

Peiling van de vaardigheid van de wereldoriëntatievakken in jaargroep 8 van het basisonderwijs in 2013 en 2014

Bas Hemker en Jan van Weerden



**Peiling van de vaardigheid van de wereldoriëntatievakken in
jaargroep 8 van het basisonderwijs in 2014**

Bas Hemker en Jan van Weerden

© Stichting Cito Instituut voor Toetsontwikkeling Arnhem (2015)

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 9

2 Vraagstelling en aanpak 11

2.1 Vraagstelling 11

2.2 Methode 11

2.2.1 Kwaliteit van de meetinstrumenten 13

2.2.2 Steekproeftrekking en analyse 16

2.2.3 Achtergrondvariabelen 18

2.3 Hoe is er geanalyseerd? 19

2.3.1 Vergelijking tussen en over de jaren 19

2.3.2 Analyses voor de overige variabelen 20

2.3.3 Interpretatie van de gekozen rapportageschaal 21

2.3.4 Significantie en effectgrootte 22

3 De resultaten van de klassieke analyses 23

3.1 Verdeling van de achtergrondvariabelen 23

3.2 Vaardigheid per variabele bij leerlingen die wel en niet de WO-toetsen hebben gemaakt 25

3.3 Weging 26

3.4 Resultaten op de drie toetsen 30

3.4.1 Jaarverschillen 30

3.4.2 Leerlingvariabelen 31

4 De resultaten van de itemresponsetheorie-analyses 35

4.1 De vergelijking over de jaren 35

4.2 Verschillen voor achtergrondvariabelen 37

4.2.1 Effecten voor de achtergrondvariabelen op leerlingniveau 38

4.2.2 Effecten voor achtergrondvariabelen op schoolniveau 42

5 Conclusies 47

5.1 Algemeen beeld 47

5.2 Invloed van achtergrondvariabelen 47

5.3 Discussie 48

Literatuur 51

Bijlagen 53

Bijlage 1 Effectschattingen voor alle vaardigheden gecategoriseerd naar achtergrondvariabele in jaargroep 8 in 2013 en 2014 54

Bijlage 2 Gemiddelden en standaarddeviaties per vaardigheid gecategoriseerd naar achtergrondvariabele in jaargroep 8 in 2013 en 2014 55

Bijlage 3 Verdeling van de achtergrondvariabelen in de onderzoekspopulatie in 2013 2014 die een eindtoets gemaakt heeft, al dan niet met de toetsen Wereldoriëntatie 56

Voorwoord

Dit rapport vormt de neerslag van de eerste meting van de wereldoriëntatievakken aardrijkskunde, geschiedenis en natuuronderwijs in groep 8 van het basisonderwijs. Hoewel dit de eerste meting betreft worden toch twee jaren met elkaar vergeleken doordat niet alleen de meting in 2014 maar ook die van 2013 in het onderzoek is betrokken. Op deze wijze kunnen we bij deze eerste meting ook al een trend aangeven.

De meting is een aanvulling op de jaarlijkse meting taal en rekenen (JMTR) die sinds 2008 uitgevoerd wordt naar aanleiding van de kwaliteitsagenda voor het Primair Onderwijs die in 2007 verscheen onder de titel 'Scholen voor morgen'. Door nu ook de wereldoriëntatievakken op te nemen worden deze eveneens betrokken in de landelijke monitor.

Bij uitvoering van dit onderzoek wordt gebruikgemaakt van data uit een toetsstelsel dat al op de meeste scholen wordt gebruikt, namelijk de Eindtoets. Naast de verplichte onderdelen voor onder andere taal en rekenen is er ook een optioneel onderdeel waarbij de leerlingen 90 vragen voorgelegd krijgen over de drie wereldoriëntatievakken. In tegenstelling tot rekenen en taal wordt dit onderdeel niet meegerekend in de standaardscore. Het aantal leerlingen dat deze toetsen maakt is lager dan het aantal bij het basisdeel van de Eindtoets Basisonderwijs, maar is met meer dan 100.000 leerlingen per jaar nog steeds aanzienlijk.

De uitkomsten van deze jaarlijkse peiling geven een indicatie van de stand van zaken met betrekking tot de wereldoriëntatievakken. Ook de effecten van verschillende achtergrondvariabelen worden hierbij meegenomen. De resultaten in dit onderzoek kunnen worden verwerkt in het jaarlijkse verslag 'De staat van het onderwijs' van de Inspectie van het onderwijs.

Bij de uitvoering van het project zijn diverse mensen betrokken geweest, waarvan er twee met name genoemd moeten worden: Loes Hiddink voor de onvolprezen scherpe pen waarmee de tekst verbeterd is, en Ron Engelen voor het wegwijs maken in de data. 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

Dit is het eerste verslag van het jaarlijks onderzoek naar de drie vaardigheden die gemeten worden in het onderdeel Wereldoriëntatie (WO) van de Eindtoets Basisonderwijs. Het betreft de vaardigheden Aardrijkskunde, Geschiedenis en Natuuronderwijs in jaargroep 8. Hiervoor is gebruikgemaakt van gegevens uit de Eindtoets van 2014 en 2013. Deze onderdelen maken weliswaar geen noodzakelijk onderdeel uit van de reguliere toetsing, maar scholen kunnen deze toetsen afnemen om ook voor deze vaardigheden een meting te hebben. Hoewel dit het eerste verslag is, betreft het onderzoek niet alleen de gegevens van 2014, maar ook die van het jaar 2013. In dat jaar zijn de vaardigheden gemeten met dezelfde toetsen, zodat een vergelijking eenvoudig is te maken.

Het onderzoek naar de vaardigheden Wereldoriëntatie is op twee manieren uitgevoerd, namelijk zowel zonder als met gebruikmaking van IRT (itemresponse theorie). Omdat bij deze toetsen, anders dan bij de meting Taal en Rekenen, sprake is van exact dezelfde itemverzameling in de verschillende jaren, kan voor een vergelijking van de resultaten met klassieke statistische technieken worden volstaan.

Ten eerste is dan ook onderzoek uitgevoerd op basis van de itemscores en de totaalscores en is de betrouwbaarheid van de opgaven onderzocht en de itemstructuur bekeken. Door naar de p-waarden te kijken is op opgavenniveau te zien wat leerlingen gemakkelijk en moeilijk vinden. Ook is op basis van de toetsscore onderzocht of groepen van elkaar verschillen. Het vergelijkend onderzoek over de jaren 2013-2014 kon op basis van de totaalscore uitgevoerd worden. Ten tweede zijn ten behoeve van de vergelijkbaarheid met de Meting Taal en Rekenen ook hier analyses uitgevoerd met IRT-methoden. Het voordeel van deze benadering is ook dat in dit geval het onderzoek direct uitgevoerd wordt op basis van de latente trek, oftewel direct op de onderliggende vaardigheid.

Doordat de toetsen voor de vaardigheden Wereldoriëntatie geen integraal onderdeel van de Eindtoets zijn en geen rol spelen in het advies voor het voortgezet onderwijs, zijn er minder scholen die deze aanvullende toetsen afnemen dan scholen die de Eindtoets afnemen. De representativiteit van de steekproef is daarmee ook minder gewaarborgd. Het eerste deel van het verslag betreft dan ook een analyse van de representativiteit van de steekproef. Hierbij is gekeken naar de verdeling van achtergrondvariabelen bij de leerlingen die de extra toetsen hebben gemaakt en bij leerlingen die de toetsen niet gemaakt hebben. Ook is er gekeken naar de verdeling van de scores op de verschillende eindtoetsonderdelen bij deze twee verschillende groepen.

