afdeling begint.

In de schoolrapporten worden gegevens van de instroom in schooljaar 2012/2013 vergeleken met landelijke cijfers. Deze laatste cijfers zijn afkomstig uit landelijke registerbestanden (BRON, gebaseerd op het onderwijsnummer) en voor dit onderzoek bewerkt door de Inspectie van het Onderwijs. Omdat de samenstelling van (brede) brugklassen nogal heterogeen is worden ook cijfers gegeven van het derde leerjaar, als de nadere uitsplitsing naar onderwijstype, ook in de landelijke gegevens, voltooid is.

Voor zover bekend uit de op de school aanwezige gegevens gaat het om de volgende kenmerken:

- percentage meisjes
- percentage leerlingen uit een armoede probleemcumulatiegebied (APCG)
- percentage leerlingen met indicatie voor leerwegondersteuning (lwoo)
- percentage leerlingen dat bij instroom vertraagd is
- percentage leerlingen met een advies dat hoger of lager was dan de gekozen afdeling
- percentage leerlingen met een score op de Cito Eindtoets basisonderwijs die hoger of lager is dan de gekozen afdeling⁵
- percentage leerlingen met een score op de NIO (of andere IQ-test) die hoger of lager is dan de gekozen afdeling⁶

Hieronder volgt een voorbeeldtabel uit een van de schoolrapporten.

⁵ Hoger of lager is gebaseerd op: Betekenis van de standaardscore op de Citotoets 2012, Arnhem: Cito

⁶ Hoger of lager is gebaseerd op: Van Dijk & Telligen, 2004, De NIO en schoolkeuze. Amsterdam, Boom. Tabel 5.6

Voorbeeldtabel 4.1 Achtergrondkenmerken leerlingen in leerjaar 1; landelijk en op uw school

landelijk eerste leerjaar	% meisjes	% APCG	% lwoo	% vertraagd bij instroom
vmbo basis en kader	45	26	49	46
vmbo-breed (incidenteel + havo)	49	18	38	39
vmbo-(g)t	52	17	5	26
vmbo-breed/havo/vwo	50	10	12	20
vmbo-(g)t/havo/vwo	50	12	1	14
havo/vwo	49	13	0	9
vwo	46	14	0	4
uw pilot afdeling				
vmbo breed leerjaar 1 2012/2013	50	99	38	53

Deze beschrijving van de leerlinginstroom en de populatie in het derde leerjaar geeft de school een beeld van hun leerlingen, vergeleken met soortgelijke afdelingen in het land. Daarmee kan duidelijk worden of er op de school sprake is van een “moeilijker” of “makkelijker” leerlingpopulatie.

Behaalde toetsresultaten

Op de pilotscholen worden verschillende toetsen gebruikt, die soms sinds kort en soms al langere tijd worden afgenomen bij de leerlingen. Het gaat om de volgende toetsen:

- Cito toetsen uit het Volgsysteem Voortgezet Onderwijs (VVO; daarvoor bekend als VAS)
- DiaTaal toetsen
- Route VO (voorheen 234VO)
- TOA toetsen
- Deviant toetsen

Deze toetsen worden uitgebreider besproken in het working paper ‘Methode-onafhankelijke toetsen in het voortgezet onderwijs’. Een belangrijk verschil is dat de eerste drie toetsen de mogelijkheid bieden om de scores van leerlingen op opeenvolgende toetsen op dezelfde schaal uit te drukken, waarmee het mogelijk wordt om vooruitgang (leerwinst) in beeld te brengen. De laatste twee toetsen zijn gekoppeld aan de referentieniveaus voor Taal en Rekenen en geven een veel globaler beeld in termen van wel/niet beheersing van een bepaald referentieniveau.

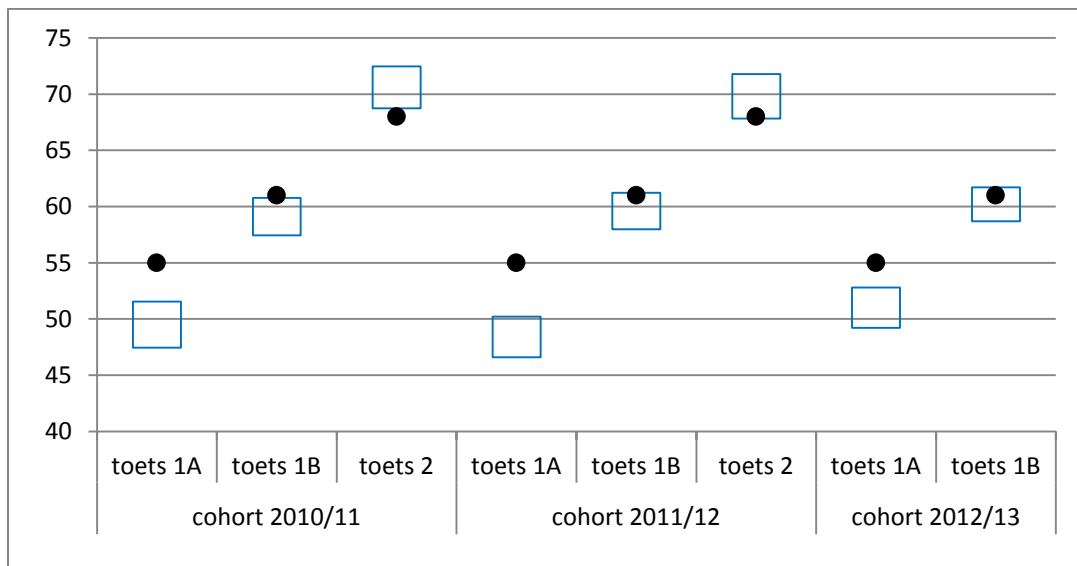
Om na te gaan of leerlingen met verschillende kenmerken en achtergronden ook een lager of hoger beginniveau hebben wordt in de schoolrapporten eerst een beschrijving gegeven van de gemiddelde scores op de in schooljaar 2012/2013 afgenomen begintoets op de school, onderscheiden naar de verschillende leerlingkenmerken. Daarmee kunnen de scholen zicht krijgen op de vraag of er bij de beoordeling van hun onderwijsopbrengsten rekening zou moeten worden gehouden met bepaalde kenmerken van hun leerlingenpubliek.

Als volgende stap in de beschrijvende analyses krijgen scholen een overzicht van de gemiddelde toetsresultaten van verschillende cohorten leerlingen in hun school. Voor leerlingen, die aan opeenvolgende toetsen hebben deelgenomen, wordt ook de gemiddelde leerwinst tussen die twee toetsafnames vermeld. Ter vergelijking zijn ook de landelijke normen van de betreffende toets in het overzicht opgenomen. Daaruit kan ook afgeleid worden hoe leerlingen zich (landelijk gemiddeld) ontwikkelen tussen twee toetsafnames.

De ontwikkeling van de gemiddelde resultaten op de toetsen in verschillende cohorten wordt ook grafisch weergegeven. Daarbij wordt tevens in beeld gebracht welke onzekerheidsmarges er rondom deze gemiddelden bestaan. De toetsen zijn immers geen perfecte meetinstrumenten: als dezelfde groep leerlingen de toets enkele dagen eerder of later zou maken dan zal daar nooit precies hetzelfde gemiddelde uitkomen. Deze onzekerheid is groter naarmate het aantal leerlingen kleiner is en naarmate er meer variatie in scores is tussen de leerlingen.

Hierna wordt een voorbeeld van deze grafieken getoond. Het schoolgemiddelde met onzekerheidsmarges wordt daarin weergegeven als een blokje en de landelijke (norm)gemiddelden als een punt. In de voorbeeldgrafiek worden scores op de DiaTaal toets voor begrijpend lezen weergegeven in drie opeenvolgende cohorten.

Voorbeeldfiguur 4.1 Score op Begrijpend lezen (Diataal) op drie toetsmomenten bij drie cohorten (de stip geeft de landelijke norm weer; het blokje geeft het schoolgemiddelde met een onzekerheidsmarge naar boven en beneden)



Eindexamencijfers

Bij de eindexamencijfers zijn twee soorten beschrijvende analyses naar de scholen gerapporteerd. Voor de leerlingen, waarbij ook een score op een begintoets bekend is, wordt de relatie gelegd tussen hun aanvangsscore en de eindexamenresultaten. Daarmee wordt duidelijk of de bij aanvang sterkere of zwakkere leerlingen ook aan het eind van de schoolperiode betere of minder goede resultaten behalen. En dat is van belang voor de vraag of er bij de beoordeling van de eindopbrengsten van scholen rekening zou moeten worden gehouden met de (cognitieve) verschillen tussen leerlingen bij instroom.

Voorbeeldtabel 4.2 Beginscores Begrijpend Lezen (toets 1A uit 2009/2010) vergeleken met het CE-cijfer Nederlands (eindexamen 2012/2013).

Kwartielen Begrijpend Lezen	% onvoldoende	% voldoende	% goed	aantal
0-25%	13%	50%	38%	8
25-50%	0%	14%	86%	7
50-75%	0%	27%	73%	11
75-100%	11%	44%	44%	9

Verder worden de eindexamencijfers in verband gebracht met de (achtergrond)kenmerken van de leerlingen. Ook dat is van belang voor de vraag of zulke verschillen in leerlingkenmerken een rol zouden moeten spelen bij het beoordelen van de opbrengsten van de school.

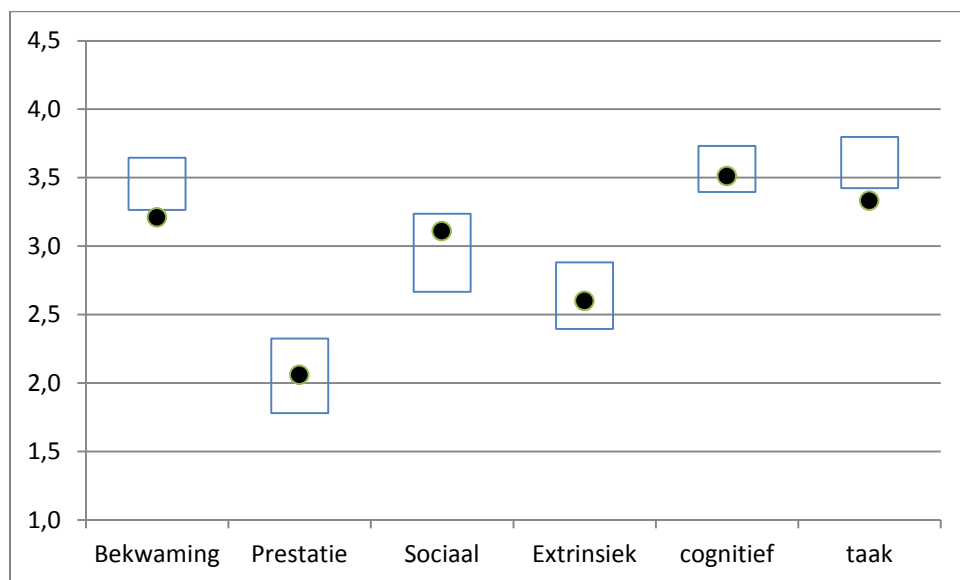
We merken wel op dat deze beschrijvende analyses alleen kunnen worden uitgevoerd voor leerlingen die ook aan het eindexamen hebben deelgenomen. Het is goed mogelijk dat tussentijds uitgevallen leerlingen, bijv. door afstroom, sterk verschillen van de leerlingen die wel tot aan het eindexamen zijn doorgestroomd in de school.

Vragenlijst Motivatie

Deze vragenlijst wordt gebruikt in het landelijke cohortonderzoek COOL⁵⁻¹⁸. Uit dat onderzoek zijn ook de landelijke gemiddelden afkomstig op de vijf schalen, die in paragraaf 4.2 zijn omschreven. In de rapportage naar de scholen worden gemiddelde scores op deze schalen gegeven, in vergelijking met deze landelijk gemiddelden uit leerjaar 3 van het voortgezet onderwijs. Ook worden verschillen naar leerlingkenmerken gepresenteerd.

In de volgende voorbeeldfiguur worden gemiddelde schoolscores (met onzekerheidsmarges) op de vijf schalen vergeleken met landelijke scores.

Voorbeeldfiguur 4.2 Score op zes motivatieschalen(leerjaar 3 vmbo tl/gl); vergeleken met landelijke gemiddelden uit leerjaar 3 vmbo tl/gl
(de stip geeft de landelijke norm weer; het blokje geeft het schoolgemiddelde met een onzekerheidsmarge naar boven en beneden)

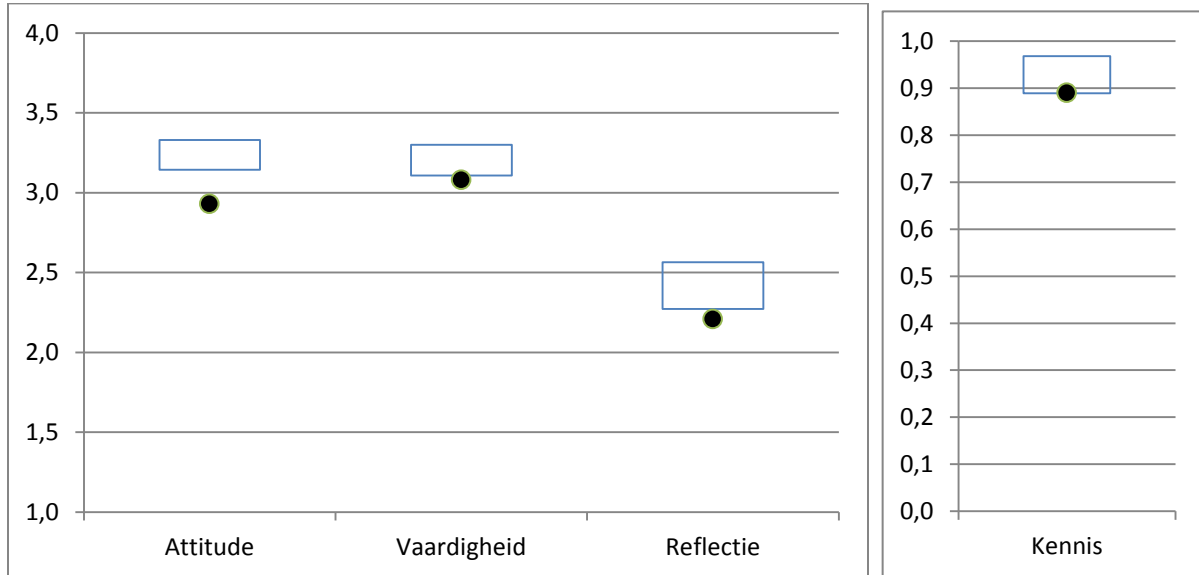


Het ligt in de bedoeling om deze vragenlijst ook in het komende schooljaar weer af te nemen, zodat ook ontwikkelingen in beeld kunnen worden gebracht.

