

METING TAAL, REKENEN EN WERELDORIËNTATIE 2015

PEILING VAN DE REKENVAARDIGHEID, DE TAALVAARDIGHEID
EN WERELDORIËNTATIEVAARDIGHEDEN IN JAARGROEP 8 VAN
HET BASISONDERWIJS IN 2015



Meting Taal, Rekenen en Wereldoriëntatie 2015

Peiling van de rekenvaardigheid, de taalvaardigheid en wereldoriëntatievaardigheden in jaargroep 8 van het basisonderwijs in 2015

De eerste peiling met behulp van de Centrale Eindtoets

Bas Hemker

© Stichting Cito Instituut voor Toetsontwikkeling Arnhem (2016)

Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van Stichting Cito Instituut voor Toetsontwikkeling worden openbaar gemaakt en/of verveelvoudigd door middel van druk, fotografie, scanning, computersoftware of andere elektronische verveelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook.

Inhoud

Voorwoord 5

Samenvatting 7

1	Inleiding: Vergelijking met eerder onderzoek	9
1.1	Historisch kader	9
1.2	Onvergelijkbaarheid van praktijk en inhoud	9
1.3	Onvergelijkbaarheid van achtergrondvariabelen	12
1.4	Referentieniveaus	13
1.5	Overzicht van het verslag Meting Taal, Rekenen en Wereldoriëntatie	14
2	Vraagstelling en aanpak	15
2.1	Vraagstelling	15
2.2	Methode	15
2.2.1	Kwaliteit van de meetinstrumenten	18
2.2.2	Steekproeftrekking en analyse	19
2.2.3	Achtergrondvariabelen	21
2.3	Analyses	22
2.3.1	Vergelijking tussen en over de jaren	22
2.3.2	Analyses voor de leerlingen in 2015	23
2.3.3	Gemiddelden op nieuwe vaardigheden	24
2.3.4	Referentieniveaus	25
2.3.5	Samenvattend: een nieuwe reeks	26
3	De vergelijking over de jaren	27
3.1	Taalvaardigheden	29
3.2	Rekenvaardigheden	30
3.3	Wereldoriëntatievaardigheden	32
4	De vergelijking van de achtergrond variabelen binnen de CET 2015	35
4.1	Effecten voor aan de leerling gerelateerde achtergrondvariabelen	36
4.2	Effecten op schoolniveau	39
5	Conclusies	45
5.1	Algemeen beeld in de vergelijking door de tijd	45
5.2	Invloed van achtergrondvariabelen	46
5.3	Discussie	48

Literatuur 51

Bijlagen 55

1	Samenvatting Flankerend Onderzoek Rekenen CET 2015	56
2	Verdeling van de achtergrondvariabelen in de onderzoekspopulatie in 2015 die een eindtoets gemaakt heeft	58
3	Gemiddelden van alle vaardigheden op de 2015-schaal per achtergrondvariabele	61
4	Gemiddelden van referentievoordigheden en Woordenschat op de 2015-schaal per achtergrondvariabele	65
5	Gemiddelden van Wereldoriëntatievaardigheden op de 2013-schaal per achtergrondvariabele	67

Voorwoord

Dit rapport is het eerste verslag van de meting van de taal- en rekenvaardigheden van leerlingen in groep 8 van het basisonderwijs met behulp van de Centrale Eindtoets (CET). Het verslag sluit echter aan bij een reeks van verslagen over de jaarlijkse metingen van vaardigheden in groep 8. Sinds 2009 zijn er verslagen gepubliceerd die betrekking hebben op de meting van de taal- en rekenvaardigheden in het jaar ervoor, waarbij in 2015 ook een additioneel rapport is verschenen waarin de metingen van de wereldoriëntatievakken zijn beschreven. Deze vaardigheden worden nu ook in dit verslag beschreven. De metingen zijn van start gegaan naar aanleiding van de kwaliteitsagenda voor het Primair Onderwijs (verschenen in 2007 onder de titel 'Scholen voor morgen') en wordt sindsdien gecontinueerd als landelijke monitor voor een aantal basisvaardigheden.

In de eerdere metingen is altijd gebruikgemaakt van de Eindtoets Basisonderwijs (EB), zoals die werd uitgegeven door Cito B.V.. Door deze keuze werd de extra toetslast voor leerlingen en leerkrachten beperkt. Dat is nog steeds het uitgangspunt van de meting van 2015 waarvan dit verslag de weergave is: we wilden geen additionele toetsen afnemen. In vergelijking met de eerdere publicaties is echter wel een ander veranderd. Deze veranderingen hebben vooral te maken met de veranderingen die hebben plaats gevonden rond de eindtoets. In 2015 is voor het eerst gebruikgemaakt van de Centrale Eindtoets (CET). Deze toets is door Stichting Cito geproduceerd in opdracht van het College voor Toetsen en Examens (CvTE).

Verschillen tussen de EB en de CET betreffen zowel praktische punten, zoals de functie van de toets en het moment van afname, als inhoudelijke punten. Een aantal vaardigheden wordt in de CET niet meer gemeten, en een aantal andere vaardigheden is toegevoegd. Een ander belangrijk verschil is dat bij de CET andere achtergrondvariabelen zijn opgevraagd dan bij de EB. Dit heeft gevolgen voor de gekozen modellen, waardoor gerapporteerde effecten ook verschillen. Al met al zijn daarmee de resultaten van de hier beschreven meting niet zonder voorbehoud te vergelijken met de resultaten in de eerder verschenen rapporten. Om die reden is er voor gekozen dit verslag als een eerste in een nieuwe reeks te beschouwen.

De uitkomsten van deze jaarlijkse peiling geven een indicatie van de stand van zaken met betrekking tot een belangrijk deel van de referentieniveaus behorende bij de doorlopende leerlijnen zoals geformuleerd in het advies van de Expertgroep Doorlopende leerlijnen: *Over de drempels met taal en rekenen*. Dit jaar zijn de gemeten vaardigheden ook voor het eerst inhoudelijk gedefinieerd door de beschrijving van de referentieniveaus. Dit betreft echter nog wel een pilot. De resultaten van dat onderzoek worden niet in dit verslag behandeld maar worden beschreven in het een online katern 'Taal en rekenen aan het einde van de basisschool' van Peil.onderwijs (Inspectie van het Onderwijs, 2016b). Dit katern verschijnt als bijlage bij De Staat van het Onderwijs 2014/2015 (Inspectie van het Onderwijs, 2016a). In het katern wordt onder andere gerapporteerd hoeveel leerlingen de referentiesniveaus voor taal en rekenen bereiken en komt aan de hand van voorbeeldopgaven uit de Centrale Eindtoets aan de orde wat het beheersen van referentieniveaus concreet betekent.

Bij de uitvoering van het project zijn diverse mensen betrokken geweest, waarvan er enkele met name genoemd moeten worden. Sylvia Walraven was bij de ondersteuning zeer belangrijk en Anne-Marie Anthonissen heeft in de coördinatie veel betekend. Jan van Weerden is de auteur van een van de bijlagen. Mijn dank gaat ook uit naar Anja de Wijs, Marleen van der Lubbe en Anja Knuver voor het kritisch doorlezen van de tekst. Ten slotte wil ik graag Loes Hiddink bedanken voor de bureauredactie.

We hopen dat dit rapport en de daarin beschreven resultaten hun weg vinden naar het onderwijsveld en de beleidsmakers.

Dr. B.T. Hemker
Senior onderzoeker

Samenvatting

In dit rapport wordt verslag gedaan van de achtste jaarlijkse peiling van de reken- en taalvaardigheid in jaargroep 8 sinds 2008 (JMTR, voorheen JPON). Hiervoor is gebruikgemaakt van gegevens uit de Centrale Eindtoets van 2015 voor Woordenschat, Begrijpend lezen, drie rekenvaardigheden en drie wereldoriëntatievaardigheden. Het is daarmee een afsluiting van een reeks waarin deze vaardigheden door de tijd gevolgd zijn.

De gemiddelde stijging in vaardigheid over de hele periode van 2008 tot en met 2014 liet zien dat er sprake was van een gemiddelde groei van ongeveer 1% van de standaarddeviatie per jaar. Al is de gemiddelde groei gelijk over de vaardigheden heen, er moet wel opgemerkt worden dat deze groei grillig was en niet voor alle vaardigheden ieder jaar op dezelfde wijze verliep. De groei van 2014 naar 2015 was vergelijkbaar, en gemiddeld over de vaardigheden zelfs iets groter dan 1%.

Wat de vergelijking tussen 2014 en 2015 echter lastig maakt, is een aantal substantiële veranderingen in de metingen. Daar waar tot en met 2014 gemeten werd met de Eindtoets Basisonderwijs, is dat in 2015 gedaan met de Centrale Eindtoets. Deze twee toetsen hebben veel overeenkomsten maar ook een aantal belangrijke verschillen. Een eerste verschil betreft de inhoud van de toetsen: er zijn vaardigheden die pas sinds 2015 gemeten worden (bijvoorbeeld interpunctie), en vaardigheden die sinds 2015 niet meer gemeten worden (kaartlezen). Een tweede verschil betreft het afnamemoment: in plaats van begin februari vindt de meting nu plaats in de tweede helft van april. Een derde belangrijk verschil is dat het doel van de toets enigszins aangepast is: het belangrijkste doel is nu een te laag gegeven schooladvies voor het voortgezet onderwijs mogelijk te kunnen aanpassen. In het verleden was het doel algemener: hulp bij keuze van een passend brugklatype.

Deze aanpassingen leiden er toe dat dit verslag ook een beginpunt is van een nieuwe reeks. De vaardigheden Rekenen, Lezen en Taalverzorging zijn opnieuw gedefinieerd op basis van de beschrijvingen van de referentieniveaus. Het nieuwe ijkpunt is eind april geworden met het nieuwe doel van de Centrale Eindtoets als kader. Wat onveranderd gebleven is, is de meting voor Woordenschat en metingen voor de wereldoriëntatievaardigheden Natuuronderwijs, Aardrijkskunde en Geschiedenis.

Een andere verandering ten opzichte van de eerder rapportages betreft de analyses van de effecten van achtergrondvariabelen op deze vaardigheden. Deze effecten worden vanaf dit verslag gegeven per variabele. In het verleden werd voor een aantal variabelen gecorrigeerd, maar dat is vanaf deze peiling niet meer het geval. Door het ontbreken van gegevens op leerlingniveau over het leerlinggewicht (geen, 0.30 of 1.20) waren vergelijkbare correcties zoals bij eerdere metingen niet meer mogelijk. Dit verslag is ook daardoor echt een nieuwe start, waarbij de gevonden effecten niet direct te vergelijken zijn met die in eerdere verslagen.

De resultaten van het onderzoek naar de achtergrondeffecten zijn verdeeld in effecten van variabelen op leerlingniveau, zoals geslacht en leertijd, en op schoolniveau zoals percentage leerlingen met een leerlinggewicht op een school.

Bij geslacht is alleen bij Woordenschat geen significant verschil gevonden tussen jongens en meisjes. Lezen werd iets beter gedaan door meisjes en Natuuronderwijs door jongens, maar de verschillen waren zeer klein. Groter waren de verschillen voor Taalverzorging dat beter gedaan werd door meisjes, en bij Rekenen, Aardrijkskunde en Geschiedenis, waarop de jongens beter scoorden.

Net als in het verleden blijkt dat vertraagde leerlingen minder goed presteren dan reguliere leerlingen, en dat leerlingen die het basisonderwijs versneld hebben doorlopen (vervroegde leerlingen) over het algemeen nog beter presteren. Dit was vooral het geval bij de vaardigheden die in het referentiekader genoemd

worden: Rekenen, Lezen, en Taalverzorging. De daling van het percentage vertraagde leerlingen en de toename van het percentage vervroegde leerlingen zette zich ook dit jaar weer voort.

Door de jaren heen is het niveau van de leerlingen die geadviseerd worden op specifieke niveaus van het voortgezet onderwijs in te stromen niet fundamenteel gewijzigd. De vaardigheid verschilt aanzienlijk tussen leerlingen met verschillende doorstroomadviezen.

Op schoolniveau blijkt dat er behoorlijke verschillen zijn tussen scholen met betrekking tot het percentage leerlingen met een leerlinggewicht. De ouders van leerlingen zonder leerlinggewicht hebben een hoger opleidingsniveau dan ouders van leerlingen met een leerlinggewicht. In het onderzoek werd gevonden dat hoe minder gewichtenleerlingen op een school zaten, hoe hoger de gemiddelde prestaties waren.

De variabelen regio, verstedelijking, schoolgrootte en denominatie hebben vrijwel geen toegevoegd effect. Als er al een effect gevonden werd, hing dat over het algemeen ook samen met hogere percentages gewichtenleerlingen op die scholen.

Er is ook uitgebreid onderzoek gedaan naar de referentieniveaus. Voor alle vaardigheden werd gevonden dat het percentage leerlingen dat het niveau 1F haalt 90% of hoger is. Het percentage leerlingen dat een hoger niveau haalt, varieert over de drie vaardigheden. Dit onderzoek staat beschreven in andere publicaties zoals *De Terugblik 2015* (College voor Toetsen en Examens, 2016), *De Staat van het Onderwijs 2015* (Inspectie van het Onderwijs, 2016a) en het online katern '*Taal en rekenen aan het einde van de basisschool. Peil. onderwijs*' (Inspectie van het Onderwijs, 2016b) dat als bijlage bij *De Staat van het Onderwijs 2015* verschijnt.

1 Inleiding: Vergelijking met eerder onderzoek

1.1 Historisch kader

Dit rapport is het eerste verslag van de meting van de taal- en rekenvaardigheden van leerlingen in groep 8 van het basisonderwijs met behulp van de Centrale Eindtoets (CET). Het verslag sluit echter aan bij een reeks van verslagen over de jaarlijkse metingen van vaardigheden in groep 8 die een langere traditie kent. In het kader van de kwaliteitsagenda 'Scholen voor morgen' is in 2008 het Jaarlijks Peilingsonderzoek naar het Onderwijsniveau (JPON) van start gegaan (Hemker & Van Weerden, 2009). Dit jaarlijkse onderzoek is gericht op het monitoren van het onderwijsniveau op het gebied van taal- en rekenvaardigheid in het basisonderwijs. In de loop van de tijd is de meting meer bekend geraakt als de Jaarlijkse Meting Taal & Rekenen (JMTR). Het uitgangspunt van het onderzoek is echter onveranderd en sluit aan op de kwaliteitsagenda 'Scholen voor morgen'. Deze kwaliteitsagenda was gericht op duurzame verbetering van het primair onderwijs en legde de prioriteit bij de verhoging van de taal/lees- en rekenopbrengsten. Dit is gecontinueerd in het beleid van daaropvolgende kabinetten en heeft onder andere geleid tot nieuwe wetgeving en voorstellen daartoe, waarin de positie van rekenen en taal en het toezicht daarop verder is verstevigd. Sinds 2014 zijn daar metingen voor drie onderdelen van het vak Wereldoriëntatie (WO) aan toegevoegd: Natuuronderwijs, Aardrijkskunde en Geschiedenis.

Het project was ondergebracht bij PPON, de periodieke peiling van het onderwijsniveau, maar kent een andere methodiek en gebruikt andere instrumenten. De overeenkomst is dat het ook hier gaat om periodiek onderzoek, zij het met een hogere frequentie. Daarnaast worden ook in dit onderzoek de resultaten op een vaste meetschaal gerapporteerd, wat maakt dat de uitkomsten over de jaren heen vergelijkbaar zijn. De belangrijkste verschillen zijn te vinden in de gebruikte instrumenten, de wijze van afnemen en de beschikbare achtergrondgegevens.

De doelstelling van het in dit verslag beschreven onderzoek is vergelijkbaar is met de eerdere onderzoeken. De veranderingen sinds de laatste twee verslagen (Hemker & Van Weerden, 2015a; Hemker & Van Weerden, 2015b) zijn echter van een zodanige aard dat er met recht gesproken kan worden over een nieuwe reeks met een beperkte vergelijkbaarheid met de eerdere reeks.

1.2 Onvergelijkbaarheid van praktijk en inhoud

Het huidige onderzoek naar de taal- en rekenvaardigheid en de vaardigheden op de WO-vakken is gebaseerd op de Centrale Eindtoets (CET). Tot en met 2014 werd daarvoor de Eindtoets Basisonderwijs (EB) gebruikt.

De voornaamste reden dat de CET en de EB niet een-op-een met elkaar te vergelijken zijn, ligt in het wettelijk kader waartoe de Centrale Eindtoets behoort. De CET is de eindtoets die door de overheid wordt aangeboden aan het basisonderwijs in Nederland. Naast de CET waren in het schooljaar 2014-2015 ook Route 8 en IEP toegelaten door de minister als eindtoets in het basisonderwijs. Het wettelijke kader brengt een aantal uitgangspunten mee voor de CET die anders zijn dan bij de EB waardoor de vergelijking van de resultaten op de twee toetsen sterk bemoeilijkt wordt. Deze verschillen zijn op te delen in praktische verschillen en inhoudelijke verschillen.

Praktische verschillen

Er is een vijftal praktische verschillen aan te geven tussen de CET en de EB:

- 1) scholen zijn sinds 2015 verplicht een toegelaten eindtoets (CET, IEP of Route 8) af te nemen bij leerlingen in groep 8;
- 2) de CET is gratis voor scholen;
- 3) het schooladvies is leidend voor toelating voortgezet onderwijs (VO); de CET kan gebruikt worden voor aanpassingen van het advies naar een hoger maar niet naar een lager niveau VO;
- 4) de CET wordt afgenomen tussen 15 april en 15 mei in plaats van begin februari;
- 5) het CvTE is verantwoordelijk voor de Centrale Eindtoets.

Wat de impact van deze praktische verschillen is, is niet direct aan te geven. In het verleden deed ongeveer 85% van de leerlingen in Nederland mee aan de EB. Een factor die een stijging zou kunnen veroorzaken is de verplichting van de CET die er niet was voor de EB. Een andere factor is dat leerlingen die in het VO leerwegondersteunend onderwijs (LWOO) gaan volgen bij de EB nog een speciale status hadden, maar nu verplicht zijn een eindtoets te maken. Een laatste factor die mogelijk tot een deelnamestijging zou kunnen leiden is het feit dat de CET, in tegenstelling tot de EB, voor scholen gratis is. Al deze factoren gelden ook voor de andere toegelaten eindtoetsen.

Een daling zou dan ook verklaard kunnen worden doordat er nu verschillende eindtoetsen zijn, waarbij een school één toetsaanbieder kiest voor alle leerlingen. Na de afname van 2015 bleek dat het aantal leerlingen dat in 2015 de CET gemaakt heeft, iets hoger is dan het aantal leerlingen dat in 2014 de EB gemaakt heeft (een netto toename van 1,8% leerlingen ten opzichte van 2014).

Verschillen die duidelijk een gevolg kunnen hebben voor de resultaten zijn de punten (3) en (4) hierboven. Het schooladvies is leidend. Hiermee is de functie van de CET enigszins verzwakt ten opzichte van de EB; op basis van de CET kan in principe alleen een te laag gegeven schooladvies in heroverweging genomen worden. De CET kan zodoende alleen een impact hebben als de prestaties op de CET beter zijn dan verwacht. In dat geval kan het advies naar boven bijgesteld worden. Bij een lagere toetsprestatie dan verwacht zal het advies niet naar beneden bijgesteld worden.

Voor leerlingen die al een relatief of absoluut hoog advies VO hebben, zou daardoor de motivatie om maximaal te presteren lager kunnen zijn dan voorheen. De leerlingen met een vwo-advies kunnen geen hoger advies krijgen, en naar beneden kan het advies niet worden bijgesteld. Leerlingen met een relatief hoog advies, die niet verwachten beter te kunnen presteren op de CET dan wat het advies aangeeft, zullen mogelijkerwijs ook minder hun best doen. Op basis hiervan zou verwacht kunnen worden dat de getoonde prestaties op de vaardigheden lager zouden kunnen liggen dan voorheen zonder dat er sprake is van een werkelijk daling in vaardigheid in de populatie. Dit zou betekenen dat een daling in geobserveerde prestaties op de eindtoets aannemelijk kunnen zijn.

Aan de andere kant is het moment van de afname aangepast. In plaats van (begin) februari worden de leerlingen nu eind april getoetst. De verandering van de periode houdt in dat de leerlingen langer op school hebben gezeten, en (idealiter) meer geleerd hebben dan bij een meting in februari. Dit zou dan juist een stijging in de te verwachten prestatie op kunnen leveren, waarbij het ook goed mogelijk is dat de impact hiervan voor verschillende subgroepen verschilt. Ook dit zou niet noodzakelijkerwijs betekenen dat de werkelijke vaardigheid van de populatie vooruit is gegaan. Wellicht was de vaardigheid in eerdere afnamejaren in april ook hoger dan in februari, en kan er dus niet van een vooruitgang van het ene jaar naar het andere jaar gesproken worden. Het moge duidelijk zijn dat er aanzienlijk meer alternatieve verklaringen voorhanden zijn dan een daadwerkelijke verandering in de vaardigheidsniveaus, zodat de gevonden resultaten niet zonder voorbehoud geïnterpreteerd kunnen worden.

Inhoudelijke verschillen

Er is ook een aantal, aan elkaar gerelateerde, inhoudelijke verschillen aan te geven tussen de CET 2015 en de EB 2014:

- 1) de inhoud van de CET is gebaseerd op de toetswijzer PO;
- 2) de Centrale Eindtoets rapporteert op referentieniveaus;
- 3) er zijn met grammatica en interpunctie twee nieuwe onderdelen in de toets;
- 4) het onderdeel studievaardigheden is komen te vervallen en deels geïntegreerd in de onderdelen Nederlandse taal en rekenen;
- 5) de aantallen opgaven en verhoudingen tussen en binnen de taal- en rekenonderdelen in de EB en de CET verschillen aanzienlijk van elkaar.

Net zoals in het verleden bij de EB, zijn de opgaven van de CET ontwikkeld door leerkrachten en vak-experts, maar de vaststelling van de CET geschiedt door en onder de verantwoordelijkheid van het CvTE. Belangrijker voor dit verslag is dat de samenstelling en inhoud aangepast is. In de toetswijzer PO spelen de referentiesets een heel belangrijke rol, aangezien er ook wordt gerapporteerd op de referentieniveaus, zij het dat dit vooralsnog een pilot betreft. Dat de referentieniveaus (mede) leidend zijn heeft ook gevolgen voor de gemeten vaardigheden en de geleverde scores. Om te beginnen is, in tegenstelling tot in de EB, Studievaardigheden geen onderdeel meer van de eindtoets. Dat betekent dat er geen score meer is voor die vaardigheid, maar ook dat de standaardscore (de "Cito-score" die van 501 tot 550 loopt) niet meer gebaseerd is op deze vaardigheid. Het onderdeel Kaartlezen van de vaardigheid Studievaardigheden is uit de CET verdwenen, maar de overige drie subvaardigheden zijn in een iets andere vorm opgenomen in de CET en ondergebracht bij taal- en rekenvaardigheden.

Dit heeft gevolgen voor de hier gerapporteerde peiling. Bij de meting Taal en Rekenen op basis van de EB waren de vaardigheden die vielen onder studievaardigheden geen onderdeel van het onderzoek. In de CET zijn de onderdelen Samenvatten en Opzoeken nu onderdeel geworden van de vaardigheid Lezen (LEZ), waarbij deze twee vaardigheden wel enige verwantschap hebben met de vaardigheden Hanteren van studieteksten en Hanteren van informatiebronnen die eerder onder studievaardigheden vielen. Dit betekent dus dat het onderdeel lezen zoals gedefinieerd bij de referentieniveaus niet alleen meer bestaat uit het onderdeel dat in de EB voorheen als Taalonderdeel Begrijpend lezen (TB) was omschreven.

Het voormalig onderdeel Lezen van schema's, tabellen en grafieken van Studievaardigheden heeft enige verwantschap met het onderdeel Verbanden, wat een toegevoegd subdomein is van het vakonderdeel Rekenen. Deze toevoegingen zijn een gevolg van de rapportage van de referentieniveaus waar deze vaardigheden een onderdeel van uitmaken.

Het rapporteren op referentieniveaus heeft ook geleid tot de toevoeging van twee extra vaardigheden, te weten Interpunctie en Grammatica. Deze onderdelen, of daarop gelijkende, waren geen onderdeel van de EB. Samen met de eerder ook gemeten vaardigheden Spellen van werkwoorden en Spellen van niet-werkwoorden vormen deze in de CET het domein Taalverzorging (TAV), waar ook referentieniveaus voor gedefinieerd zijn.

De vaardigheden Schrijven van teksten en Woordenschat die in de EB waren opgenomen zijn ook nog steeds onderdeel van de CET. In 2008 is besloten dat Schrijven van teksten geen onderdeel zou uitmaken van de jaarlijkse peiling (JPON; nu MTR), en in de CET is het aantal opgaven voor dit onderdeel ook verminderd. Woordenschat is wel in eerdere peilingen meegenomen en zal ook in dit rapport meegenomen worden.

Behalve de veranderingen in de samenstelling van de toets voor wat de betreft de onderdelen, hebben ook de verschillende onderdelen binnen de CET niet meer exact hetzelfde aantal opgaven als in de EB. Het gevolg is dat voor de vaardigheden waar vooral de aandacht op ligt bij de CET, te weten de vaardigheden waarvoor referentieniveaus gerapporteerd worden, een jaarvergelijking met de eerdere

metingen niet mogelijk is. Alleen op subschalen kan inhoudelijk een vergelijking gemaakt worden, hoewel daarbij nog steeds de praktische verschillen die hierboven aangegeven zijn, gelden. In hoofdstuk 2, waarin de gemeten schalen worden beschreven, wordt verder beschreven hoe de CET- en de EB-schalen van elkaar verschillen.