Hoewel dit in feite de eerste meting betreft en deze dus als nulmeting kan worden beschouwd, is er toch al een vergelijking over jaren mogelijk voor de drie vaardigheidsschalen, namelijk voor de jaren 2013 en 2014. We vinden dan weliswaar kleine verschuivingen in de schaalwaarden, maar die zijn te klein om betekenisvol te kunnen worden genoemd. Daarbij lijkt er voor Geschiedenis, en in mindere mate voor Natuuronderwijs, een lichte vooruitgang te zijn en voor Aardrijkskunde en kleine achteruitgang. Geen van deze verschillen halen echter het significantieniveau en ze zijn ook zeker niet betekenisvol te noemen.

De conclusie is dan ook dat er tussen 2013 en 2014 niets is veranderd in het niveau van de leerlingen op deze vaardigheden, niet in positieve zin, maar ook niet in negatieve zin.

Uit de analyse op achtergrondvariabelen blijkt dat er wel enige verschillen zijn tussen de drie onderscheiden vaardigheden, maar er zijn ook grote overeenkomsten.

Zo behalen jongens op alle drie de vaardigheden hogere scores dan meisjes, waarbij het verschil bij Aardrijkskunde het grootst is. Vertraagde leerlingen presteren minder goed dan reguliere leerlingen, maar dat verschil is bij Natuuronderwijs het kleinst. Leerlingen met een formatiegewicht behalen lagere scores dan leerlingen zonder een gewicht, maar alleen bij Natuuronderwijs is er ook een verschil tussen leerlingen met een laag gewicht en een hoog gewicht, ten nadele van die laatste groep. De achtergrondvariabele die

de grootste verschillen laat zien in prestaties is het advies VO. Tussen leerlingen met een vmbo bb-advies en leerlingen met vwo-advies is een verschil variërend van 2,46 tot 2,86 standaarddeviaties aangetroffen. Dat zijn in termen van de educatieve onderzoeksliteratuur enorme verschillen. De verschillen zijn het kleinst bij Natuuronderwijs en het grootst bij Aardrijkskunde. Wel moet hier bij aangetekend worden dat de hier gevonden verschillen kleiner zijn dan verschillen die doorgaans worden gevonden bij de meeste taal- en rekenvaardigheden.

Ook is er een verschil van circa een hele standaarddeviatie tussen leerlingen die de Eindtoets Niveau en de leerlingen die de Eindtoets regulier hebben gemaakt.

Een variabele die er ook toe doet, is de taal die bij de leerlingen thuis wordt gesproken. Gegevens over die variabele zijn alleen beschikbaar voor 2014. We zien het grootste verschil in prestaties bij het contrast 'alleen een andere taal dan Nederlands' versus 'alleen Nederlands' en dan vooral bij Natuuronderwijs. Meer specifiek blijkt dat Surinaams/Antilliaans en vervolgens Turks en Marokkaans als thuistaal het grootste verschil opleveren.

De achtergrondvariabelen op schoolniveau laten weinig significante, laat staan betekenisvolle, effecten zien. Alleen bij de variabele Stratum zien we een betekenisvol effect, vooral bij Natuuronderwijs, maar gezien de sterke relatie met de variabele formatiegewicht ligt dat voor de hand. De andere variabelen, regio, mate van verstedelijking en schoolgrootte, doen er echter niet toe.

Het viel overigens ook op dat de leerlingen die de WO-toetsen maken iets beter presteerden op de reguliere eindtoets dan de leerlingen die dat niet deden. De populatie leerlingen verschilt ook enigszins. Bij het analyseren is daarvoor gecorrigeerd, zodat de gepresenteerde resultaten een correcte weerspiegeling geven van de landelijke verschillen.

Er is geen onderzoek gedaan naar mogelijke oorzaken voor de uitkomsten over de jaren heen. Zo is, anders dan bij regulier peilingsonderzoek (PPON), geen zicht op het onderwijsaanbod of veranderingen daarin, laat staan dat dergelijke gegevens direct gekoppeld kunnen worden aan de leerlingprestaties. Dat maakt het lastig om aan deze uitkomsten een interpretatie te verbinden. We weten dus niet op welke wijze bijvoorbeeld het gebruik van andere lesmethoden of het besteden van meer tijd aan bepaalde onderwerpen te maken heeft met deze resultaten. Om een relatie te leggen tussen eventuele waargenomen veranderingen of het uitblijven daarvan met wijzigingen in onderwijsbeleid is nader onderzoek nodig naar onderliggende factoren.

1 Inleiding

Sinds 2008 wordt een jaarlijkse peiling verricht die gericht is op het monitoren van het onderwijsniveau op het gebied van de taal- en rekenvaardigheid in het basisonderwijs. Dit onderzoek is uitgevoerd in het kader van de kwaliteitsagenda 'Scholen voor morgen'. In de eerste jaren werd dit onderzoek verricht in de jaargroepen 4 en 8 onder de naam Jaarlijks Peilingsonderzoek naar het Onderwijsniveau (JPON). Sinds 2013 is deze meting de Jaarlijkse Meting Taal & Rekenen (JMTR) genoemd en richt het onderzoek zich alleen op jaargroep 8.

Het uitgevoerde onderzoek heeft veel overeenkomsten met de periodieke peiling van het onderwijsniveau (PPON). Zo betreft dit ook periodiek onderzoek met als doel trends in het niveau van leerprestaties waar te nemen, zij het met een hogere frequentie. Een andere overeenkomst is dat ook hier resultaten op een vaste meetschaal worden gerapporteerd, die 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. Daar waar bij PPON de onderwerpen met veel verschillende opgaven gemeten worden, waardoor op verschillende subschalen in detail uitspraken te doen zijn, is dat bij de JMTR minder het geval. Ook is het aantal onderwerpen bij PPON breder. Niet alleen taal en rekenen, maar ook andere vaardigheden zijn daar gemeten, variërend van praktische vaardigheden in techniek tot burgerschaps- en sociale competenties.

Voor het eerst worden nu ook andere vakken betrokken bij het jaarlijkse onderzoek. Dit betreft vakken die vallen onder de noemer "Wereldoriëntatie". Deze worden traditioneel aangeduid als aardrijkskunde, geschiedenis en natuuronderwijs. Dat laatste vak betreft vooral onderwerpen die onder de biologie vallen, maar er worden ook enkele vragen betreffende natuurkunde en techniek gesteld.

Een verschil met de taal- en rekenvaardigheden is dat de wereldoriëntatievakken niet in de kwaliteitsagenda zijn opgenomen: daarin ligt de focus vooral op taal en rekenen. In het kabinetsbeleid na 2008 is deze focus blijven bestaan en dat heeft onder andere geleid tot nieuwe wetgeving en wetsvoorstellen waarin de positie van rekenen en taal en het toezicht daarop verder is verstevigd. Dit heeft onder andere geleid tot de invoering van de referentieniveaus voor taal en rekenen. Deze zijn ontwikkeld op basis de aanbevelingen van de Expertgroep Doorlopende Leerlijnen Taal en Rekenen (EGDLL, ofwel de commissie Meijerink, SLO, 2008). Daarin is vastgelegd wat leerlingen moeten kennen en kunnen voor taal en rekenen bij het verlaten van de basisschool en op de daaropvolgende drempelmomenten. Wereldoriëntatie is tot nu toe niet in dit beleid meegenomen. De laatste jaren zijn er echter ontwikkelingen die er op duiden dat ook deze vakken als belangrijk worden gezien. Zo zijn er toetsen ontwikkeld om ook het niveau voor deze vakken bij pabo-studenten te evalueren en te borgen (Mathijssen, 2014). Ook het verzoek om het huidige onderzoek uit te voeren kan gezien worden als een onderdeel van die ontwikkeling. Op termijn is het niet ondenkbaar dat ook voor deze vakken referentieniveaus ontwikkeld worden.