Vragenlijst Burgerschapscompetenties

Ook deze vragenlijst wordt gebruikt in het landelijke cohortonderzoek COOL⁵⁻¹⁸. Het instrument levert scores op vier schalen, die in vorige paragraaf zijn omschreven. In de rapportage naar de scholen worden gemiddelde scores op deze schalen gegeven, in vergelijking met landelijk gemiddelden uit leerjaar 3 van het voortgezet onderwijs. Hieronder staat weer een voorbeeldfiguur.

Voorbeeldfiguur 4.3 Score op vier burgerschapsschalen (afd. vwo); vergeleken met landelijke gemiddelden uit leerjaar 3 vwo
(de stip geeft de landelijke norm weer; het blokje geeft het schoolgemiddelde met een onzekerheidsmarge naar boven en beneden)



Verder worden in het schoolrapport verschillen naar leerlingenkenmerken gepresenteerd. Ook deze vragenlijst wordt het komende schooljaar nogmaals afgenomen, om ontwikkelingen in beeld te brengen.

4.4 Bespreking schoolrapporten met de scholen

Voor elke school die deelneemt aan de pilot is een schoolrapport gemaakt, dat begin oktober 2013 naar de school is verstuurd. De ondersteuners van Oberon hebben vervolgens de rapporten op de scholen besproken. Van deze gespreken is door Oberon de volgende samenvatting gemaakt.

Op 25 scholen⁷ is met schoolleiders en andere betrokkenen een gesprek gehouden over de eerste schoolrapportage van de pilot leerwinst en toegevoegde waarde. In deze gesprekken hebben de schoolbegeleiders van Oberon de inhoud van de rapportages besproken. Daarnaast is ook gekeken wat er nog verbeterd kan worden aan de rapportages, welke meerwaarde het rapport heeft voor de school en of het rapport een handig hulpmiddel is bij opbrengstgericht werken. Tenslotte is met de schoolleiders besproken met wie zij de uitkomsten uit de rapportage willen delen en op welke manier zij verder ondersteund willen worden. Hieronder bespreken we de belangrijkste uitkomsten uit de gesprekken op de pilotscholen.

Suggesties voor verbetering schoolrapport

Ongeveer helft van de schoolleiders vond het schoolrapport goed te interpreteren terwijl anderen meer behoefte hadden aan uitleg. Bijvoorbeeld door aan te geven hoe de tabellen en figuren gelezen moeten worden en waarom bepaalde gegevens wel en niet zijn weergegeven in de rapportage. De meeste schoolleiders vonden het nuttig om de rapporten te bespreken, omdat op die manier de gegevens in het rapport verduidelijkt werden. Schoolleiders hadden soms hele specifieke verbeterpunten voor hun schoolrapport, maar over het algemeen is behoefte aan:

- het plaatsen van legenda's bij tabellen en figuren;
- het opnemen van (eind)conclusies;
- het noemen van absolute aantallen bij tabellen;
- het benoemen van de landelijke leerwinst;
- het weergeven van leerwinst (landelijk en van de school) in een lijngrafiek;
- het vergelijken van de resultaten met scholen met een vergelijkbare leerlingpopulatie.

⁷ Op twee scholen is geen gesprek gevoerd, omdat er nog geen of onvoldoende data was aangeleverd voor het maken van een eerste schoolrapportage. Bij het schrijven van deze rapportage waren van drie scholen de verslagen nog niet gereed. We verwachten niet dat deze drie verslagen tot zeer grote veranderingen van de samenvattende bevindingen leiden.

Meerwaarde van het schoolrapport ten opzichte van de eigen rapportages

Voor veel schoolleiders bevestigt het rapport dat wat ze eigenlijk al wisten, maar in sommige gevallen niet hard kunnen maken. Voor een vijfde van de schoolleiders heeft de rapportage echter geen meerwaarde. Zij beschikken zelf over uitgebreide rapportages. De meerwaarde van het rapport zit volgens de schoolleiders vooral in:

- het weergeven van de leerwinst van de leerlingen;
- de vergelijking van de resultaten op basis van de verschillende leerlingkenmerken;
- de vergelijking met landelijke gegevens.

Opbrengstgericht werken

Voor de meeste schoolleiders is de rapportage een handig middel om opbrengstgericht mee te kunnen werken. Met name doordat de leerwinst inzichtelijk wordt gemaakt kunnen schoolleiders zelf zien waar nog meer winst behaald kan worden. Een op de vijf schoolleiders ziet weinig of geen meerwaarde van de rapportage. Dit komt omdat zij al over rapportages beschikken die soortgelijke of uitgebreidere informatie bevatten.

Delen van informatie uit de rapportage

De schoolleiders gaan de uitkomsten uit de rapportage in eerste instantie vooral bespreken met de betreffende sectie- of kernteamleiders. Vervolgens zullen ze de resultaten ook bespreken met hun docententeam. Iets meer dan een tiende van de schoolleiders wil de rapportage ook delen met hun bestuur of (vergelijkbare) scholen uit hun bestuur. Een op de vijf schoolleiders geeft aan de resultaten niet te delen met ouders omdat ze de rapportage daarvoor niet geschikt vinden. Tenslotte geven twee schoolleiders aan dat de rapportage niet zonder extra uitleg verspreid zal worden onder docenten, omdat deze te ingewikkeld is voor de docenten.

Ondersteuningsbehoefte

Schoolleiders konden tijdens het gesprek zelf aangegeven aan welke vorm van ondersteuning zij behoefte hebben. Vooral de workshop opbrengstgericht werken voor docenten en de bijeenkomst over afstroom werden veel genoemd. Ook gaven twee schoolleiders aan behoefte te hebben aan ondersteuning bij de bespreking van de rapportage met het team of de vakgroepen. Daarnaast waren er twee schoolleiders die graag geholpen worden bij het uitvoeren van extra analyses. Tenslotte gaven twee schoolleiders aan dat zij op dat moment geen behoefte hadden aan ondersteuning.

5. Prestatieniveau en leerwinst vergeleken

De opbrengsten van scholen kunnen beoordeeld worden naar het gemiddelde prestatieniveau, dat de leerlingen op een bepaald moment halen of naar de leerwinst, die de leerlingen over de afgelopen periode hebben behaald. Het is aan te bevelen om beide aspecten te gebruiken bij het beoordelen van onderwijsopbrengsten (Roeleveld, Veen & Ledoux, 2009; Bosker, 2012). Op die manier kunnen scholen in het volgende kwadrant ingedeeld worden:

laag niveau veel leerwinst	hoog niveau veel leerwinst
laag niveau weinig leerwinst	hoog niveau weinig leerwinst

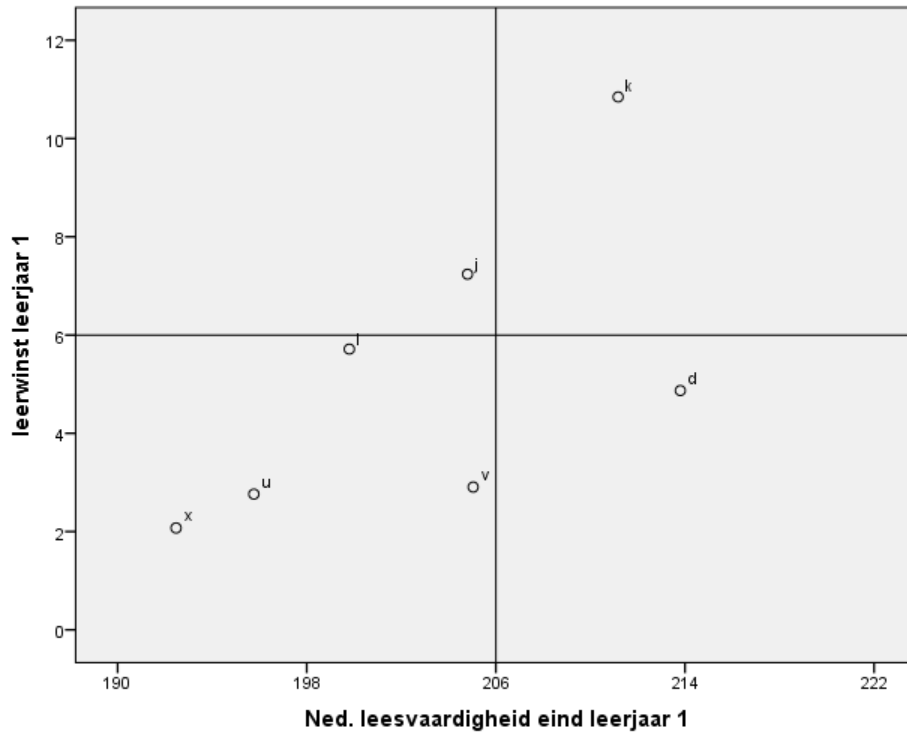
Met de gegevens uit de pilot wordt het mogelijk om de deelnemende scholen op een dergelijke manier in te delen. Wij doen dit hierna per type afdeling en voor de score op Nederlandse leesvaardigheid uit de Cito-VVO toetsen voor het eerste leerjaar.

Op de horizontale as van de volgende figuren staat de gemiddelde score aan het eind van het eerste leerjaar. Op de verticale as staat de leerwinst, behaald tussen het begin en het eind van dat eerste leerjaar. De gekruiste lijnen door het midden geven de landelijke gemiddelden uit het normeringsonderzoek van Cito. De afzonderlijke scholen worden in deze figuren aangeduid met een letter.

Voor deze figuren gebruiken we zoveel mogelijk gegevens over de vooruitgang van leerlingen in het eerste leerjaar, namelijk van zowel schooljaar 2012/2013 als van eerder schooljaren (voor zover bekend). De schoolgemiddelden worden als een punt (met letteraanduiding) weergegeven.

Rondom deze gemiddelden zit echter ook een (statistische) onzekerheidsmarge. Onder de figuren vermelden we welke scholen significant van de landelijke gemiddelde normen afwijken. Daaruit blijkt dat het vaak niet heel scherp mogelijk is om de scholen in de vier kwadranten in te delen: de onzekerheid in de metingen verhindert dit dan.

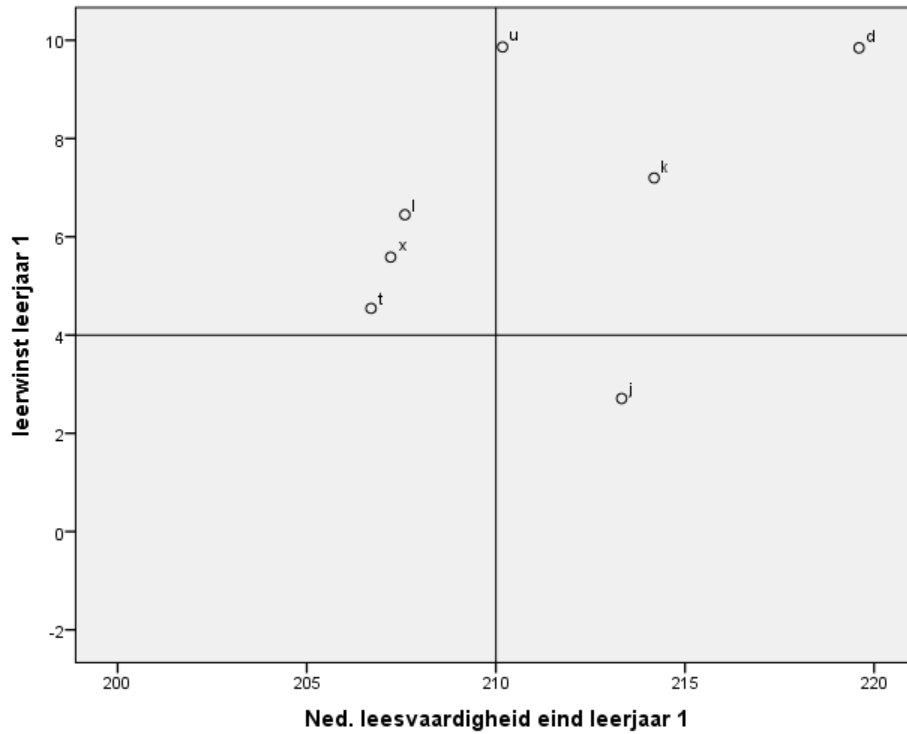
Figuur 5.1 Score leesvaardigheid eind leerjaar 1 versus leerwinst in leerjaar 1; voor vmbo-basis (normering Toets 0 en Toets 1: 200 resp. 206; vooruitgang 6)



school:	d	j	k	l	u	v	x
eind leerjaar 1	hoger	n.s.	hoger	lager	lager	n.s.	lager
leerwinst	n.s.	n.s.	hoger	n.s.	n.s.	lager	n.s.