Wereldoriëntatie: beter vergelijkbaar met vorige jaren

Bij de CET is in 2015 voor de meting van Natuuronderwijs, Aardrijkskunde en Geschiedenis gebruikgemaakt van de facultatieve toets Wereldoriëntatie (WO). De vaardigheid die gemeten wordt met de toets voor Natuuronderwijs betreft vooral biologie onderwerpen. Deze aanvullende taken zijn in 2015 inhoudelijk onveranderd ten opzichte van 2013 en 2014 en sluiten daarmee aan op het verslag uit 2015 (Hemker & Van Weerden, 2015b). Dat betekent zodoende dat de veranderingen bij WO minder groot zijn dan bij de andere onderdelen. Inhoudelijk is er niets veranderd: de vaardigheden, en zelfs de opgaven, zijn identiek. Ook de praktische verschillen zijn kleiner dan in het geval van de overige vaardigheden. Net als bij de EB is het onderdeel WO bij de CET niet verplicht. Ook is onveranderd dat de meting voor WO geen formele rol heeft in het schooladvies: deze toetsen waren eerder ook al geen onderdeel van de standaardscore. Een verschil met de vorige jaren is echter wel dat in 2015 deze toetsen zijn afgenomen tussen 15 april en 15 mei in plaats van begin februari. Het percentage leerlingen dat de WO-toetsen maakt in verhouding tot de leerlingen die de toetsen maken die gebruikt worden voor de bepaling van de standaardscore is in 2015 iets lager dan in de twee jaren daarvoor. Hierover meer in de beschrijving van de resultaten van WO.

1.3 Onvergelijkbaarheid van achtergrondvariabelen

Een andere belangrijke reden waarom de resultaten in dit verslag slecht te vergelijken zijn met die van de eerdere verslagen ligt in de beschikbaarheid van gegevens over de achtergrondvariabelen. Een variabele die in eerdere analyses een cruciale rol speelde was leerlinggewicht, ook wel formatiegewicht genoemd. In deze variabele was het opleidingsniveau van de ouders verdisconteerd. Deze variabele had bij vrijwel alle vaardigheden in alle afgelopen jaren een invloed: er was sprake van (aanzienlijke) verschillen tussen de prestaties van leerlingen die geen leerlinggewicht hadden (leerlinggewicht = 0.00) en de leerlingen die wel een leerlinggewicht hadden (Hemker & Van Weerden, 2015a, 2015b). Het verschil tussen een laag (0.30) en een hoog (1.20) leerlinggewicht was in veel gevallen een stuk minder prominent.

In eerdere onderzoeken is de variabele leerlinggewicht een van de variabelen die bij de vergelijkingen binnen een jaar¹ altijd in het basisanalysemodel opgenomen waren. Dat betekent dat in de beschreven analyses altijd voor de invloed van die variabele is gecorrigeerd. Door het corrigeren van het effect van het leerlinggewicht was het mogelijk om te onderzoeken wat het additioneel effect was van het al dan niet hebben van leerlingen met een leerlinggewicht. Hierbij was een driedeling gemaakt van scholen met minder dan 10% gewichtenleerlingen (stratum 1), vanaf 10% tot en met 25% gewichtenleerlingen (stratum 2) en meer dan 25% gewichtenleerlingen (stratum 3). Door nu niet meer voor het individuele effect van het leerlinggewicht te corrigeren zal het stratumeffect veel groter worden en niet meer vergelijkbaar zijn met de eerdere rapportages. Dit zal mogelijk ook voor andere variabelen gelden, zoals het doorstroomniveau van de leerlingen.

Een gevolg van het noodzakelijke verschil in keuze van het model is dat de er verschillen kunnen ontstaan met de eerder gerapporteerde effecten. Het ontbreken van deze variabele heeft uiteraard ook gevolgen voor de interpretatie: de gerapporteerde effecten zijn geen gecorrigeerde, maar ruwe effecten en het betreft dus niet meer het additioneel effect van de variabele maar het geobserveerde verschil tussen groepen.

¹ Bij de vergelijkingen tussen jaren is met ongecorrigeerde modellen gewerkt. Een rationale voor die keuze is gegeven in Hemker en Van Weerden (2008).

In hoofdstuk 2 zal verder beschreven worden hoe de analyses nu uitgevoerd worden en wat de verschillen zijn met de eerdere werkwijze, evenals de rationale voor deze wijziging.

Een andere variabele die niet meer opgevraagd wordt, is thuistaal. Het ontbreken van deze variabele heeft een aanzienlijk minder grote impact dan die van het ontbreken van leerlinggewicht, aangezien thuistaal niet in het basismodel opgenomen was. Deze variabele ontbrak al eerder: in 2012 en 2013 was deze ook niet opgevraagd. De variabele was in 2014 weer opgenomen, omdat dat het mogelijk maakt te onderzoeken of er sprake is van vraagpartijdigheid (*item bias; differential item functioning*) of toetspartijdigheid (Bügel & Sanders, 1998). Met een dergelijk onderzoek wordt de vraag beantwoord of er specifieke opgaven zijn die anderstaligen benadelen. Bij de beoordelingen van tests en toetsen door de COTAN worden dergelijke analyses als zeer belangrijk geacht om de kwaliteit van een test of toets te kunnen evalueren (Evers, Lucassen, Sijtsma & Meijer, 2010).

Doordat in het verleden thuistaal opgenomen was, was het mogelijk te zien dat, gegeven de opleiding van de ouders zoals deze verwerkt is in het leerlinggewicht, thuistaal geen invloed had op de rekenvaardigheid of spelling, maar dat bij Woordenschat de invloed nog wel sterk was. Om die reden is het interessant om de vaardigheid Woordenschat mee te nemen in de analyses. Door de samenhang met thuistaal kan juist door het ontbreken van die variabele mogelijk via Woordenschat toch iets gezegd worden over die impact.

1.4 Referentieniveaus

In de aanbevelingen van de Expertgroep Doorlopende Leerlijnen Taal en Rekenen (EGDLL, ofwel de commissie Meijerink, SLO, 2008) werd vastgelegd wat leerlingen moeten kennen en kunnen op het gebied van taal en rekenen bij het verlaten van de basisschool: de referentieniveaus voor taal en rekenen. Dit is bij wet geregeld in 2010. Toen in 2008 de Jaarlijkse Meting Taal & Rekenen (voorheen JPON) van start ging, werd al gedacht aan het meenemen van de referentieniveaus. Doordat deze echter in 2008 nog niet inhoudelijk gedefinieerd waren, konden toentertijd de vaardigheden in de EB nog niet afgestemd worden op de referentieniveaus. De officiële grenswaarden voor 1F of 1S waren logischerwijs in 2008 ook nog niet vastgesteld. Wel waren er in 2008 omschrijvingen en veronderstellingen met betrekking tot de referentieniveaus. Op basis van die omschrijvingen is er bij de peilingen in 2008 voor gekozen om op de geobserveerde vaardigheidsverdeling vijf verschillende typische leerlingen te definiëren (zie tabel 1.1). Aan de schaalwaarde horende bij de laag-vaardige leerling werd in 2008 een speciale betekenis toegekend. Dat was het punt op de vaardigheidsschaal dat door minstens 75% van de leerlingen wordt bereikt. In het advies van de Expertgroep Doorlopende Leerlijnen Rekenen en Taal (SLO, 2008) werd dit punt regelmatig genoemd als het referentieniveau voor 1F, zowel bij taal als bij rekenen. In de loop van de tijd is de ontwikkeling van de referentieniveaus verder uitgekristalliseerd, zowel voor wat betreft de inhoud van de schaal als de beoogde prestatie van de leerlingen. Hierdoor komen de definities van 2008 niet meer geheel overeen met die in 2015. Het grenspunt P75 – gebaseerd op de gegevens van 2008 (Hemker & Van Weerden, 2009) – correspondeert zodoende ook niet met het definitieve niveau 1F. Het 'echte' niveau 1F is inmiddels op basis van inhoudelijke kenmerken vastgesteld.

Tabel 1.1 Definiëring typische leerlingen in de vaardigheidsverdeling

Aanduiding	Afkorting	Percentiel 2008	% kandidaten 2008 van minstens dit niveau	Schaalwaarde
Zeer laag vaardig	ZLV	P10	90%	186
Laag vaardig	LV	P25	75%	216
Gemiddelde/standaard	G/S	P50	50%	250
Hoog vaardig	HV	P75	25%	284
Zeer hoog vaardig	ZHV	P90	10%	314

In de loop der jaren is weliswaar meer bekend geworden over de definitieve referentieniveaus, maar tot en met 2014 was er nog geen inhoudelijke koppeling tussen de Eindtoets en (de inhoudelijke kenmerken van) de referentieniveaus. In eerdere verslagen (onder andere Hemker & Van Weerden, 2015a) kon dan ook geen goede relatie gelegd worden tussen prestaties op de Eindtoets van 2014 enerzijds en de referentieniveaus anderzijds.

In de CET is er een pilot uitgevoerd waarbij de gemeten vaardigheden ook voor het eerst inhoudelijk gedefinieerd zijn door de beschrijving van de referentieniveaus. Het is hierdoor mogelijk de resultaten op de toetsen aan de referentieniveaus te relateren. De resultaten van dat onderzoek worden niet in dit verslag behandeld maar zijn meegenomen in de jaarlijkse publicatie *De Staat van het Onderwijs* (Inspectie van het Onderwijs, 2016a) en het bijbehorend online katern '*Taal en rekenen aan het einde van de basisschool. Peil. onderwijs*' (Inspectie van het Onderwijs, 2016b) dat als bijlage bij De Staat van het Onderwijs verschijnt. Doordat de gebruikte schalen over de verschillende jaren inhoudelijk verschillen, zijn deze resultaten niet direct te vergelijken met de eerder gerapporteerde resultaten op de vaardigheden uit de jaren voor 2015. Enige voorzichtigheid bij de interpretatie van de resultaten van 2015 zou betracht moeten worden, aangezien het vooralsnog een pilot betreft.

1.5 Overzicht van het verslag Meting Taal, Rekenen en Wereldoriëntatie

Het verslag begint met een beschrijving van de opzet van het onderzoek, de gebruikte toetsen en de wijze waarop deze toetsen met elkaar vergeleken kunnen worden. Vervolgens wordt de onderzoekspopulatie beschreven en worden de definities van de achtergrondvariabelen gegeven (hoofdstuk 2). Uit de inleiding bleek al dat de jaarvergelijking met de eerdere peilingen beperkt mogelijk is.

In hoofdstuk 3 wordt voor de vaardigheden waarvoor dat mogelijk is een vergelijking gemaakt met de eerdere peilingen, waarbij 2008 als referentiepunt gebruikt wordt. Deze vergelijking betreft twee taalvaardigheden (Begrijpend lezen en Woordenschat), drie rekenvaardigheden en drie vaardigheden met betrekking tot Wereldoriëntatie .

In het vierde hoofdstuk wordt dieper ingegaan op de resultaten van 2015. De gemeten vaardigheden zijn hier voor taal en rekenen deels anders dan in voorgaande jaren aangezien de focus nu ligt op de vaardigheden die cruciaal zijn voor de referentieniveaus. Dat betekent dat hier de vaardigheden Lezen (waar Begrijpend lezen een onderdeel van is), Taalverzorging en Rekenen (zonder onderscheid in subschalen) de kern vormen. Daarnaast worden de resultaten gepresenteerd voor de vaardigheid Woordenschat en de drie schalen voor Wereldoriëntatie (Natuuronderwijs, Aardrijkskunde en Geschiedenis). In dit hoofdstuk worden de effecten beschreven van de achtergrondvariabelen op deze vaardigheden. Het laatste hoofdstuk bevat een samenvatting, conclusie en discussie.

2 Vraagstelling en aanpak

2.1 Vraagstelling

Bij de Jaarlijkse Meting Taal en Rekenen (JMTR) is het vaststellen van veranderingen in de taal- en rekenvaardigheden van leerlingen in jaargroep 8 van het basisonderwijs over de jaren heen een belangrijke vraag. Zoals aangegeven in hoofdstuk 1 zorgt de overgang van de EB naar de CET ervoor dat – om uiteenlopende redenen – de interpretatie van de resultaten van deze jaarvergelijking lastiger is dan voorheen.

De taalvaardigheden waarvoor een jaarvergelijking gegeven wordt zijn Begrijpend lezen (TB) en Woordenschat (TW). De rekenvaardigheden waarvoor dat mogelijk is zijn Getallen en Getalsbewerkingen (RG), Breuken, Procenten en Verhoudingen (RB) en Meten, Meetkunde, Tijd en Geld (RM). Tot slot wordt de vergelijking gegeven voor de WO-vakken Natuuronderwijs (WN), Aardrijkskunde (WA) en Geschiedenis (WG). Deze acht vaardigheden zijn ook in eerdere metingen meegenomen. Het zijn gedetailleerde subschalen die door de tijd heen een gelijke inhoudelijke definitie hebben. Echter, door de praktische verschillen tussen de EB en de CET zal de conclusie dat leerlingen vaardiger of juist minder vaardig geworden zijn door de tijd, niet goed te geven zijn: het verschil kan ook door de verschillen in de afnamecondities veroorzaakt worden.

De referentieniveaus hebben betrekking op de vaardigheden Lezen (LEZ), Taalverzorging (TAV) en Rekenen (REK). Voor deze vaardigheden is het dit jaar niet goed mogelijk een goede jaarvergelijking te maken aangezien bij deze schalen inhoudelijke onderdelen opgenomen zijn die niet in eerdere EB-toetsen zaten (zie ook paragraaf 2.2). Deze situatie zal eenmalig zijn: in dit verslag van 2015 zullen de metingen voor LEZ, TAV en REK als startpunt genomen worden. In het verslag van 2016 kan wel een vergelijking gemaakt worden tussen de resultaten van de CET 2015 en de CET 2016 op deze schalen.

Naast de vergelijking van prestaties in de verschillende afnamejaren worden in dit verslag ook onderzoeksresultaten gerapporteerd over de prestatieverschillen tussen groepen van leerlingen. Deze groepen leerlingen verschillen van elkaar op belangrijke achtergrondvariabelen. Deze effecten van de achtergrondvariabelen worden berekend voor de vaardigheden waar referentieniveaus voor gedefinieerd zijn: REK, LEZ en TAV. Voor de vaardigheden Woordenschat en de drie WO-vaardigheden worden zowel de jaarvergelijkingen gegeven als de effecten van de achtergrondgegevens. De vaardigheid TB (nu onderdeel van LEZ) en de subvaardigheden voor Rekenen (RG, RB en RM; nu alle onderdeel van REK) worden behalve voor de jaarvergelijking niet meer² apart onderzocht op de effecten van de achtergrondvariabelen.

2.2 Methode

Zoals eerder aangegeven is bij deze metingen gebruik gemaakt van instrumenten die eerder in een ander kader werden ingezet. Het zijn dus geen nieuwe toetsen die afzonderlijk voor dit project zijn ontwikkeld. De gegevens die hier weergegeven worden voor 2008 tot en met 2014 betreffen de toetsen die vallen onder de EB. De gegevens van 2015 betreffen toetsen die vallen onder de CET. Deze toetsen bevatten een aantal onderdelen voor rekenen en taal die we goed voor dit doel kunnen gebruiken, maar die elk jaar volledig worden vernieuwd. Omdat het elk jaar andere opgaven (items) betreft, kunnen de resultaten niet zomaar van jaar tot jaar met elkaar worden vergeleken, en zeker niet voor wat betreft de overgang naar 2015 zoals uitgebreid is beschreven in hoofdstuk 1. Hoe de resultaten over de jaren heen vergeleken worden, wordt verderop in dit hoofdstuk beschreven.

² In de eerdere rapporten, tot en met 2014, worden effecten gegeven van achtergrondvariabelen op deze subschalen.

De verschillende EB- en CET-toetsen zijn in aantallen opgaven en mogelijk te onderscheiden onderwerpen lang niet zo gedifferentieerd als in het reguliere onderzoek van PPON tot nu toe gebruikelijk was. Het aantal vaardigheden is daardoor beperkt terwijl er in de laatste rekenpeiling voor deze groep (Scheltens, Hemker, & Vermeulen, 2013) alleen al onderscheid gemaakt werd tussen meer dan 20 subvaardigheden binnen het reken-wiskundeonderwijs. Bij PPON was het daardoor bijvoorbeeld mogelijk de vooruitgang op schattend rekenen waar te nemen, en tegelijkertijd de verminderde vaardigheid in vermenigvuldigen en delen. Ook werden bij PPON meer en andere itemvormen ingezet dan alleen de vaste vierkeuzevraag die in de EB en de CET gebruikt wordt. Door middel van flankerend onderzoek is gekeken naar de impact van gesloten in vergelijking met open vragen in een CET-context. Zie bijlage 5 voor een samenvatting van dat onderzoek. Zoals beschreven zijn er verschillen tussen de EB en de CET in de afnameconditie: zowel het doel van de toets als het afnamemoment in het schooljaar is enigszins gewijzigd. De verschillen tussen de EB/CET en PPON zijn echter nog groter. Bij PPON zijn alle toetsen en taken afgenomen door een getrainde toetsleider. In het geval van de hier beschreven metingen is de eigen leerkracht de toetsleider, aangezien deze de EB en de CET afneemt. Dat geeft een mogelijke aantasting van de objectiviteit van de afname. Een verschil dat mogelijk een positief effect heeft op de in dit onderzoek gerapporteerde resultaten is het verschil in impact van de verschillende toetssituaties. Zeker de EB, maar ook de CET kunnen worden beschouwd als een 'high-stake'-toets. Het zijn toetsen waar leerlingen hun uiterste best op zullen doen, niet alleen vanwege de gevolgen van de uitslag voor hun verdere toekomst, maar ook vanwege de grote mate van aandacht en publiciteit eromheen. Voor de afnames voor PPON was dat niet het geval. Deze verschillen in conditie hebben hun weerslag op de hoogte van de toetsscores, zo is inmiddels uitgezocht (zie verder Hemker, 2013).

Voor een vergelijking over de jaren heen waren voorheen de risico's bij de vergelijking van de EB van minder groot belang, omdat de condities niet over de jaren heen varieerden. Het was dus geen probleem om een trend in beeld te brengen. Met de veranderde condities bij de CET is het lastiger geworden een trend te interpreteren.

Uit overwegingen van efficiëntie en kosten zijn voor de jaarlijkse niveaupeilingen dus overwegend gegevens gebruikt die toch al door scholen worden verzameld. Voor de meting van de taal- en rekenvaardigheden in jaargroep 8 zijn dat de verzamelde gegevens uit campagne van de CET 2015³. Dit gaat niet om een enkele toets. In deze campagne wordt een aantal toetsen afgenomen. Om te beginnen is dat de reguliere CET (CET Basis) die door veruit de meeste leerlingen wordt gemaakt. Dat is een toets op papier die in een vaste periode van drie dagen in april gemaakt wordt. In 2015 was dat van 21 tot en met 23 april.

Daarnaast is gebruikgemaakt van gegevens die verzameld zijn met twee toetsen die in exact dezelfde periode worden afgenomen en een identieke functie hebben: de CET Niveau en de CET ankertoets. De CET Niveau is bedoeld voor leerlingen met een advies op (laag) vmbo-niveau: voor die leerlingen is de reguliere toets vaak te moeilijk. De CET ankertoets wordt gebruikt om de gegevens te kunnen vergelijken met de EB. De methode om de verschillende toetsen met elkaar te vergelijken wordt beschreven in paragraaf 2.3. De resultaten van 2008 tot en met 2014 zijn ook nog met elkaar verbonden door middel van de resultaten op de toets Basisvaardigheden, maar die wordt vanaf 2015 niet meer afgenomen.

Naast de genoemde toetsen zijn er nog een tweetal digitale toetsen die ook in het kader van de CET worden afgenomen. Deze worden niet binnen de vaste periode afgenomen. De resultaten op deze digitale toetsen vallen daardoor ook buiten dit onderzoek. Het hier gerapporteerde onderzoek heeft betrekking op de resultaten op de toetsen die op papier zijn uitgegeven en in dezelfde periode zijn afgenomen: CET Basis, CET Anker en CET Niveau.

³ Voor de EB-gegevens taal en rekenen betreft het de jaren 2008 tot en met 2014; voor wereldoriëntatie betreft het 2013 en 2014.

Voor de metingen van de WO-vaardigheden is gebruikgemaakt van de toetsen die eveneens in dezelfde periode zijn afgenomen als de CET. Dit zijn geen verplichte toetsen en deze worden ook niet door alle leerlingen gemaakt. Deze toetsen zijn wel onderdeel van het onderzoek omdat deze ons in de gelegenheid stellen om de vaardigheden te meten die verkregen zijn bij het Natuuronderwijs, Aardrijkskunde en Geschiedenis.

De EB en de CET verschillen voor wat betreft de aantallen opgaven: de EB bestond uit 200 opgaven terwijl de CET 220 opgaven bevat. Zoals in hoofdstuk 1 al beschreven is, zijn er ook verschillende vaardigheden gemeten. In tabel 2.1 worden de gemeten vaardigheden in de CET aangegeven. Hierbij is onderscheid gemaakt tussen de referentieniveauvaardigheden (LEZ, TAV, REK) en de subschalen die onderdeel uitmaken van deze vaardigheden. Bij ieder van deze vaardigheden is aangegeven hoeveel opgaven er in de CET (alle varianten) zitten. Dit aantal opgaven is per vaardigheid aangegeven in de kolommen "#its(CET)". De eerste kolom #its(CET) betreft het aantal opgaven dat voor de analyses van 2015 is opgenomen. In de tweede kolom #its(CET) is het aantal opgaven in de CET per subschaal weergegeven. Een deel van de subschalen kon gebruikt worden voor de jaarvergelijking (aantallen in zwart) en een deel kon daar niet voor gebruikt worden (aantallen in rood). De kolom #its(EB) geeft het aantal opgaven per subvaardigheid dat per EB opgenomen was. Een aantal vaardigheden is niet eerder in de EB afgenomen. Bij die vaardigheden staat een punt bij #its(EB). De aantallen opgaven die gebruikt konden worden voor de jaarvergelijking zijn gegeven in zwart; de opgaven in rood konden niet gebruikt worden voor de jaarvergelijking. De schalen waar een jaarvergelijking voor mogelijk was zijn gegeven in de kolom EB/CET.

De 40 opgaven Studievaardigheden die onderdeel uitmaakten van de EB zijn niet in tabel 2.1 gegeven aangezien deze geen deel van de eerdere metingen uitmaakten. In tabel 2.1 is terug te vinden dat het aantal opgaven per subschaal kan verschillen tussen de CET en de EB. De toetsen WO bestaan uit 90 opgaven. Die zijn gelijk gebleven in 2015.

Tabel 2.1 Schalen (inclusief afkortingen) en aantal opgaven (#its) in de CET en de EB

meting 2015			Subschaal			
CET	Referentieniveau	#its(CET)	EB/CET	volledige omschrijving	#its(EB)	#its(CET)
REK	Rekenen	85	RG	getallen en bewerkingen	25	30
			RB	verhoudingen, breuken en procenten	20	20
			RM	meten, meetkunde, tijd en geld	15	20
			.	verbanden	.	15
LEZ	Lezen	45	TB	begrijpend lezen	30	25
			.	opzoeken	.	10
			.	samenvatten	.	10
TAV	Taalverzorging	50	.	spellen van werkwoorden	10	10
			.	spellen van niet-werkwoorden	10	10
			.	interpunctie	.	15
			.	grammatica	.	15
TW	geen referentieniveau	20	TW	woordenschat	20	20
			.	schrijven van teksten	30	20
WN	geen referentieniveau	30	WN	natuur en techniek	30	30
WA	referentieniveau	30	WA	aardrijkskunde	30	30
WG		30	WG	geschiedenis	30	30

2.2.1 Kwaliteit van de meetinstrumenten

De CET is een nieuw samengestelde toets, die bestaat uit opgaven die al in proeftoetsen in voorgaande jaren zijn uitgeprobeerd. Dat geldt ook voor de CET Niveau en de ankertoets die voor het grootste gedeelte overeenkomt met de CET. Dit was in het verleden ook het geval bij de EB Basis, EB Anker en EB Niveau. De psychometrische kenmerken zijn door de proeftoetsen van tevoren goed in te schatten en blijken telkens van hoog niveau te zijn. De eigenschappen van de CET zijn gepubliceerd in een afzonderlijk document (CvTE, 2015), net zoals bij de EB het geval was⁴. We mogen hier stellen dat de psychometrische kwaliteit bij de Eindtoets gewaarborgd is door de gehanteerde procedures.

De kwaliteit van de WO-toetsen is beschreven in het verslag waarin de vorige metingen van WO-vaardigheden (Hemker & Van Weerden, 2015). Deze toetsen zijn onveranderd van 2013 tot en met 2015. Het is belangrijk hier te wijzen op het feit dat de wijze waarop deze toets en de versies daarvan in dit onderzoek worden ingezet wel een andere is dan waarvoor zij zijn bedoeld. Het primaire doel van de EB en de CET is namelijk het opleveren van een goede prognose voor de kans van slagen in verschillende vormen van voortgezet onderwijs voor individuele leerlingen. In deze studie gaat het echter niet om de totaalscore op alle onderdelen, de standaardscore, maar om de uitkomsten op delen van de toets. In de jaarlijkse meting Taal, Rekenen en WO worden verschillende onderwerpen van de toets afzonderlijk gebruikt om een beeld te schetsen van de kwaliteit van het onderwijs op systeemniveau. Voor dat doel zijn de psychometrische eisen die aan een toets worden gesteld anders. Met dat fenomeen is rekening gehouden in de analyses, op een vergelijkbare wijze als in voorgaande jaren (zie verder Hemker & Van Weerden, 2009; Hemker, Kuhlemeier & Van Weerden, 2010; Hemker, Kordes en Van Weerden, 2011; Hemker en Van Weerden, 2012; Van Weerden, Hemker, Straat & Mulder, 2013; Van Weerden, Hemker, Straat & Mulder, 2014; Hemker & Van Weerden, 2015a, 2015b). In tabel 2.2 worden de betrouwbaarheden gegeven van de referentieniveauvaardigheden. Deze waarden zijn ruimschoots voldoende voor systeemevaluatiemetingen, maar ook voor metingen op individueel (subschaal-)niveau.