In aansluiting op de JMTR wordt ook bij dit onderzoek gebruikgemaakt van de uitkomsten van reeds afgenomen toetsen. Scholen beschikken doorgaans voor taal- en rekenvaardigheden over voldoende toetsgegevens, gebaseerd op leerlingvolgsystemen en eind- of entreetoetsen. Deze toetsen hebben vooral een functie in het ondersteunen van leerlingen. Ze kunnen voor leerkrachten ook dienen als input voor verbeterplannen op leerling- en groepsniveau, of voor de schoolleiding voor verbeteringen op school- en bestuursniveau. Onder bepaalde condities kunnen deze gegevens ook worden ingezet voor een landelijke niveaubepaling. Dat wordt gedaan in het project JMTR, maar kan ook toegepast worden voor de analyses van de vaardigheden bij wereldoriëntatie, hoewel het aantal beschikbare volgtoetsen voor die vakken zeer beperkt is.

Ook bij de meting voor wereldoriëntatie wordt dus gebruikgemaakt van bestaande toetsgegevens, en evenals bij de JMTR betreft het gegevens die zijn ingewonnen met de Cito Eindtoets Basisonderwijs (EB). In tegenstelling tot de taal- en rekenonderdelen is het onderdeel wereldoriëntatie in deze toets optioneel.

Het wordt niet meegerekend in de standaardscore die leidt tot een advies voor het VO. De school kan er ook voor kiezen dit onderdeel niet af te nemen.

Bij de peilingen van de JMTR is het startjaar 2008 als het uitgangspunt voor de vergelijking over de jaren beschouwd. Dat betekent dat de vaardigheid over de jaren heen vergeleken wordt met die van het jaar 2008, die is geïndexeerd op 250 (gemiddelde), met een schaling die zo gedaan is dat de standaarddeviatie gelijk is aan 50. Dit is overeenkomstig de wijze waarop dit in PPOON gerapporteerd wordt. Ondanks het ontbreken van referentieniveaus is het hiermee toch mogelijk goed interpreteerbare vergelijkingen over de jaren heen te maken.

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). Ook wordt beschreven hoe we de resultaten van de onderzoekspopulatie op de gemeten vaardigheden kunnen vergelijken met voorgaande jaren. Vervolgens beschrijven we de resultaten van de leerlingen in jaargroep 8 op een klassieke wijze waarbij de scores op de opgaven en de toetsen gebruikt worden (hoofdstuk 3). In dat hoofdstuk wordt ook de representativiteit van de steekproef nader onderzocht. In het vierde hoofdstuk worden de resultaten weergegeven op de vaardigheidsschalen. Daarbij worden de effecten van verschillende achtergrondvariabelen onderzocht. Hoofdstuk 5 bevat een samenvatting, conclusie en discussie.

2 Vraagstelling en aanpak

2.1 Vraagstelling

De belangrijkste vraag bij deze eerste peiling is: wat is het niveau van de leerlingen in jaargroep 8 van het basisonderwijs bij de vaardigheden van de vakken geschiedenis, aardrijkskunde en natuuronderwijs? Doordat we in deze peiling direct twee afnamejaren betrekken is het ook mogelijk een jaarvergelijking te maken en zo te zien of er een trend waar te nemen is in het niveau van de leerprestaties. Daarnaast is het doel van deze peiling na te gaan in hoeverre er verschillen in prestaties tussen bepaalde groepen leerlingen gevonden kunnen worden.

2.2 Methode

Voor deze jaarlijkse meting maken we gebruik van instrumenten die reeds in een ander kader zijn ingezet. Het zijn dus geen nieuwe toetsen die voor dit project zijn ontwikkeld. We maken gebruik van de toetsen die vallen onder de Eindtoets Basisonderwijs (EB) die wordt afgenomen aan het eind van jaargroep 8. Bij de EB hebben scholen de mogelijkheid het onderdeel wereldoriëntatie te toetsen. Hoewel dit onderdeel niet meetelt bij de eindscore die gebruikt wordt voor het advies voor het voortgezet onderwijs, geeft het wel een goede indruk van hoe breed een leerling zich ontwikkeld heeft op andere gebieden dan taal, rekenen en studievoordigheden. Hoewel het onderdeel wereldoriëntatie optioneel is, wordt het, met meer dan 100.000 afnamen per jaar, door een aanzienlijk aantal leerlingen in jaargroep 8 gemaakt.

Het onderdeel wereldoriëntatie verschilt op een aantal punten van het 'reguliere' deel (taal, rekenen en studievoordigheden) van de EB. Een eerste verschil is dat er slechts één variant van deze toets bestaat. Bij het reguliere deel van de EB zijn er twee varianten. Er is een aparte variant ontwikkeld, de Eindtoets Niveau (EN), voor leerlingen waarvan verwacht wordt dat ze hun voortgezet onderwijs zullen vervolgen in het praktijkonderwijs of de basisberoepsgerichte of kaderberoepsgerichte leerweg in het vmbo. Het voordeel van deze toetsvariant is dat in deze toets de opgaven meer op het niveau zijn van deze leerlingen. Overigens kunnen leerlingen op basis van hun score op deze variant wel een advies krijgen dat hoger ligt dan kaderberoepsgericht. De EN, die door ongeveer 15% van de leerlingen gemaakt wordt, heeft dan ook een overlap in opgaven met de EB Basis die door de overige leerlingen gemaakt wordt.

Van de EB zijn er ook twee digitale varianten beschikbaar – ook weer een niveautoets en een basistoets – die door een beperkt aantal leerlingen gemaakt worden: in totaal maakt minder dan 3% van de leerlingen een van de twee digitale toetsen.

Een tweede verschil met het reguliere deel van de EB is dat de toets Wereldoriëntatie niet ieder jaar vernieuwd wordt. Omdat de toets geen deel uitmaakt van de eindscore die gerelateerd is aan het advies voor het voortgezet onderwijs, worden het belang van de toets als minder groot gezien: de toets is hier mee minder *high stakes* dan het reguliere van de EB. De kans op fraude wordt gering geacht: de leerkracht of de school heeft daar namelijk geen baat bij, net zo min als de leerlingen daar baat bij zouden kunnen hebben. Om die reden wordt deze toets niet ieder jaar vervangen. Een nadelig effect daarvan zou kunnen zijn dat leerlingen de toets minder serieus maken dan de het reguliere deel. Die kans lijkt echter klein, omdat de toetsonderdelen afgenomen worden in exact dezelfde periode als het reguliere deel. Nadat in de ochtend de toetsen van het reguliere deel zijn afgenomen, worden in de middags de toetsen van wereldoriëntatie voorgelegd. Bovendien verschijnen de scores op de wereldoriëntatietoetsen op dezelfde rapportage als die van het reguliere deel van de EB. In welke mate er door de kandidaten anders op de verschillende onderdelen gereageerd wordt is lastig in te schatten.

Omdat de toets niet ieder jaar vernieuwd wordt, is het ook niet nodig is om voor de jaarvergelijking een aparte afname te organiseren. Voor de reguliere afname wordt op een beperkt aantal scholen een speciaal

ankerboekje ingezet waarin voor een deel (geheime) opgaven zitten die in een jaar daarvoor ook al in een apart ankerboekje zaten. Met deze methode van ankerboekjes is het mogelijk om de EB-gegevens over de jaren heen te vergelijken. Voor de toets Wereldoriëntatie was geen ankerboekje nodig aangezien in beide jaren dezelfde opgaven gebruikt zijn.