School k valt in het meest gunstige kwadrant: de prestaties aan het eind van leerjaar 1 zijn bovengemiddeld, evenals de leerwinst die in dat jaar behaald is. Voor scholen x en u geldt juist het tegendeel: de twee indicatoren zijn ondergemiddeld. Voor beide scholen geldt echter ook dat de afwijking van leerwinst van het landelijke gemiddelde niet significant is.

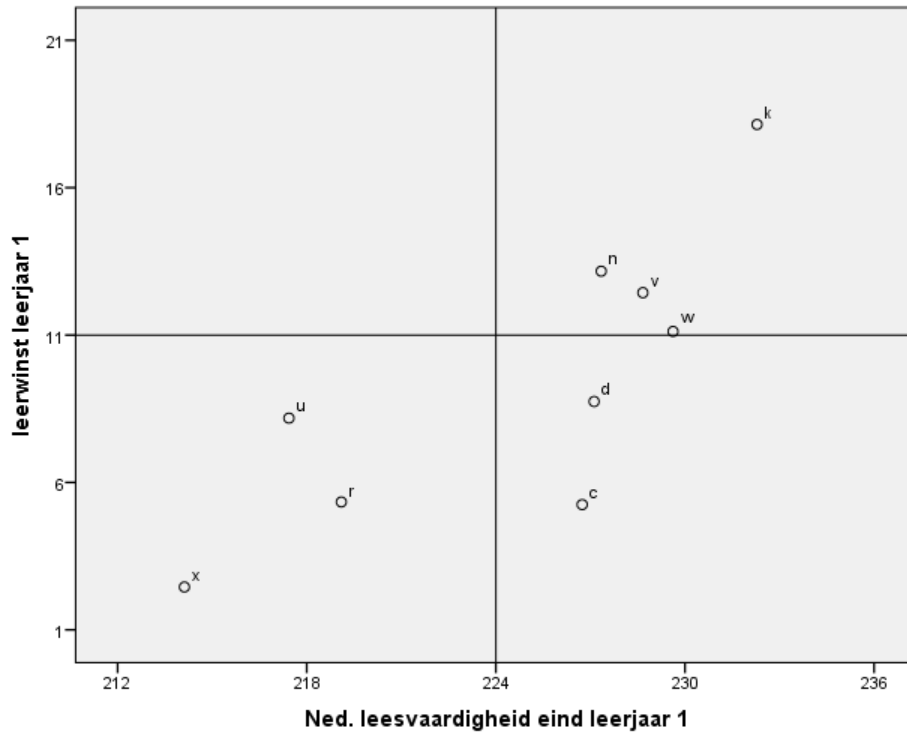
Figuur 5.2 Score leesvaardigheid eind leerjaar 1 versus leerwinst in leerjaar 1; voor vmbo-kader (normering Toets 0 en Toets 1: 206 resp. 210; vooruitgang 4)



school:	d	j	k	l	t	u	x
eind leerjaar 1	hoger	n.s.	hoger	n.s.	lager	n.s.	lager
leerwinst	hoger	n.s.	hoger	n.s.	n.s.	hoger	hoger

Hier valt vooral school d op: zowel de eindscore als de leerwinst liggen flink boven het gemiddelde. Op deze zelfde school behaalde de vmbo-basis leerlingen ook een bovengemiddelde score, maar was de leerwinst ongeveer gemiddeld (zie Figuur 5.1). Er valt hier geen enkele school in het meest ongunstige kwadrant linksonder.

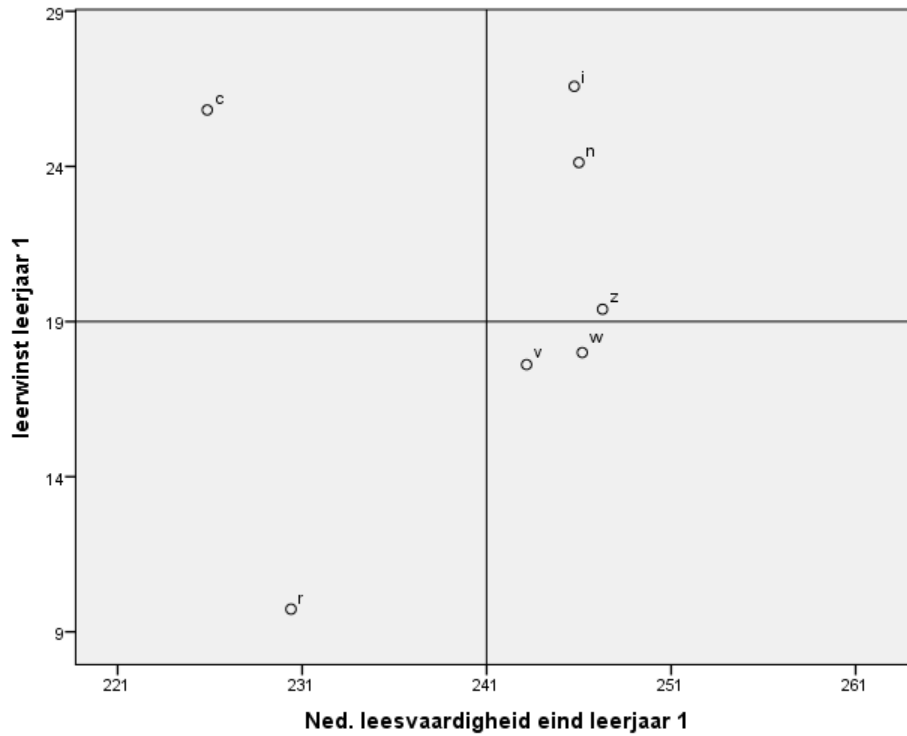
Figuur 5.3 Score leesvaardigheid eind leerjaar 1 versus leerwinst in leerjaar 1; voor vmbo-tl/gl (normering Toets 0 en Toets 1: 213 resp. 224; vooruitgang 11)



school:	c	d	k	n	r	u	v	w	x
eind leerjaar 1	n.s	n.s.	hoger	hoger	lager	lager	hoger	hoger	lager
leerwinst	n.s.	n.s.	hoger	n.s.	lager	n.s.	n.s.	n.s.	lager

De meeste scholen vallen in de kwadranten rechtsboven (zowel eindscore als leerwinst gunstig) of linksonder (beide ongunstig), hoewel de verschillen met de landelijke normen niet altijd significant blijken te zijn. Scholen d en c zitten in het kwadrant rechtsonder, maar hun scores wijken niet significant af van de landelijke normen.

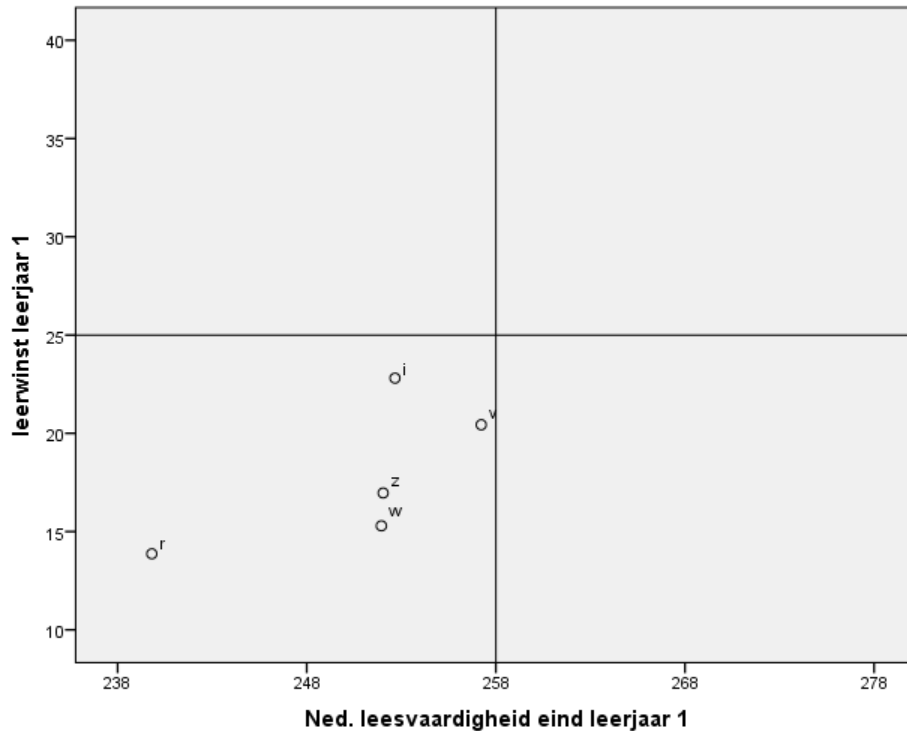
Figuur 5.4 Score leesvaardigheid eind leerjaar 1 versus leerwinst in leerjaar 1; voor havo (normering Toets 0 en Toets 1: 222 resp. 241; vooruitgang 19)



school:	c	i	n	r	v	w	z
eind leerjaar 1	lager	hoger	hoger	lager	n.s.	hoger	hoger
leerwinst	hoger	hoger	hoger	lager	n.s.	n.s.	n.s.

Hier valt vooral school c in het kwadrant linksboven, op. Bij deze school is het niveau aan het eind van leerjaar 1 lager dan landelijk gemiddeld, maar tegelijk wordt er wel meer leerwinst behaald dan landelijk. Kennelijk zijn hier leerlingen met lage scores aan het begin van leerjaar 1 ingestroomd, die weliswaar relatief veel leerwinst hebben behaald, maar aan het eind van leerjaar 1 nog steeds achterblijven.

Figuur 5.5 Score leesvaardigheid eind leerjaar 1 versus leerwinst in leerjaar 1; voor vwo (normering Toets 0 en Toets 1: 233 resp. 258; vooruitgang 25)



school:	i	r	v	w	z
eind leerjaar 1	lager	lager	n.s.	lager	lager
leerwinst	n.s.	lager	n.s.	lager	lager

Opvallend is dat alle 5 scholen in het kwadrant linksonder liggen: laag niveau en weinig leerwinst. Overigens zijn de verschillen tussen schoolgemiddelde en landelijk cijfers niet steeds significant. Maar voor scholen r, w en z geldt dat zij op beide indicatoren voor schoolopbrengst lager scoren dan landelijk gemiddeld.

Tot slot van deze paragraaf merken we op dat de indeling van scholen in deze vier kwadranten laten zien waar de scholen staan, in vergelijking met landelijke normen voor de gebruikte toetsen. Maar de verschillen, die we tussen de scholen zien, kunnen niet rechtstreeks en eenvoudig worden toegeschreven aan verschillen in kwaliteit van de scholen. Ze kunnen ook te maken hebben met een hoog of laag instroomniveau van de leerlingen en/of met allerlei andere kenmerken van leerlingen en de context van de school, die het leren en vooruitgaan belemmeren of juist bevorderen.

Daarmee komen we bij het begrip toegevoegde waarde: welk deel van de vooruitgang van leerlingen kan beschouwd worden als unieke bijdrage van de school zelf en niet aan de meer of minder gunstige kenmerken van de leerlingen of de context van de school.

6 Toegevoegde waarde

6.1 Inleiding

Bij het bepalen van toegevoegde waarde wordt rekening gehouden met verschillen tussen scholen in hun leerlingenpubliek. Bij sommige leerlingen is het makkelijker om vooruitgang te boeken dan bij andere leerlingen. Bij leerlingen met ‘ongunstige’ achtergrondkenmerken zullen de verwachtingen omtrent de te behalen leerwinst lager zijn dan bij andere leerlingen. En een school met veel leerlingen met ‘ongunstige’ achtergrondkenmerken zal meer moeite moeten doen om de landelijke normen, gebaseerd op ‘doorsnee’ leerlingen, te behalen dan een school met veel leerlingen met juist ‘gunstige’ achtergrondkenmerken.

Voor de bepaling van welke verwachtingen we bij welke leerlingen mogen hebben, zijn geen rechtstreekse landelijke vergelijkingsgegevens voorhanden. Deze moeten gehaald worden uit onderzoek waarin de resultaten van scholen met verschillende leerlingpopulaties onderling worden vergeleken.

In het meeste onderzoek naar de toegevoegde waarde van scholen wordt uitgegaan van een dataverzameling met gestandaardiseerde meetinstrumenten op een representatieve steekproef van scholen. In de analyses worden de opbrengsten van scholen (in termen van bereikt eindniveau of van leerwinst; zonder en met uitgebreide correcties voor verschillen in leerlingpopulatie) onderling vergeleken: zijn er scholen die met hun leerlingen gemiddeld duidelijk betere of juist mindere resultaten boeken dan de andere scholen in de steekproef? De statistische modellen kunnen daarbij eenvoudiger of complexer zijn, maar in de uitkomsten gaat het steeds om een relatieve vergelijking: er zijn scholen die bovengemiddeld scoren en scholen die achterblijven ten opzichte van het gemiddelde van alle scholen in de steekproef (zie ook paragraaf 4.1 over indicatoren op basis van toegevoegde waarde). De representatieve steekproef levert dus de *benchmark*, waartegen de toegevoegde waarde van afzonderlijke scholen kan worden afgezet.

Voorbeelden van dergelijk onderzoek in Nederland zijn verschillende deelstudies in het proefschrift van Anneke Timmermans op data van het VOCL-cohort (Timmermans, 2012), onderzoek naar groeicurven van leerlingen uit het PRIMA-onderzoek (Guldmond & Bosker, 2006; Roeleveld, 2011) en onderzoek op COOL-data naar correctiefactoren in het basisonderwijs (Roeleveld e.a., 2011; Mooij e.a. 2012).