Tabel 2.2 *Betrouwbaarheden op basis van interitemrelaties (Guttman's lambda-2 en Cronbach's alpha) voor de referentieniveauvaardigheden*

	Lambda-2		Alpha	
	CET	CET-N	CET	CET-N
LEZ	0,76	0,81	0,76	0,81
TAV	0,84	0,82	0,84	0,81
REK	0,92	0,92	0,91	0,92

De betrouwbaarheden zoals gevonden bij de ankertoets zijn niet apart gegeven omdat deze toets voor meer dan 80% overeenkomt met de CET. De subschalen van de taal- en rekenvaardigheden die voor de jaarvergelijking zijn gebruikt zijn lager dan die van de referentieniveauvaardigheden, maar alle betrouwbaarheden zijn (veel) groter dan 0,60. Daarmee zijn ze hoog genoeg voor systeemevaluatiemetingen. De betrouwbaarheden van de drie WO-toetsen zijn eerder gerapporteerd (Hemker & Van Weerden, 2015b). Die betrouwbaarheden zijn allen ruim groter dan 0,70.

⁴ Deze publicatie is verzorgd door het College voor Toetsen en Examens (CvTE), in samenwerking met Cito. De publicaties met betrekking tot de eerder afgenomen EB werden geheel door Cito verzorgd.

2.2.2 Steekproeftrekking en analyse

In 2015 waren er 164.228 leerlingen die een Centrale Eindtoets gemaakt hebben. Er waren 5541 leerlingen die buiten de onderzoekspopulatie vielen. Dit waren onder andere leerlingen van scholen in het buitenland, het speciaal onderwijs en het voortgezet onderwijs. In de dataset zijn deze onder andere te herkennen aan het ontbreken van een BRIN-code (buitenland), het ontbreken van gegevens over de verdeling van de formatie binnen een school (kenmerkend voor speciaal onderwijs en voortgezet onderwijs), of aan het feit dat in de database al specifiek is aangegeven dat het geen school uit het basisonderwijs betreft. Als additionele controle zijn de scholen afgezet tegen de lijst met scholen voor primair onderwijs van DUO. Leerlingen die een digitale versie van de Eindtoets hebben gemaakt zijn ook buiten beschouwing gelaten, omdat deze leerlingen de toets op een ander moment gemaakt hebben. De basis voor de analyses is dientengevolge een dataset met 158.687 leerlingen.

De hierboven genoemde exclusiecriteria zijn ook in eerdere onderzoeken van JPON/JMTR toegepast. In 2014 waren er 161.192 leerlingen, maar vielen er meer leerlingen buiten de onderzoekspopulatie waardoor er toen 148.314 leerlingen in de analyses opgenomen waren.

Uit het CET 2015-databestand met 158.687 leerlingen zijn twee typen steekproeven getrokken: één ten behoeve van de schaalconstructie en één voor de analyse met achtergrondvariabelen. Het eerste bestand is de "kalibratiesteekproef" die gebruikt wordt om de opgaven van verschillende jaren op dezelfde schaal te kunnen plaatsen en de toetsresultaten daarmee over de jaren heen vergelijkbaar te maken. Het is een gestratificeerde steekproef met als belangrijkste doel de koppeling tussen de verschillende jaren zo sterk mogelijk tot stand te brengen. Dat leidt tot een gerichte selectie van groepen leerlingen waarbij sommige leerlingen een grotere kans hebben in deze steekproef te zitten. Zo zitten alle 3259 leerlingen die de CET anker gemaakt hebben in deze steekproef. Daarnaast zijn er steekproeven genomen van 17.127 leerlingen die de reguliere CET Basis (CET-B) gemaakt hebben (ongeveer 1 op de 7 CET-B leerlingen) en 10.211 leerlingen die de CET Niveau (CET-N) gemaakt hebben (ongeveer 1 op de 3 CET-N leerlingen).

Bij het schalen van de opgaven – het schatten van de itemparameters – is de representativiteit van de steekproef van minder groot belang, vanwege populatie-onafhankelijkheid van de itemparameterschattingen bij de gebruikte analysemethode (item response theorie, zie verder paragraaf 2.3). Merk op dat het voor de steekproef voor de analyse met achtergrondvariabelen wel van belang is dat de steekproef representatief is. Om die reden is een tweede steekproef getrokken voor die analyses waarbij wel gelet is op de representativiteit.

Voor de toetsen WO was het niet noodzakelijk om een aparte steekproef te trekken voor het schalen van de opgaven. De toetsen in 2015 zijn gelijk aan die in 2013 en 2014 waardoor de itemparameters bekend waren. Wel is gecontroleerd of deze schattingen nog steeds correct waren. Alle leerlingen die de WO-toetsen maken, maken dezelfde toetsen binnen een afnamejaar. Er zijn geen aparte Basis- of Niveau-varianten, wat betekent dat in tegenstelling tot de CET-toetsen er ook geen toetsen binnen een jaar aan elkaar gerelateerd hoeven te worden.

De WO-toetsen verschillen op meer punten van de CET. De WO toetsen zijn niet verplicht en hebben geen enkele impact op de advisering naar het VO. Deze toetsen worden dan ook door minder leerlingen gemaakt dan de CET-toetsen: 110.937 leerlingen hebben de WO-toetsen gemaakt. De groep leerlingen die de WO-toetsen gemaakt heeft, verschilt ook op een aantal kenmerken van de groep die de toetsen wel gemaakt heeft. Om te beginnen is het percentage dat niet in de onderzoekspopulatie hoort bij de groep die geen WO-toetsen gemaakt hebben groter (9,1%) dan bij de groep die de WO-toetsen wel gemaakt hebben (0,6%). Zodoende is ook het percentage dat de WO-toetsen gemaakt heeft in de onderzoeksgroep aanzienlijk groter (69,5% heeft daarvan de WO-gemaakt) dan bij de groep die niet tot de onderzoekspopulatie hoort (12,8%). Al met al zijn er in de onderzoekspopulatie 110.228 leerlingen die de WO-toetsen gemaakt hebben en de 48.459 die dat niet gedaan hebben.

De kwaliteit van de gebruikte schalen is niet anders dan die in voorgaande jaren (zie voor nadere toelichting de vorige JPON-rapportages). Dit geldt voor de schalen voor de jaarvergelijking, maar ook voor de nieuwe referentieniveauschalen.

Na de schaalanalyse is met de geconstrueerde schalen met vrijwel dezelfde psychometrische kenmerken als in voorgaande jaren de analyse gedaan op achtergrondvariabelen. Hierbij is onderscheid gemaakt tussen de schalen die alleen gebruikt zijn voor een jaarvergelijking (TB en de reken-subschalen) en schalen die gebruikt zijn voor de analyses binnen 2015 (de referentieniveauschalen LEZ, TAV en REK). Alleen de schalen voor Woordenschat en de WO-schalen zijn voor beide typen analyses gebruikt.

Zoals hierboven al gesteld is, is bij deze analyses de representativiteit van de steekproef wel van groot belang. Daartoe is een nieuwe steekproefprocedure uitgevoerd, zodanig dat van ieder van de drie te onderscheiden groepen (CET, CET-anker, en CET-Niveau) 20% van de leerlingen in de steekproef van ruim 31.744 leerlingen terecht kwam. Van bijna iedere school die een van deze toetsen heeft gemaakt zijn leerlingen in deze steekproef opgenomen. Deze steekproef verschilt op geen van de scores significant van de 126.943 leerlingen die niet in de steekproef zitten, zowel wat betreft de gemiddelden als wat betreft de standaardafwijkingen. Deze steekproef kan zodoende als representatief gezien worden.

In eerdere analyses is bij robuustheidsonderzoek gevonden dat er weinig steekproeffluctuatie was bij additionele analyses op andere steekproeven die op vergelijkbare wijze getrokken waren. De kenmerken van de groepen op achtergrondvariabelen zijn volkomen vergelijkbaar met de totale groep. We mogen dus stellen dat deze steekproeven een representatief beeld geven van de Nederlandse populatie in jaargroep 8. De verdeling van de achtergrondvariabelen is gegeven in bijlage 6.

Eenzelfde werkwijze is uitgevoerd bij de groep leerlingen die de WO-toetsen gemaakt hebben. Op deze steekproef heeft wel nog een weging plaatsgevonden aangezien de groepen die wel en geen WO-toetsen maken van elkaar verschillen in vaardigheid voor wat betreft de gemeten vaardigheden in de CET-toetsen. De leerlingen die ook de WO-toetsen hebben gemaakt, maken gemiddeld ruim 3 opgaven⁵ meer goed dan de groep die geen WO-toetsen heeft gemaakt. De leerlingen die wel of geen WO-toetsen gemaakt hebben, verschillen ook wat betreft de verdeling van de achtergrondvariabelen. De groep die de WO-toetsen gemaakt heeft, is zodoende niet representatief voor de gehele onderzoekspopulatie, waardoor er bij de analyses gewogen wordt om de landelijke verdeling te verkrijgen. Dit was ook in de eerdere rapportage over de WO-vaardigheden het geval (Hemker & Van Weerden, 2015b). De getrokken steekproef van 20% uit de ruim 110 duizend leerlingen verschilt niet van de 80% die niet in de steekproef zitten.

De analyses van de effecten van de achtergrondvariabelen zijn gedaan op de vaardigheidsscores, zodat de gerapporteerde resultaten vergelijkbaar zijn met die van de eerdere rapportages. Naast de geschatte gemiddelden (en andere verdelingseigenschappen) van de latente variabelen is er bij de analyse gebruikgemaakt van het programma SAUL (*Structural Analysis of a Univariate Latent Variable*; Verhelst en Verstralen, 2002). Hiermee zijn de marginale effecten en de effecten berekend. In tegenstelling tot de eerdere rapporten betreft dit voor alle gevallen de ongecorrigeerde effecten. Meer hierover in paragraaf 2.3. Bij de WO-toetsen was het daarnaast ook mogelijk analyses direct op de scores uit te voeren omdat er per vaardigheid maar een toets is gebruikt bij alle leerlingen in dit onderzoek.

⁵ Gerelateerd aan het aantal opgaven (220) is dat 1,4% opgaven meer goed; gerelateerd aan de standaardafwijking van de totaalscores is dat ongeveer 10% van de standaardafwijking.

2.2.3 Achtergrondvariabelen

Behalve de genoemde taal-, reken- en WO-vaardigheden is ook een aantal achtergrondkenmerken in het onderzoek betrokken. Deze zijn te onderscheiden in leerlingkenmerken en schoolkenmerken. Deze lijst verschilt meer dan andere jaren van die in de vorige meting. Het aantal leerlingkenmerken is duidelijk kleiner geworden, terwijl het aantal gebruikte schoolkenmerken groter is dan voorheen. Dit jaar is een specifiek onderdeel van het onderzoek de vergelijking van de verschillende definities van achtergrondvariabelen. Verschillende definities zien we vooral bij de verschillende indelingen voor de variabele "stratum". Dat is een variabele die betrekking heeft op het percentage leerlingen op een school met een leerlinggewicht: in Nederland worden verschillende indelingen gehanteerd die in dit onderzoek met elkaar vergeleken worden: vier mogelijke opdelingen worden hier onderzocht.

Ook bij de variabelen leeftijd, regio en schoolgrootte is in dit onderzoek gewerkt met verschillende varianten. Bij deze variabelen betrof dat een verdere detaillering van de te onderscheiden niveaus.

Op basis van de resultaten in dit onderzoek zal waarschijnlijk bij volgende metingen een keuze gemaakt worden voor een specifieke variant.

Leerlingkenmerken:

- Geslacht: jongens-meisjes.
- Leertijd: variabele die gerelateerd is aan leeftijd van de leerling op 1 oktober 2015; twee varianten:
 - L2: tweedeling van leertijd: 1 = regulier/vervroegd (t/m 11 jaar); 2 = vertraagd (12 jaar en ouder)
 - L3: dreedeling van leertijd: 1 = regulier (11 jaar); 2 = vertraagd (12 en ouder); 3 = vervroegd (jonger dan 11)
- Advies VO: door leerkracht ingeschat niveau vervolgonderwijs voor de leerling; 8-deling
 - Vijf hoofd niveaus, en de drie⁶ tussenliggende combinatieniveaus: 1 = vmbo-basisberoeps (BB), 2 = BB/KB; 3 = vmbo-kaderberoeps (KB), 4 = vmbo-theoretische of gemengde leerweg (GT), 5 = GT/ha; 6 = havo, 7 = ha/vw; en 9 = vwo
- Gemaakte toets: CET-Basis of ankertoets; of CET-Niveau
- Ontheffing: Leerling krijgt wel⁷ of geen ontheffing

Behalve bij "gemaakte toets" en "ontheffing" zijn bij al deze variabelen ook ontbrekende waarden mogelijk.

Schoolkenmerken:

- Stratum: schoolindeling op basis van het percentage leerlingen met een leerlinggewicht (percentages tussen haakjes). Er zijn vier varianten van deze variabele opgenomen:
 - S3: 1 = weinig (<10%), 2 = matig (10-25%) en 3 = veel (>25%)
 - S5O: 1 = geen (0), 2 = weinig (>0% -<10%), 3 = matig(10-25%), 4 = veel (>25%-50%); 5 = zeer veel (>50%)
 - S5L: 1 = geen (0), 2 = weinig (>0% -<10%), 3 = matig(10-25%), 4 = veel (>25%-40%); 5 = zeer veel (>40%)
 - S4: 1 = weinig (<12%), 2 = matig (12-<25%), 3 = veel / 0.30 (25% of meer; meer 0.30 dan 1.2), 4 = veel / 1.20 (25% of meer gewichtenleerlingen; meer leerlingen met gewicht 1.20 dan met 0.30)
- Regio van de locatie van de school: twee varianten
 - R4: Vierdeling regio: noord, oost, west, zuid (indeling CBS);
 - R12: Verdeling op basis van provincie

⁶ De tussenliggende combinatie vmbo-KB en vmbo-GT (KB/GT) wordt niet gebruikt.

⁷ Zeer moeilijk lerende leerlingen; Meervoudig gehandicapte leerlingen voor wie het zeer moeilijk lerend zijn een van de handicaps is; Leerlingen die vier jaar of korter in Nederland zijn en om die reden de Nederlandse taal onvoldoende beheersen.

- Urbanisatiegraad van de locatie van de school: drie varianten
 - U5: Vijf groepen; van zeer sterk stedelijk tot en met niet stedelijk
 - G42: G4-tweedeling: 1 = Niet in een van de 4 grote steden; 2 = Wel in een van de 4 grote steden
- Schoolgrootte: gebaseerd op aantal leerlingen op de school: twee varianten
 - SG2: Tweedeling: 1= klein (tot en met 200 leerlingen); 2 = groot (meer dan 200 leerlingen)
 - SG5: Vijfdeling: 1 = 1-100; 2 = 101-200; 3 = 201-300; 4 = 301-400; 5 = meer dan 400 leerlingen
- Denominatie: Katholiek; Openbaar; Protestants-Christelijk; Algemeen Bijzonder; Overig/Onbekend
- Nieuwe school sinds 2015: 1 = eerder ook in 2013, 2014 of beiden; 2 = niet in 2013 of 2014
- WO-toetsen gedaan: Wel WO-toetsen gedaan; geen WO-toetsen gedaan

Bij geen van de variabelen zijn er ontbrekende waarden.

De leerlingkenmerken geslacht, leertijd (L2), advies VO en gemaakte toets waren bij eerdere metingen ook al opgenomen. Ook over het onderscheid bij leertijd tussen reguliere leerlingen en vervroegde leerlingen is eerder gerapporteerd (Hemker & Van Weerden, 2015a). Een ander belangrijk verschil met vorige metingen betreft het ontbreken van een tweetal leerlingkenmerken, te weten leerlingegewicht (formatiegewicht) en thuistaal. Zoals in paragraaf 1.3 is uitgelegd had vooral het ontbreken van leerlingegewicht ook (aanzienlijke) consequenties voor de analyses en de vergelijkbaarheid van de resultaten.

Er is een variabele die niet eerder opgenomen is geweest in dit onderzoek meegenomen, en dat is de variabele "nieuwe school". Bij deze variabele wordt onderscheid gemaakt tussen scholen die in een van de twee voorgaande jaren (2013, 2014 of beiden) de EB afgenomen hebben en de scholen die dat niet hebben gedaan. Dit laatste type scholen kan gezien worden als de groep "nieuwe scholen" die mogelijk door de verplichtstelling van een eindtoets de CET hebben afgenomen.

2.3 Analyses

Hieronder geven we beknopt weer hoe de resultaten van 2015 geanalyseerd zijn. In hoofdstuk 1 is al aangegeven dat de jaarvergelijking beperkingen kent. Voor een aantal subvaardigheden heeft deze toch plaatsgevonden. De vergelijkingen binnen het jaar 2015 zijn ook fundamenteel anders dan in de vorige metingen gebruikelijk was. De resultaten voor 2015 zijn daarmee ook niet te vergelijken met die in de eerdere rapportages. Naast de verandering van de CET, zowel praktisch als inhoudelijk, rechtvaardigt dit verschil de aanname dat deze meting daadwerkelijk als een nieuwe reeks te zien is.

2.3.1 Vergelijking tussen en over de jaren

Een voorwaarde om vaardigheden in de verschillende jaren direct met elkaar te kunnen vergelijken, is dat de vaardigheden op dezelfde schaal gemeten zijn. Aan deze voorwaarde is voldaan als de leerlingen in de steekproef van 2015 precies dezelfde toetsen hebben gemaakt als de leerlingen in eerdere steekproeven. Dat is bij de WO-toetsen het geval maar de CET-toetsen in 2015 zijn duidelijk anders dan de EB-toetsen in 2014 en daarvoor. Daar waar de EB-toetsen voor de metingen van jaargroep 8 al eerder ieder jaar volledig vernieuwd werden, waren de verandering van EB naar CET in 2015 nog groter (zie hoofdstuk 1).

Een directe vergelijking tussen 2014 en 2015 is daarmee onmogelijk geworden. Voor een aantal subschalen in de EB en de CET hebben we echter enige vergelijkbaarheid kunnen realiseren. Het betreft de twee taalvaardigheden Begrijpend Lezen (TB; niet te verwarren met referentieschaal Lezen LEZ waar TB slechts een onderdeel van uitmaakt) en Woordenschat (TW), en drie van de vier rekenschalen. Door middel van de ankertoets, waarin opgaven die horen bij deze schalen zowel in 2015 als 2014 zijn afgenomen, is het mogelijk een relatie te leggen zodat de opgaven op een schaal te plaatsen zijn. Deze CET-anker heeft een zeer grote overlap met de reguliere CET (CET-basis; of ook wel CET-B). Deze CET-basis heeft ook opgaven die overlappen met de CET-Niveau (CET-N) waardoor de vaardigheden van deze toetsen binnen een jaar, maar ook over de jaren met elkaar te vergelijken zijn.

Met behulp van een speciale analysetechniek met bijbehorend programma: het *One Parameter Logistic Model* (OPLM; Verhelst, 1993; Verhelst en Glas, 1995), een variant van het item response model, zijn schalen geconstrueerd voor alle afzonderlijke onderwerpen. Dit is al in 2008 gedaan, waarbij de toetsen alle op dezelfde schaal werden gezet, met een gemiddelde van 250 punten en een standaardafwijking van 50. Dit is vergelijkbaar met de aanpak die bij PPOON werd gehanteerd. Op deze wijze kunnen de prestaties op de verschillende toetsen over de jaren heen met elkaar worden vergeleken (vgl. Hemker, Kordes & Van Weerden, 2011).

Voor het maken van de schalen is de representativiteit van de steekproef van minder groot belang, maar voor het vergelijken van prestaties in jaargroep 8 over de jaren heen moet bij de samenstelling van de responsgroep wel degelijk rekening gehouden worden met de representativiteit. Als verschillende jaren van elkaar verschillen in samenstelling, moet onderzocht worden waar die verandering vandaan zou kunnen komen. Veranderingen in de samenstelling van een responsgroep kunnen het gevolg zijn van zogenaamde steekproeffluctuaties, maar ook van 'echte' veranderingen in de samenstelling van de populatie door de tijd.

Het probleem van steekproeffluctuaties zal zich voor jaargroep 8 niet zo gauw voordoen. Er is namelijk in elk jaar op vergelijkbare wijze een zeer grote aselechte steekproef uit een bestand van vele honderdduizenden leerlingen getrokken (telkens ongeveer 85% van de populatie). Wel kan er zich een wijziging in de samenstelling van de populatie hebben voorgedaan. Het gevolg zou kunnen zijn dat er op basis van de analyses die daar geen rekening mee houden onjuiste conclusies getrokken zouden kunnen worden. Aan de andere kant zou het onterecht corrigeren van de gegevens ook tot onjuiste conclusies kunnen leiden, bijvoorbeeld als de eigenschappen van de groepen door de tijd veranderen zonder dat het label van die groep verandert. Bijvoorbeeld, een werkelijke toename in vaardigheid in de populatie zou tot een toename van vwo-leerlingen kunnen leiden. Als nu voor deze toename gecorrigeerd zou worden, wordt in het model de toename in vaardigheid "weg gecorrigeerd". Om die reden is ervoor gekozen om voor de jaarvergelijkingen de ongecorrigeerde resultaten te gebruiken. Bij de conclusies moet dan wel gekeken worden of er zich grote wijzigingen voordoen in de achtergrondvariabelen, die wellicht een nieuwe kijk geven op de gevonden resultaten. De berekeningen worden uitgevoerd met het programma SAUL (Verhelst & Verstralen, 2002).

2.3.2 Analyses voor de leerlingen in 2015

Een model met ongecorrigeerde effecten

Voor het vergelijken van de resultaten over de jaren heen wordt uiteraard gebruikgemaakt van data waarin alle jaren vertegenwoordigd zijn. Voor de analyses van de effecten van de achtergrondgegevens in 2015 wordt alleen gebruikgemaakt van gegevens uit 2015. In het verleden is bij de analyses binnen een jaar altijd gebruikgemaakt van een gecorrigeerd model. In dat model waren altijd variabelen geslacht, leertijd (L2), stratum (S3) en leerlingewicht opgenomen. Dit model was het basismodel en werd voor de overige analyses altijd aangevuld met een extra achtergrondvariabele.

Op basis van het gecorrigeerde model kon nagegaan worden wat het additioneel effect was van een variabele. Dat betekent dat voor een leerling gesteld werd dat, als het geslacht, de leertijd, stratum en leerlingewicht gelijk waren, het toegevoegd effect van de extra variabele kon worden onderzocht. Als bijvoorbeeld meer leerlingen met een hoog leerlingewicht in het westen van het land wonen, kon onderzocht worden of verschillen tussen het westen van het land en andere delen nu veroorzaakt werden door de locatie op zich (kwaliteit van de scholen in het westen), of door de samenstelling van de groep leerlingen die daar woont. Ook konden zo de additionele effecten onderzocht worden van het gegeven of er veel of weinig leerlingen met een leerlingewicht op een school zitten, los van het specifieke gewicht van de leerling zelf: doet een leerling met leerlingewicht 1.20 het anders op een school met veel of met weinig andere gewichtenleerlingen?

Aangezien het lastig te voorzien is of, dan wel wanneer, andere achtergrondvariabelen niet meer opgevraagd worden, is ervoor gekozen in dit verslag geen gecorrigeerde modellen te gebruiken zodat ongeacht het wegvallen van een of meer variabelen de resultaten vergelijkbaar blijven over de jaren heen. Er is dus in dit verslag gewerkt met modellen waarin telkens alleen de gerapporteerde variabele is opgenomen. De gerapporteerde effecten zijn zodoende ongecorrigeerde effecten.

Significantie en effectgrootte

De (ongecorrigeerde) effecten waarover gesproken wordt, zijn relevant om de verschillen tussen groepen te kunnen interpreteren. Of een gemiddeld vaardigheidsverschil tussen twee jaren statistische significantie oplevert, hangt in belangrijke mate af van de steekproefgrootte. Hoe groter de steekproef, hoe eerder een verschil statistisch significant is. De steekproef voor jaargroep 8 is dusdanig groot dat verschillen tussen groepen al heel snel significant zijn. Om de resultaten toch zinvol met elkaar te kunnen vergelijken rapporteren wij behalve de statistische significantie ook de zogeheten effectgrootte. De effectgrootte wordt in ons geval berekend als het verschil tussen de gemiddelden van subgroepen (bijvoorbeeld jongens en meisjes) gedeeld door de (gepoolde) standaardafwijking van de twee groepen die onderling worden vergeleken. Bij de interpretatie van de effectgrootte hanteren we de vuistregel van Cohen (1988) die is afgebeeld in tabel 2.1. Alles met een effectgrootte boven de 0,20 noemen we hier betekenisvol.