Omdat er geen verschillende varianten van de WO-toetsen bestaan en de toetsen over de jaren 2013 en 2014 niet van elkaar verschilden, is het, anders dan bij de JMTR, hier mogelijk om de toetsscores rechtstreeks met elkaar te vergelijken. Zelfs op opgavenniveau is hiermee dus een jaarvergelijking mogelijk. Naast een rechtstreekse vergelijking van scores is het echter nog steeds aan te raden ook analyses uit te voeren met behulp item-responsetechnieken. Hoe dat is uitgevoerd en wat daar de voordelen van zijn, wordt verder besproken in paragraaf 2.3 ("Hoe is er geanalyseerd?").

De toets Wereldoriëntatie is wat betreft het aantal items en de mogelijk te onderscheiden eenheden lang niet zo gedifferentieerd als in peilingsonderzoek van PPON gebruikelijk is. De toets omvat negentig opgaven: 30 voor natuuronderwijs, 30 voor aardrijkskunde en 30 voor geschiedenis. Een diepgaander onderscheid dan tussen deze drie subschalen is daarom moeilijk te maken. Bij reeds eerder uitgevoerd PPON-onderzoek was een fijner onderscheid wel mogelijk. Zo is voor natuuronderwijs bij de vierde peiling een tweetal rapporten (Balansen) uitgebracht waarbij onderscheid gemaakt is naar vier subvaardigheden voor biologie en een aparte subschaal voor voeding (Thijssen, Van der Schoot, & Hemker, 2011). Daarnaast was er een afzonderlijke peiling met een subschaal voor natuurkunde, een voor techniek en een voor praktische techniekopgaven (Kneepkens, Van der Schoot, & Hemker, 2011). Bij aardrijkskunde is er bij PPON onderscheid gemaakt tussen drie verschillende kennisgebieden voor topografie (Nederland, Europa en de Wereld) en drie inhoudelijke vaardigheden (Notté, Van der Schoot, & Hemker, 2011). Bij geschiedenis is bij PPON apart gerapporteerd voor zeven onderdelen, waarbij het onderdeel 'geschiedenis van de 20^e eeuw' ook nog eens opgedeeld was in 'kennis over de Tweede Wereldoorlog' en 'overige kennis van de 20^e eeuw' (Wagenaar, Van der Schoot, Hemker, 2010). Deze manier van onderscheid maken is met de gegeven aantallen opgaven voor Wereldoriëntatie in de EB niet mogelijk. Per 'PPON'-onderdeel is het aantal opgaven in de EB te beperkt.

Een ander onderscheid tussen PPON en de WO-toets in de EB is de variatie in itemvormen. De EB heeft als vaste vraagvorm een vierkeuze format, in te vullen op een afzonderlijk antwoordblad, terwijl in PPON een variatie aan itemvormen wordt gehanteerd: meerkeuzevragen (2-, 3-, 4-keuze), open vragen (kort antwoord – lang antwoord), alles doorgaans aangeboden in de vorm van in te vullen toetsboekjes. Ook is er bij PPON (bijvoorbeeld Techniek) sprake van praktische opdrachten vergezeld van een scorevoorschrift.

Een verder verschil met PPON is de afnameconditie. Bij PPON worden alle toetsen en taken afgenomen door een getrainde toetsleider. In het geval van de EB is de eigen leerkracht de toetsleider. Dat zou mogelijk een aantasting van de objectiviteit van de afname kunnen vormen. Omdat de toets wordt afgenomen onder dezelfde condities als de rest van de EB zou men kunnen stellen dat de motivatie van de leerlingen groter is dan bij de reguliere afname van PPON. Die afname is doorgaans in de periode na de eindtoets in mei of juni, en leerlingen weten dat er voor hen persoonlijk niets van afhangt. Dat verschil in conditie kan bij rekenen en taal tot aanzienlijke prestatieverschillen leiden (zie Hemker, 2013). In hoeverre dat bij de toets WO ook het geval is, staat nog te bezien. Immers enerzijds kan gesteld worden dat bij de leerlingen wellicht ook bekend is dat het WO-deel voor hun VO-advies niet van belang is. Anderzijds zal de school wel graag zien dat leerlingen ook hierop hun best doen, want de uitkomsten worden op de school-rapportage vermeld en de school zal zich hiervoor ook moeten verantwoorden.

Voor een vergelijking over de twee jaren heen speelt variatie in afnameconditie waarschijnlijk geen rol, omdat die condities immers in de meeste gevallen niet over de jaren heen zullen variëren. Het is dus geen probleem om een trend in leerprestaties in beeld te brengen. Voor een vergelijking van deze uitkomsten met die van andere onderzoeken zijn de eventuele verschillen in afnameconditie echter wel relevant.

Een belangrijk aandachtspunt in de vergelijking over de tijd is wel de samenstelling van de groep leerlingen. Indien die samenstelling verandert, kan dat gevolgen hebben voor de uitkomsten op de toets.

De samenstelling van de leerlingengroep in de verschillende jaren zal dus gecontroleerd moeten worden op representativiteit.

Zoals gesteld zijn de gegevens verzameld tijdens de Eindtoetscampagne. Deze liep in 2013 van dinsdag 5 tot en met donderdag 7 februari. In 2014 was dat van dinsdag 11 tot en met donderdag 13 februari. In beide jaren werd dinsdag de toets Natuuronderwijs afgenomen, op woensdag Aardrijkskunde en op donderdag Geschiedenis.

Behalve de item- en toetsscores zijn tijdens de EB-afname ook achtergrondgegevens verzameld. Omdat de onderzoekspopulatie de leerlingen in het reguliere basisonderwijs in Nederland betreft, zijn de leerlingen die niet aantoonbaar hun onderwijs op een reguliere basisschool in Nederland gevolgd hebben niet in de data opgenomen. Dat betekent bijvoorbeeld dat leerlingen die in het speciaal onderwijs zitten, leerlingen die de toets in het buitenland hebben gemaakt of in het voortgezet onderwijs zitten, niet zijn meegenomen in het onderzoek. Een exclusie criterium is ook de periode waarin de EB is gemaakt. Leerlingen die een digitale variant gemaakt hebben kunnen dat ook op een ander tijdstip hebben gedaan. Bij het reguliere deel van de EB is van ongeveer 93% van de leerlingen vastgesteld dat ze zeker in het reguliere basisonderwijs zitten en de toets in de vastgestelde periode gemaakt hebben. Bij de leerlingen die naast de EB ook de toetsen Wereldoriëntatie gemaakt hebben, was dit percentage hoger: meer dan 96%. Het blijkt dat geen van de leerlingen die op een ander moment de EB gemaakt hebben, de toetsen Wereldoriëntatie gemaakt hebben. Ook leerlingen in het speciaal onderwijs en leerlingen in het buitenland hebben minder vaak toetsen Wereldoriëntatie gemaakt. In paragraaf 2.2.2 en hoofdstuk 3 gaan we verder in op de steekproef en de representativiteit.

2.2.1 Kwaliteit van de meetinstrumenten

De eigenschappen van de Eindtoets worden elk jaar gepubliceerd in een afzonderlijk document dat op de website van Cito geplaatst wordt. Zowel voor 2013 als voor 2014 is er een rapportage verschenen (Cito, 2013; Cito, 2014). We mogen hier stellen dat de psychometrische kwaliteit bij de Eindtoets onder andere gewaarborgd is door de gehanteerde procedures. Nadere informatie hierover is te vinden in de uitgebreide verantwoording van de Eindtoets van 2010 (Van Boxtel e.a., 2012; zie ook www.toetswijzer.nl). Deze verantwoording is aangeboden aan de COTAN (Commissie Test Aangelegenheden Nederland) en is daar op alle relevante criteria als voldoende of goed beoordeeld.