Bij deze pilots rondom leerwinst en toegevoegde in het voortgezet onderwijs is van een dergelijke representatieve steekproef geen sprake. Er doet een beperkt aantal scholen mee, met afdelingen van verschillend niveau (van de basisberoepsgerichte leerweg tot aan gymnasium) en met uiteenlopende, onderling niet vergelijkbare meetinstrumenten.

Deze situatie levert twee problemen op voor analyses rond toegevoegde waarde. Allereerst is er, als gezegd, geen goede landelijke steekproef waarmee vergeleken kan worden. Als een school in de analyses weinig toegevoegde waarde zou laten zien kan dat evengoed komen doordat de school

vergeleken wordt met enkele zeer goede scholen, die toevallig aan de pilot deelnemen, en niet met een representatieve groep scholen.

Maar daarnaast is er, door het gebruik van verschillende meetinstrumenten en scholen met verschillende niveaus van voortgezet onderwijs, ook steeds maar een beperkt aantal scholen beschikbaar voor analyse. Uit de onderzoeksliteratuur blijkt dat de schattingen van toegevoegde waarde een behoorlijk onzekerheidsmarge kennen en dat het vaak alleen mogelijk is om scholen met de allerlaagste en allerhoogste uitkomsten van elkaar te onderscheiden. Het gros van de scholen (soms 60 tot 80 procent) verschilt niet significant van elkaar in toegevoegde waarde (Raudenbush, 2004; zie ook verschillende figuren van schoolresiduals en betrouwbaarheidsintervallen bij Timmermans, 2012).

Bij de kleine aantallen scholen met eenzelfde afdeling en eenzelfde toets in deze pilot verwachten we nauwelijks scholen significant van elkaar te kunnen onderscheiden. Dit betekent dat in het onderzoek een onderlinge vergelijking van afdelingen en scholen maar beperkt zinvol is.

Om deze reden gaan we in dit hoofdstuk daarom meer beschrijvend verder in op de grondgedachte achter het idee van de toegevoegde waarde van scholen: scholen verschillen in hun leerlingenpubliek en bij de beoordeling van hun opbrengsten moet rekening worden gehouden met het feit dat sommige leerlingen makkelijker en andere moeilijker vooruitgang zullen boeken. We doen dat niet door te corrigeren voor verschillen in leerlingpubliek tussen de scholen, maar door meer beschrijvend te laten zien of en in hoeverre de beschikbare leerlingkenmerken samenhangen met zowel het instroomniveau van de leerlingen, als met hun leerwinst. Waar in onderzoek naar toegevoegde waarde een reeks correctiefactoren (dat kunnen zowel leerling- als contextkenmerken zijn) in een complex statistisch model gecombineerd worden, knippen we hier als het ware de vraag naar toegevoegde waarde op in stukjes, per leerlingkenmerk.

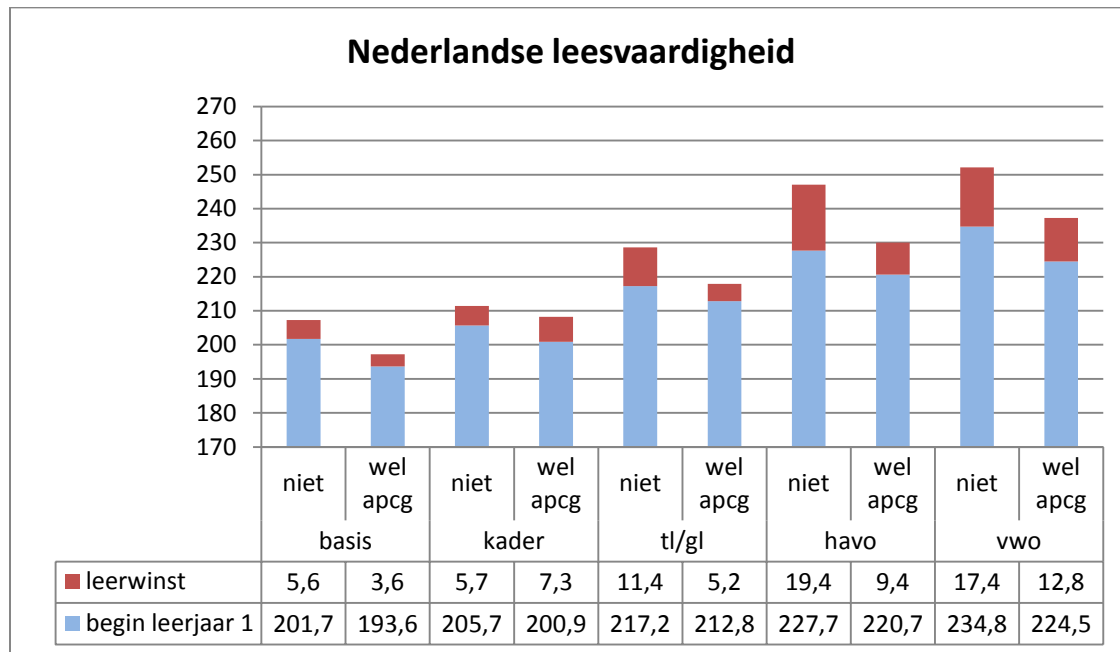
In de volgende paragrafen gaan we in op de leerwinst in het eerste leerjaar op scholen, die de Cito-VVO-toetsen af hebben genomen. We kiezen voor die toetsen, omdat daar de meeste gegevens over beschikbaar zijn. We laten steeds de beginscores zien van leerlingen in de verschillende onderwijstypen en de leerwinst die zij in dat eerste leerjaar boeken⁸. We gebruiken zoveel mogelijk gegevens over dat eerste leerjaar: niet alleen van schooljaar 2012/2013, maar ook van eerdere schooljaren.

⁸ Dergelijke analyses kunnen natuurlijk ook uitgevoerd worden voor de leerjaren twee en drie.

6.2 Beginniveau en leerwinst naar afkomst uit APCG-gebied

In het voortgezet onderwijs krijgen scholen, in het kader van het Leerplusarrangement, extra middelen als er veel leerlingen afkomstig zijn uit een armoede probleem cumulatie gebied (APCG). De gedachte daarachter is dat er bij die leerlingen een grote kans is dat er sprake is van onderwijsachterstanden en een weinig stimulerend thuismilieu (Claassen, Driesen, Aarntzen & Mulder, 2005). In de navolgende figuren laten we zien of leerlingen uit een APCG-gebied met lagere toetscores in de school instromen en of hun leerwinsten anders zijn dan bij de overige leerlingen. We onderzoeken de vier scores uit de Cito-toetsen T0 en T1.

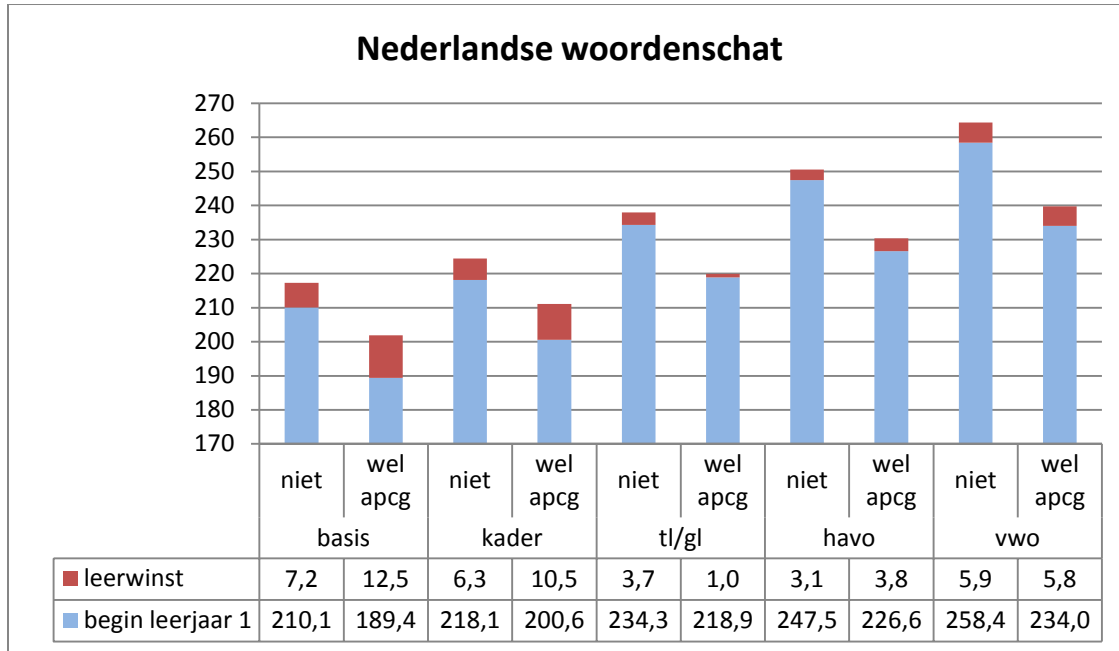
Figuur 6.1 Instroomniveau en leerwinst eerste leerjaar bij Nederlandse leesvaardigheid; naar afkomst uit APCG-gebied



Bij alle vijf schooltypen hebben leerlingen uit APCG-gebieden gemiddeld lagere scores bij binnenkomst dan de overige leerlingen. De verschillen zijn behoorlijk: effectgroottes⁹ van 0.4 tot 0.7. Verder blijkt dat ook de leerwinst doorgaans lager uitvalt. Alleen bij vmbo-kader behalen APCG-leerlingen wat meer leerwinst.

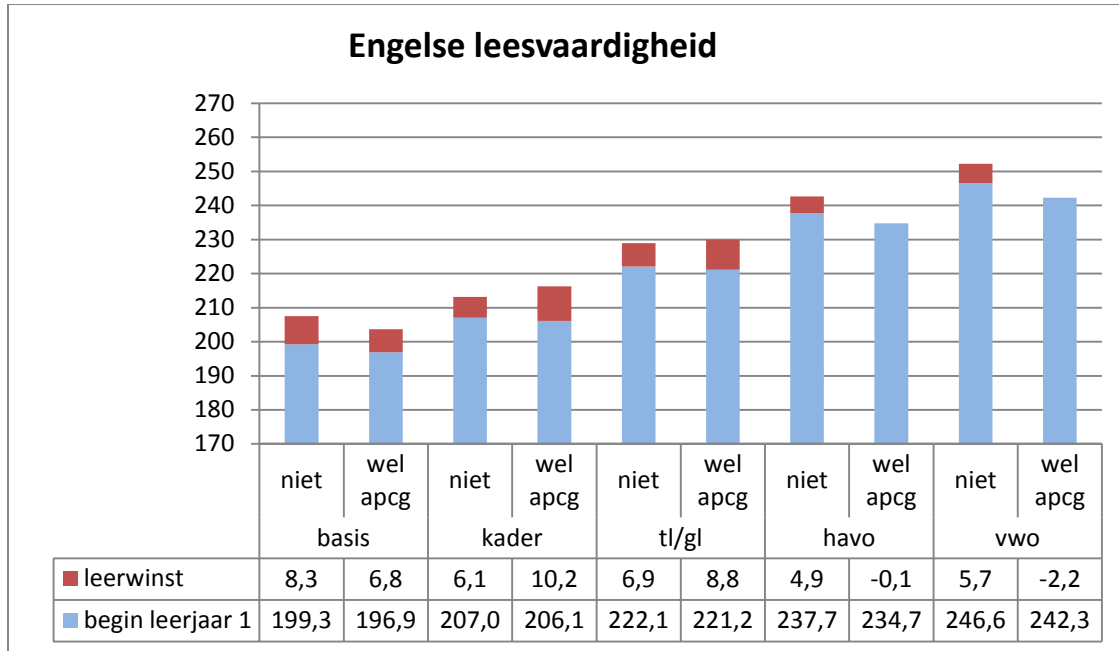
⁹ De effectgrootte geeft het verschil in gemiddelde gedeeld door de standaarddeviatie. Een verschil van 0.2 wordt in de literatuur beschouwd als klein; 0.5 als middelgroot en 0.8 en meer als groot (Cohen, 1988).

Figuur 6.2 Instroomniveau en leerwinst eerste leerjaar bij Nederlandse woordenschat; naar afkomst uit APCG-gebied



Ook hier scoren APCG-leerlingen fors lager bij instroom dan de overige leerlingen (effectgrootte van 0.7 tot 1.0). De verschillen bij leerwinst zijn veel kleiner en zowel bij basis als bij kader boeken de APCG-leerlingen wat meer leerwinst in het eerste leerjaar.