Tabel 2.2 Kwalificatie van effectgrootten

Effectgrootte (absolute waarde)	Kwalificatie
0,00 tot 0,20	geen effect
0,20 tot 0,50	klein effect
0,50 tot 0,80	matig effect
0,80 of groter	groot effect

2.3.3 Gemiddelden op nieuwe vaardigheden

Behalve natuurlijk de noodzakelijke verandering van het model maakt ook de verandering van de gemeten vaardigheden de vergelijking met de eerdere peilingen niet goed mogelijk. In 2015 is de aandacht niet zozeer gericht op de subschalen voor Taal en Rekenen, maar is men vooral geïnteresseerd in de referentieniveauvaardigheden. De schalen zijn zodoende representaties van deze vaardigheden, die weliswaar overeenkomsten hebben met, maar niet volledig gelijk zijn aan, in eerdere jaren gerapporteerde vaardigheden. De schalen waar nu over gerapporteerd wordt zijn Lezen (LEZ; met daarin deels TB), Taalverzorging (TAV; met daarin deels spelling) en Rekenen (REK, met daarin deels RG, RB en RM). Daarnaast zijn er opgaven voor "schrijven van teksten" en "woordenschat" (TW). De eerste was eerder al geen onderdeel van de jaarlijkse metingen maar TW was dat wel. Deze schaal zal ook bij deze meting weer worden meegenomen.

Behalve de effecten zijn ook de gegevens van de gemiddelden van de te onderscheiden groepen vergeleken op de vaardigheidsschalen. Deze zijn direct vergelijkbaar met de ongecorrigeerde effecten. In eerdere rapportages lagen de vaardigheden van de verschillende jaren op dezelfde schaal, maar doordat er nu gewerkt wordt met nieuwe vaardigheden is dat niet het geval.

Hierbij is er wel een onderscheid bij de jaarvergelijking: bij een aantal vaardigheden wordt nog met de oude schalen gewerkt en bij andere met de nieuwe schalen. Bij de jaarvergelijking voor TB, TW, RG, RB, RM en de WO-vaardigheden WN, WA en WG wordt de oude schaal gebruikt. Bij de nieuwe schalen LEZ, TAV en REK wordt een nieuwe schaal gebruikt.

De gekozen rapportageschaal

Elke vaardigheid in dit onderzoek is getransformeerd naar een schaal met een gemiddelde van 250 en een standaarddeviatie van 50. Dit is conform de werkwijze bij PPON (zie bijv. Janssen, Van der Schoot & Hemker, 2005). Bij de schalen waar een jaarvergelijking mogelijk is, is de startwaarde voor elke schaal het gemiddelde dat we in 2008 hebben aangetroffen. Dat gemiddelde is op 250 gesteld (zie verder Hemker & Van Weerden, 2009). De transformatie heeft als voordeel dat we de prestaties voor verschillende vaardigheden en voor verschillende jaren naast elkaar kunnen zetten op een en dezelfde schaal.

Voor de referentieniveauvaardigheden LEZ, TAV en REK zijn er geen gegevens van 2008 bekend. Dat betekent dat de schaal opnieuw gedefinieerd moet worden. Ook hier is ervoor gekozen om deze op 250 te zetten met een standaardafwijking van 50. Er was bij deze vaardigheden toch ook al geen jaarvergelijking mogelijk. Bij de presentatie van de resultaten binnen 2015 is dat dan ook geen probleem.

Voor de vaardigheden die in het verleden ook gemeten zijn en waarvoor ook in het rapport van 2015 resultaten worden gepresenteerd, te weten TW en de WO-vaardigheden, is de vraag of de oude schaal gebruikt moet worden of niet. Voor TW zou dat de schaal van 2008 zijn, terwijl voor WN, WA en WG die van 2013 gebruikt zou worden. Er is in dit verslag voor gekozen om wanneer de resultaten over jaren heen vergeleken worden de oude schalen te gebruiken voor TW en de WO-vaardigheden. Bij de vergelijkingen binnen 2015 wordt echter alleen de nieuwe schaal gebruikt die gebaseerd is op de gegevens van 2015. Dit zal ook duidelijk in de tabellen aangegeven worden.

In het verleden werden de prestaties ook gedefinieerd in termen van typische leerlingen (zie paragraaf 1.4). Dat zal in dit verslag voor de schalen die sinds 2008 over de jaren heen gevolgd kunnen worden voor de laatste keer gerapporteerd worden, omdat die schalen in de toekomst niet meer gebruikt worden. Voor het online katern '*Taal en rekenen aan het einde van de basisschool. Peil. onderwijs*' (Inspectie van het Onderwijs, 2016b) is op de schalen Lezen (LEZ), Taalverzorging (TAV) en Rekenen (REK) onderzoek gedaan naar de leerlingen die een vaardigheid hebben die precies op het 1F-niveau zit en de leerlingen met een vaardigheid precies op het niveau daarboven (REK 1S, LEZ 2F, TAV 2F).

Wegingen voor de wereldoriëntatieschalen

De leerlingen die de WO-toetsen maakten, presteerden beter op taal en rekenen dan de leerlingen die deze toetsen niet maakten. Het is zodoende noodzakelijk om tot een landelijke vaardigheidsverdeling te komen door de onderzoekspopulatie voor deze (zelf)selectie te corrigeren. In eerder onderzoek (Hemker & Van Weerden, 2015b) zijn verschillende wegingsmethodes met elkaar vergeleken. Vanwege de vergelijkbaarheid van de resultaten, het feit dat de toetsen onveranderd zijn, en dat de weging op basis van stratum (S3) in de eerdere analyses de beste resultaten opleverde, in die zin dat dit de meest eerlijke weging was, is ook bij de analyses in dit rapport voor de weging op basis van stratum gekozen. In de meting was stratum 1 oververtegenwoordigd bij de leerlingen die de WO-toetsen gemaakt hadden. De stratum 1-leerlingen behaalden ook de betere resultaten. Om nu de vaardigheden in de gehele onderzoekspopulatie te kunnen bepalen is gewogen naar de percentages van de strata in de gehele populatie.

2.3.4 Referentieniveaus

De referentieniveauvaardigheden LEZ, TAV en REK zijn naast een weergave op een vaardigheidsschaal ook weer te geven in termen van referentieniveaus. In de toetswijzer PO spelen de referentiesets een zeer belangrijke rol aangezien er ook wordt gerapporteerd op de referentieniveaus, zij het nu vooralsnog in een pilot-opzet. Dat de referentieniveaus (mede) leidend zijn heeft ook gevolgen voor de gemeten vaardigheden en de behaalde scores.

In een rapportage van het CvTE met betrekking tot de terugblik (CvTE, 2016) worden enige resultaten gegeven van de prestaties gerelateerd aan deze niveaus. De onderzoekspopulatie in die terugblik verschilt

iets van die van het hier beschreven onderzoek. Zo zijn de resultaten van de digitale CET onderdeel van het onderzoek in de terugblik, en zijn er ook resultaten gegeven voor het speciaal basisonderwijs. Op basis van die verschillen zijn er wel enige verschillen tussen de resultaten van de twee onderzoeken, maar deze verschillen zijn niet erg groot.

Voor de vaardigheden waarvan referentieniveaus bekend zijn (LEZ, TAV en REK), zijn de gevonden referentieniveaus ook te koppelen aan de hier gerapporteerde vaardigheidsniveaus. Dat zijn de vaardigheidsniveaus op de schaal met een gemiddelde van 250 en een standaardafwijking van 50. Dit onderzoek is verder beschreven in publicaties van de Inspectie van het Onderwijs (2016a, 2016b).

2.3.5 Samenvattend: een nieuwe reeks

In vergelijking met de eerder geschreven rapporten over de meting Taal en Rekenen zijn de aanpassingen in dit rapport zodanig dat we met recht kunnen spreken over een nieuwe reeks. Een directe vergelijking van de resultaten met de eerdere rapportages wordt sterk bemoeilijkt door de verschillen. Voor een groot deel ligt dat in het feit dat er een aantal belangrijke vaardigheden anders gemeten wordt. De jaarvergelijking wordt daarnaast bemoeilijkt door de verschillen in afnamemoment en afnamecondities. De effecten van de achtergrondvariabelen binnen een jaar zijn weer lastig te vergelijken door het ontbreken van de variabele leerlinggewicht.

3 De vergelijking over de jaren

Zoals in de inleiding al is aangegeven is de vergelijking over de jaren voor de vaardigheden een uitdaging. Inhoudelijk zijn de toetsen aanzienlijk veranderd, wat de vergelijkbaarheid bemoeilijkt. Het is wel mogelijk een deel van de oude schalen uit de EB terug te vinden in de nieuwe CET. Voor de toetsen WO is het nog gemakkelijker: die zijn onveranderd. De grootste uitdaging ligt echter in de interpretatie van de resultaten gegeven de eerder besproken verschillen tussen de EB en de CET.

In tegenstelling tot de analyses binnen een afnamejaar zijn de jaarvergelijkingen ook in het verleden altijd gebaseerd op de ongecorrigeerde effecten. De rationale daarachter is dat een verandering in de samenstelling van de populatie tussen twee jaren ook een echte verbetering in vaardigheid kan representeren die bij een correctie op achtergrondvariabelen gemist zou kunnen worden (zie ook Hemker, Kuhlemeier & Van Weerden, 2010). Een belangrijke reden hiervoor is dat de steekproef dusdanig groot is dat de verschuivingen in achtergrondvariabelen werkelijke veranderingen representeren. De gevolgen van de scholen die erbij gekomen zijn omdat een eindtoets nu verplicht is, en af zijn gevallen omdat er alternatieve eindtoetsen beschikbaar zijn, lijken minimaal⁸.

In tabel 3.1 zijn de uitkomsten voor de acht opeenvolgende jaren weergegeven als gemiddelde op de vaste schaal. Het gemiddelde voor 2008 is per definitie 250 en de standaarddeviatie is 50. De schaalwaarde van 250 is vastgelegd in 2008 en vormt het startpunt waarmee we de uitkomsten van 2009 en later vergelijken (zie verder Hemker & Van Weerden, 2009).

Tabel 3.1 Gemiddelde schaalwaarden jaargroep 8*

Vaardigheden	2008	2009	2010	2011	2012	2013	2014	2015
T: Woordenschat	250	249	257	251	252	254	253	253
T: Begrijpend lezen (anker)	250	252	254	257	253	253	252	258
RW: Getallen en bewerkingen	250	249	252	252	255	252	252	254
RW: Breuken, procenten en verhoudingen	250	250	254	254	257	254	254	255
RW: Meten, meetkunde, tijd en geld	250	250	254	254	256	255	254	256
WO: Natuuronderwijs	250	251	254
WO: Aardrijkskunde	250	249	250
WO: Geschiedenis	250	251	252

*Merk op dat er in tabel 3.1 kleine verschillen met eerder gerapporteerde tabellen kunnen ontstaan. Dit is te wijten aan de herberekeningen die effect kunnen hebben op afrondingen. Deze hebben geen effect op de eerder getrokken conclusies. De verschillen in de controledata waarin meer (mogelijk speciaal onderwijs) scholen zaten, zijn zo klein dat deze niet tot een andere duiding van tabel 3.1 zouden leiden.

Vergelijken we de uitkomsten voor 2015 met 2014 in tabel 3.1 (gehele getallen) dan zien we dat de (afgeronde) schaalwaarde in 2015 voor geen enkele vaardigheid lager is dan in 2014. Alleen voor Woordenschat is de geobserveerde waarde in 2015 in decimalen 0,1 punt lager dan in 2014. De grootste stijging die gevonden wordt, is bij Begrijpend lezen (5,9 schaalpunten hoger in 2015 dan in 2014). In vergelijking met eerdere peilingen zaten er nu minder opgaven in de ankertoets, wat mogelijk een effect

⁸ In beide gevallen zijn er scholen met relatief laagvaardige leerlingen bijgekomen en afgevallen; hierover meer in hoofdstuk 4.

zou kunnen hebben. Het gemiddelde verschil was twee punten op de hier gegeven vaardigheidsschalen. Standaarddeviaties variëren van 49 tot 53; in 2008 zijn deze per definitie gelijk gesteld aan 50.

In de eerdere peilingen was van 2008 tot en met 2014 voor alle vaardigheden over het algemeen een licht stijgende trend waargenomen. De trends die beschreven zijn voor de gemiddelden kunnen ook worden beschreven in termen van effectschattingen. De verschillen van opvolgende jaren staan als effectgrootten in tabel 3.2. Bovendien geven we de effectgrootte voor het contrast 2015-2008, dat wil zeggen het verschil in uitkomst van de meting in 2015 met die van de eerste meting in 2008, dus over de afgelopen 7 jaar. Als het verschil tussen twee jaren statistisch significant is, is het effect vetgedrukt.

Tabel 3.2 Effectgroottes van de jaarverschillen jaargroep 8*

Vaardigheden	2009 -2008	2010 -2009	2011 -2010	2012 -2011	2013 -2012	2014 -2013	2015 -2014	2015 -2008
T: Woordenschat	-0,02	0,15	-0,11	0,02	0,04	-0,03	0,00	0,06
T: Begrijpend lezen (anker)	0,04	0,04	0,07	-0,09	0,00	-0,01	0,11	0,16
RW: Getallen en bewerkingen	-0,01	0,05	0,00	0,07	-0,06	-0,01	0,05	0,09
RW: Breuken, procenten en verhoudingen	-0,01	0,08	0,01	0,06	-0,07	0,01	0,03	0,11
RW: Meten, meetkunde, tijd en geld	0,01	0,07	0,00	0,05	-0,04	-0,01	0,04	0,12
						2014 -2013	2015 -2014	2015 -2013
WO: Natuuronderwijs	0,01	0,06	0,08
WO: Aardrijkskunde	-0,01	0,01	0,00
WO: Geschiedenis	0,03	0,01	0,04

*Vet = significant verschil met voorgaand jaar of, in laatste kolom, met 2008 ($\alpha < 0,01$); de gekleurde cellen geven de effecten met het jaartal 2015; de lichte kleuren geven het effect ten opzichte van het afgelopen jaar; de donkere kleuren met de eerst gerapporteerde peiling.

Bij de meeste schalen zie we significante positieve effecten. De interpretatie van het verschil tussen 2015 en de andere jaren is zoals gezegd lastiger dan in de voorgaande jaren. Behalve een daadwerkelijk toegenomen vaardigheid van 2014 naar 2015 kan het positieve jaareffect ook veroorzaakt worden door het latere afnamemoment. Hieruit kan overigens niet de conclusie getrokken worden dat de leerlingen door dit latere afnamemoment vaardiger geworden zijn. Het zou kunnen zijn dat in eerdere jaren deze leerlingen ook deze vaardigheid hadden in april. Het feit dat de effecten positief zijn, doet vermoeden dat de impact van het veranderde doel van de CET ten opzichte van de EB niet dusdanig is dat dit een grote invloed heeft op de prestaties. Dat het resultaat op de CET voor een deel van de leerlingen geen of amper invloed heeft op het definitieve doorstroomadvies, lijkt er vooralsnog niet toe te leiden dat de leerlingen als gehele groep minder presteren – het weegt in ieder geval niet op tegen het latere meetmoment. Of dat ook het geval is in latere jaren is daarmee nog niet gezegd. Het kan zijn dat het effect van het net iets andere doel van de eindtoets pas op langere termijn een effect heeft omdat de toetsen vooralsnog gezien worden als de EB. Het is bekend dat de ervaren gevolgen van de belangen van de toets wel aanzienlijke impact kunnen hebben op de geleverde prestatie, onafhankelijk van de vaardigheid (Hemker, 2012). Dat is een resultaat dat ook in ander internationaal onderzoek gevonden is (Keskpaik & Rocher, 2013; Van Barneveld, Pharand, Ruberto & Haggarty, 2013).

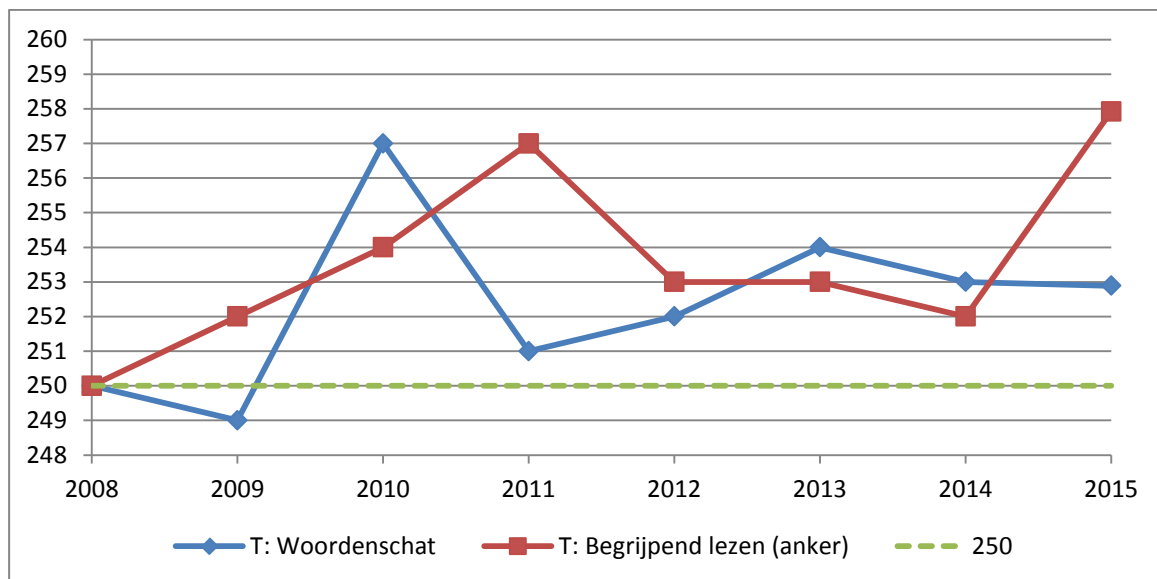
In de volgende paragrafen gaan we iets dieper in op de drie groepen van vaardigheden: Taal, Rekenen en WO.

3.1 Taalvaardigheden

De ontwikkeling in schaalscores is afzonderlijk in beeld gebracht voor de twee taalvaardigheden die we konden volgen vanaf 2008 tot en met 2015 (figuur 3.1). We zien dat de trendlijnen redelijk van elkaar verschillen. Ook valt het bij de taalvaardigheden op dat de resultaten van jaar tot jaar grote verschillen op kunnen leveren. Bij de stap van 2014 naar 2015 is daar een aantal duidelijke verklaringen te geven, maar als de gehele periode in ogenschouw genomen wordt, is te zien dat ook in eerdere jaren soms plotselinge stijgingen en dalingen zijn waargenomen. Bij kleine ankers, zoals bij Woordenschat in het verleden is dat iets aannemelijker, maar nu zien we bij een iets kleiner anker Begrijpend lezen ook een relatief groot verschil. Overigens is de grootte van het anker bij Woordenschat de afgelopen jaren niet aangepast terwijl de resultaten daar redelijk stabiliseren. Daarnaast is het anker bij Begrijpend lezen ook nog steeds groter dan bij Woordenschat, dus de grootte van het anker is zodoende zeker niet de belangrijkste verklaring van de verschillen.

Een verklaring voor de stijging van de vaardigheid bij Begrijpend lezen van de EB naar de CET zou kunnen liggen in de inrichting van de lessen. Als in de lessen meer aan Begrijpend lezen gedaan wordt dan aan Woordenschat zou een verlengde lesperiode ook meer effect hebben op de vaardigheid begrijpend lezen dan op Woordenschat. Er zijn aanwijzingen, waarbij gekeken is naar de invloed van thuistaal op de verschillende vaardigheden (Hemker en Van Weerden, 2015), dat er bij Woordenschat minder dan bij andere vaardigheden invloed is van het schoolse leren en relatief veel van het buitenschoolse leren.

Figuur 3.1 Trends over de jaren voor de taalonderwerpen in jaargroep 8



In eerdere rapportages zijn ook de percentages per leerlinggroep aangegeven. Deze dienden bij de eerdere rapportages, toen de referentieniveaus nog niet bepaald waren, als alternatieve indicator van de gevonden trends. In dit rapport worden deze nu voor de laatste keer gegeven voor de vaardigheden die we vanaf 2008 hebben kunnen volgen.

Deze gegevens komen overeen met de trends zoals gegeven door de weergaven van de gemiddelden en de effecten. In het geval van Woordenschat waarbij het gevonden verschil tussen 2014 en 2015 geheel te verwaarlozen is, is de verdeling van de leerlingen in 2015 ook gelijk aan die in 2014. Het percentage leerlingen dat het niveau van een zeer zwakke leerling (definitie 2008; zie Hemker en Van Weerden, 2009) behaalt is in beide jaren 91%. Het percentage leerlingen dat het niveau van een zeer vaardige leerling

behaalt is in beide jaren 11%. Deze percentages zijn in de eerdere peilingen gerelateerd aan de opgaven die de leerlingen waren voorgelegd (Hemker en Van Weerden, 2009; Hemker et al., 2010). Aan deze opgaven waren echter geen standaarden opgelegd. Dat is veranderd met de komst van de referentieniveaus. De resultaten die gevonden worden bij het onderzoek naar de referentieniveaus worden elders gegeven, te weten in publicaties van het College voor Toetsen en Examens (2016) en van de Inspectie van het Onderwijs (2016b). De tabellen zoals gegeven in tabel 3.3 en tabel 3.4 hebben daardoor ook aan belang ingeboet, en worden hier alleen nog gegeven voor de vergelijkbaarheid met de eerdere rapportages.

Tabel 3.3 *Vergelijking over de jaren heen in percentage per leerlingengroep voor taalvaardigheden in jaargroep 8*

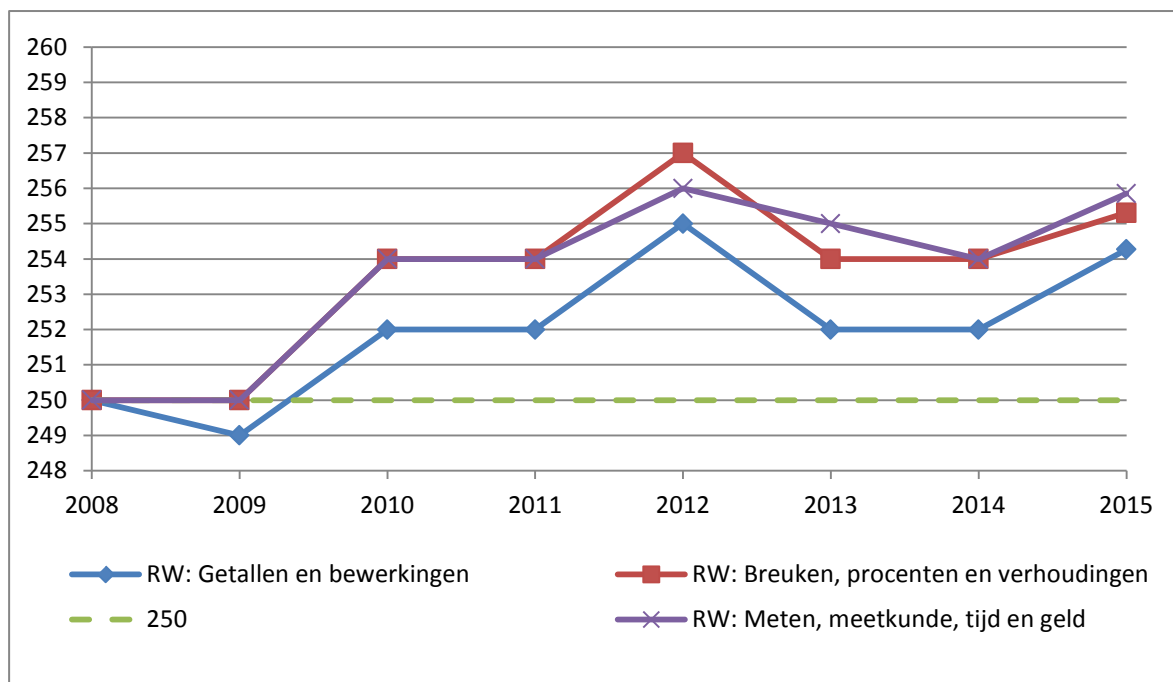
Leerlinggroep*	Percentielgrens	Woordenschat: % leerlingen minstens op niveau leerlinggroep							
		2008	2009	2010	2011	2012	2013	2014	2015
> dan ZLV	P10	90	90	92	90	91	91	91	91
> LV	P25	75	74	79	76	76	78	77	77
> Standaard 2008	P50	50	49	55	51	52	53	52	52
> HV	P75	25	24	29	26	26	28	27	27
> ZHV	P90	10	10	13	10	11	12	11	11

Leerlinggroep*	Percentielgrens	Begrijpend lezen: % leerlingen minstens op niveau leerlinggroep							
		2008	2009	2010	2011	2012	2013	2014	2015
> dan ZLV	P10	90	91	91	92	91	91	91	93
> LV	P25	75	76	77	79	77	77	77	80
> Standaard 2008	P50	50	52	53	56	52	52	52	56
> HV	P75	25	26	27	30	27	27	27	30
> ZHV	P90	10	11	11	13	11	11	11	13

3.2 Rekenvaardigheden

Bij de jaarvergelijking voor Rekenen wordt gerapporteerd over drie subschalen. Anders dan bij de taalvaardigheden is bij de rekenvaardigheden de samenhang tussen de losse rekenvaardigheidsschalen groter. Over het algemeen zijn de jaartrends – zeker in richting – vergelijkbaar: de vaardigheden stijgen en dalen samen over de tijd. In eerdere rapportages zijn daarom ook resultaten voor een totaalschaal voor Rekenen gegeven.

Figuur 3.2 Trends over de jaren voor de rekenonderwerpen in jaargroep 8



Om verwarring te voorkomen met de schaal Rekenen zoals gedefinieerd voor de referentieniveaus (REK), die ook opgaven bevatten over de rekenvaardigheid “verbanden” die eerder niet opgenomen was in de EB, is ervoor gekozen die oude samenvattende rekenschaal nu niet te rapporteren, maar alleen de subschalen weer te geven. Bij de resultaten van alleen 2015 en de effecten van de achtergrondvariabelen (hoofdstuk 4) wordt gerapporteerd over de nieuwe rekenschaal waaraan de referentieniveaus zijn gerelateerd.