Het deel van de EB dat betrekking heeft op Wereldoriëntatie bestaat uit een vaste set van opgaven die door alle leerlingen in het onderzoek gemaakt is. Als we kijken naar de toetseigenschappen, dan is te zien dat de betrouwbaarheid van de drie afzonderlijke schalen (Natuuronderwijs, Aardrijkskunde, Geschiedenis) hoog genoeg is om volgens de COTAN-normen als voldoende aangemerkt te worden voor subschalen bij individuele metingen. Van alle schalen is de betrouwbaarheid (Cronbach's alpha en Lambda-2) namelijk hoger dan 0,70. Hierbij kan opgemerkt worden dat in de rapportage van de EB vooral ook gerapporteerd wordt over de score op de drie WO-toetsen samen. De betrouwbaarheid van de gehele WO-schaal is 0,90, wat hoog genoeg is voor belangrijke beslissingen om een COTAN-beoordeling "goed" te krijgen. In tabel 2.1 staan o.a. de betrouwbaarheidsmaten Cronbach's alpha en Lambda-2 gegeven. Die statistische maten zijn bij deze schalen (vrijwel) aan elkaar gelijk.

Tabel 2.1 Toetseigenschappen WO-schalen (2013 en 2014)

Betrouwbaarheid	Vaardigheid			
	Natuuronderwijs	Aardrijkskunde	Geschiedenis	WO (samen)
Cronbach's Alpha	0,73	0,78	0,77	0,90
Lambda-2	0,73	0,78	0,77	0,90
Scoreverdeling				
Gemiddelde score	19,85	20,02	21,16	61,03
Standaarddeviatie	4,60	5,00	4,83	12,61
Scheefheid	-0,37	-0,38	-0,60	-0,43
Kurtosis	-0,26	-0,41	-0,01	-0,26

In tabel 2.1 zijn ook de overige toetseigenschappen gegeven voor de subschalen en de gehele WO-schaal. Deze eigenschappen gelden voor de geobserveerde verdelingen op basis van de 234.396 observaties in 2013 en 2014 samen. Verderop in dit verslag zullen de verschillen tussen de jaren gegeven worden.

Voor wat betreft de betrouwbaarheid zijn er geen verschillen tussen de jaren gevonden. Op basis van deze toetseigenschappen is op te merken dat de toets Geschiedenis de gemakkelijkste van de drie is en de toets Natuuronderwijs de moeilijkste, maar dat het verschil in moeilijkheid niet heel erg groot is (1,3 punt).

De scoreverdeling bij de toets Geschiedenis is iets scheef, wat samenhangt met het feit dat deze toets ook relatief aan de gemakkelijke kant is. De geobserveerde scoreverdeling en de geschatte verdeling op grond van veronderstelde normaliteit liggen niet erg uit elkaar.

In tabel 2.2 zijn enkele eigenschappen van de opgaven gegeven. Aangezien deze toets ongewijzigd ook in 2015 afgenomen wordt, gaan we niet op de individuele opgaven in. Wel kunnen algemene gegevens, zoals de verdeling van de p-waarden en de item-testcorrelaties gerapporteerd worden. De p-waarde varieert van 0,38 tot en met 0,88, terwijl de item-testcorrelaties variëren tussen de 0,08 en de 0,44. De gemiddelde item-testcorrelatie r_{it} ligt voor alle vaardigheden tussen 0,20 en 0,30, wat door de COTAN als voldoende wordt gekwalificeerd.

Tabel 2.2 Itemeigenschappen per vaardigheidsschaal: Natuuronderwijs (NA), Aardrijkskunde (AA) en Geschiedenis (GE)*

Itemeigenschap	P-waarde			Item-testcorrelatie (r_{it})		
	NA	AA	GE	NA	AA	GE
gemiddelde	0,66	0,67	0,71	0,25	0,29	0,28
SD	0,13	0,13	0,11	0,05	0,07	0,05
minimum	0,40	0,38	0,47	0,08	0,10	0,15
maximum	0,88	0,86	0,86	0,34	0,44	0,38

*In de analyse zijn de gegevens voor 2013 en 2014 samengenomen.

In tabel 2.3 wordt de verdeling van de p-waarden en de item-testcorrelaties in gedetailleerde vorm weergegeven. Uit die tabel valt af te leiden dat rond de 50% van de opgaven p-waarden heeft onder de 0,70 (maar niet lager dan 0,38) en 50% boven de 0,70 (maar niet hoger dan 0,88). Bij de item-testcorrelaties zien we dat minder dan 10% van de r_{it} -waarden van alle WO-schalen onder de 0,20 komt, al is er enige variatie over de schalen. Bij Natuuronderwijs (NA) zien we dat het iets meer dan 10% van de opgaven is, terwijl het bij Geschiedenis (GE) slechts 3% van de opgaven betreft. Meer dan 90% van opgaven heeft een r_{it} -waarde groter dan 0,20 en meer dan 30% een r_{it} -waarde groter dan 0,30 (variërend van 17% bij NA tot 40% bij GE).

Tabel 2.3 Verdeling van p-waarden en item-test-correlaties (r_{it}) voor de vaardigheden Natuuronderwijs (NA), Aardrijkskunde (AA) en Geschiedenis (GE) in 2013 en 2014*

P tussen	aantal opgaven			P-waarde	% opgaven			P-waarde	% opgaven		
	NA	AA	GE		NA	AA	GE		NA	AA	GE
0,3 0,4	1	1	0	<,4	3	3	0	>=,3	100	100	100
0,4 0,5	4	4	2	<,5	17	17	7	>,4	97	97	100
0,5 0,6	6	3	1	<,6	37	27	10	>,5	83	83	93
0,6 0,7	5	6	13	<,7	53	47	53	>,6	63	73	90
0,7 0,8	8	14	8	<,8	80	93	80	>,7	47	53	47
0,8 0,9	6	2	6	<,9	100	100	100	>,8	20	7	20

r_{it} tussen	aantal opgaven			r_{it} -waarde	% opgaven			r_{it} -waarde	% opgaven		
	NA	AA	GE		NA	AA	GE		NA	AA	GE
0,0 0,1	1	1	0	<,1	3	3	0	>,0	100	100	100
0,1 0,2	3	2	1	<,2	13	10	3	>,1	97	97	100
0,2 0,3	21	16	17	<,3	83	63	60	>,2	87	90	97
0,3 0,4	5	10	12	<,4	100	97	100	>,3	17	37	40
0,4 0,5	0	1	0	<,5	100	100	100	>,4	0	3	0

* De gegevens zijn voor deze twee jaren samengenomen.

Ook is onderzocht of de opgaven wel op de juist schaal liggen. Zo is gekeken of de hoogste correlatie van een item ook daadwerkelijk gevonden wordt met de schaal waar dit item bij hoort. Bij alle opgaven was dit het geval.

De geobserveerde correlatie tussen de schaal Natuuronderwijs en de andere twee schalen ligt iets hoger dan 0,60. De correlatie tussen Aardrijkskunde en Geschiedenis is nog hoger: 0,70. Als we deze correlaties bekijken na attenuatiecorrectie, dat wil zeggen na rekening te hebben gehouden met het feit dat de schalen geen perfecte betrouwbaarheid hebben, dan is te zien dat de correlatie op vaardigheidsniveau varieert van 0,81 tot 0,90. De correlatie in 2013 en 2014 tussen deze drie schalen verschillen bijzonder weinig van elkaar (de verschillen liggen tussen 0,007 en 0,002) zodat ervoor gekozen is in tabel 2.4 de gemiddelde correlaties over de twee jaren te rapporteren.