Figuur 6.3 Instroomniveau en leerwinst eerste leerjaar bij Engelse leesvaardigheid; naar afkomst uit APCG-gebied

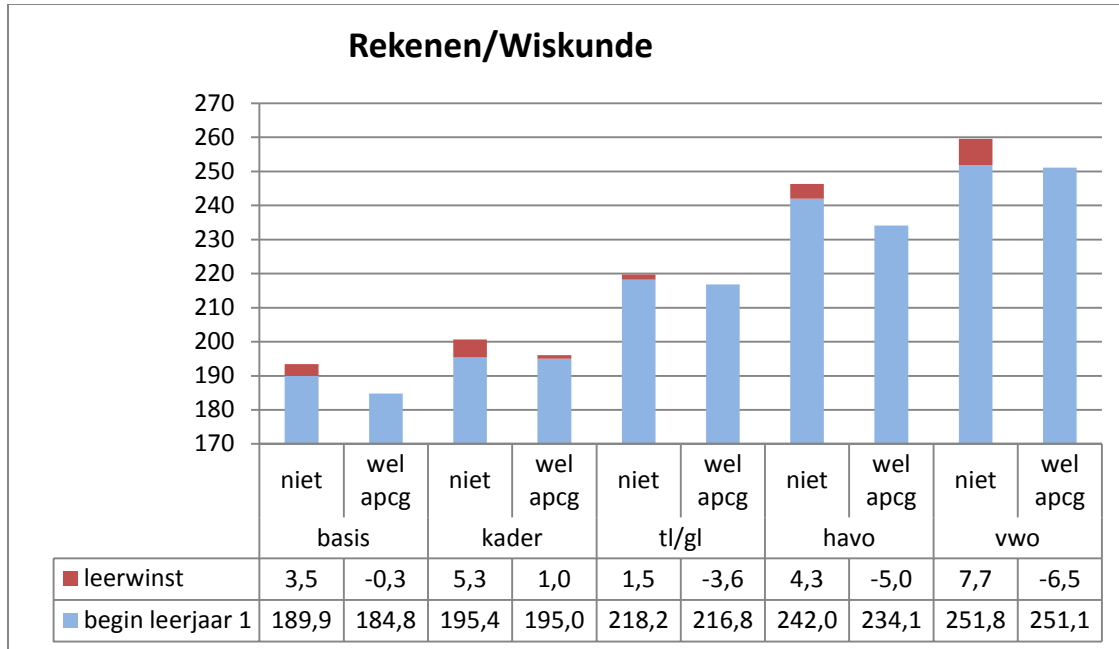


Bij Engelse leesvaardigheid scoren APCG-leerlingen wat lager, maar de verschillen zijn hier beperkt (het grootste verschil, bij vwo, heeft een effectgrootte van 0.16).

Bij havo en vwo is er bij de APCG-leerlingen geen leerwinst, maar juist een lichte daling. De niet APCG-ers gaan juist wel vooruit: de verschillen in leerwinst met APCG-ers hebben een effectgrootte van 0.24 tot 0.33.

Bij kader en tl/gl gaan de APCG-ers iets meer vooruit dan de overige leerlingen (effectgrootte rond de 0.15).

Figuur 6.4 Instroomniveau en leerwinst eerste leerjaar bij Rekenen/wiskunde; naar afkomst uit APCG-gebied



Bij rekenen/wiskunde beginnen leerlingen uit APCG-gebieden doorgaans met een (lichte) achterstand. Alleen bij havo en basis zijn de verschillen vrij groot, met effectgroottes van resp. 0.5 en 0.3.

Opvallende genoeg gaan de leerlingen uit APCG-gebieden (behalve bij kader) gedurende het eerste leerjaar achteruit in vaardigheid, terwijl de overige leerlingen wel leerwinst boeken. Bij de drie vmbo-niveaus ligt de effectgrootte van het verschil in leerwinst rond de 0.3, maar bij havo en vwo duidelijk hoger: 0.7 tot 0.8.

Samenvattend:

Leerlingen uit een APCG-gebied behalen bij instroom lagere scores; de verschillen met overige leerlingen zijn het kleinst bij Engelse leesvaardigheid.

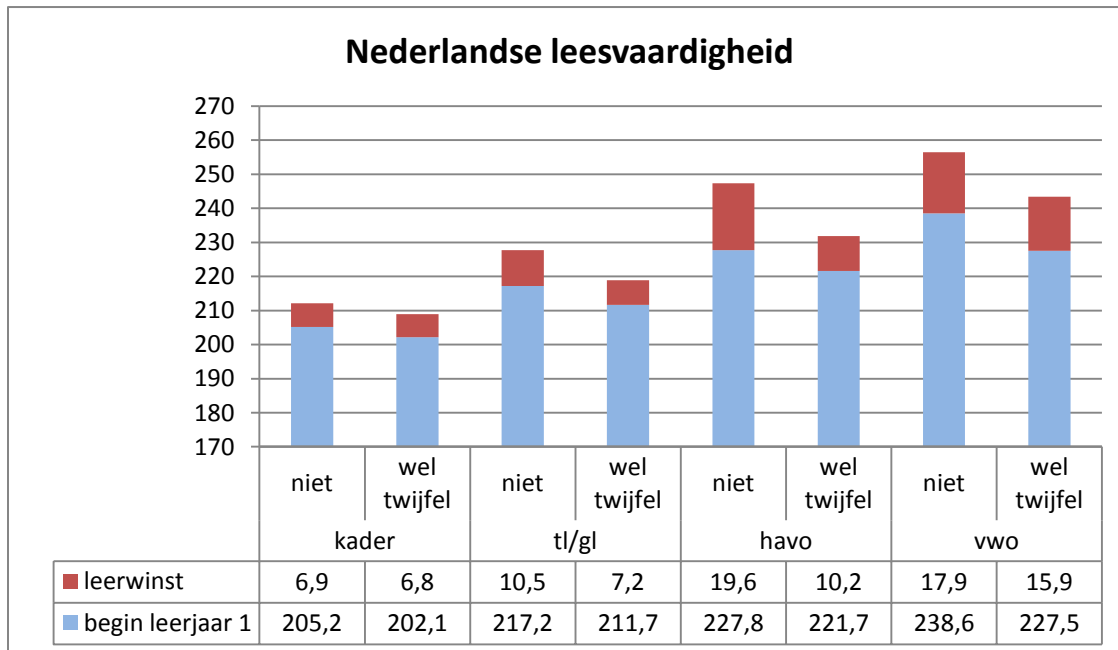
Bij leerwinst is het beeld meer divers. Bij sommige schooltypen en bij sommige vaardigheden halen APCG-leerlingen wat meer leerwinst; bij andere juist minder. Bij Rekenen/wiskunde is de leerwinst van APCG-leerlingen steeds lager: behalve bij vmbo-kader scoren ze aan het eind van leerjaar 1 lager dan aan het begin.

6.3 Beginniveau en leerwinst naar voordeel van de twijfel

Sommige scholen geven in hun toelatingsbeleid meer leerlingen het voordeel van de twijfel dan andere scholen. De vraag is hoe beginniveau en leerwinst van zulke leerlingen zich verhouden tot dat van de overige leerlingen.

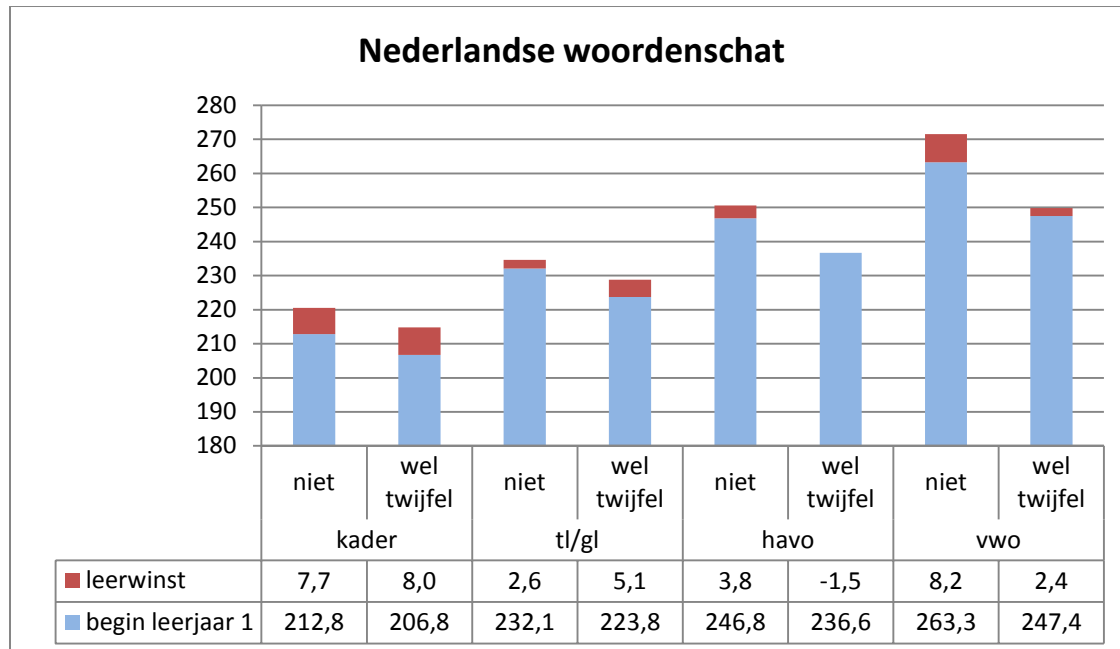
We zeggen dat een leerling het voordeel van de twijfel heeft gehad als het advies lager was dan het schooltype van instroom en/of de score op de Eindtoets basisonderwijs van Cito en/of de score op de NIO lager was dan gebruikelijk bij dit schooltype. (Omdat vmbo basis als laagste niveau wordt beschouwd zijn daar geen leerlingen met twijfel.)

Figuur 6.5 *Instroomniveau en leerwinst eerste leerjaar bij Nederlandse leesvaardigheid; naar voordeel van de twijfel*



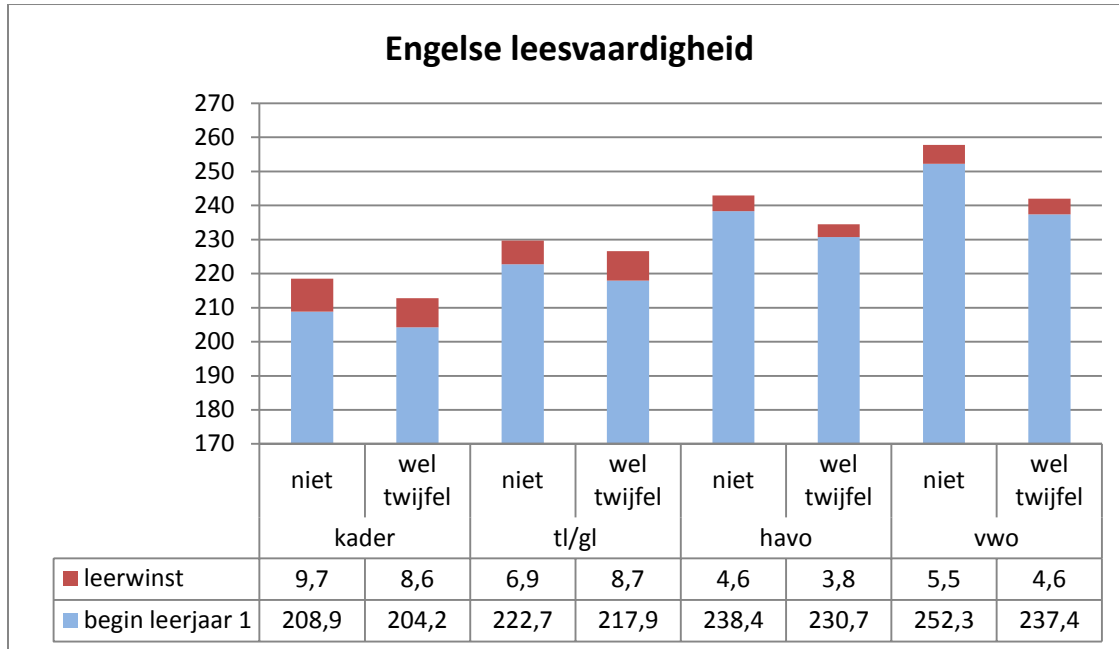
Bij Nederlandse leesvaardigheid zijn de beginscores van leerlingen met twijfel duidelijk lager dan die van de overige leerlingen (effectgroottes van 0.3 bij kader tot 0.8 bij vwo). Verder boeken ze, behalve bij kader, ook minder leerwinst.

Figuur 6.6 *Instroomniveau en leerwinst eerste leerjaar bij Nederlandse woordenschat; naar voordeel van de twijfel*



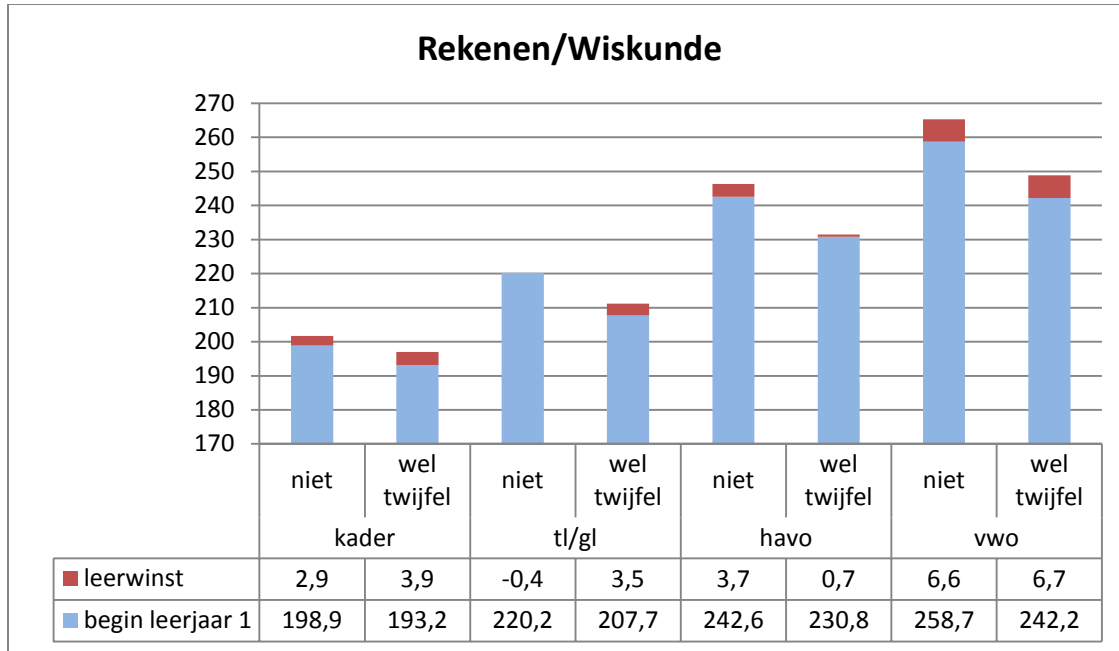
Ook bij Nederlandse woordenschat zijn de beginscores van leerlingen die het voordeel van de twijfel hebben gekregen lager (effectgroottes van 0.3 tot 0.6). Bij beide vmbo-typen behalen ze vervolgens wel wat meer leerwinst, maar bij havo en vwo weer minder dan de overige leerlingen.