Bij de vergelijking van 2014 met 2015 is te zien dat alle drie de rekenvaardigheden stijgen, en ook in ongeveer dezelfde mate (effectgrootte van gemiddeld 0,04). De verschillen hierin zijn miniem. Overigens kan opgemerkt worden dat de gevonden stijgingen niet de grootste waren die de afgelopen periodes gevonden zijn. De stijgingen van 2009 naar 2010 en van 2011 naar 2012 waren groter. Het is ook moeilijk aan te geven wat de invloed van de praktische verschillen tussen de EB (2014) en de CET (2015) is geweest. Het zou kunnen zijn dat de stijging nog groter was geweest als de eindtoets praktisch en inhoudelijk onveranderd was geweest, maar dat is op basis van deze gegevens niet te achterhalen.

De interpretatie van tabel 3.4 is vergelijkbaar met die van tabel 3.3 bij de taalvaardigheden. Ook de opmerkingen die hierover gemaakt zijn in paragraaf 3.1 zijn vergelijkbaar.

Tabel 3.4 Vergelijking over de jaren heen in percentage per leerlingengroep voor de rekenonderwerpen in jaargroep 8

Leerlingengroep*	Percentielgrens	Getallen en bewerkingen: % leerlingen							
		2008	2009	2010	2011	2012	2013	2014	2015
> dan ZLV	P10	90	90	91	91	92	91	91	91
> LV	P25	75	74	76	76	78	76	76	78
> Standaard 2008	P50	50	49	52	51	54	52	52	53
> HV	P75	25	24	26	26	29	26	26	28
> ZHV	P90	10	10	11	11	12	11	11	12

Leerlingengroep*	Percentielgrens	Breuken, procenten en verhoudingen: % leerlingen							
		2008	2009	2010	2011	2012	2013	2014	2015
> dan ZLV	P10	90	90	91	91	92	91	91	92
> LV	P25	75	75	77	78	79	77	77	78
> Standaard 2008	P50	50	50	53	53	56	53	53	54
> HV	P75	25	25	27	28	30	27	28	28
> ZHV	P90	10	10	11	12	13	11	11	12

Leerlingengroep*	Percentielgrens	Meten, meetkunde, tijd en geld: % leerlingen							
		2008	2009	2010	2011	2012	2013	2014	2015
> dan ZLV	P10	90	90	91	91	92	92	91	92
> LV	P25	75	75	77	77	79	78	78	79
> Standaard 2008	P50	50	50	53	53	55	54	53	55
> HV	P75	25	25	27	28	29	28	28	29
> ZHV	P90	10	10	11	11	12	12	12	12

3.3 Wereldoriëntatievaardigheden

De resultaten voor WO worden pas gevolgd sinds 2013. Aangezien de toetsen voor WO onveranderd zijn, is het mogelijk om, in tegenstelling tot de taal- en rekenvaardigheden, bij de WO-vaardigheden wel rechtstreeks de aantallen goed op de drie toetsen te vergelijken. Een ander verschil met de taal- en rekenvaardigheden is dat de leerlingen die deze toetsen maken niet geheel representatief zijn voor de populatie. De verdeling van de achtergrondvariabelen verschilt enigszins voor de twee groepen. Zo zijn de VO-adviesniveaus hoger bij de leerlingen die de WO-toetsen maken dan bij de leerlingen die de toetsen niet maken, en ook voor wat betreft de verdeling over de strata zijn er verschillen tussen de twee groepen. De groep leerlingen die de WO toetsen maakte, scoorde ook op de EB- en de CET-toetsen iets hoger dan de groep die de WO-toetsen niet maakt.

Met deze verschillen is rekening gehouden door te wegen voor deze verschillen zodat het mogelijk is landelijk representatieve scores te verkrijgen. Een beschrijving van die werkwijze en hoe tot die werkwijze gekomen is, is beschreven in de eerdere rapportage over de metingen bij WO (Hemker & Van Weerden, 2015). Deze werkwijze is hier nu herhaald, met dezelfde definitie voor stratum (S3; zie hoofdstuk 2). De verschillen met de overige definities zijn echter miniem.

Tabel 3.5 Gemiddelde score per toets, ongewogen en gewogen

Scores	ongewogen			gewogen: Advies VO			gewogen: Stratum		
	2013	2014	2015	2013	2014	2015	2013	2014	2015
Toetsonderdeel									
Natuuronderwijs	19,84	19,86	20,20	19,72	19,76	20,14	19,76	19,78	20,11
Aardrijkskunde	20,03	20,01	20,21	19,91	19,91	20,14	19,95	19,94	20,13
Geschiedenis	21,09	21,23	21,39	20,97	21,13	21,32	21,02	21,16	21,31

Net zoals bij de vaardigheidsschattingen en de effectschattingen is te zien dat de scores in 2015 hoger liggen dan in de voorgaande jaren. De verschillen tussen 2015 en 2014 zijn bij Geschiedenis vergelijkbaar met die van 2013 en 2014, maar zijn bij de andere twee vaardigheden een stuk groter.

Er is gecontroleerd of deze verschillen door specifieke opgaven veroorzaakt werden of dat deze verschillen voor alle opgaven ongeveer gelijk waren. Als de verschillen voor een enkele opgave, of een klein aantal opgaven, veel groter zijn dan is die opgave wellicht niet meer representatief voor de gehele vaardigheid. Die opgave kan bijvoorbeeld bekend zijn geraakt bij een deel van de leerlingen waardoor die opgaven op zich niet meer (alleen) als representatieve maat voor de vaardigheid functioneert. Dit onderzoek naar vraagpartijdigheid, dat ook bekend is als “item bias” of “differential item functioning” (DIF), hoort zeker ook bij systeemonderzoek bij de analyses die gedaan worden voor schaalconstructie. In het geval van de WO toetsen was er sprake van een drietal opgaven waarbij de vooruitgang uitzonderlijk was. De p-waarden in de drie jaren zijn gegeven in tabel 3.6.

Tabel 3.6 P-waarden van opgaven met een uitzonderlijk grote groei vergeleken met de overige opgaven

P-waarden		Jaar		
Toetsonderdeel	Itemnummer	2013	2014	2015
Natuuronderwijs	19	0,67	0,67	0,76
Aardrijkskunde	9	0,39	0,38	0,50
Geschiedenis	25	0,61	0,63	0,70

Op individueel niveau maken deze verschillen weinig uit, maar voor systeemevaluatie waarbij kleine verschillen tussen jaren wel indicatief zijn, kunnen ze verstorend werken bij de interpretatie van resultaten. Bij analyses op systeemniveau, zoals het PISA-onderzoek, is het gebruikelijk deze opgaven te verwijderen uit de vaardigheidsschalen. Als het aannemelijk is dat deze opgaven anders functioneren om redenen die niet direct met de vaardigheid te maken hebben, kan dat een acceptabele procedure zijn.

Het al dan niet opnemen van deze drie opgaven heeft wel enig effect op gevonden verschillen tussen de resultaten van 2015 en de twee andere afnamejaren. Er is weliswaar nog steeds sprake van een toename in scores (zie tabel 3.5), maar deze is wel kleiner dan bij de vaardigheidsschaal mét deze opgaven. In de rapportage over de vaardigheidsschaal en de effectgroottes is ervoor gekozen het onderzoek te rapporteren zonder deze drie opgaven. Deze resultaten zijn gegeven in tabel 3.7 (gelijk aan de resultaten van tabel 3.1 en 3.2). Als de drie opgaven wel meegenomen zouden zijn, dan zouden de vaardigheidsschattingen van 2015 gemiddeld 1 punt hoger uitvallen en de effectschattingen gemiddeld 0,02 punten hoger.

Tabel 3.7 Resultaten voor de jaarvergelijking op de WO-vaardigheden

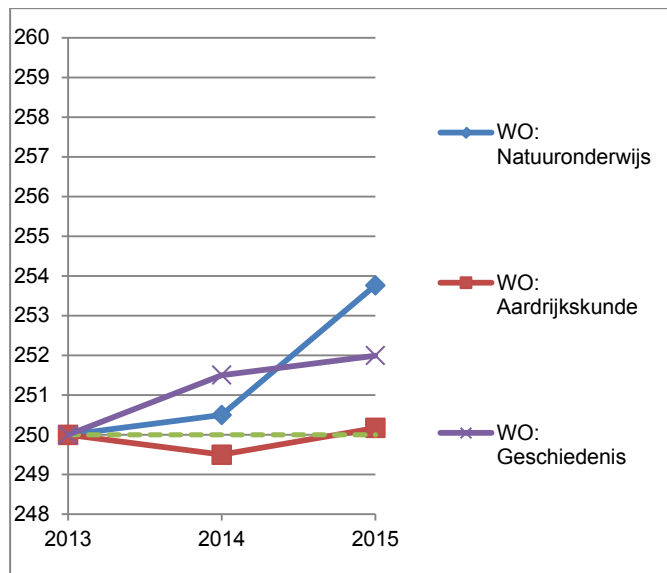
	Vaardigheidsscores			Effecten		
	2013	2014	2015	2014-2013	2015-2014	2015-2013
Natuuronderwijs	250	251	254	0,01	0,06	0,08
Aardrijkskunde	250	249	250	-0,01	0,01	0,00
Geschiedenis	250	251	252	0,03	0,01	0,04

In tegenstelling tot de taal- en rekenvaardigheden is het doel van de WO-toetsen niet aangepast. Ook gemeten vaardigheden zijn niet veranderd. Het enige verschil is het moment van afname dat wat later in het schooljaar ligt. Als we de stijging van 2015 ten opzichte van 2014 vergelijken met de verschillen tussen 2013 en 2014 zien we dat de gemiddelde stijging wel groter is, maar dat dit niet voor alle vaardigheden geldt. Wat exact de groei is in vaardigheid en wat de impact van het latere afnamemoment is, is hier lastig uit af te leiden.

Deze gegevens op de vaardigheidsschaal zijn ook afgebeeld als trendfiguur (figuur 3.3) met een drietal jaren. Er is voor gekozen om de as die vaardigheid vertegenwoordigt met hetzelfde bereik af te beelden als in de figuren voor Taal en Rekenen (figuur 3.1 en 3.2) en ook de jaren op een gelijke afstand te houden zodat er niet bij een oppervlakkige vergelijking van de figuren de indruk zou ontstaan dat de veranderingen bij de WO-schalen veel groter zou zijn.

De groene onderbroken lijn in figuur 3.3 geeft de waarde 250 aan. Die waarde is het gemiddelde van de basismetings in 2013 voor de drie WO-schalen. Alle schalen beginnen zodoende ook met een gemiddelde vaardigheid van 250 (en een standaardafwijking van 50) in 2013.

Figuur 3.3 Trends over de jaren voor de wereldoriëntatie-onderwerpen in jaargroep 8



4 De vergelijking van de achtergrondvariabelen binnen de CET 2015

In het voorgaande hoofdstuk is duidelijk gemaakt waarom de jaarvergelijking lastig te interpreteren is. Het zijn vooral de praktische verschillen tussen de CET in 2015 en de EB in de jaren ervoor die dat lastig maken. Als we de resultaten voor wat betreft de achtergrondvariabelen voor alleen de CET in 2015 rapporteren, dan spelen deze verschillen geen rol. Voor alle leerlingen die de papieren toetsen gemaakt hebben, en dat zijn de leerlingen die we hier onderzoeken, zijn de condities vergelijkbaar. Voor de leerlingen die verschillende toetsen gemaakt hebben, zijn de resultaten met behulp van IRT vergelijkbaar gemaakt waardoor het mogelijk is te rapporteren op de volgende schalen:

Afkorting	Omschrijving	Opmerking
REK	Rekenen	referentieniveauschaal
LEZ	Lezen	referentieniveauschaal
TAV	Taalverzorging	referentieniveauschaal
WS	Woordenschat	ook in eerdere peilingen
NA	Natuuronderwijs	WO-schaal
AA	Aardrijkskunde	WO-schaal
GE	Geschiedenis	WO-schaal

De samenstelling van deze schalen is beschreven in hoofdstuk 2. In dat hoofdstuk is ook beschreven dat een inhoudelijke vergelijking niet goed mogelijk is van de referentieniveauschalen met schalen waarover gerapporteerd is in de eerdere onderzoeken Meting Taal en Rekenen met de Eindtoets Basisonderwijs¹⁰. Voor de WS-schaal en de drie WO-schalen is dat inhoudelijk wel mogelijk.

Ondanks de inhoudelijke overeenkomst tussen een deel van de schalen is geen van de gerapporteerde effecten in dit hoofdstuk direct te vergelijken met eerder gerapporteerde resultaten. De vergelijking van de effecten is niet goed te maken doordat sommige achtergrondvariabelen in 2015 niet opgevraagd zijn. Een variabele die in eerdere analyses een cruciale rol speelde was leerlingewicht. Deze variabele, waarin het opleidingsniveau van de ouders is verdisconteerd, had bij vrijwel alle vaardigheden in alle afgelopen jaren een invloed.

Daar waar het ontbreken van de variabele leerlingewicht geen gevolgen had voor de jaarvergelijking – daar werden altijd al ongecorrigeerde effecten gegeven – heeft dit wel een impact op de vergelijkbaarheid van de rapportage met de eerdere peilingen voor wat betreft de effecten van achtergrondvariabelen binnen een jaar. Dit is uitvoerig beschreven in paragraaf 1.3 en 2.3.2. Er is in dit rapport voor gekozen om alle effecten als ongecorrigeerde effecten weer te geven. Het is duidelijk dat dit ook gevolgen heeft voor de interpretatie van de effecten. Zoals eerder gemeld betekent dit dus ook dat de resultaten van 2015 niet te vergelijken zijn met de resultaten zoals die eerder gerapporteerd waren.

In hoofdstuk 4 rapporteren we de verschillen tussen groepen leerlingen binnen 2015. De verschillende groepen zijn gebaseerd op de categorisering naar achtergrondvariabele. We presenteren hier alleen de resultaten van 2015. In de teksten worden de effecten beschreven. In de bijlagen worden de gemiddelden op de vaardigheidsschaal gegeven. In bijlage 3 betreft dat de gemiddelden van alle schalen waarbij het gemiddelde in 2015 gelijk gesteld is aan 250 (met een standaardafwijking van 50). In bijlage 4 worden op deze schaal de gemiddelden en standaardafwijkingen gegeven voor de vaardigheden waar referentieniveaus bekend voor zijn (REK, LEZ en TAV) en voor Woordenschat.

¹⁰ Zie de Literatuurlijst voor het volledig overzicht van publicaties.

In bijlage 5 zijn voor de WO-vaardigheden de vaardigheidsschalen gerapporteerd op de schalen zoals die in 2013 en 2014 vastgesteld zijn. Dat is bij deze vaardigheden mogelijk omdat bij de WO-toetsen in vergelijking met de andere vaardigheden relatief weinig is veranderd. Bij de referentieniveauschalen is er geen verwijzing naar oude schalen, aangezien deze schalen nieuw zijn in 2015.

Meer dan in de eerdere peilingen ligt de focus in de analyses op de aan school gerelateerde achtergrondvariabelen; in het verleden werd de meeste aandacht gegeven aan de persoonlijke kenmerken. Er zijn nu ook wat minder persoonlijke kenmerken beschikbaar, zoals de "IJK" variabele. Deze variabele is nu wel vervangen door de variabele "onthefing" die gebruikt wordt voor leerlingen die om verschillende redenen niet mee hoeven te tellen voor de schoolscore. Het percentage leerlingen waar dat voor geldt, is overigens zeer laag (ongeveer een half procent van de leerlingen). De andere persoonsvariabelen die in de analyses meegenomen worden in 2015 zijn geslacht, leertijd (twee opdelingen), gemaakte toets (CET basis tegenover CET niveau) en advies Voortgezet onderwijs.

Daar waar er bij de persoonsvariabelen minder gegevens beschikbaar zijn, zijn dat er bij de schoolvariabelen meer. Dat is mede te danken aan de keuze om meer opdelingen van bestaande variabelen te hanteren. Zo is de variabele stratum nu beschikbaar in vier varianten, terwijl bij de meeste eerdere peilingen er maar één definitie gebruikt werd¹¹. Alleen in de peiling van 2010 (Hemker et al., 2011) zijn er meer opdelingen gerapporteerd. Andere additionele opdelingen betreffen schoolgrootte (naast een tweedeling ook een vijfdeling), regio (nu ook op provincieniveau beschikbaar) en verstedelijking (met speciale aandacht voor de vier grote steden). Mede op basis van dit onderzoek zijn er keuzes gemaakt over de te rapporteren stratumdefinities in het online katern '*Taal en rekenen aan het einde van de basisschool. Peil. onderwijs*' (Inspectie van het Onderwijs, 2016b). In dit hoofdstuk zullen de resultaten voor alle mogelijke definities gerapporteerd worden. Een laatste toegevoegde schooleigenschap is de deelname van aan de EB in het recente verleden (in 2013, 2014 of beiden).

De resultaten worden in dit hoofdstuk afgebeeld als effectschattingen. De waarden op de vaardigheidsschaal laten in alle gevallen eenzelfde beeld zien, aangezien het hier ongecorrigeerde effecten betreft. De gemiddelden van de vaardigheidsschattingen voor alle vaardigheden en alle achtergrondvariabelen zijn gegeven in bijlage 1 voor alle schalen op de 2015-schaal. De gemiddelden en standaardafwijkingen voor de referentieschalen en Woordenschat zijn gegeven in bijlage 2 (2015-schaal) en voor de WO-vaardigheden in bijlage 3 (2013-schaal).

4.1 Effecten voor aan de leerling gerelateerde achtergrondvariabelen

De ongecorrigeerde effecten van de persoonsvariabelen zijn gegeven in tabel 4.1. Voor de interpretatie van de effecten is in tabel 2.2 een kwalificatie gegeven van de effectgroottes. In tabel 4.1 zijn de classificaties van de grootte van de effectschattingen ook in een kleurenschema weergegeven. Dat geldt overigens ook voor de overige tabellen met geschatte effectgroottes in dit hoofdstuk. De niet gekleurde cellen geven aan dat het effect als te klein voor interpretatie wordt gezien. De meest licht gekleurde groene cellen worden geïnterpreteerd als kleine effecten (effectgrootte in absolute waarde tussen 0,20 en 0,50); iets donkerder gekleurde cellen betreffen matige effecten (0,50 tot 0,80, of -0,50 tot -0,80); en tot slot de meest donker-groene cellen betreffen grote effecten (effectgrootte van -0,80 of kleiner, of 0,80 of groter). De effecten die significant ($\alpha = 0,01$) afwijken van 0 zijn vet gedrukt.

¹¹ Ook in de peiling van 2010 (gepubliceerd in 2011) is er meer dan één definitie van Stratum onderzocht omdat rond die tijd de definitie van leerlinggewicht veranderde voor leerlingen in groep 8.

Geslacht

Voor het verschil in resultaten tussen jongens en meisjes zien we bij zes van zeven gevallen significante effecten. Alleen bij Woordenschat is er geen significant verschil gevonden. De zes significante verschillen zijn echter niet alle betekenisvol¹². De verschillen bij de referentieschaal Lezen (LEZ) en WO-schaal Natuuronderwijs (NA) zijn weliswaar significant, maar de verschillen zijn in absolute waarde minder groot dan 0,20, wat betekent dat de verschillen weinig impact hebben. Een betekenisvol klein verschil wordt gevonden bij de referentieschalen Rekenen (REK), waar jongens beter presteren, en Taalverzorging (TAV) waar de meisjes beter presteren. Ook bij de WO-schalen Aardrijkskunde (AA) en Geschiedenis (GE) is het effect klein, maar betekenisvol. Op beide vaardigheden worden betere prestaties geleverd door jongens. Het grootste verschil zien we bij AA, waar het verschil bijna als matig geclassificeerd kon worden (door afronding op 0,50, de echte waarde ligt daar net iets onder).

Tabel 4.1 Ongecorrigeerde effecten op leerling gerelateerde achtergrondvariabelen in jaargroep 8 in 2015

Variabele	Cluster	Rekenen				Taalvaardigheden				Wereldoriëntatie		
		REK	LEZ	TAV	WS	NA	AA	GE				
Geslacht												
Meisjes - Jongens												
		-0,24	0,14	0,34	0,02	-0,13	-0,50	-0,28				
Leertijd (dichotoom)												
Regulier - Vertraagd												
		0,77	0,74	0,80	0,60	0,47	0,54	0,52				
Leertijd (trichotoom)												
Vervroegd - Regulier												
		0,31	0,29	0,36	0,26	0,13	0,19	0,15				
Regulier - Vertraagd												
		0,73	0,71	0,77	0,58	0,46	0,52	0,51				
Vervroegd - Vertraagd												
		1,05	0,99	1,13	0,84	0,59	0,71	0,66				
Toets												
Basis - Niveau												
		1,57	1,83	1,45	1,26	1,26	1,29	1,37				
Advies VO												
vmbo-KB - vmbo-BB												
		0,98	1,07	0,79	0,72	0,59	0,71	0,64				
vmbo-GT - vmbo-KB												
		0,50	0,73	0,51	0,44	0,49	0,53	0,46				
havo - vmbo-GT												
		0,95	0,89	0,79	0,75	0,63	0,66	0,61				
vwo - havo												
		1,29	1,17	1,22	1,05	0,98	0,95	0,83				
vwo - vmbo-BB												
		3,72	3,86	3,31	2,95	2,69	2,84	2,54				
Ontheffing												
Geen - Wel ontheffing												
		1,51	1,70	1,36	1,59	1,49	1,30	1,52				

REK = referentieniveau schaal rekenen; LEZ = referentieniveau schaal lezen; TAV = referentieniveau schaal taalverzorging, WS = woordenschat; NA = natuuronderwijs; AA = aardrijkskunde; GE = geschiedenis

Hoewel de waarden van de effecten niet te vergelijken zijn met de eerdere peilingen, komen de richtingen overeen. Zo is bijvoorbeeld bij de vorige meting WO ook gevonden dat het verschil tussen jongens en meisjes het kleinst was bij NA en het grootst bij AA.

¹² Zie Tabel 2.2 voor de interpretatie van de effectgroottes.

De waarden op de vaardigheidsschaal laten eenzelfde beeld zien. Deze zijn gegeven in bijlage 1 voor alle schalen op de 2015-schaal. De gemiddelden en standaardafwijkingen voor de referentieschalen en Woordenschat zijn gegeven in bijlage 2 (2015-schaal) en voor de WO-vaardigheden in bijlage 3 (2013-schaal). Dat is ook het geval voor alle overige achtergrondvariabelen.

Leertijd

In het verleden bij de JPON- en MTR-rapportages is voor de variabele leertijd alleen de opdeling Regulier tegenover Vertraagd gebruikt. De vertraagde leerlingen waren op 1 oktober in het jaar dat zij in groep 8 zaten ouder dan 11 jaar. Deze werden vertraagd genoemd omdat ze gezien hun leeftijd het basisonderwijs al afgerond hadden kunnen hebben. De vertraagde leerlingen hebben een verlengde leertijd in het basisonderwijs gehad ten opzichte van de reguliere leerlingen. In welk leerjaar ze de verlengde leertijd hebben gekregen is daarbij niet bekend. Een relatief groot deel zal dit mogelijk al in de kleuterperiode (groep 1 of 2) hebben opgelopen. Ook is er een relatief grote groep leerlingen geboren in de maanden augustus en september die dit overkomt. De reguliere leerlingen waren 11 of jonger op 1 oktober. In de vorige peiling (Hemker & Van Weerden, 2015) werd al opgemerkt dat de groep vertraagde leerlingen relatief afneemt, van iets meer dan 20% van de leerlingen in 2008 naar minder dan 16% van de leerlingen in 2014. Dat percentage is zelfs verder afgenomen in 2015, en is nu 14,36% van de leerlingen waar de leeftijd van bekend is. De aantallen reguliere leerlingen nemen dus toe en deze toename zit daarbij vooral in de toename van het aantal vervroegde leerlingen. Dat zijn de leerlingen op de 1 oktober 10 jaar of jonger zijn en in groep 8 zitten. De meeste van hun leeftijdsgenoten zitten in groep 7, maar deze leerlingen volgen een versneld traject. In het verleden was dit een relatief zeer kleine groep, maar die is in de loop van de tijd alleen maar groter geworden. Het gaat bij deze groep om 9,45% van de leerlingen, ruimschoots groot genoeg om als aparte categorie op te nemen in de analyses.

De tweedeling (dichotoom) en de driedeling (trichotoom) laat in alle gevallen significante verschillen zien. In al deze gevallen is de vertraagde groep de minst presterende groep, en als er een onderscheid gemaakt wordt tussen de reguliere en de vervroegde leerlingen, dan is de vervroegde groep de best presterende. Bij de WO-schalen is dat verschil niet betekenisvol, maar bij de referentieschalen en Woordenschat is dat een klein betekenisvol verschil. De verschillende met de vertraagde leerlingen zijn in de meeste gevallen als matig te classificeren en tussen de vertraagde en de vervroegde leerlingen bij de referentieschalen en Woordenschat zelfs als groot.