Tabel 2.4 De correlaties tussen de schalen wereldoriëntatie en de schalen taal en rekenen, gemiddeld voor 2013 en 2014

Geobserveerde correlaties	Wereldoriëntatie			Taal*			Rekenen*		
	NA	AA	GE	SP	BL	WO	GG	BPV	MMTG
Natuuronderwijs (NA)		0,61	0,62	0,26	0,57	0,55	0,41	0,45	0,46
Aardrijkskunde (AA)	0,61		0,70	0,34	0,56	0,51	0,50	0,54	0,53
Geschiedenis (GE)	0,62	0,70		0,32	0,60	0,54	0,46	0,49	0,48

***Taal:** SP=spelling; BL=begrijpend lezen; WO=Woordenschat; **Rekenen:** GG=getallen en getalsbewerkingen, BPV=breuken, procenten, verhoudingen; MMTG=meten, meetkunde, tijd en geld.

Er is ook gekeken naar de correlatie van de WO-schalen met de schalen rekenen en taal van de Jaarlijkse meting Taal en Rekenen (JMTR). Ook hier verschillende de correlaties niet heel erg veel over de jaren. Met verschillen tussen de -0,02 en de 0,04 is dat wel iets meer dan tussen de WO-schalen onderling, maar de verschillen lijken klein genoeg om ook hier het gemiddelde van de correlaties te kunnen nemen. We zien dan dat de correlaties tussen de WO-schalen onderling hoger zijn dan de correlaties met de andere schalen. Daarbij is de correlatie tussen Begrijpend lezen (BL) en Woordenschat enerzijds en de WO-

schalen anderzijds relatief het hoogst. Dat alle correlaties aan de hoge kant zijn is te begrijpen als we ons rekenschap geven van het feit dat het om een toets gaat met als functie het geven van een schooladvies.

Tabel 2.5 De gemiddelde correlaties over de twee jaren 2013 en 2014 tussen de WO-schalen en de schalen taal en rekenen bij alle leerlingen in de onderzoekspopulaties van 2013 en 2014 met een vmbo-gt/havo-advies

	Wereldoriëntatie			Taal			Rekenen		
	NA	AA	GE	SP	BL	WO	GG	BPV	MMTG
Natuuronderwijs (NA)		0,43	0,45	-0,11	0,33	0,36	0,03	0,09	0,14
Aardrijkskunde (AA)	0,43		0,55	-0,03	0,26	0,25	0,14	0,19	0,21
Geschiedenis (GE)	0,45	0,55		-0,06	0,33	0,32	0,07	0,12	0,13

*zie onderschrift tabel 2.4.

We geven hier het voorbeeld van de groep leerlingen met het advies vmbo-gt/havo. Deze groep heeft een gemiddelde standardscore van 535, dezelfde score die als gemiddelde voor de hele populatie geldt. Bij deze meer homogene groep van 25882 leerlingen is te zien dat de correlaties tussen de schalen lager zijn dan in de gehele populatie, zie tabel 2.5. De grootste daling in hoogte van de correlatie is te vinden tussen de WO-schalen enerzijds en de rekenschalen en spelling anderzijds. Daar is nog amper sprake van correlatie in deze subgroep. Ook bij de andere twee taalschalen zien we een aanzienlijke daling in correlatie (tussen 0,20 en 0,30). De kleinste achteruitgang in hoogte van de correlatie is te vinden bij de samenhang tussen de WO-schalen onderling: daar is de samenhang nog steeds behoorlijk.

Vergelijkbare resultaten worden ook gevonden bij andere homogene groepen met eenzelfde doorstroomadvies. Dit toont aan dat de WO-schalen afzonderlijk een eigen betekenis hebben en onderling meer samenhang vertonen dan met de overige schalen. Ondanks het feit dat deze toetsen in dit onderzoek worden ingezet met een ander doel dan waarvoor zij zijn bedoeld, namelijk het geven van een rapportage op het niveau van de drie te onderscheiden WO-schalen in plaats van op een enkele WO-schaal, voldoen zelfs deze subschalen aan eisen van de COTAN. Merk ook op dat de scores van de WO-schaal geen directe invloed hebben op de standardscore (501-550) die gebruikt wordt bij de prognose voor de kans van slagen in verschillende vormen van voortgezet onderwijs voor individuele leerlingen.

In deze studie gaat het dus, net als in de JMTR, niet om een totaalscore op meerdere onderdelen op de Eindtoets, de standardscore, maar om de uitkomsten op delen van de toets. In dit rapport worden verschillende onderdelen van de toets afzonderlijk gebruikt om een beeld te schetsen van de kwaliteit van het onderwijs op systeemniveau. 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 & Mulder, 2014).

2.2.2 Steekproeftrekking en analyse

Er is sprake van een steekproefprocedure, zodat de resultaten als representatief mogen worden gezien voor het niveau in Nederland op de beide meetmomenten. Voordat deze steekproef getrokken werd, is eerst een schifting gemaakt. Uit de steekproef zijn leerlingen weggelaten 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 (speciaal onderwijs en voortgezet onderwijs), of aan een specifieke aanduiding in het bestand dat het geen school uit het basisonderwijs betreft. Tevens zijn de leerlingen die de digitale versie van de Eindtoets hebben gemaakt verwijderd, omdat deze leerlingen de toets op een ander moment gemaakt hebben. Daarnaast zijn de scholen afgezet tegen de lijst met scholen voor basisonderwijs van DUO. Als wel de hoofdvestiging (00) in de lijst staat, maar niet de specifieke vestiging, dan is de school

alsnog opgenomen (dit betreft een zeer beperkt aantal scholen). Bovengenoemde exclusiecriteria zijn ook in de eerdere onderzoeken van JMTR gehanteerd.

Van de 161.192 leerlingen die in 2014 een van de eindtoetsen gemaakt heeft, vallen er hierdoor in totaal 9.679 leerlingen uit de dataset, omdat zij niet tot de onderzoekspopulatie behoren. Als we daarbij alleen de scholen willen opnemen die ook in 2013 in de data voorkwamen dan vallen er 3.199 extra leerlingen af, zodat er in totaal 12.878 buiten de onderzoeksgroep vallen. We willen met name richten een vergelijkbare dataset om zo een goede jaarvergelijking te kunnen maken. We streven dus naar zoveel mogelijk overlap met de dataset van 2013. De basis voor de analyses is om die reden de dataset met 148.314 leerlingen in 2014 (zie tabel 2.6). Ter controle zijn overigens ook analyses uitgevoerd met de iets grotere dataset, met 151.513 leerlingen. Het percentage uitval van leerlingen in beide datasets is vergelijkbaar met de percentages in de datasets van 2013, waarbij de uitval bij de kleinere dataset 1% groter is en bij de grotere dataset 1% kleiner dan in 2013.

Tabel 2.6 Aantal leerlingen in de analyses in 2013 en 2014

	2013			2014		
	Geen WO	Wel WO	Totaal	Geen WO	Wel WO	Totaal
Niet in onderzoekspopulatie	6698	4732	11430	8082	4796	12878
Wel in onderzoekspopulatie	30244	120512	150756	34430	113884	148314
Totaal	36942	125244	162186	42512	118680	161192

*WO = toets wereldoriëntatie.