Figuur 6.7 Instroomniveau en leerwinst eerste leerjaar bij Engelse leesvaardigheid; naar voordeel van de twijfel



Bij Engelse leesvaardigheid beginnen de leerlingen met twijfel lager. Dat geldt in het bijzonder voor het vwo (effectgrootte 0.6). De daarna behaalde leerwinsten verschillen niet sterk tussen beide groepen leerlingen (alle effectgroottes kleiner dan 0.1).

Figuur 6.8 Instroomniveau en leerwinst eerste leerjaar bij Rekenen/wiskunde; naar voordeel van de twijfel



Ook bij Rekenen/wiskunde is de start van leerlingen met twijfel lager bij alle onderwijstypen. De verschillen zijn het kleinst bij kader (effectgrootte 0.3) maar bij de andere onderwijstype fors (effectgroottes van 0.7 tot 0.9).

Het beeld bij de leerwinsten wisselt nogal: bij kader en vwo maar kleine verschillen, bij havo blijft de leerwinst van twijfel-leerlingen achter, maar bij tl/gl boeken ze juist meer leerwinst (beide verschillen hebben een effectgrootte van rond de 0.25).

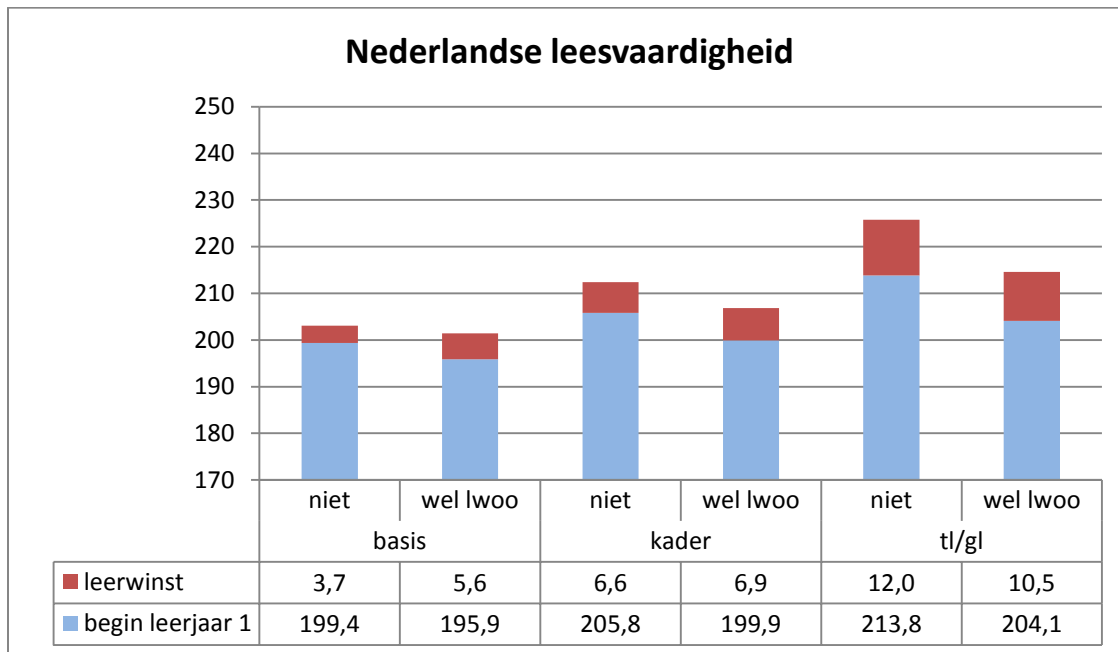
Samenvattend

Leerlingen die het voordeel van de twijfel hebben gekregen stromen in met, zoals verwacht kon worden, lagere beginscores. De leerwinsten laten echter een wisselend beeld zien, per vaardigheid en per onderwijstype: soms blijven de twijfel-leerlingen verder achter, soms halen ze juist meer leerwinst en soms is er weinig verschil.

6.4 Beginniveau en leerwinst naar leerwegondersteuning

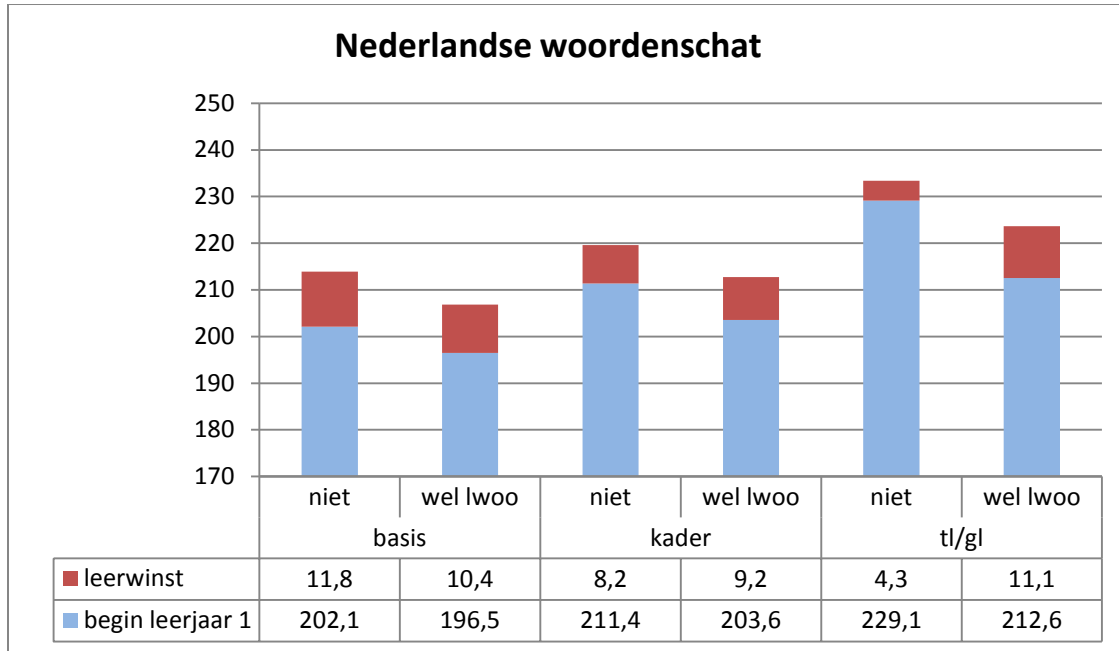
Binnen het vmbo zijn er leerlingen die in aanmerking komen voor leerwegondersteuning (lwoo). Ook hier gaan we weer na in hoeverre het instroomniveau en de leerwinst in het eerste leerjaar bij deze leerling verschilt met die van de overige leerlingen zonder lwoo.

Figuur 6.9 *Instroomniveau en leerwinst eerste leerjaar bij Nederlandse leesvaardigheid; naar leerwegondersteuning*



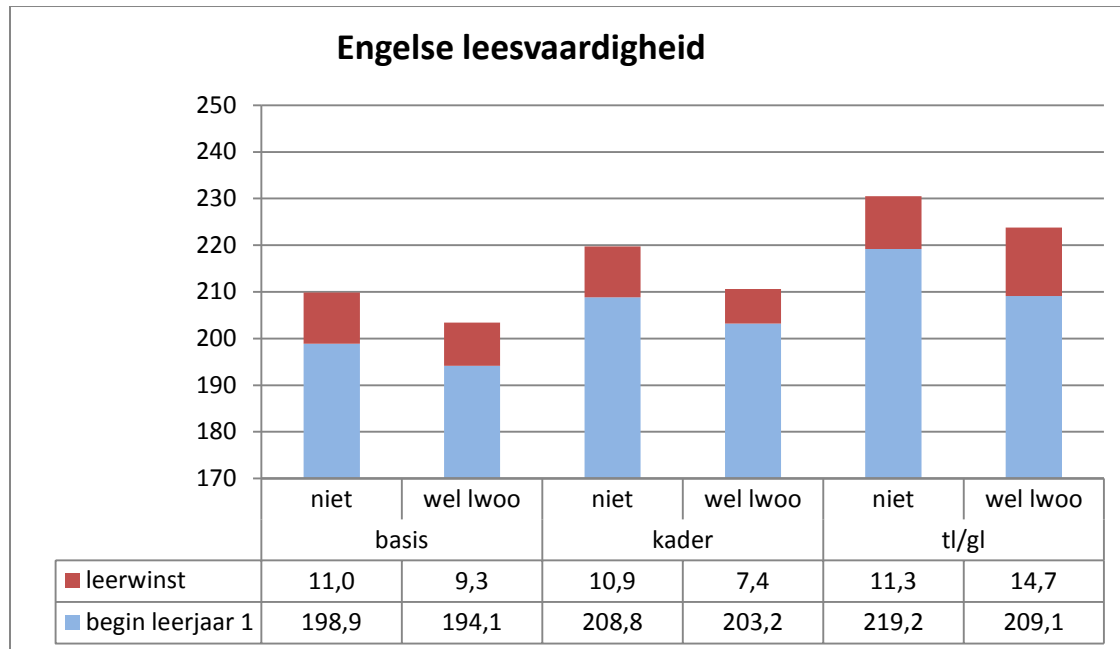
Bij alle drie typen vmbo komen de leerlingen met leerwegondersteuning binnen met lagere scores dan de overige leerlingen. De effectgroottes van het verschil lopen op van 0.3 bij basis, via 0.5 bij kader tot 0.8 bij tl/gl. Bij de leerwinsten is het beeld gunstiger: bij basis en kader lopen de lwoo-ers iets in, maar bij tl/gl is ook de leerwinst van de lwoo-ers wat lager.

Figuur 6.10 Instroomniveau en leerwinst eerste leerjaar bij Nederlandse woordenschat; naar leerwegondersteuning



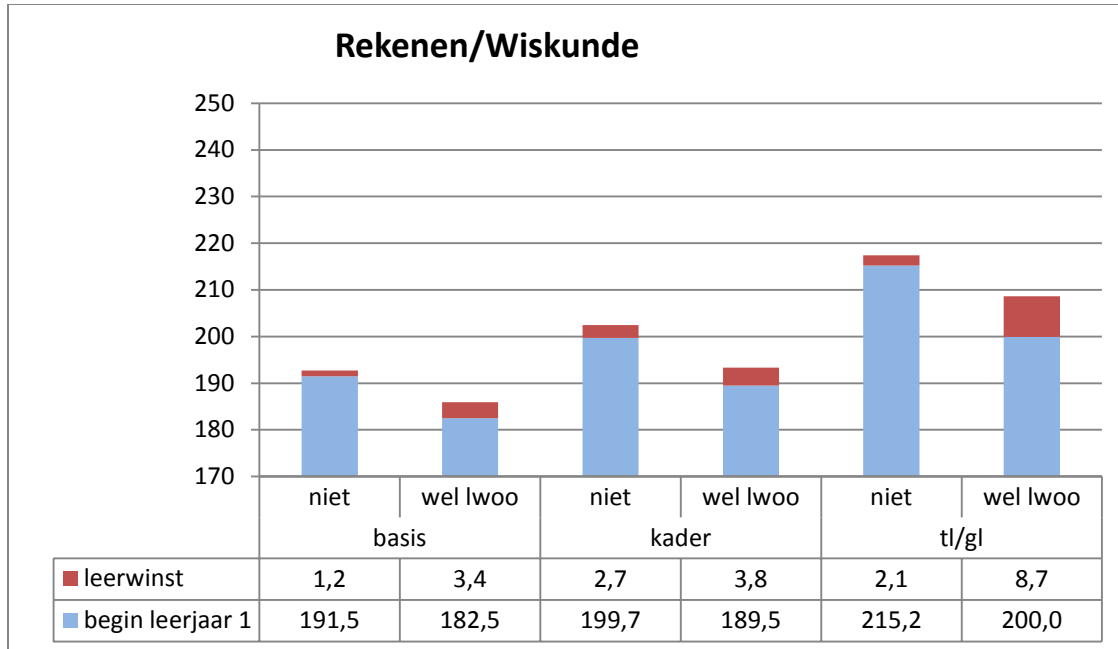
Ook bij Nederlandse woordenschat stromen lwoo-leerlingen in met lagere scores. De verschillen zijn opnieuw het grootst bij tl/gl (effectgrootte 0.7). De leerwinsten van lwoo-ers bij kader en basis verschillen niet veel van die van de overige leerlingen. Bij tl/gl lopen leerlingen met lwoo een flink stuk in op de overige leerlingen: ze behalen duidelijk meer leerwinst (effectgrootte 0.3).

Figuur 6.11 Instroomniveau en leerwinst eerste leerjaar bij Engelse leesvaardigheid; naar voordeel van de twijfel



Opnieuw beginnen bij Engels de lwoo-leerlingen lager. En bij basis en kader zijn ook hun leerwinsten wat lager; bij tl/gl behalen ze juist wat meer leerwinst. De verschillen in leerwinst zijn echter steeds bescheiden (effectgrootte maximaal 0.15).