Toets

Ook de toets die is gemaakt kan als variabele worden gezien. Naast de reguliere eindtoets (CET-basis) is er een alternatieve, gemakkelijkere eindtoets beschikbaar (CET-niveau). Dit onderscheid wordt gemaakt sinds 2013. Tot 2013 was deze gemakkelijkere toets, aangeduid als Niveautoets (NT), alleen digitaal beschikbaar en bedoeld voor leerlingen met een leerachterstand van anderhalf jaar of meer. Vanaf 2013 is er een papieren alternatief ingevoerd, zodat de leerlingen gewoon in de klas met de andere leerlingen de toets kunnen maken. De CET-niveau is bestemd voor leerlingen van wie de leerkracht verwacht dat ze doorstromen naar vmbo-BB of vmbo-KB. Deze groep is dus groter dan de groep die voor 2013 de digitale Niveautoets maakte en bij deze groep leerlingen is er per definitie een duidelijk verband met het VO-advies. Zoals te verwachten zijn de gevonden effectgroottes als groot te classificeren: de groepen die de CET-basis en de CET-niveau maken verschillen op alle vaardigheden meer dan 1 standaardafwijking van elkaar. Bij de schaal LEZ is deze het grootst en is het verschil bijna twee standaardafwijkingen groot.

VO-advies

Voor deze variabele onderscheiden we een vijftal VO-adviezen: vmbo-BB, vmbo-KB, vmbo-GT, havo en vwo. Gegevens over de tussenliggende adviezen zijn beschikbaar, maar de verschillen liggen doorgaans halverwege de naastliggende adviezen. Bij de gegevens op vaardigheidsschalen (Bijlage 1 tot en met 3) worden de drie tussenschalen die onderscheiden worden, wel gegeven. Deze tussencategorieën zijn vmbo-BB/KB, vmbo-GT/havo en havo/vwo. De categorie vmbo-KB/GT ontbreekt als aparte groep.

Alle gevonden verschillen zijn significant en betekenisvol. De kleinste verschillen worden gevonden tussen vmbo-GT en vmbo-KB met een gemiddeld effect van 0,52 (matig effect) en effecten variërend van 0,44 (bij Woordenschat) tot en met 0,73 (bij Lezen). Het verschil tussen de het laagst gegeven advies (vmbo-BB) en het hoogste advies (VWO) is gemiddeld ruim drie standaardafwijkingen op de vaardigheidsschaal. Dat is een zeer groot effect. Bij de drie vaardigheden waar referentieniveaus voor beschikbaar zijn, zijn de effecten alle groter dan 3, met bijna 4 bij Lezen en Rekenen. Bij de overige vaardigheden zijn de effecten kleiner dan 3. De effecten bij Wereldoriëntatie zijn kleiner dan bij Woordenschat, maar met een minimum effect van ruim 2,5 bij Geschiedenis nog steeds zeer groot te noemen.

Ontheffing

Ook de verschillen tussen leerlingen met en zonder ontheffing zijn zoals te verwachten groot. Leerlingen die een ontheffing krijgen, (zie paragraaf 2.2.3) presteren aanzienlijk minder goed dan de leerlingen zonder ontheffing.

4.2 Effecten op schoolniveau

In deze peiling zijn meer variabelen op schoolniveau opgenomen dan in eerdere peilingen. Deze variabelen zijn in een aantal groepen op te delen. Ten eerste de variabelen die alle betrekking hebben op de verhouding leerlingen met een leerlinggewicht op een school, de zogenaamde stratumvariabelen. Daar zijn vier verschillende definities toegepast. Daarnaast zijn er de variabelen die te maken hebben met de locatie van de school, waarbij provincie en mate van verstedelijking een rol spelen. Het derde type variabele betreft schoolgrootte waar twee varianten zijn opgenomen. Tot slot is er een restcategorie, waarin denominatie is opgenomen, en of de school de afgelopen twee jaar ervaring heeft opgedaan met de EB. Scholen die in 2013 noch in 2014 de EB hadden gedaan worden als *nieuw* geclassificeerd. De interpretatie van de effecten is vergelijkbaar met die van tabel 4.1, en ook hetzelfde kleurenschema is daarvoor gebruikt. Ook hier verschillen de vetgedrukte effecten significant van 0 ($\alpha < 0,01$).

De prestaties worden in dit hoofdstuk afgebeeld als effectschattingen. De waarden op de vaardigheidsschaal laten in alle gevallen eenzelfde beeld zien, aangezien het hier ongecorrigeerde effecten betreft. De gemiddelden van de vaardigheidsschattingen voor alle vaardigheden en alle achtergrondvariabelen zijn gegeven in bijlage 1 voor alle schalen op de 2015-schaal. De gemiddelden en standaardafwijkingen voor de referentieschalen en Woordenschat zijn gegeven in bijlage 2 (2015-schaal) en bijlage 3 (2013-schaal) voor de WO-vaardigheden.

Stratum

Stratum is een classificatie van scholen waarbij een school in een specifiek stratum ingedeeld wordt op basis van het percentage leerlingen met een leerlinggewicht. Bij dit percentage wordt er geen onderscheid gemaakt tussen leerlingen met een gewicht van 0.30 en 1.20. Het onderscheid tussen deze twee gewichten speelt slechts bij een van de vier definities een rol. In dit verslag is er speciale aandacht voor de resultaten voor de verschillende definities. Op basis daarvan kunnen keuzes gemaakt worden voor de rapportages in toekomstige metingen. Mede op basis van dit onderzoek is er ook voor een van de definities gekozen in de rapportage van de Inspectie van het Onderwijs (2016b).

In de eerste rapportages over de jaarlijkse peilingen (Hemker & Van Weerden, 2009; Hemker, Kuhlemeier & Van Weerden, 2010) waren de strata ingedeeld op basis van oude definities van leerlinggewichten. In het verslag over de peiling in 2010 is vanwege de invoering van de huidige leerlinggewichten in groep 8 een nieuwe indeling van de strata gemaakt. Het onderzoek naar drie mogelijk definities is in het betreffende verslag beschreven (Hemker, Kordes en Van Weerden, 2011). De keuze die daar gemaakt is, is de jaren daarna ook gehanteerd. In het huidige verslag wordt dit de "traditionele" driedeling S3 genoemd, waarbij er drie categorieën zo zijn verdeeld dat het eerste stratum (S3.1) de groep scholen betreft met 0% tot 10%

gewichtenleerlingen; het tweede (S3.2) 10% tot en met 25%, en het derde stratum (S3.3) meer dan 25% gewichtenleerlingen. Bij de twee alternatieve stratumdefinities met vijf niveaus zijn deze drie strata verder opgedeeld.

De vijfdelingen worden S5O en S5L genoemd: de O staat voor de Onderwijsinspectie-opdeling; de L staat voor LOVS-opdeling. Het verschil tussen deze twee definities is zeer klein. In beide definities is S3.1 opgedeeld in twee groepen: de groep met scholen zonder gewichtenleerlingen (S5O1 en S5L1: 0% gewichtenleerlingen) en de groep scholen met leerlingen met een leerlinggewicht, waarbij die groep minder dan 10% van de schoolpopulatie betrof (S5O2 en S5L2: >0% tot 10% gewichtenleerlingen). De definitie van zowel S5O3 als S5L3 is gelijk aan die van S3.2: 10%-25% gewichtenleerlingen. Bij S5O wordt de grens tussen de hoogste twee categorieën gelegd bij scholen met 50% of meer gewichtenleerlingen: S5O4 is >25% tot en met 50% gewichtenleerlingen; S5O5 is meer dan 50% gewichtenleerlingen. Bij S5L ligt die tweedeling bij 40%: S5L4 is >25% tot en met 40%, en S5L5 meer dan 40% gewichtenleerlingen.

De opdeling in vier groepen bij S4 is anders: het grenspunt 10% is daarbij 12% geworden (S4.1: 0 tot 12% gewichtenleerlingen; S4.2: 12% tot en met 25% gewichtenleerlingen). De hoogste categorie, met meer dan 25% leerlingen met een leerlinggewicht op een school, is opgedeeld in twee groepen gebaseerd op de aantallen leerlingen met leerlinggewicht 0.30 en 1.20. In categorie S4.3 vallen de scholen met leerlingen met meer dan 25% gewichtenleerlingen en meer 0.30 leerlingen dan 1.20-leerlingen. In groep S4.4 vallen de scholen met evenveel of meer 1.20-leerlingen in vergelijking met 0.30-leerlingen.

De resultaten voor het onderzoek voor de verschillen in de effecten bij deze verschillende definities is gegeven in tabel 4.2. Merk op dat de resultaten voor de eerste twee contrasten van S5O en S5L (te weten S5.2 – S5.1 en S5.3 – S5.2) per definitie aan elkaar gelijk zijn. Het verschil tussen het eerste contrast van S3 en S4 wijken ook niet sterk van elkaar af; het verschil tussen S3.1 en S3.2 en S4.1 en S4.2 is dan ook klein.

Het belangrijkste verschil tussen S3 en S4 is of er in de categorie met meer dan 25% gewichtenleerlingen onderscheid gemaakt wordt tussen scholen met minder 0.30-leerlingen dan 1.20-leerlingen. Dat onderscheid heeft bij twee vaardigheden geen impact, te weten REK en TAV. De vaardigheid REK was in het verleden ook een vaardigheid waar de thuistaal weinig impact op had. TAV bevat onder andere de vaardigheid Spelling, waar in het verleden thuistaal ook weinig effect op had. Een vaardigheid waar thuistaal juist wel een effect had was Woordenschat. Ook in het huidige onderzoek is een behoorlijk effect gevonden tussen de twee typen strata (S4.3 en S4.4) bij deze vaardigheid. Alleen bij Natuuronderwijs – waarbij in het verleden ook een aanzienlijk effect voor thuistaal gevonden werd – was het verschil groter. Bij de andere twee vaardigheden en bij LEZ is het onderscheid tussen S4.3 en S4.4 iets hoger dan 0,20 (klein effect).

Bij alle vaardigheden is het duidelijk dat er een effect voor stratum te vinden is. Bij de vaardigheden TAV en REK is dit relatief het kleinst. Het effect tussen de categorie met de meeste (hoge) gewichtenleerlingen in vergelijking met die met het laagste aantal is klein (bij S3 en S4) tot matig (bij S5O en S5L). Als per categoriecontrast gekeken wordt (maximum effect gedeeld door het aantal stappen, dus bij S5O REK $0,71/4 = 0,18$, of bij TAV $0,32/2 = 0,16$) is te zien dat die effecten gemiddeld liggen rond de 0,17 bij REK en 0,14 bij TAV. Bij die vaardigheden is het verschil redelijk klein en worden de grootste verschillen gevonden tussen de laagste categorie met de minste gewichtenleerlingen, en de categorie daar net boven.

Tabel 4.2 Ruwe effecten op schoolniveau (stratum) in jaargroep 8 in 2015

Variabele	Cluster	Rekenen	Taalvaardigheden			Wereldoriëntatie		
		REK	LEZ	TAV	WS	NA	AA	GE
Stratum (S3)								
S3.1 (0-10%) – S3.2 (10-25%)		0,20	0,26	0,17	0,24	0,22	0,25	0,22
S3.2 (10-25%) – S3.3 (>25%)		0,21	0,46	0,15	0,50	0,68	0,42	0,46
S3.1 (0-10%) – S3.3 (>25%)		0,41	0,72	0,32	0,74	0,90	0,67	0,69
Stratum (S5O)								
S5O1(0%) – S5O2 (>0-10%)		0,24	0,29	0,20	0,27	0,21	0,19	0,17
S5O2 (>0-10%) – S5O3 (10-25%)		0,18	0,24	0,15	0,23	0,21	0,24	0,21
S5O3 (10-25%) – S5O4 (25-50%)		0,20	0,41	0,14	0,45	0,62	0,38	0,42
S5 O4 (25-50%) – S5O5 (>50%)		0,09	0,36	0,11	0,36	0,43	0,27	0,29
S5O1 (0%) – S5O5 (>50%)		0,71	1,30	0,61	1,30	1,47	1,08	1,09
Stratum (S5L)								
S5L1(0%) – S5L2 (>0-10%)		0,24	0,29	0,20	0,27	0,21	0,19	0,17
S5L2 (>0-10%) - S5L3 (10-25%)		0,18	0,24	0,15	0,23	0,21	0,24	0,21
S5L3 (10-25%) - S5L 4(25-40%)		0,16	0,34	0,12	0,37	0,56	0,36	0,37
S5L4 (25-40%) - S5L5 (>40%)		0,13	0,32	0,10	0,36	0,37	0,19	0,27
S5L1 (0%) - S5L5 (>40%)		0,71	1,19	0,57	1,23	1,35	0,98	1,02
Stratum (S4)								
S4.1(0-12%)-S4.2(12-25%)		0,20	0,29	0,17	0,28	0,28	0,27	0,25
S4.2 (12-25%) – S4.3(>25%;0.3)		0,16	0,23	0,14	0,10	0,23	0,22	0,26
S4.3(>25%;0.3) – S4.4(>25%;1,2)		0,06	0,26	0,01	0,47	0,52	0,23	0,23
S4.1(0-12%) - S4.4(>25%;1,2)		0,41	0,77	0,31	0,85	1,03	0,72	0,74

Bij de WO-vaardigheden Aardrijkskunde (AA) en Geschiedenis (GE), en bij Lezen (LEZ) en Woordenschat (WS) zijn de effecten tussen hoogste en de laagste stratumcategorie gemiddeld rond 0,71 (bij S3) en rond de 0,77 (bij S4). Alleen bij WS in S4 is er een groot effect (0,85). Bij S5O en S5L zijn de effecten tussen de meest extreme categorieën altijd groot bij deze vaardigheden. De gemiddelde effecten per categoriestap zijn bij S3 het grootst (gemiddeld 0,35). Bij de andere stratumdefinities is dat gemiddeld 0,28 voor deze vier vaardigheden. Bij GE en AA liggen de gemiddelde effecten over de stratumdefinities iets onder de 0,30 en bij LEZ en WS daar iets boven. De grootste effecten worden gevonden bij Natuuronderwijs (NA) bij S5O. De extreme verschillen hebben daar een effect van 1,47. Dat is een zeer groot verschil.

Samenvattend kan gesteld worden dat de variabele stratum, en daarmee dus het percentage gewichtenleerlingen, het minste effect heeft bij REK en TAV. Er worden wel significante resultaten gevonden en ook het effect is betekenisvol, maar dat geldt niet voor alle opvolgende categorieën bij deze twee vaardigheden. De effecten zijn duidelijk groter bij AA en GE. De extreme stratumcategorieën bij een vijfdeling laten zelfs een effect van meer dan 1 zien (de twee extreme groepen verschillen meer dan 1 standaarddeviatie van elkaar). Alle opvolgende categoriestappen laten een significant verschil zien, maar het verschil is niet altijd groter dan 0,20, een klein verschil dus. Bij LEZ en WS zijn de effecten nog groter. Het gemiddelde effect van opvolgende stratumcategorieën is over de verschillende definities heen groter dan 0,30. Het grootste effect wordt gevonden bij NA. Voor die vaardigheid doet het er het meest toe wat de samenstelling van de schoolpopulatie is in termen van gewichtenleerlingen. Het gemiddelde effect per opvolgende categorie is hier bij alle stratumdefinities groter dan 0,30, met een gemiddelde van 0,37.

Bij de evaluatie van de verschillende definities van stratum is te zien dat de traditionele definitie de grootste effecten per opvolgende categorie laat zien (gemiddeld effect van 0,32). Van de alternatieve definities is de S5O-definitie de meest onderscheidende (gemiddeld effect van 0,27). Deze laatste opdeling wordt verder ook gebruikt in de publicaties van Inspectie van het Onderwijs (2016a, 2016b).

Regio: windroos en provincies

Het gebruikelijke onderscheid op regio bij het jaarlijks onderzoek is een vierdeling op basis van de CBS-verdeling naar de provincie waar de school zich in bevindt. De categorie Noord bevat scholen uit de provincies Friesland, Groningen en Drenthe, Oost betreft Overijssel, Gelderland en Flevoland, West betreft Noord- en Zuid-Holland, Utrecht en Zeeland. Zuid betreft Noord-Brabant en Limburg. In tabel 4.3 is te zien dat op basis van deze indeling er weliswaar significante effecten gevonden worden bij alle vaardigheden voor enkele contrasten, maar geen van deze contrasten levert een effectgrootte op van enig formaat: alle effectgroottes blijven tussen de -0,20 en de 0,20.

Tabel 4.3 Ongecorrigeerde effecten op schoolniveau (Locatie) in jaargroep 8 in 2015

Variabele	Cluster	Rekenen		Taalvaardigheden		Wereldoriëntatie		
		REK	LEZ	TAV	WS	NA	AA	GE
Regio	Vaardigheid							
Oost - Noord		-0,01	0,03		-0,06	-0,04	0,04	-0,03
West - Noord		-0,01	-0,01		-0,13	-0,06	0,09	0,05
Zuid - Noord		0,00	-0,03		-0,06	0,01	0,09	0,02
West - Oost		0,06	0,05		-0,07	-0,02	0,05	0,08
Zuid - Oost		0,07	0,02		-0,01	0,05	0,05	0,05
Zuid - West		0,07	0,05		0,06	0,07	0,00	-0,03
Verstedelijking								
Sterk - Zeer sterk		0,03	0,11		0,17	0,21	0,07	0,1
Matig - Sterk		0,04	0,06		0,06	0,09	0,04	0,07
Weinig - Matig		-0,03	-0,03	-0,05	0,00	0,04	-0,02	-0,04
Niet - Weinig		0,01	0,00	-0,01	0,06	0,04	0,02	-0,03
Wel of niet Grote 4								
Niet G4 - Wel G4		0,06	0,18	0,03	0,27	0,37	0,13	0,18

De effecten zijn ook berekend voor een indeling waarbij alle provincies als afzonderlijke categorie zijn opgenomen. Die tabel is hier niet weergegeven aangezien dat per vaardigheid 66 contrasten betreft. Uit de analyses volgde dat voor de referentivaardigheden slecht bij 3 van de 198 contrasten (1,5%) een significant (bij een alpha van 1%) betekenisvol effect (absoluut groter dan 0,20) gevonden werd. Twee van deze contrasten betroffen het verschil tussen Flevoland en Limburg (TAV en REK), waarbij Limburg beter presteerde. Bij Woordenschat waren er drie van dergelijke effecten, waarbij Zeeland iets beter presteerde dan de drie minst presterende provincies (Flevoland, en Noord- en Zuid-Holland). Bij de WO-schalen waren er wat meer systematische resultaten: daar presteerde Flevoland op bijna alle vaardigheden slechter dan alle 11 andere provincies: dit betrof in alle gevallen een matig effect. In termen van een aantal goed op de toetsen betekent dat in dit geval dat bij alle drie de WO-vaardigheden de leerlingen in Flevoland ongeveer 1 scorepunt (van de 30) lager scoorden dan in de overige provincies. Verder bleek dat de leerlingen in de provincie Noord-Holland voor NA significant en betekenisvol lager scoorden dan in 6 van de 11 overige provincies. Het verschil tussen Noord-Holland met de overige provincies was echter wel kleiner dan in het geval van Flevoland (gemiddeld ongeveer een halve scorepunt) en betrof alleen NA.

Verstedelijking: een vijfdeling en de grote vier steden

Voor de mate van verstedelijking is een vijfdeling gebruikt. In de meeste gevallen levert dit geen betekenisvol effect op. De enige uitzondering wordt gevonden bij NA tussen zeer sterk stedelijke gebieden en sterk stedelijke gebieden. Dit wordt ook gevonden als we ons specifiek richten op de vier grote steden in Nederland. Ook daarbij is te zien dat bij WS en NA de prestaties in de grotere steden slechter zijn. Woordenschat (WS) is een vaardigheid waar bij eerdere peilingen een sterk effect van thuistaal gevonden was. De natuuronderwijsstoetsen betreffen vooral biologie-onderwerpen.

Als de resultaten vergeleken worden met eerdere peilingen heeft het ontbreken van de variabele leerlinggewicht hier waarschijnlijk wel enige invloed. Daar waar in het verleden gecorrigeerd werd voor deze variabele is dat nu niet het geval. In de G4-steden is het gemiddelde percentage leerlingen zonder leerlinggewicht ongeveer 10% lager dan buiten de G4 (81% om 91% vergelijkbaar met zeer sterk stedelijk versus niet zeer sterk stedelijk). Het percentage leerlingen met gewicht 0.30 is vergelijkbaar op G4- en niet-G4-scholen. Het grote verschil zit in het percentage leerlingen op de scholen met gewicht 1.20, dat op G4-scholen een kleine 10% hoger ligt. In het verleden werden dan ook kleinere verschillen gevonden voor de mate van verstedelijking.

Schoolgrootte

Voor schoolgrootte zijn twee opdelingen geëvalueerd. De eerste is een tweedeling die in eerdere rapportages ook is gebruikt: de grens tussen grote en kleine scholen ligt bij 200 leerlingen. De alternatieve opdeling is een vijfdeling, een gedetailleerde variant van de tweedeling. De groep van kleine scholen is opgedeeld in de groep tot en met 100 leerlingen per school in vergelijking tot 101 tot en met 200 leerlingen. De groep grote scholen is opgedeeld in drie groepen: 201 tot en met 300 leerlingen, 301 tot en met 400 leerlingen en meer dan 400 leerlingen. Bij beide opdelingen wordt bij de vergelijking van de opvolgende categorieën nergens een effect gevonden betekenisvol is: alle waarden in Tabel 4.4 liggen tussen -0,20 en 0,20 in.

Als er gekeken wordt naar de vaardigheidsschattingen is de gevonden trend wel dat de grote scholen tenderen wat beter te presteren dan de kleine scholen. Hierbij valt op dat de effecten wat groter zijn dan in eerdere peilingen, wat zeer wel mogelijk kan komen doordat er niet gecorrigeerd wordt voor een variabele als leerlinggewicht. Zo is er een kleine, maar significante correlatie voor schoolgrootte en percentage leerlingen zonder een leerlinggewicht.

Tabel 4.4 Ongecorrigeerde effecten op schoolniveau (Schoolgrootte) in jaargroep 8 in 2015

Variabele	Cluster	Rekenen	Taalvaardigheden			Wereldoriëntatie		
		REK	LEZ	TAV	WS	NA	AA	GE
Schoolgrootte (2cat)								
Groot(>200) - Klein (1-200)		0,11	0,12	0,12	0,05	0,06	0,16	0,16
Schoolgrootte (5cat)								
(101-200) - (1-100)		-0,04	-0,03	0,02	-0,11	-0,12	0,00	-0,01
(201-300) - (101-200)		0,08	0,08	0,09	0,05	0,10	0,12	0,14
(301-400) - (201-300)		0,03	0,03	0,01	0,01	-0,06	0,02	-0,01
(>400) - (301-400)		0,07	0,10	0,07	0,05	0,06	0,09	0,09

Denominatie en bekendheid met de eindtoets

De laatste twee schoolvariabelen die vergeleken worden zijn denominatie en de variabele die hier 'bekendheid met de Eindtoets' genoemd wordt. Die laatste variabele geeft aan of de leerling op een school zat waar in het recente verleden (2013, 2014 of beide) de Eindtoets Basisonderwijs (EB) van Cito is afgenomen. Ongeveer 6,8% van de leerlingen was afkomstig van een school die niet recent de EB had afgenomen.

Bij denominatie worden er significante betekenisvolle effecten gevonden. Dat is vooral het geval bij de vergelijking tussen algemeen bijzondere (AB) scholen en openbare scholen. Ook hier zou een correctie op basis van formatie waarschijnlijk impact hebben: de openbare scholen hebben relatief meer gewichtenleerlingen dan de AB-scholen. De verschillen tussen de openbare scholen, de rooms-katholieke (RK) en protestants-christelijke scholen (PC) laten nergens betekenisvolle effecten zien.

Tabel 4.5 Ongecorrigeerde effecten op schoolniveau (overige variabelen) in jaargroep 8 in 2015

Variabele	Cluster	Rekenen		Taalvaardigheden		Wereldoriëntatie		
		REK	LEZ	TAV	WS	NA	AA	GE
Denominatie								
RK - Openbaar	Vaardigheid	0,09	0,09	0,08	0,05	0,11	0,07	0,07
PC - Openbaar		0,06	0,06	0,04	0,02	0,10	0,08	0,12
PC - RK		-0,03	-0,03	-0,04	-0,03	-0,01	0,01	0,05
AB - RK		0,13	0,22	0,11	0,20	0,19	0,26	0,26
AB - Openbaar		0,22	0,31	0,19	0,25	0,30	0,33	0,33
AB - PC		0,15	0,25	0,15	0,23	0,20	0,26	0,21
EB gedaan (2013 en/of 2014)								
Wel - Niet eerder		0,24	0,19	0,24	0,10	0,18	0,20	0,17

** Alle vetgedrukte verschillen zijn significant ($p=0,01$)

Op de scholen die recent geen EB hadden afgenomen lieten leerlingen lagere scores zien dan op de scholen die recent wel de EB hadden afgenomen. De effecten zijn significant en bij REK, TAV en AA ook betekenisvol, zij het dat de effecten klein zijn. De verhouding gewichtenleerlingen ligt in de twee groepen nagenoeg gelijk. Er is ook gekeken naar de scholen die in 2015 niet de EB afnamen, maar eerder wel de EB hadden afgenomen. Dat betrof rond 5% van de scholen, waarbij die scholen in 2014 ook lagere scores hadden dan de scholen die zowel 2014 als 2015 hebben meegedaan. De effectschattingen lijken iets lager te liggen, maar zijn wel vergelijkbaar, waardoor dit op het jaareffect waarschijnlijk zeer weinig invloed heeft.

5 Conclusies

Het jaarlijks onderzoek naar de vaardigheden voor Taal, Rekenen en Wereldoriëntatie is in 2015 een nieuwe fase ingegaan. Dit verslag geeft een weergave van de overgang van de oude metingen met de Eindtoets Basisonderwijs (EB) naar de nieuwe metingen met de Centrale Eindtoets (CET). Deze overgang betreft niet alleen een overgang naar een nieuwe naam, maar ook meer fundamentele verschillen. Een heel belangrijk verschil is de verandering van de inhoud waardoor de jaarvergelijking niet meer voor alle vaardigheden die eerder gemeten werden te volgen zijn. De nieuwe vaardigheden vormen een nieuw begin. Dat nieuwe begin betreft vooral de vaardigheden waar de referentieniveaus voor gedefinieerd zijn: Rekenen (REK), Lezen (LEZ) en Taalverzorging (TAV). Uiteraard werden deze vaardigheden in enige vorm in de EB ook gemeten, maar de definities zijn dusdanig veranderd dat de meting in 2015 echt een nieuw startpunt vormt.