Binnen deze selectie zijn er leerlingen die de toetsen voor Wereldoriëntatie wel gemaakt hebben en leerlingen die dat niet hebben gedaan. Bij de leerlingen die buiten de onderzoekspopulatie vallen, maakt slechts 41,4% in 2013 en 37,2% in 2014 de WO-toetsen, terwijl dit percentage bij de leerlingen die in de onderzoekspopulatie vallen aanzienlijk hoger is: 79,9% in 2013 en 76,8% in 2014.

De leerlingen in de onderzoekspopulatie die de Eindtoets hebben gemaakt, kunnen worden gezien als een landelijk representatieve steekproef. Het is nu van belang te controleren of de set leerlingen die daarvan de WO-toetsen hebben gemaakt ook gezien kan worden als een willekeurige steekproef uit die populatie. We onderzoeken of deze groep hogere of lagere standaardscores heeft dan de groep die geen WO-toetsen gemaakt heeft. Ook onderzoeken we of de verdeling van de achtergrondvariabelen voor de verschillende groepen vergelijkbaar is. Dit onderzoek wordt beschreven in hoofdstuk 3. Deze analyse betreft de gehele groep van kandidaten. Als uit de analyse van de achtergrondvariabelen blijkt dat de WO-groep geen representatieve steekproef is, zal bij de rapportage van de uitkomsten in schaalscores daarvoor een correctie moeten worden toegepast. Daarvoor zullen wegingsprocedures moeten worden ingezet. Dit onderzoek zal voor zowel de leerlingen uit 2013 als uit 2014 worden uitgevoerd. De referentiegroepen zijn de gehele onderzoekspopulaties van 2013 en 2014.

In hoofdstuk 4 worden analyses uitgevoerd met behulp van de item-responsetheorie. Hoewel bij de JMTR twee verschillende steekproeven getrokken moesten worden, respectievelijk voor de kalibratie van de schalen en de analyse op jaareffecten en achtergrondvariabelen, is dat voor het onderzoek naar de WO-schalen niet noodzakelijk. Bij de WO-schalen is de overlap in toetsvarianten al optimaal omdat alle leerlingen in beide jaren dezelfde opgaven hebben gemaakt. Het is daarom ook niet nodig om voor de schaling en de analyse afzonderlijke steekproeven te trekken. Vanwege de uniformiteit in analyses, vergeleken met de JTMR, is dat echter toch gedaan. Bij de OPLM-analyses zijn vijf verschillende steekproeven uit de (ongewogen) populatie getrokken om de itemkenmerken te schatten.

Er bleek weinig steekproeffluctuatie te zijn. De kenmerken van deze groepen op achtergrondvariabelen zijn volkomen vergelijkbaar met de totale groep en die zijn terug te vinden in het document *Terugblik bij de Eindtoets 2014* (Cito, 2014). We mogen dus stellen dat deze steekproeven een representatief beeld geven van de populatie leerlingen in Nederland in jaargroep 8. In hoofdstuk 4 wordt ingegaan op de kwaliteit van

de drie WO-schalen in OPLM. Er zal daarbij niet ingegaan worden op de eigenschappen van de afzonderlijke opgaven aangezien de toetsen nogmaals worden afgenomen in 2015. De opgaven moeten dus nog geheim blijven om een objectieve afname te waarborgen.

Het is vanwege de populatieafhankelijkheid van de parameterschattingen niet noodzakelijk dat de gehanteerde steekproeven bij de kalibratie van de schalen representatief zijn. Voor de analyses van de effecten van de achtergrondvariabelen is dit echter wel noodzakelijk. Op basis van de resultaten in hoofdstuk 3 is gekozen voor een gewogen steekproef. Deze steekproef wordt verder beschreven in hoofdstuk 3.

2.2.3 Achtergrondvariabelen

Ten behoeve van een nadere analyse van de drie genoemde vaardigheden voor de WO-schalen zijn er ook een aantal achtergrondkenmerken in het onderzoek betrokken. De eerste achtergrondvariabele waarop analyses zijn uitgevoerd is het jaar van afname. Er wordt onderzocht of de resultaten van 2013 en 2014 per vaardigheid van elkaar verschillen. Dat geldt ook voor alle andere analyses die uitgevoerd worden met betrekking tot de achtergrondvariabelen. Deze variabelen zijn te onderscheiden in de volgende leerlingkenmerken en schoolkenmerken.

Leerlingkenmerken

- Geslacht: jongens-meisjes;
- Leertijd: leerlingen die al of niet eens hebben gedoubleerd, resp. regulier en vertraagd;
- Formatiegewicht: gewicht van de leerling voor de formatieregeling op grond van opleiding en herkomst (alleen bij de oude regeling) van de ouders; verdeeld in drie categorieën, geen gewicht (0.00), laag gewicht (0.30) en hoog gewicht (1.20)¹;
- Thuis taal: algemeen - drie niveaus (1) Leerling spreekt thuis alleen Nederlands, (2) Leerling spreekt naast Nederlands ook een andere taal, of (3) Leerling spreekt thuis geen Nederlands, maar alleen een andere taal;
- Thuis taal: detail: alleen Nederlands gesproken versus verschillende talen: Turks, Marokkaans, Surinaams/Antilliaans, en andere, al dan niet in combinatie met Nederlands;
- Advies VO: ingeschat niveau vervolgonderwijs door de leerkracht (5 hoofdniveaus van vmbo-bb tot en met vwo)²;
- Gemaakte toets: Eindtoets Basis of Eindtoets Niveau.

¹ Bij de opleidingen van de ouders wordt onderscheid gemaakt tussen een drietal categorieën:

- Categorie 1: > de ouder heeft maximaal (speciaal) basisonderwijs gehad.
- Categorie 2: > de ouder heeft maximaal lbo/vbo, praktijkonderwijs, vmbo basis- of kaderberoepsgerichte leerweg gedaan, of de ouder heeft maximaal twee leerjaren van een andere opleiding in het VO gedaan.
- Categorie 3: > de ouder heeft meer dan twee jaar mavo, havo, vwo dan wel vmbo gemengde of theoretische leerweg gevolgd.

Op basis van deze categorieën worden de volgende formatiegewichten toegekend:

- 0.00 (geen gewicht) voor leerlingen met (minstens) één ouder uit categorie 3;
- 0.30 voor leerlingen die niet gewicht 0.00 en niet gewicht 1.20 hebben;
- 1.20 voor leerlingen die geen ouder hebben in categorie 3 en (minstens) één uit categorie 1.

² Alleen de hoofdniveaus worden gepresenteerd, de waarden voor de vier tussenliggende combinatieniveaus liggen daar ook echt ongeveer halverwege tussenin.

Schoolkenmerken

- Stratum: schoolindeling op basis van het percentage leerlingen met een formatiegewicht, verdeeld in drie categorieën; 1 = weinig (<10%), 2 = matig (10-25%) en 3 = veel (>25%);
- Regio: locatie van de school in Nederland: noord, oost, west, zuid (volgens opdeling CBS);
- Urbanisatiegraad: locatie van de school (5 categorieën; van zeer sterk stedelijk tot en met niet stedelijk);
- Schoolgrootte (200 of minder leerlingen versus meer dan 200 leerlingen).

In de loop van de onderzoeksperiode van de JMTR (voorheen JPON) vanaf 2008 tot 2013 zijn de definities van een aantal variabelen veranderd, zoals die van het formatiegewicht en van stratum, maar tussen 2013 en 2014 waren er geen veranderingen. In tegenstelling tot de JMTR kan zodoende wel met een vaste definitie voor formatiegewicht en stratum gewerkt worden. De definitie van formatiegewicht staat in de 'Brochure nieuwe gewichtenregeling basisonderwijs', april 2008, van het cfi (zie ook www.cfi.nl) en is in vereenvoudigde vorm gegeven in voetnoot 1.