Figuur 6.12 Instroomniveau en leerwinst eerste leerjaar bij Rekenen/wiskunde; naar voordeel van de twijfel



Bij rekenen/wiskunde tenslotte starten de lwoo-ers ook weer op een flink lager niveau (effectgroottes van 0.5 bij basis tot 0.8 bij tl/gl). Daarna behalen ze overal wat meer leerwinst; vooral bij tl/gl is het verschil met de overige leerlingen aanzienlijk (effectgrootte 0.4).

Samenvattend

Leerlingen met een beschikking voor leerwegondersteuning starten, niet onverwacht, bij alle drie vmbo-typen en bij alle onderzochte domeinen, met een achterstand ten opzichte van de overige leerlingen.

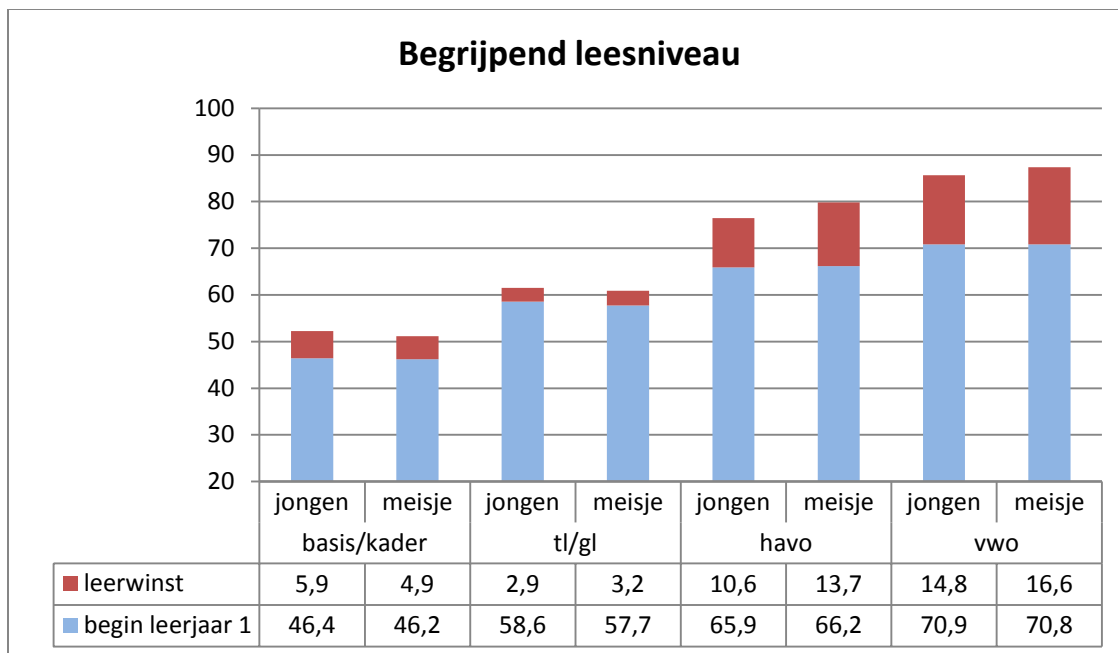
Voorals bij vmbo-tl/gl is het verschil vaak fors. Tegelijk behalen deze leerlingen juist bij tl/gl vaak meer leerwinst in het eerste leerjaar (uitzondering is Nederlandse leesvaardigheid). Het lijkt er daarmee op dat de tl/gl-afdelingen in het onderzoek er in slagen de leerlingen een deel van hun achterstand te laten inlopen.

Bij basis en kader zijn de verschillen in leerwinst doorgaans kleiner en soms ook wisselend.

6.5 Beginniveau en leerwinst naar geslacht

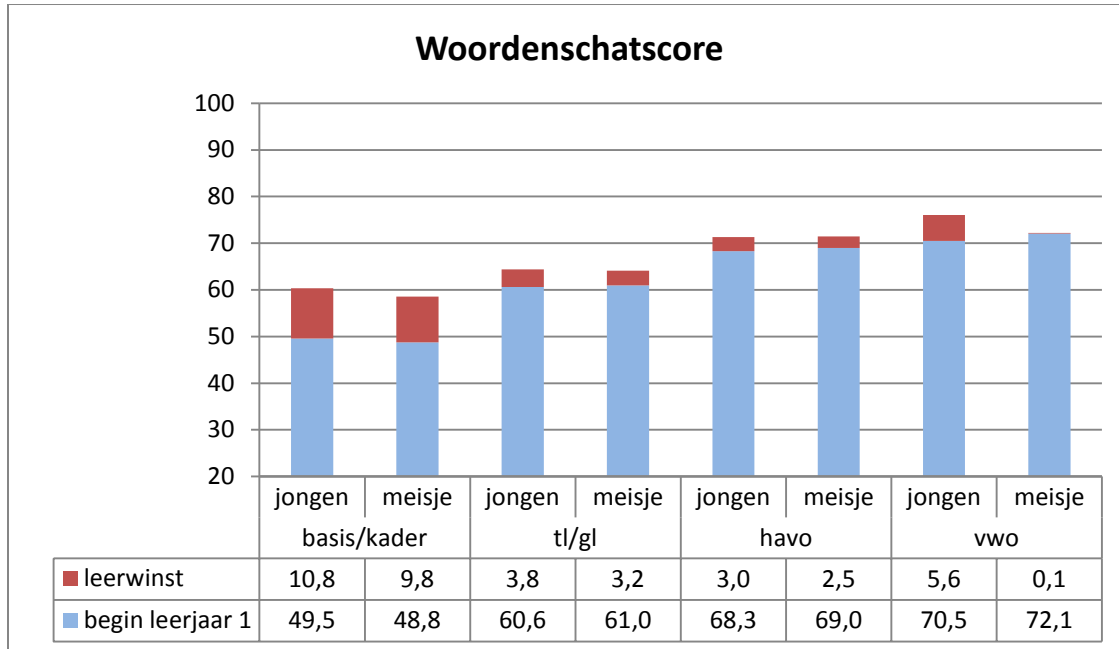
Als laatste laten we in deze paragraaf nog verschillen zien naar het geslacht van de leerling. We gebruiken dit keer de scores op andere toetsen, die geschikt zijn om leerwinst mee te bepalen, namelijk de Diataal-toetsen. Het instroomniveau betreft hier de score op Toets 1A, de leerwinst is de vooruitgang die bij Toets 1B aan het eind van het schooljaar is geboekt. Omdat we maar over weinig scores van leerlingen uit vmbo-basis beschikken zijn deze hier samengevoegd met vmbo-basis.

Figuur 6.13 *Instroomniveau en leerwinst eerste leerjaar bij Begrijpend lezen; naar geslacht*



De instroomverschillen tussen jongens en meisjes zijn maar klein (effectgrootte maximaal 0.1). Bij de twee vmbo-groepen verschilt ook de leerwinst niet erg veel. Bij havo en vwo behalen meisjes meer leerwinst (effectgrootte respectievelijk 0.3 en 0.2).

Figuur 6.14 Instroomniveau en leerwinst eerste leerjaar bij Woordenschat; naar geslacht



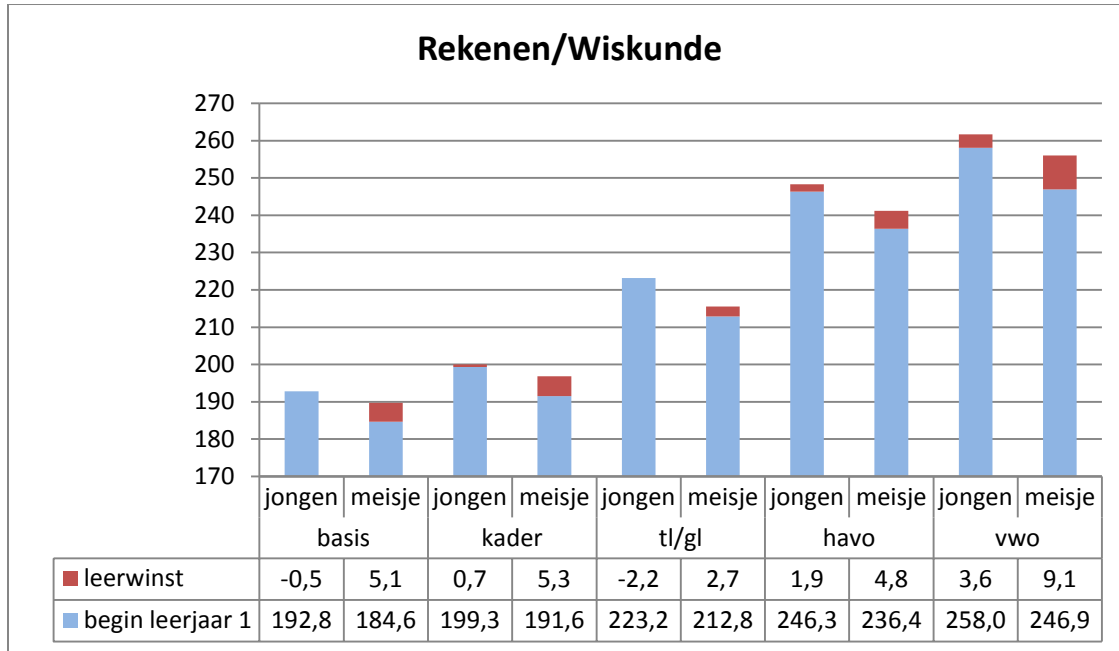
Ook bij woordenschat zijn de instroomverschillen beperkt. Het grootste verschil vinden we nog bij vwo, waar de meisje met een hogere score starten (effectgrootte 0.14). Ook de verschillen in leerwinst zijn beperkt. Uitzondering is het vwo, waar jongens beduidend meer leerwinst behalen dan meisjes (effectgrootte 0.6).

Samenvattend

Bij deze twee schaalscores uit de toets DiaTaal zien we maar beperkte verschillen tussen jongens en meisjes. Het grootste verschil vinden we bij Woordenschat op het vwo: hier komen de jongens lager binnen, maar ze halen de meisjes in het eerste leerjaar duidelijk in en eindigen hoger.

Omdat verschillen tussen jongens en meisjes op het gebied van taal mogelijk beperkt zijn kijken we tot slot ook nog naar de scores voor Rekenen/wiskunde op de Cito-toetsen.

Figuur 6.15 Instroomniveau en leerwinst eerste leerjaar bij Rekenen/wiskunde; naar geslacht



Bij alle onderwijstypen scoren meisjes bij instroom duidelijk lager dan jongens; effectgroottes lopen van 0.4 tot 0.6. Tegelijk blijkt dat de meisjes in het eerste leerjaar een deel van die achterstand inlopen: hun leerwinsten zijn groter (effectgrootte rond de 0.3). Maar aan het eind van het eerste leerjaar blijven ze nog steeds achter bij de jongens.

6.6 Verschillen tussen scholen en cohorten

In de vorige paragrafen is geïllustreerd welke verschillen er tussen leerlingen met diverse (achtergrond)kenmerken bestaan in instroomscores én in leerwinst in het eerste leerjaar. Bij het bepalen van de toegevoegde waarde van scholen wordt geprobeerd te corrigeren voor verschillen tussen scholen in zulke kenmerken van hun leerlingen. Bij sommige leerlingen zal het immers makkelijker zijn om vooruitgang te boeken dan bij andere leerlingen. Bij leerlingen met ‘ongunstige’ achtergrondkenmerken zullen de verwachtingen omtrent de te behalen leerwinst lager zijn dan bij andere leerlingen. En een school met veel leerlingen met ‘ongunstige’ achtergrondkenmerken zal meer moeite moeten doen om de landelijke normen, gebaseerd op ‘doorsnee’ leerlingen, te behalen dan een school met veel leerlingen met juist ‘gunstige’ achtergrondkenmerken.

Zoals eerder opgemerkt (paragraaf 6.1) lijkt het weinig zinvol, gezien het beperkte aantal scholen dat aan de pilots deelneemt en gezien het feit dat zij samen geen representatieve steekproef vormen, om deze scholen onderling te gaan vergelijken op hun via statistische modellen geschatte toegevoegde waarde. Wel hebben we geprobeerd een schatting te krijgen van de omvang van de verschillen in onderwijsresultaten tussen de beperkte groep pilotscholen, voor en na statistische correctie voor enkele relevante leerlingkenmerken.

Doel van deze analyses is allereerst om een beeld te krijgen van de omvang van de systematische verschillen tussen deze scholen. Pas wanneer dergelijke verschillen ook substantieel zijn heeft het zin om na te gaan of de kwaliteit van (het onderwijs op) de scholen een verklaring kan bieden voor deze verschillen. Wanneer de onderwijsopbrengsten vooral individueel bepaald zijn (binnen iedere school zijn er ongeveer evenveel leerlingen die het beter en die het slechter doen) presteren alle scholen ongeveer even goed.

Daarnaast leveren de analyses ook informatie op over de stabiliteit van de gevonden schoolverschillen over de jaren. Doordat we over gegevens van meerdere instroomcohorten beschikken kan zowel worden nagegaan in hoeverre er systematische verschillen tussen scholen bestaan en in hoeverre er sprake is van fluctuaties van cohort tot cohort binnen dezelfde school.

Voor de schattingen van zulke verschillen tussen scholen en cohorten hebben we analyses uitgevoerd met een *multilevel random effect* model. Met zo’n model wordt onderzocht in hoeverre de verschillen tussen leerlingen in toetsscores samenhangen met hun eerdere prestaties en met hun achtergrondkenmerken en kan vervolgens worden nagegaan of de dan resterende (‘onverklaarde’) verschillen in toetsscores systematisch samenhangen met de school die de leerlingen bezoeken. In het model wordt er expliciet rekening mee gehouden dat leerlingen ‘genest’ zijn binnen een bepaalde school.¹⁰ Klassieke literatuur over deze modellen zijn de publicaties van Aitkin en Longford (1986), Raudenbush en Bryk (1986) en Goldstein (1987). Een aantal recente Nederlandse toepassingen zijn te vinden in Timmermans (2012) en Roeleveld, Mooij, Fettelaar en Ledoux (2011).