Een andere belangrijke reden voor een nieuw startpunt is dat er een aantal praktische verschillen is tussen de metingen met de EB en met de CET. Een belangrijk verschil is het toetsdoel. Daar waar de EB een adviserende rol had voor wat betreft de doorstroom naar het VO, heeft de CET meer een corrigerende rol. Op basis van de uitslag van de CET kan een te laag ingeschat niveau voor een leerling gecorrigeerd worden naar een hoger niveau. Omgekeerd is niet mogelijk: op basis van de CET kan een advies niet naar beneden bijgesteld worden. Een ander verschil is het moment van afname. De afname van de op papier uitgegeven CET-toetsen vond plaats in de tweede helft van april, terwijl EB altijd begin februari werd afgenomen. Daarnaast waren er ook verschillen in de bekostiging en opgenomen achtergrondvariabelen.

Naast de taal- en rekenvaardigheden zijn in dit verslag ook de prestaties op de toetsen Wereldoriëntatie opgenomen. Dat betreft het Natuuronderwijs (NA; betreft vooral biologisch gerelateerde onderwerpen), Aardrijkskunde (AA) en Geschiedenis (GE). Vorig jaar is het eerste verslag over de prestaties op deze vaardigheden verschenen, waarin de resultaten van 2013 en 2014 beschreven waren. Deze vaardigheden zijn met aan de EB en CET gerelateerde toetsen afgenomen. Deze spelen in geen van de afnamejaren een rol in het schooladvies, en zijn afgenomen bij ongeveer twee derde van de leerlingen die ook deelnamen aan de CET of de EB. De toetsen Wereldoriëntatie waren in 2015 inhoudelijk gelijk aan die in de twee jaren ervoor. Het belangrijkste praktisch verschil was het moment van afname, dat net als voor de CET later in het schooljaar lag dan de EB.¹³

5.1 Algemeen beeld in de vergelijking door de tijd

Voor de variabelen waarvoor de vaardigheden door de tijd te volgen waren, zijn jaarvergelijkinganalyses uitgevoerd. Dat betrof de taalonderdelen Begrijpend Lezen (BL; is een subschaal van LEZ), Woordenschat (WS), de rekenonderdelen Getallen en Getalsbewerkingen (RG), Breuken, Procenten en Verhoudingen (RB) en Meten, Meetkunde, Tijd en Geld (RM) die allen onderdeel uitmaken van de vaardigheid REK (dat

¹³ In dit hoofdstuk worden naast de effectschatting ook resultaten weergegeven als theoretisch aantal opgaven verschil per 100 opgaven. Deze verschillen worden berekend door de effectschattingen met de standaarddeviatie van de geobserveerde score te vermenigvuldigen. Deze zijn voor alle vaardigheden geschaald naar een theoretische toets van 100 (vergelijkbare) opgaven zodat de verschillen niet afhangen van het aantal gebruikte opgaven bij een schaal, waardoor de verschillen bij verschillende schalen te vergelijken zijn. Er moet hierbij ook opgemerkt worden dat het echt om een theoretisch verschil gaat, omdat hierbij geen rekening gehouden wordt met plafondefecten. Het plafondefect houdt in dat er nooit meer vragen goed beantwoord kunnen worden dan het maximum aantal, waardoor de (herschaalde) geobserveerde verschillen vaak kleiner uit zullen vallen dan de theoretische verschillen die hier gerapporteerd worden.

aangevuld is met de rekenvaardigheid Verbanden), en de drie wereldoriëntatievaardigheden Natuuronderwijs (NA), Aardrijkskunde (AA) en Geschiedenis (GE). Bij het merendeel van deze vaardigheden is een significante stijging in vaardigheid gevonden. Deze stijging was echter wel zeer klein. Het grootste effect werd gevonden bij BL, met een effectgrootte van 0,11. Dat is op 100 BL-vragen een verschil van minder dan 2 vragen. Alle overige vaardigheden lieten een kleiner verschil zien (minder dan 1 vraag op de 100). Bij de vaardigheden WS, AA en GE waren de verschillen niet significant en hadden in absolute waarde een effectgrootte van 0,01 of lager.

Hoe deze verschillen geïnterpreteerd moeten worden, is lastig te zeggen. Het is moeilijk te stellen dat de vaardigheid op de onderdelen waar een stijging gevonden is echt toegenomen is. Het moment van afname kan wellicht voor een deel het verschil verklaren: als de EB in de tweede helft van april was afgenomen waren wellicht vergelijkbare prestaties gevonden. Het enigszins aangepaste doel van de CET had in ieder geval dit jaar geen dusdanige impact op de motivatie van de leerlingen dat een daling in getoonde vaardigheid gevonden werd.

5.2 Invloed van achtergrondvariabelen

Net als in voorgaande verslagen die de jaarlijkse meting Taal, Rekenen en Wereldoriëntatie betroffen, zijn ook in dit verslag de effecten gegeven van achtergrondvariabelen op de gemeten vaardigheden. Dit betreft de effecten binnen het afnamejaar 2015, voor de referentieniveauvaardigheden REK, LEZ en TAV. Daar zijn de wereldoriëntatievaardigheden en WS aan toegevoegd. Een verschil met eerdere metingen is dat er in dit verslag alleen gewerkt wordt met ongecorrigeerde modellen. Daar waar in het verleden effecten geschat werden waarbij gecorrigeerd werd voor een aantal belangrijke achtergrondvariabelen, is dat in dit verslag niet het geval. De reden daarvoor is dat in tegenstelling tot in het verleden de variabele leerlinggewicht (of formatiegewicht) niet meer op leerlingniveau beschikbaar was. Zodoende worden alleen de ruwe effecten weergegeven.

De achtergrondvariabelen zijn te verdelen in persoonsgebonden variabelen en de aan school gerelateerde variabelen. De persoonsgebonden variabelen zijn geslacht, leertijd (een aan leeftijd gerelateerde variabele), de gemaakte toets, het VO-advies en of de leerling een ontheffing kreeg of niet.

Geslacht

Alleen bij WS is geen significant verschil gevonden tussen jongens en meisjes. Bij twee vaardigheden was er wel een verschil, maar was de effectgrootte niet betekenisvol (absolute waarde van het effect was kleiner dan 0,20). Dat betrof LEZ, dat beter gedaan werd door meisjes, en NA dat beter gedaan werd door jongens. De effecten waren betekenisvol voor TAV, REK, AA en GE. Van deze vier werd TAV beter gedaan door meisjes, en de overige drie vaardigheden beter door jongens. Het grootste gevonden verschil was bij AA. De effectschatting was zodanig dat geschat wordt dat op 100 van dergelijke opgaven jongens er ruim 8 meer goed zouden hebben dan meisjes. Bij de vaardigheid waar de meisjes het best op presteerden, TAV, zouden zij bij het gevonden effect op 100 TAV-vragen er tussen de 5 en 6 meer goed maken dan jongens.

Leertijd

Bij alle vaardigheden wordt gevonden dat vertraagde leerlingen beduidend slechter presteren dan reguliere leerlingen. Het effect bij de drie referentieniveauvaardigheden ligt tussen de 0,70 en de 0,80 en is iets lager bij WS. Dat is bij deze vaardigheden een verschil tussen de 10 en 13 punten op een schaal van 100 opgaven per vaardigheid. De effecten bij de WO-schalen liggen wat lager met effectgroottes van rond de 0,50 wat 8 à 9 opgaven op de 100 items betekent.

De verschillen tussen vervroegde leerlingen en reguliere leerlingen zijn aanzienlijk kleiner. Bij de referentieniveauschalen en WS zijn deze nog wel betekenisvol (effectgroottes rond de 0,30), maar bij de WO-schalen al niet meer.

De trend dat het percentage vertraagde leerlingen door de tijd afneemt, en het percentage vervroegde leerlingen toeneemt, zet nog steeds door. Dit lijkt ook nog steeds geen daling in de vaardigheid tot gevolg te hebben, waarmee de efficiëntie van het onderwijs nog steeds toe lijkt te nemen.

Toetsen

De leerlingen die de CET-basis maken presteren aanzienlijk beter dan de leerlingen die de CET-niveau maken. Bij de referentieniveauschalen is dat verschil meer dan anderhalve standaardafwijking: wat neerkomt op een inschatting van gemiddeld meer dan 25 opgaven meer goed op 100 bij deze schalen. Bij WS en de WO-schalen zijn de verschillen iets kleiner, maar met effectgroottes van gemiddeld 1,3 nog steeds meer dan aanzienlijk. Dit resultaat is naar verwachting aangezien de CET-niveau zich richt op leerlingen van wie de school verwacht dat zij het best op hun plaats zijn in de basisberoepsgerichte of kaderberoepsgerichte leerweg van vmbo.

Advies VO

De leerlingen met verschillende VO-adviezen verschillen ook enorm van elkaar. De stap naar een advies een niveau hoger heeft bij REK en LEZ een effectgrootte van meer dan 0,90. Bij TAV is het iets boven en bij WS wat onder de 0,80 – wat de grens voor een groot effect is. De effecten bij de WO-schalen liggen een fractie lager dan bij WS, rond de 0,70. De kleinste stap zit tussen vmbo-kaderberoeps en vmbo-gemengd/theoretisch, maar betreft nog steeds een gemiddeld effect van rond de 0,50 (uitschieter is hier LEZ met een effect van 0,73). Het verschil tussen vmbo-basisberoeps en vwo is gigantisch. Bij REK en LEZ bijna 4 standaardafwijkingen, wat bij deze schalen in theorie neer zou komen op een verschil van rond de 60 opgaven meer goed op de 100. Merk overigens op dat in de praktijk dergelijke verschillen niet direct gevonden worden vanwege plafondeffecten in het maximum aantal goede antwoorden. Bij de andere schalen zijn de effecten iets kleiner maar ook nog steeds enorm (tussen de 2,5 en 3,5 standaardafwijking).

Ontheffingsleerlingen

Leerlingen die ontheffing krijgen, presteren duidelijk op een wat lager niveau dan de andere leerlingen. Dit resultaat was uiteraard te verwachten aangezien het zeer moeilijk lerende leerlingen betreft of leerlingen die vier jaar of korter in Nederland zijn en om die reden de Nederlandse taal onvoldoende beheersen.

Bij de aan de school gerelateerde achtergrondvariabelen is voor een aantal variabelen additioneel onderzoek gedaan. Daarbij zijn verschillende definities van de onderscheiden niveaus onderzocht. Zo waren er vier varianten van de variabele stratum, en is regio naast een vierdeling op basis van de windroos ook meer gedetailleerd uitgevoerd op basis van de provincies.

Stratum

Bij de variabele stratum worden scholen verdeeld in categorieën op basis van het percentage leerlingen met een leerlingewicht. Onafhankelijk van de gekozen definitie van stratum kan over het algemeen gesteld worden dat hoe hoger de stratumcategorie (en dus hoe meer gewichtenleerlingen), hoe lager de prestaties zijn. De verschillen waren het kleinst voor REK en TAV. De grootste effecten worden gevonden bij NA, gevolgd door WS en LEZ. Hoe groot het gevonden effect was, hangt af van de definitie van stratum. Het onderzoek naar een specifieke definitie wijst op twee definities die het best functioneren. De eerste is de traditionele driedeling waarbij de grenzen tussen de categorieën gezet worden bij 10% en 25% gewichtenleerlingen. Deze geeft per opvolgende categorieën de duidelijkste verschillen, variërend van 2,6 items per 100 verschil bij TAV tot 7,6 items per 100 bij NA. De tweede is een verfijning van deze opdeling, waarbij twee extra grenzen gesteld worden, te weten bij 0% gewichtenleerlingen en 50% gewichtenleerlingen. Het verschil tussen de twee extreme categorieën (0% versus >50% gewichtenleerlingen) is een theoretisch verschil van 10 (TAV) tot 25 (NA) opgaven.

Het is duidelijk dat dit niets zegt over de kwaliteit van de scholen: de verschillen worden vooral veroorzaakt door de samenstelling van de schoolpopulatie: het is evident dat scholen uit de hogere strata meer gewichtenleerlingen hebben die over het algemeen minder hoge prestaties leveren, zoals bleek uit eerdere

onderzoeken. Binnen de strata zijn er overigens nog wel aanzienlijke verschillen tussen scholen (Inspectie van het Onderwijs, 2016b)

Regio

Als naar de verschillen tussen Noord-, Oost-, West- en Zuid-Nederland gekeken wordt, zijn die zeer klein. Bij alle vaardigheden zien we dat het absolute verschil in scores op een schaal van 100 opgaven gemiddeld lager dan 1 opgave is. Als er meer gedetailleerd naar de provincies gekeken wordt, is het enige resultaat dat systematisch lijkt, dat de provincie Flevoland op alle WO-schalen lager scoort dan de overige provincies (gemiddeld 5 opgaven op 100).

Verstedelijking

De mate van verstedelijking heeft ook een zeer kleine impact op de resultaten. Het gemiddelde absolute verschil bij de opvolgende categorieën is kleiner dan 1. De meest opvallende resultaten betreffen het verschil tussen de zeer sterk stedelijke gebieden en de overige gebieden bij WS en NA, waarbij de zeer sterk stedelijke gebieden tot 5 (WS) en 6 (NA) opgaven op de 100 minder goed scoren. Voor Woordenschat zou dat kunnen komen doordat er niet gecorrigeerd wordt voor leerlinggewicht en de NA-opgaven vooral over de natuur gaan, waar de leerlingen in zeer stedelijke gebieden mogelijk minder mee in aanraking komen. Bij een nadere beschouwing van de vier grote steden, die ook relatief veel leerlingen hebben met een leerlinggewicht (vooral 1.20), worden vergelijkbare resultaten gevonden.

Schoolgrootte

De effecten die gevonden zijn bij schoolgrootte liggen allemaal tussen de -0,20 en 0,20, waardoor de effecten als betekenisloos geclassificeerd worden. De absolute verschillen in aantallen opgaven zijn ook nergens groter dan 3 op de 100 opgaven, met een gemiddelde van 1.

Denominatie

De denominatie heeft weinig impact. Alleen het algemeen bijzonder onderwijs, dat relatief weinig leerlingen met een leerlinggewicht heeft, scoort hoger dan de openbare, rooms-katholieke en protestants-christelijke scholen. De verschillen zijn overal kleiner dan 2 op de 100 opgaven.

Eerder EB gedaan

De scholen die eerder aan de EB meegedaan hebben, presteren iets beter dan de scholen die de afgelopen twee jaar geen EB hebben afgenomen. Voor WS is dat verschil verwaarloosbaar; voor de andere vaardigheden zit het gemiddelde rond een klein effect. In termen van aantal opgaven per 100 is het verschil 3 opgaven. Opgemerkt kan worden dat de scholen die in 2014 wel de EB afnamen maar in 2015 niet meer, in 2014 lager scoorden dan scholen die in 2015 wel de CET afnamen. De impact op het jaareffect is zodoende minimaal.

5.3 Discussie

De resultaten liggen in dezelfde lijn als in eerdere rapportages gevonden werd. Doordat er nu niet gecorrigeerd is voor leerlinggewicht zijn de groottes van de effecten over het algemeen iets groter dan eerder gerapporteerd.

Behalve het onderzoek naar deze effecten is er ook onderzoek uitgevoerd naar het behalen van de verschillende referentieniveaus. Dat was mogelijk omdat er een pilot is uitgevoerd waarbij de referentieniveauvaardigheden (REK, LEZ en TAV) aan de referentieniveaus gekoppeld werden. Het was hierdoor te achterhalen hoeveel procent van de leerlingen onder het 1F-niveau uitkwam of boven het 1S (REK) dan wel het 2F (LEZ en TAV) uitkwam. Het 1F-niveau werd door 90% (bij REK) tot meer dan 95% (bij TAV) van de leerlingen gehaald. De percentages leerlingen dat het hogere niveau haalden varieerden van 35% (bij TAV) tot 65% (bij LEZ).

Uitgebreide resultaten van dat onderzoek zijn gegeven in de publicaties *Terugblik 2015* (College voor Toetsen en Examens, 2016) en het online katern '*Taal en rekenen aan het einde van de basisschool. Peil. onderwijs*' (Inspectie van het Onderwijs, 2016b). De resultaten liggen in de lijn van de hierboven gepresenteerde resultaten. Daarnaast is specifiek ook onderzoek uitgevoerd waarbij de school de meeteenheid was en niet zozeer de leerling. Tevens zijn voorbeelden van opgaven gegeven gerelateerd aan de beheersing van de referentieniveaus.

De verwachting is dat de hier beschreven meting van 2015 aanzienlijk minder afwijken van de meting met 2016 dan met de meting in 2014. De CET zal in 2016 op een zelfde moment en met dezelfde doelstelling worden afgenomen als in 2015. In de die meting kunnen de vaardigheden op de referentie-niveauvaardigheden ook gevolgd worden en zal het mogelijk zijn te achterhalen of de percentages leerlingen die de verschillende niveaus halen, veranderen.

Literatuur

- Boxtel, H. van, R. Engelen, A. de Wijs (2012). *Wetenschappelijke verantwoording van de Eindtoets Basisonderwijs 2010*. Arnhem: Cito.
- Bügel, K. & Sanders, P.F. (1998). *Richtlijnen voor de ontwikkeling van onpartijdige toetsen*. Arnhem: Cito.
- Cito (2012). *Terugblik en resultaten 2012. Eindtoets basisonderwijs jaargroep 8*. Arnhem: Cito.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (second ed.)*. Lawrence Erlbaum Associates.
- Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2008). *Over de drempels met taal en rekenen*. Enschede: SLO.
- Hemker, B.T. (2012). *The impact of motivation: modeling motivation in educational measurement. Presentation presented July 4, 2012 at the ITC conference, Amsterdam*.
- Hemker, B.T. & J.J. van Weerden (2009). *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2008 - Jaarlijks Peilingsonderzoek van het Onderwijsniveau - Technische rapportage*. Arnhem: Cito. (<http://www.minocw.nl/documenten/133682d.pdf>)
- Hemker, B.T. & J.J. van Weerden (2015a). *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 in 2014 - Jaarlijks Peilingsonderzoek van het Onderwijsniveau*. Arnhem: Cito. (zie www.cito.nl, Onderzoek en wetenschap, PPON).
- Hemker, B.T. & J.J. van Weerden (2015b). *Peiling van de vaardigheid van de wereldoriëntatievakken in jaargroep 8 van het basisonderwijs in 2013 en 2014*. Arnhem: Cito. (zie www.cito.nl, Onderzoek en wetenschap, PPON).
- College voor Toetsen en Examens (2016). *Terugblik 2015: Resultaten Centrale Eindtoets 2015*. https://www.centraleeindtoetspo.nl/media/uploads/files/Terugblik_Centrale_Eindtoets_2015_CvTE_jan_2016.pdf
- Inspectie van het Onderwijs (2012). *Monitor verbetertrajecten taal en rekenen 2008/2009, 2009/2010 en 2010/2011*. Utrecht: Inspectie van het Onderwijs/OCW.
- Inspectie van het Onderwijs (2016a). *De Staat van het Onderwijs. Onderwijsverslag 2014/2015*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2016b). *Taal en rekenen aan het einde van de basisschool. Peil. onderwijs*. Utrecht: Inspectie van het Onderwijs.
- Janssen, Jan, Frank van der Schoot, Bas Hemker (2005). *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 4. Uitkomst van de vierde peiling in 2004*. PPON-reeks nummer 32. Arnhem: Cito. (http://www.cito.nl/po/ppon/rekwisk/eind_fr.htm)
- Keskaik, S. & Rocher, T. (2015). *La motivation des élèves français face à des évaluations à faibles enjeux, éducation & formations n° 86-87* (http://cache.media.education.gouv.fr/file/revue_86-87/58/2/depp-2015-EF-86-87-motivation-eleves-francais-face-evaluations-faibles-enjeux_424582.pdf)

- Sarwar, G.S., Zepa, C., Hachey, K., Simon, M., & van Barneveld, C. (2012). Teaching Practices and Student Motivation That Influence Student Achievement on Large-Scale Assessments. *International Journal of Education*, Vol.4, No.3 (<http://dx.doi.org/10.5296/ije.v4i3.2087>).
- Schellens, F., Hemker, B.T., & Vermeulen, J. (2013). *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 5. Uitkomst van de vijfde peiling in 2011*. Arnhem: Cito.
- Van Barneveld, C., Pharand, S.L., Ruberto, L., Haggarty, D. Student motivation in large-scale assessments. In: *Improving large-scale assessment in education: theory, issues and practice*. Edited by: Simon, M., Ercikan, K., Rousseau, M. Oxfordshire: Routledge; 2013:43–61.
- Verhelst, N.D. (1993). Itemresponstheorie. In: T.J.H.M. Eggen & P.F. Sanders (red.). *Psychometrie in de praktijk*. Arnhem: Cito (p. 83-178).
- Verhelst, N.D., C.A.W. Glas & H.H.F.M. Verstralen (1995). *OPLM: One Parameter Logistic Model. Computer program and manual*. Arnhem: CITO.
- Verhelst, N.D. & H.H.F.M. Verstralen (2002). *Structural Analysis of a Univariate Latent Variable (SAUL); Theory and a Computer Program*. Arnhem: Cito.

Eerder verschenen rapporten over de jaarlijkse metingen (chronologische volgorde)

Meting 2008 – schooljaar 2007/2008:

Hemker, B.T. & J.J. van Weerden (2009). *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2008 - Jaarlijks Peilingsonderzoek van het Onderwijsniveau - Technische rapportage*. Arnhem: Cito. (zie www.cito.nl, Onderzoek en wetenschap, PPON)

Meting 2009 – schooljaar 2009/2009:

Hemker, B.T., J.B. Kuhlemeier & J.J. van Weerden (2010). *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2009 - Jaarlijks Peilingsonderzoek van het Onderwijsniveau*. Arnhem: Cito. (zie www.cito.nl, Onderzoek en wetenschap, PPON)

Meting 2010 – schooljaar 2009/2010:

Hemker, B.T., J. Kordes & J.J. van Weerden (2011). *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2010 - Jaarlijks Peilingsonderzoek van het Onderwijsniveau*. Arnhem: Cito Arnhem. (zie www.cito.nl, Onderzoek en wetenschap, PPON)

Meting 2011 – schooljaar 2010/2011:

Hemker, B.T. & J.J. van Weerden (2012). *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2011 - Jaarlijks Peilingsonderzoek van het Onderwijsniveau*. Arnhem: Cito. (zie www.cito.nl, Onderzoek en wetenschap, PPON)

Meting 2012 – schooljaar 2011/2012:

Van Weerden, J.J., B.T. Hemker, H. Straat & K. Mulder (2013). *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2012*. Arnhem: Cito (zie www.cito.nl, Onderzoek en wetenschap, PPON).

Meting 2013 – schooljaar 2012/2013:

Van Weerden, J.J., B.T. Hemker, & K. Mulder (2014). *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2013*. Arnhem: Cito (zie www.cito.nl, Onderzoek en wetenschap, PPON).

Metingen 2014 – schooljaar 2013/2014:

Hemker, B.T. & J.J. van Weerden (2015a). *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 in 2014 - Jaarlijks Peilingsonderzoek van het Onderwijsniveau*. Arnhem: Cito. (zie www.cito.nl, Onderzoek en wetenschap, PPON).

Hemker, B.T. & J.J. van Weerden (2015b). *Peiling van de vaardigheid van de wereldoriëntatievakken in jaargroep 8 van het basisonderwijs in 2013 en 2014*. Arnhem: Cito. (zie www.cito.nl, Onderzoek en wetenschap, PPON).

Relevante websites:

www.cito.nl

www.hetcvte.nl

www.owinsp.nl

www.minocw.nl

www.toetswijzer.nl

<http://ppon.cito.nl>

Bijlagen

Bijlage 1 Samenvatting Flankerend Onderzoek Rekenen CET 2015

Auteur: Jan van Weerden

Deze bijlage betreft een beknopt verslag van een onderzoek bij de Centrale Eindtoets (CET) van 2015 met een extra boekje met rekenopgaven 'Vermenigvuldigen en delen'. In dit onderzoek zijn resultaten van opgaven in de vorm van kale sommen vergeleken met sommen met context.

Inleiding

Gedurende de afname van de Centrale Eindtoets (CET) in 2015 is in het kader van flankerend onderzoek bij een deel van de scholen een extra boekje afgenomen. Dat gebeurde bij 18 scholen die niet deelnamen aan de toets Wereldoriëntatie en vond plaats op de laatste toetsdag (april). Het extra rekenboekje bevatte opgaven over vermenigvuldigen en delen. Anders dan in de reguliere afname betrof het geen meerkeuzevragen, maar open opgaven, dat wil zeggen dat de leerlingen het antwoord zelf moesten opschrijven. De vraag was o.a. hoe een dergelijk onderzoek zou vallen bij de scholen. Het extra boekje werd door 505 leerlingen gemaakt. De belangrijkste onderzoeksvragen waren:

- Hoe reageren scholen en leerlingen op een aanvullend boekje bij de CET?
- Hoe reageren scholen en leerlingen op vragen van een ander type dan de standaard 4-keuzevraag van de CET?
- Is er een verandering in vaardigheid bij het onderwerp Basisoperaties Vermenigvuldigen en delen vergeleken met de resultaten van PPON in 2011?
- Is er een verschil in het meten van deze vaardigheid met kale opgaven of met contextopgaven?