¹⁰ Zie ook het working paper *Indicatoren voor schoolkwaliteit op basis van onderwijsopbrengsten*.

In het model zijn de leerlingen genest binnen scholen. Maar omdat we ook de verschillen tussen de opeenvolgende instroomcohorten wilden verdisconteren hebben we gebruik gemaakt van een drieniveau model: leerlingen, genest in cohorten, genest in scholen. Voor zulke analyses kwamen de scholen met Cito- en Diataal-toetsen in aanmerking. We hebben gekozen voor de toetsen, waarbij we over de meeste gegevens beschikken, namelijk de toetsen van Cito.

Voor elk van de vier vaardigheidsscores, die we binnen de pilot gebruiken, hebben we analyses uitgevoerd op de scores van leerlingen bij Toets 0 en bij Toets 1. De analyses van Toets 0 geven informatie over het instroomniveau van de leerlingen in de school. Bij de analyses van Toets 1, waarbij rekening wordt gehouden met de eerdere score op Toets 0, krijgen we informatie over de vooruitgang van de leerlingen in het eerste schooljaar. Bij beide typen analyses hebben we, als gezegd, onderzocht welke systematische verschillen er, na controle voor het schooltype en enkele achtergrondkenmerken van de leerlingen, overblijven tussen cohorten binnen eenzelfde school en tussen de pilotscholen onderling.

Idealiter zouden we in de analyses ook willen controleren voor verschillen tussen leerlingen in het opleidingsniveau van hun ouders. Uit tal van onderzoek blijkt immers dat dat een belangrijke voorspeller is van succes in het onderwijs. Maar scholen voor voortgezet onderwijs beschikken niet over gegevens met betrekking tot het ouderlijke onderwijsniveau en we kunnen slechts gebruik maken van een beperkt aantal andere achtergrondkenmerken, namelijk¹¹:

- geslacht
- woonachtig in een APCG-gebied
- advies (geschaald van 1 = praktijkonderwijs, 2 = vmbo-basis tot 6 = vwo)
- indicatie voor ‘voordeel van de twijfel’ (zie paragraaf 6.3)

Zoals gezegd hebben we ook verschillen tussen instroomcohorten op eenzelfde school geanalyseerd. Niet alle scholen hebben de toetsen de afgelopen vier jaar ook gebruikt, zodat het aantal cohorten wisselt per school. In totaal beschikten we voor de analyses over de gegevens van 3770 leerlingen in 38 instroomcohorten op 13 scholen. De analyses zijn uitgevoerd met MIWin (Rasbash e.a., 2005).

Uitkomsten

De pilotscholen bieden onderwijs aan van vmbo-basis tot aan vwo. En uiteraard bestaan er dan grote verschillen in de toetscores die de leerlingen op die scholen behalen. De analyses van de behaalde scores op de verschillende onderdelen van de Cito-toetsen laten zien dat het onderwijstype en de achtergronden van de leerlingen een groot deel van deze verschillen tussen leerlingen verklaren. Van de resterende onverklaarde verschillen in toetsscores blijkt maar een klein deel systematisch samen te hangen met de school die de leerlingen bezoeken: in de orde van enkele procenten variantie. Verschillen

¹¹ We gebruiken hier geen leerwegondersteuning, omdat dat alleen bij vmbo-scholen relevant is.

tussen deze scholen in toegevoegde waarde zullen alleen betrekking hebben op deze beperkte hoeveelheid onverklaarde variantie.

Verder blijkt dat er ook verschillen bestaan tussen de opeenvolgende cohorten binnen de scholen; deze verschillen zijn van eenzelfde orde als die tussen de scholen. Dat betekent dat op dezelfde school de leerlingen het in het ene jaar wat beter doen en in het andere jaar wat slechter. En dat, voor zover een school in een bepaald jaar meer toegevoegde waarde zou laten zien dan de andere scholen, dat bij een volgend jaar weer niet zo hoeft te zijn.

Zoals gezegd hebben deze analyses de nodige beperkingen. Het gaat om een klein aantal scholen, die ook niet representatief zijn voor alle scholen voor voortgezet onderwijs in Nederland. Door het kleine aantal zullen de onderlinge verschillen ook niet snel statistisch significant zijn. Een verdere beperking is dat op lang niet alle onderzochte scholen ook steeds gegevens van vier opeenvolgende cohorten leerlingen beschikbaar zijn. Ook is de statistische controle voor verschillen in leerlingpubliek beperkt. Als gezegd zou vooral het opleidingsniveau van de ouders een belangrijke verbetering kunnen opleveren. Het is goed mogelijk dat bij een betere controle voor verschillen in leerlingpubliek er grotere onderlinge verschillen tussen de scholen worden gevonden.

Het is daarom van belang dat voor een meer gefundeerde analyse van de toegevoegde waarde van scholen voor voortgezet onderwijs gegevens beschikbaar komen van een groter aantal scholen, representatief voor het hele voortgezet onderwijs. Daarbij zouden ook betere gegevens over de sociale en culturele achtergrond van de leerlingen verzameld moeten worden. Onze uitkomsten laten verder zien dat zulke analyses in elk geval meer dan één leerlingcohort moeten omvatten.

7 Vooruitblik

Inmiddels is het pilotproject het tweede jaar ingegaan. In dat jaar zal vanuit Oberon verdere 'ondersteuning op maat' geboden worden aan de scholen. Op de tweede landelijke scholenbijeenkomst van 1 november 2013 werden daarbij genoemd:

- workshops opbrengstgericht werken voor teams
- workshop opbrengstgericht werken management
- bijeenkomst afstroom
- bijeenkomst 'succes' in vervolgonderwijs
- schoolspecifieke vragen
- website www.pilotsltwvo.nl
- ondersteuning dataverzameling
- bespreken tweede schoolrapportage

Voor wat betreft het onderzoek zal er in het tweede jaar naar gestreefd worden om de dataverzameling op de scholen (zowel van gegevens uit de leerlingenadministraties als de scores op de toetsen en andere instrumenten) te versnellen. Daardoor zouden de schoolrapporten korter na de zomervakantie van 2014 gereed moeten kunnen zijn.

In deze nieuwe schoolrapporten zal zoveel als mogelijk blijkt worden tegemoet gekomen aan de opmerkingen en wensen, die de scholen hebben geuit bij de bespreking van de eerdere rapporten (zie paragraaf 4.4). Vanuit de verdere contacten van Oberon met de scholen kunnen daar mogelijk nog aanvullingen op komen.

Een belangrijk nieuw element is in ieder geval dat er vervolgmetingen beschikbaar zullen komen met de instrumenten voor motivatie en burgerschap, waarmee ook de ontwikkeling van leerlingen op niet-cognitief gebied verder in beeld kan komen.

Referenties

- Aitkin, M. & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society*, 149, 1-43
- Ali, J., & McInerney, D.M. (2004). *Multidimensional assessment of school motivation*. Paper presented at the 3th SELF Research Conference, Berlin.
- Blok, H., Otter, M.E. & Roeleveld, J. (2002). Coping with conflicting demands: student assessment in Dutch primary schools. *Studies in Educational Evaluation*, 28, pp. 177-188
- Bosker, R. (2012). De toegevoegde waarde van een school: begripsbepaling, meting en causale attributie. In: Dijkstra, A.B. & Janssens, F.J.G. (red) (2012). *Om de kwaliteit van het onderwijs. Kwaliteitsbepaling en kwaliteitsbevordering*. Den Haag: Boom
- Claassen, A., Driesen, G., Aarntzen, D. & Mulder, L. (2005). Voorstel voor een indicator voor de toedeling van middelen voor het Leerplusarrangement in het voortgezet onderwijs. Nijmegen: ITS
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates
- Dam, G. ten, Geijsel, F., Reumerman, R., & Ledoux, G. (2010). Burgerschapscompetenties: de ontwikkeling van een meetinstrument. *Pedagogische Studiën*, 87(5) 313-333.
- Driessen, G., Mulder, L., & Roeleveld, J. (2012). *Cohortonderzoek COOL⁵⁻¹⁸. Technisch rapport basisonderwijs, tweede meting 2010/11*. Nijmegen: ITS / Amsterdam: Kohnstamm Instituut
- Inspectie van het Onderwijs (2012). *Projectplan pilot leerwinst/toegevoegde waarde in het voortgezet onderwijs*. Utrecht: IvHO
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Griffin
- Guldemond, H. & Bosker, R. (2006). *Onderwijsachterstanden en ontwikkelingen in leerprestaties*. Groningen: GION
- Hacquebord, H., Stellingwerf, B, Linthorst, R. & Andringa, S. (2005) *Diataal. Verantwoording en normering*. Groningen: Rijksuniversiteit Groningen
- Inspectie van het onderwijs (2010). *Opbrengstgericht werken in het basisonderwijs*. Utrecht: Inspectie van het Onderwijs.
- Ledoux, G., Blok, H., & Boogaard, M.; m.m.v. Krüger, M. (2009). *Opbrengstgericht werken. Over de waarde van meetgestuurd onderwijs*. Amsterdam: SCO-Kohnstamm Instituut.
- McInerney, D.M. & Ali, J. (2006). Multidimensional and hierarchical assessment of school motivation: cross-cultural validation. *Educational Psychology*, 26 (6), 717-734.

- Meijer, J., Ledoux, G., & Elshof, D. (2011). *Gebruikersvriendelijke leerlingvolgsystemen in het primair onderwijs*. Amsterdam: Kohnstamm Instituut.
- Midgley, C., Maehr, M.L., Huda, L.Z., Anderman, E., Anderman, L., Freeman, K.E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M.J., Nelson, J., Roeser, R., & Urda, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS)*. Ann Arbor, MI: University of Michigan.
- Ministerie van OCW (2011). *Actieplan Basis voor Presteren. Naar een ambitieuze leercultuur voor alle leerlingen*. Den Haag: OCW.
- Ministerie van OCW (2011). *Actieplan Beter Presteren: opbrengstgericht en ambitieus*. Den Haag: OCW.
- Ministerie van OCW (2011). *Actieplan Leraar 2020. Een krachtig beroep*. Den Haag: OCW.
- Mooij, T., Roeleveld, J., Fettelaar, D. & Ledoux, G. (2012). Kwaliteitsbeoordeling van scholen primair onderwijs: Het correctiemodel van de inspectie vergeleken met alternatieve modellen. *Pedagogische Studiën*, 89, pp. 272-287
- OECD (2008). *Measuring Improvements in Learning Outcomes. Best Practices to Assess the Value-Added of Schools*. Parijs: OECD
- Oomens, M., Aarsen van, E. Hulsen, M. (2012). *Gebruik van en ervaringen met leerlingvolgsystemen in het voortgezet onderwijs*. Utrecht: Oberon
- Peschar, J.L. (2007). *Over leerwinst als stelselindicator*. Haren
- Rasbash, J., Charlton, C., Browne, W.J., Healy, M. & Cameron, B. (2005). *MLwiN Version 2.02*. Centre for Multilevel Modelling, University of Bristol.
- Raudenbusch, S. W. (2004). *Schooling, statistics, and poverty: Can we measure school improvement?* Princeton, NJ: Educational Testing Service. Opgehaald op 13 maart 2012, van www.ets.org/Media/Education_Topics/pdf/angoff9.pdf
- Raudenbush, S. & Bryk, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17
- Roeleveld, J. (2003). *Herkomstkenmerken en begintoets. Secundaire analyse op het PRIMA-cohortonderzoek*. Studie voor de Onderwijsraad.
- Roeleveld, J. (2011). De cognitieve ontwikkeling van doelgroep leerlingen gedurende het basisonderwijs. In: Roeleveld, J., Driessen, G., Ledoux, G., Cuppen, J. & Meijer, J. (2011). *Doelgroep leerlingen in het basisonderwijs. Historische ontwikkeling en actuele situatie*. pp. 61-88 Amsterdam/Nijmegen: Kohnstamm Instituut/ITS.
- Roeleveld, J., Veen, I. van der & Ledoux, G. (2009). *Verkenning leerwinst als indicator voor onderwijskwaliteit*. Onderzoek voor het Ministerie van Onderwijs, Cultuur en Wetenschappen. Amsterdam: SCO-Kohnstamm Instituut

Roeleveld, J., Mooij, T., Fettelaaar, D. & G. Ledoux (2011). *Correctiefactoren bij opbrengstmaten in het primair onderwijs*. Amsterdam / Nijmegen: Kohnstamm Instituut / ITS

SCP/CBS(2007). *Armoedemonitor 2007*, Sociaal en Cultureel Planbureau, Den Haag.

Seegers, G., Putten, C.M. van, & Brabander, C.J. de (2002). Goal orientation, perceived task outcome and task demands in mathematics tasks: Effects on students' attitude in actual task settings. *British Journal of Educational Psychology*, 72(3), 365-384.

Timmermans, A. (2012). *Value added in educational accountability: Possible, fair and useful?* Groningen, Proefschrift RUG

Vreeburg, B. (2008). *Absoluut versus relatief beoordelen van de opbrengsten van scholen*. Paper Onderwijs Researchdagen 2008.

Zijsling, D., Keuning, J., Naayer, H. & Kuyper, H. (2012). *Cohortonderzoek⁵⁻¹⁸. Technisch rapport meting VO-3 in 2011*. Groningen / Arnhem: GION / Cito.