Daarnaast werden nog enkele aanvullende vragen opgenomen betreffende gebruik van kladpapier en de relatie van de uitkomsten op deze toets met die op de CET. Het onderzoek was voor een deel een herhaling van een vergelijkbaar onderzoek bij de Eindtoets van Cito in 2013, en voor een deel nieuw onderzoek vanwege de expliciete constructie van extra opgaven om de nieuwe inhoudelijke onderzoeksvragen te kunnen beantwoorden.

Theoretisch kader/inhoudelijke verkenning

Het flankerend onderzoek van 2015 richt zich op een studie met rekenopgaven. Flankerend onderzoek bij de CET is in principe gericht op mogelijkheden om de toets te verbeteren. De gelegenheid werd ook te baat genomen om de mogelijkheid te onderzoeken de CET te koppelen aan systeemevaluatie. Daarom werd een vaardigheid onderzocht die ook in het periodiek peilingsonderzoek aan de orde was geweest en waarbij de uitkomst vragen had opgeroepen, t.w. Basisoperaties: Vermenigvuldigen en delen. Dat onderwerp bleek goed te combineren met een verder onderzoek naar de kwestie of kale opgaven anders (beter) de rekenvaardigheid kunnen meten dan sommen met context. In de literatuur zijn daar aanwijzingen voor te vinden, gekoppeld aan variabelen als geslacht, taalvaardigheid en gebruik van kladpapier.

Onderzoeksofzet en uitvoering

Alleen scholen die niet deelnamen aan de toets Wereldoriëntatie werden gevraagd deel te nemen. Die scholen hadden immers in principe nog ruimte in het rooster. Uit de lijst van deze scholen werd een aselecte steekproef van 100 getrokken, in die zin dat scholen die al meededen aan een ander onderzoek voor de CET, werden buitengesloten. Om de gang van zaken op de reguliere toetsen niet te beïnvloeden, werd gekozen voor afname op de laatste dag van de CET onmiddellijk na de laatste toets van die dag. Uiteindelijk deden 18 basisscholen mee aan dit onderzoek (respons van 18%, evenwichtig verdeeld over de regio's). In totaal was er respons van 505 leerlingen.

Het extra boekje bevatte 10 ankeropgaven (uit de PPON-peiling 2011) en 38 nieuw geconstrueerde opgaven, zodat er voldoende opgaven waren voor een goede analyse. De 48 items werden evenredig verdeeld over vier boekjes. Elk boekje bestond uit 12 contextuele rekenopgaven en 12 kale rekenopgaven

die niet met elkaar overeenkwamen. Er stonden nooit meer dan twee opgaven van dezelfde soort achter elkaar. Op deze manier werd gewinning uitgesloten. Elk boekje bevatte 12 vermenigvuldigopgaven en 12 deelopgaven. Elke leerling kreeg één boekje te maken en had daar 30 minuten de tijd voor. De afname vond plaats door de eigen leerkracht, die daartoe een instructie van CVTE had ontvangen.

Resultaten

De betrouwbaarheid van de toets bleek goed te zijn ($\alpha = 0.86-0.88$ per boekje). De gemiddelde moeilijkheidsgraad was $p: 0,65$. Dat is conform de conventies die voor een toets als de CET gelden. Er zijn tien opgaven met een lagere p -waarde dan $.50$, de overige 38 opgaven wordt door meer dan de helft van de leerlingen goed gemaakt. Drie opgaven worden door minder dan 10% van de leerlingen fout gemaakt. Ook de kalibratie met OPLM leverde een goede meetschaal op.

Vergelijking 2015-2011

Een vergelijking met de resultaten van PPON van 2011, mogelijk dank zij de aanwezigheid van ankeritems, laat zien dat er wel een significant verschil is in het voordeel van 2015, maar ook dat dat verschil niet betekenisvol is (effectgrootte 0.16).

Analyse van context versus kaal

Een toetsing op het verschil tussen de twee subschalen blijkt geen statistisch significant verschil op te leveren tussen de contextuele rekenopgaven en kale rekenopgaven ($t = \text{toets: } t(24) = .18, p = .857$). Met een effectgrootte van $.02$ kunnen we concluderen dat de opgavenvorm geen tot nauwelijks effect heeft op het goed kunnen beantwoorden van een opgave. Opgemerkt kan ook worden dat bij vergelijking van de twee versies van een som het bepaald niet zo is dat de kale versie altijd makkelijker is dan de contextversie. Het omgekeerde komt ook dikwijls voor en meestal is er nauwelijks verschil. Verder analyses laten zien dat de variabelen geslacht, leesvaardigheid en gebruik van kladpapier daar geen invloed op lijken te hebben.

Conclusies en discussie

Scholen (en leerlingen) lijken geen probleem te hebben met het maken van een extra boekje met andersoortige opgaven dan in de reguliere CET zitten. De vaardigheid in 'Basisoperaties Vermenigvuldigen en delen' blijkt vergeleken met de resultaten van PPON in 2011 nauwelijks gewijzigd te zijn. Het meten van deze vaardigheid met contextopgaven levert geen ander beeld op dan met kale opgaven. Deze uitkomsten zijn niet in lijn met de verwachtingen gebaseerd op literatuurstudie voorafgaand aan dit onderzoek (zie Olthuis, 2015).

Er zijn enkele kanttekeningen te plaatsen bij deze uitkomsten:

- De resultaten zijn verkregen in een andere context dan bij peilingsonderzoek. Dat maakt een vergelijking over de tijd lastig. De afnamecondities kunnen de uitkomsten sterk beïnvloeden.
- De vergelijking van sommen op basis van de opgavevorm, kaal of met context, is niet eenvoudig. In deze studie zijn er strakke regels voor operationalisering gehanteerd die maken dat een contextsom inhoudelijk dicht bij de originele kale som blijft en andersom. Contexten kunnen het oplossen van de som vergemakkelijken, maar ook bemoeilijken, afhankelijk van de operationalisering.
- Het toetsen van rekenvaardigheid lijkt het meest gediend met de opname van beide varianten. Er zijn leerlingen die beter scoren op de ene vorm en leerlingen die beter scoren op de andere vorm.

Aanbevelingen

Dit onderzoek vond plaats bij een relatief kleine groep leerlingen in jaargroep 8 en betrof alleen het onderwerp Vermenigvuldigen en delen. Verder onderzoek is nodig om te zien of deze conclusies ook opgaan voor andere onderwerpen en domeinen en bij leerlingen van een andere leeftijd. Vooralsnog lijkt er geen aanleiding te veronderstellen dat de meting van de rekenvaardigheid met contextsommen andere uitkomsten oplevert dan met kale sommen. Met een combinatie kan men tegemoet komen aan de verschillen die er bij leerlingen zijn wat betreft de reactie op deze opgavevormen.

Bijlage 2a Verdeling van de achtergrondvariabelen in de onderzoekspopulatie in 2015 die een eindtoets gemaakt heeft (individuele kenmerken)

Totaal aantal leerlingen is 158.687.

Individuele kenmerken		aantal	%	% excl miss
Geslacht	Jongen	78269	49,3	49,7
	Meisje	79209	49,9	50,3
	Onbekend	1209	0,8	
Leeftijd (tweedeling)	10-12 jaar (regulier)	135082	85,1	85,6
	13-14 jaar (oud)	22644	14,3	14,4
	Onbekend	961	0,6	
Leeftijd (driedeling)	10-11 jaar (jong)	14910	9,4	9,5
	12 jaar (regulier)	120172	75,7	76,2
	13-14 jaar (oud)	22644	14,3	14,4
	Onbekend	961	0,6	
Toets	CET B (en anker)	128054	80,7	
	CET N	30633	19,3	
Advies VO	BB	9466	6	7,7
	BB-KB	3493	2,2	2,8
	KB	15241	9,6	12,4
	GT	25055	15,8	20,4
	GT-HA	10926	6,9	8,9
	HA	23369	14,7	19,1
	HA-VW	11491	7,2	9,4
	VW	23619	14,9	19,3
Onbekend	36027	22,7		

Bijlage 2b Verdeling van de achtergrondvariabelen in de onderzoekspopulatie in 2015 die een eindtoets gemaakt heeft (schoolkenmerken)

Totaal aantal leerlingen is 158.687.

Schoolkenmerken

		aantal	%
Stratumdefinities			
S3 (traditioneel)	0% t/m 10% gew IIn	108616	68,4
	>10% t/m 25% gew IIn	34958	22
	>25% gew IIn	15113	9,5
			0
S5 - onderwijsinspectie	0% gew IIn	7882	5
	>0% t/m 10% gew IIn	100734	63,5
	>10% t/m 25% gew IIn	34958	22
	>25% t/m 50% gew IIn	13040	8,2
	>50% gew IIn	2073	1,3
			0
S5 -LOVS	0% gew IIn	7882	5
	>0% t/m 10% gew IIn	100734	63,5
	>10% t/m 25% gew IIn	34958	22
	>25% t/m 40% gew IIn	9620	6,1
	>40% gew IIn	5493	3,5
			0
S4 - onderwijsinspectie	>88% P000 IIn	117728	74,2
	>75-88% P000 IIn	25740	16,2
	0-75% P000 IIn EN meer P030 dan P120	3890	2,5
	0-75% P000 IIn EN P120 >= P030	11329	7,1
Schoolgrootte			
SG tweedeling	Klein (.-200)	48372	30,5
	Groot (>200)	110315	69,5
SG vijfdeling	0-100	10568	6,7
	101-200	37804	23,8
	201-300	45673	28,8
	301-400	30334	19,1
	>400	34308	21,6
Regio			
vierdeling (windroos)	Noord	14885	9,4
	Oost	39880	25,1
	West	70083	44,2
	Zuid	33839	21,3

Bijlage 2b vervolg **Verdeling van de achtergrondvariabelen in de onderzoekspopulatie 2015 die een eindtoets gemaakt heeft (schoolkenmerken)**

Totaal aantal leerlingen is 158.687.

Regio (vervolg)

12-deling (provincie)	Groningen	5211	3,3
	Friesland	5638	3,6
	Drenthe	4036	2,5
	Overijssel	12656	8
	Flevoland	4559	2,9
	Gelderland	22665	14,3
	Utrecht	13727	8,7
	Noord-Holland	22144	14
	Zuid-Holland	30729	19,4
	Zeeland	3483	2,2
	Noord-Brabant	24453	15,4
	Limburg	9386	5,9

Verstedelijking

vijfdeling	Zeer sterk stedelijk	23739	15
	Sterk stedelijk	41527	26,2
	Matig stedelijk	33889	21,4
	Weinig stedelijk	38826	24,5
	Niet stedelijk	20706	13

Grote-vier (detail)	Geen G4	140002	88,2
	Amsterdam	6660	4,2
	Rotterdam	5189	3,3
	Den Haag	3757	2,4
	Utrecht	3079	1,9

Grote-vier (tweedeling)	Geen G4	140002	88,2
	G4	18685	11,8

Overige schoolkenmerken

Denominatie	Rooms-katholiek	52400	33
	Openbaar	43743	27,6
	Protestants-christelijk	34385	21,7
	Algemeen bijzonder	6784	4,3
	Overig (<5%)	8372	5,3
	Missing	13003	8,2
Nieuwe CET-school	In 2013 en/of 2014 EB gedaan	147850	93,2
	Afgelopen 2 jaar geen EB gedaan	10837	6,8

Bijlage 3a Gemiddelden van alle vaardigheden op de 2015-schaal per achtergrondvariabele (leerlingkenmerken)

Algemeen gemiddelde van de 2015-schaal is 250 voor alle schalen.

Leerlingkenmerken	<u>REK</u>	<u>LEZ</u>	<u>TAV</u>	<u>WS</u>	<u>NA</u>	<u>AA</u>	<u>GE</u>
Geslacht							
Jongens	256	247	242	250	254	263	258
Meisjes	244	253	258	251	246	238	243
Leertijd (2 cat)							
Regulier & Vroeg	255	255	256	254	253	254	254
Vertraagd	218	219	217	225	230	227	228
Leertijd (3 cat)							
Vervroegd	268	267	271	265	259	262	260
Regulier	254	254	254	253	253	253	253
Vertraagd	218	219	217	225	230	227	228
Toets							
CET-basis	263	264	262	261	260	261	261
CET-niveau	196	190	199	204	204	203	200
Advies VO							
BB	177	171	185	187	191	192	186
BB/KB	198	192	200	207	203	203	200
KB	212	209	213	215	217	215	214
GT	230	234	232	232	236	235	235
GT/HA	243	247	245	244	246	246	248
HA	264	265	261	261	261	260	261
HA/VW	280	279	276	275	271	273	275
VW	310	307	309	304	296	301	300
Ontheffing							
Nee	250	251	250	250	250	250	250
Ja	176	166	183	171	176	186	175

Bijlage 3b Gemiddelden van alle vaardigheden op de 2015-schaal per achtergrondvariabele (schoolkenmerken - stratum)

Algemeen gemiddelde van de 2015-schaal is 250 voor alle schalen.

Stratum	<u>REK</u>	<u>LEZ</u>	<u>TAV</u>	<u>WS</u>	<u>NA</u>	<u>AA</u>	<u>GE</u>
S3							
S1 (0-10%)	254	256	253	256	256	256	256
S32_1025	244	244	245	244	246	244	245
S33_>25%	234	222	237	220	213	223	222
S5 (O en L)							
S1(0% ;0&L)	265	269	263	269	266	265	263
S2(>0-10% ;0&L)	253	255	253	255	256	255	255
S3(>10-25% ;0&L)	244	244	245	244	246	244	245
S4(>25-50%;O)	235	224	238	223	216	225	225
S5(>50%;O)	230	206	232	205	195	212	210
S4(>25-40%;L)	236	227	239	226	219	226	227
S5(>40%;L)	230	212	234	209	201	216	213
S4							
S1(0-12%)	253	255	253	256	256	255	255
S2(12-25%)	244	242	245	242	243	242	243
S3(>25%;meer 0.30)	236	231	238	237	232	231	231
S4(>25%;meer 1.20)	233	218	237	214	207	220	219

Bijlage 3c Gemiddelden van alle vaardigheden op de 2015-schaal per achtergrondvariabele (schoolkenmerken - schoollocatie)

Algemeen gemiddelde van de 2015-schaal is 250 voor alle schalen.

Locatie	<u>REK</u>	<u>LEZ</u>	<u>TAV</u>	<u>WS</u>	<u>NA</u>	<u>AA</u>	<u>GE</u>
Regio							
Noord	250	249	247	254	251	246	249
Oost	249	250	249	251	249	249	248
West	249	249	250	248	249	251	252
Zuid	253	251	252	250	252	251	250
Provincie							
Groningen	250	252	250	254	254	248	253
Friesland	250	248	246	253	249	246	245
Drenthe	249	247	246	256	249	244	247
Overijssel	249	249	249	251	249	248	247
Flevoland	244	248	245	246	234	236	234
Gelderland	250	252	250	252	252	250	250
Utrecht	254	256	253	254	253	253	256
Nood- Holland	247	249	248	248	242	248	247
Zuid-Holland	249	247	251	245	251	253	253
Zeeland	249	248	251	256	258	252	254
Noord- Brabant	252	251	250	250	252	250	250
Limburg	255	252	256	253	253	251	250
Verstedelijking							
Zeer Sterk	248	245	250	241	238	247	245
Sterk	249	250	249	249	248	250	250
Matig	252	252	252	252	252	251	253
Weinig	250	251	250	252	254	250	251
Niet	250	251	249	255	256	251	250
Grote 4 steden							
Nee	250	251	250	251	252	251	251
Ja	247	243	249	239	234	245	242

Bijlage 3d Gemiddelden van alle vaardigheden op de 2015-schaal per achtergrondvariabele (schoolkenmerken – schoolgrootte en overige schoolkenmerken)

Algemeen gemiddelde van de 2015-schaal is 250 voor alle schalen.

Schoolgrootte	<u>REK</u>	<u>LEZ</u>	<u>TAV</u>	<u>WS</u>	<u>NA</u>	<u>AA</u>	<u>GE</u>
2-delig (aantal lln)							
1-200	246	246	246	248	248	244	245
>200	252	252	252	251	251	252	252
5-delig (aantal lln)							
1-100	248	247	245	252	252	244	245
101-200	246	245	246	247	247	244	245
201-300	250	249	250	250	252	251	251
301-400	251	251	251	250	249	252	251
>400	254	256	254	252	251	256	255
Overige kenmerken	REK	LEZ	TAV	WS	NA	AA	GE
Denominatie							
Rooms-Katholiek	252	252	252	251	252	251	250
Openbaar	248	248	249	249	246	248	247
Protestants Christelijk	251	251	250	249	251	252	253
Algemeen Bijzonder	259	263	258	261	261	264	263
Eerder EB gedaan							
In 2013 en/of 2014	251	251	251	250	250	250	250
Niet in 2013 of 2014	239	241	239	245	241	240	242

Bijlage 4 Gemiddelden van de referentiewaardigheden en Woordenschat op de 2015-schaal per achtergrondvariabele

Algemeen gemiddelde van de 2015-schaal is 250 voor alle schalen.

Leerlingkenmerken	gem	sd	gem	sd	gem	sd	gem	sd
Geslacht								
Jongens	256	51	247	50	242	50	250	49
Meisjes	244	48	253	49	258	48	251	51
Leertijd (2 cat)								
Regulier & Vroeg	255	49	255	49	256	49	254	49
Vertraagd	218	46	219	48	217	43	225	48
Leertijd (3 cat)								
Vervroegd	254	49	254	49	254	49	253	49
Regulier	218	46	219	48	217	43	225	48
Vertraagd	268	46	267	45	271	49	265	46
Toets								
CET-basis	263	43	264	41	262	45	261	45
CET-niveau	196	39	190	39	199	36	204	44
Advies VO								
BB	177	34	171	35	185	34	187	38
BB/KB	198	32	192	32	200	31	207	39
KB	212	30	209	31	213	29	215	37
GT	230	27	234	28	232	29	232	33
GT/HA	243	28	247	27	245	29	244	32
HA	264	28	265	27	261	32	261	34
HA/VW	280	29	279	27	276	33	275	31
VW	310	34	307	30	309	43	304	39
Ontheffing								
Nee	250	50	251	50	250	50	250	50
Ja	176	53	166	50	183	49	171	44
Stratum								
S3								
S1 (0-10%)	254	49	256	48	253	50	256	48
S32_1025	244	50	244	50	245	50	244	50
S33_>25%	234	52	222	54	237	50	220	51
S5 (O en L)								
S1(0% ;0&L)	265	50	269	47	263	53	269	50
S2(>0-10% ;0&L)	253	49	255	48	253	49	255	48
S3(>10-25% ;0&L)	244	50	244	50	245	50	244	50
S4(>25-50%;O)	235	52	224	54	238	51	223	52
S5(>50%;O)	230	48	206	50	232	48	205	47
S4(>25-40%;L)	236	49	227	53	239	50	226	50
S5(>40%;L)	230	49	212	54	234	51	209	51
S4								
S1(0-12%)	253	49	255	48	253	50	256	48
S2(12-25%)	244	51	242	51	245	50	242	50
S3(>25%;meer 0.30)	236	53	231	53	238	51	237	50
S4(>25%;meer 1.20)	233	51	218	53	237	50	214	51

Locatie	Rekenen		Lezen		Taalverzorging		Woordenschat	
	gem	sd	gem	sd	gem	sd	gem	sd
Regio								
Noord	250	49	249	50	247	49	254	50
Oost	249	49	250	49	249	49	251	48
West	249	51	249	51	250	51	248	52
Zuid	253	50	251	48	252	49	250	48
Provincie								
Groningen	250	50	252	50	250	50	254	51
Friesland	250	49	248	49	246	48	253	50
Drenthe	249	49	247	50	246	49	256	49
Overijssel	249	50	249	50	249	50	251	49
Flevoland	244	47	248	48	245	50	246	47
Gelderland	250	49	252	48	250	49	252	48
Utrecht	254	51	256	51	253	51	254	51
Nood-Holland	247	51	249	53	248	52	248	54
Zuid-Holland	249	51	247	51	251	51	245	51
Zeeland	249	50	248	49	251	50	256	48
Noord-Brabant	252	49	251	49	250	49	250	48
Limburg	255	50	252	48	256	49	253	47
Verstedelijking								
Zeer Sterk	248	53	245	56	250	52	241	56
Sterk	249	51	250	51	249	50	249	51
Matig	252	49	252	49	252	49	252	48
Weinig	250	49	251	48	250	49	252	47
Niet	250	49	251	48	249	49	255	48
Grote 4 steden								
Nee	250	50	251	49	250	50	251	49
Ja	247	53	243	57	249	52	239	57
Schoolgrootte								
2-deling (aantal lln)								
1-200	246	50	246	49	246	50	248	49
>200	252	50	252	50	252	50	251	50
5-deling (aantal lln)								
1-100	248	49	247	47	245	49	252	48
101-200	246	50	245	50	246	50	247	50
201-300	250	50	249	51	250	50	250	51
301-400	251	49	251	51	251	50	250	50
>400	254	50	256	50	254	50	252	50
Denominatie								
Rooms-katholiek	252	49	252	49	252	49	251	48
Openbaar	248	51	248	52	249	51	249	53
Protestantschristelijk	251	50	251	48	250	50	249	48
Algemeen bijzonder	259	50	263	49	258	52	261	50
Eerder EB gedaan								
In 2013 en/of 2014	251	50	251	50	251	50	250	50
Niet in 2013 of 2014	239	50	241	50	239	49	245	50

Bijlage 5 Gemiddelden van wereldoriëntatievaardigheden op de 2013-schaal per achtergrondvariabele

Vergelijkbaar met eerdere WO-rapportage (Hemker & Van Weerden, 2015) waarbij het gemiddelde van 2013 op is 250 gesteld

Leerlingkenmerken	Natuurkunde		Aardrijkskunde		Geschiedenis	
	Gem	sd	Gem	sd	Gem	sd
Geslacht						
Jongens	258	53	263	52	260	54
Meisjes	250	46	238	45	245	44
Leertijd (2 cat)						
Regulier & Vroeg	257	49	254	50	256	49
Vertraagd	234	49	227	44	230	48
Leertijd (3 cat)						
Vervroegd	256	49	253	50	255	49
Regulier	234	49	227	44	230	48
Vertraagd	263	49	262	50	262	50
Toets						
CET-basis	264	46	261	46	263	45
CET-niveau	207	40	203	36	202	40
Advies VO						
BB	194	38	192	34	188	38
BB/KB	206	39	203	34	202	39
KB	221	38	215	34	216	38
GT	239	36	235	36	237	36
GT/HA	250	37	246	36	250	35
HA	265	38	260	37	263	35
HA/VW	275	39	273	40	277	35
VW	300	43	301	43	302	39
Ontheffing						
Nee	254	50	250	50	252	50
Ja	180	46	186	41	177	42
Stratum						
S3						
S1 (0-10%)	260	47	256	49	257	48
S32_1025	250	50	244	49	247	50
S33_>25%	217	50	223	47	224	52
S5 (O en L)						
S1(0% ;0&L)	270	47	265	50	265	47
S2(>0-10% ;0&L)	259	47	255	49	257	48
S3(>10-25% ;0&L)	250	50	244	49	247	50
S4(>25-50%;O)	220	51	225	48	227	53
S5(>50%;O)	199	45	212	45	212	47
S4(>25-40%;L)	223	52	226	48	229	54
S5(>40%;L)	205	45	216	46	215	48
S4						
S1(0-12%)	260	48	255	49	257	48
S2(12-25%)	247	50	242	49	245	49
S3(>25%;meer 0.30)	236	51	231	46	233	53
S4(>25%;meer 1.20)	211	49	220	47	221	52

Locatie	Natuuronderwijs		Aardrijkskunde		Geschiedenis	
	<u>Gem</u>	<u>sd</u>	<u>Gem</u>	<u>sd</u>	<u>Gem</u>	<u>sd</u>
Regio						
Noord	255	45	246	47	251	48
Oost	253	50	249	50	250	49
West	253	52	251	52	254	52
Zuid	256	49	251	48	252	49
Provincie						
Groningen	257	48	248	49	255	48
Friesland	253	44	246	48	247	47
Drenthe	253	43	244	45	249	49
Overijssel	253	48	248	50	249	47
Flevoland	238	48	236	49	236	49
Gelderland	255	50	250	50	252	50
Utrecht	257	50	253	51	258	50
Nood-Holland	246	53	248	52	249	54
Zuid-Holland	254	52	253	52	255	51
Zeeland	262	48	252	49	256	50
Noord-Brabant	255	49	251	48	252	49
Limburg	257	48	251	48	252	48
Verstedelijking						
Zeer Sterk	242	56	247	55	247	55
Sterk	251	50	250	50	252	50
Matig	256	49	251	49	255	49
Weinig	258	48	250	48	253	49
Niet	259	48	251	48	252	49
Grote 4 steden						
Nee	256	49	251	49	253	49
Ja	238	56	245	55	244	55
Schoolgrootte						
2-deling						
1-200	252	49	244	48	247	50
>200	255	50	252	51	254	50
5-deling						
1-100	256	45	244	47	247	47
101-200	250	50	245	49	247	51
201-300	256	51	251	50	253	50
301-400	253	51	252	52	253	51
>400	255	48	256	50	257	48
Denominatie						
Rooms-katholiek	255	49	251	49	252	49
Openbaar	250	52	248	50	249	52
Protestants-christelijk	255	49	252	51	255	49
Algemeen bijzonder	265	49	264	50	265	50
Eerder EB gedaan						
In 2013 en/of 2014	254	50	251	50	252	50
Niet in 2013 of 2014	245	47	240	49	244	50

info@centraleeindtoetspo.nl
www.centraleeindtoetspo.nl