De achtergrondvariabelen voor de leerlingen worden bij de Eindtoets verzameld via het antwoordblad dat bij deze toets hoort en waarop de leerling en de leerkracht een aantal gegevens invoeren. In 2013 (en 2012) werd de achtergrondvariabele thuistaal niet opgevraagd. Dat was in 2014 wel weer het geval.

2.3 Hoe is er geanalyseerd?

Hieronder geven we beknopt weer hoe we de resultaten van 2014 hebben vergeleken met die van 2013. Dit wordt gevolgd door een beschrijving van de wijze waarop de effecten van de leerling- en schoolkenmerken zijn geanalyseerd. Bij deze analyse is het van belang of er al dan niet sprake is van interactie tussen afnamejaar en de andere achtergrondvariabelen. Vervolgens worden de rapportageschalen beschreven. Deze paragraaf wordt afgesloten met de beschrijving van de interpretatie van de effecten.

2.3.1 Vergelijking tussen en over de jaren

Een voorwaarde om vaardigheden over de verschillende jaren direct met elkaar te kunnen vergelijken, is dat de vaardigheden op dezelfde schaal gemeten zijn. Aan deze voorwaarde is in dit geval voldaan omdat alle leerlingen precies dezelfde toetsen hebben gemaakt. Ook de achtergrondvariabelen zijn op dezelfde wijze gemeten. Een mogelijke andere voorwaarde is het ontbreken van itempartijdigheid (*itembias*) als gevolg van afnamefouten. Aangezien fysiek dezelfde toetsen gebruikt worden, is dat ook niet het geval.

Dit betekent, zoals eerder gemeld, dat een analyse rechtstreeks op de totaalscore mogelijk is. Daarnaast wordt ook gebruikgemaakt van een andere analysetechniek om een meetschaal te verkrijgen: het One Parameter Logistic Model (OPLM) (Verhelst, 1993; Verhelst en Glas, 1995). Dit is een variant van het item responsemodel en daarmee zijn meetschalen geconstrueerd voor de drie afzonderlijke onderwerpen: Natuuronderwijs, Aardrijkskunde en Geschiedenis.

Er is een drietal redenen om deze schaling toe te passen. Ten eerste is het hiermee mogelijk om de scores om te zetten naar een normaal verdeelde vaardigheidsverdeling. De scores volgen bijna een normaalverdeling, zoals aangegeven is in tabel 2.1, maar niet geheel. Vooral de schaal voor Geschiedenis is, omdat deze relatief gemakkelijk is, iets scheef verdeeld met een plafondeffect voor de hogere scores. Dit kan met OPLM getransformeerd worden naar een normaalverdeling.

Ten tweede is het op deze wijze mogelijk direct te meten op de vaardigheidsschaal. Dat houdt in dat gecorrigeerd wordt voor het feit dat de metingen met schalen niet perfect betrouwbaar zijn. Er wordt zodoende gecorrigeerd voor extra variantie door meetfouten.

De derde reden is van praktische aard. Op deze wijze realiseren we een goede vergelijkbaarheid met de resultaten van de andere jaarlijkse peilingen. In de JMTR wordt namelijk ook gewerkt met deze schalingstechniek, zij het daar uit noodzaak, omdat daar niet alle kandidaten dezelfde toets hebben

gemaakt. Zo zijn de gevonden effecten in het onderzoek naar de WO-schalen ook te vergelijken met de resultaten uit de JMTR.

Bij de schaling moet een keuze gemaakt worden betreffende de transformatie. Binnen het peilingsonderzoek PPO is het tot dusver gebruikelijk geweest de scores voor de populatie van het eerste peilingsjaar te transformeren naar een vaardigheidsschaal met een gemiddelde van 250 en een standaardafwijking van 50. Op deze wijze kunnen de prestaties op de verschillende toetsen over de jaren heen met elkaar worden vergeleken. In dit geval is het eerste peilingsjaar 2013, en zal het gemiddelde van dat jaar als referentie dienen.

Voor een uitgebreide technische verantwoording van de statistische analyse en de rapportagemethodiek die hiervoor is gebruikt wordt verwezen naar de technische rapportages van de peilingen JMTR in eerdere jaren (Hemker & Van Weerden, 2009; Hemker, Kuhlemeier en Van Weerden, 2010; Hemker, Kordes & Van Weerden, 2011; Van Weerden, Hemker, Straat & Mulder, 2012).

Voor het ontwikkelen van de meetschalen is het geen probleem, maar voor het vergelijken van prestaties in groep 8 over de jaren heen moet ook rekening gehouden worden met de samenstelling van de responsgroep. Als de samenstelling gelijk is, dan is ook een directe vergelijking van de prestaties mogelijk, maar in het geval van verschillen in samenstelling moet onderzocht worden waar die wijziging vandaan zou kunnen komen. Veranderingen in de samenstelling van een responsgroep kunnen het gevolg zijn van zogenaamde steekproeffluctuaties en van 'echte' veranderingen in de samenstelling van de populatie. Het probleem van steekproeffluctuaties zal zich voor jaargroep 8 niet zo gauw voordoen. Er is namelijk in elk jaar 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. We moeten dan onderzoeken of dit ook geldt voor de selectie van leerlingen die de WO-toetsen gemaakt hebben. Dit wordt verder beschreven in hoofdstuk 3.

Opgemerkt moet worden dat correcties voor veranderingen in de verdeling die werkelijke veranderingen in de populatie representeren, vreemde resultaten op kunnen leveren. Stel dat de leerlingen door de tijd heen meer naar het vwo verwezen worden omdat zij vaardiger geworden zijn, en voor deze achtergrondvariabele wordt gecorrigeerd, dan zal de toename in vaardigheid onterecht gecorrigeerd worden. Een ander punt van aandacht is de verandering van de verdeling van leertijd over de jaren heen, waarbij correctie ook lastig uit te voeren is. Vandaar dat wij voor jaargroep 8 ongecorrigeerde gegevens verstrekken. In dit geval wordt gebruikgemaakt van een speciaal programma voor regressieanalyse van vaardigheidsscores geschat met behulp van OPLM, te weten SAUL: Structural Analysis of Univariate Latent variabels (Verhelst & Verstralen, 2002). Hiermee zijn de marginale effecten en de gecorrigeerde effecten berekend. 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. Dit zijn resultaten gecorrigeerd voor variantie van mogelijke meetfouten van de toetsen.

2.3.2 Analyses voor de overige variabelen

Als er geen interactie is tussen het afnamejaar en andere achtergrondvariabelen, bijvoorbeeld als het verschil tussen jongens en meisjes in 2013 en 2014 gelijk aan elkaar is, dan kunnen de verschillen op deze variabelen over de jaren heen rechtstreeks gerapporteerd worden. Of dat mogelijk is, kan al goed onderzocht worden met behulp van de analyses op de totaalscore zoals wordt beschreven in hoofdstuk 3. Als er wel sprake is van interactie worden de resultaten voor 2013 en 2014 voor de verschillende achtergrondvariabelen afzonderlijk gegeven.

Naast de effecten per vaardigheid worden ook de effecten binnen een jaar gecorrigeerd geschat, waarbij de jaarvariabele niet wordt meegenomen. Door deze gecorrigeerde effecten is het mogelijk een inschatting te maken van het additieve effect van een variabele, zoals het aanvullende effect van de gesproken thuistaal, naast het effect dat al gevonden wordt op basis van de vooropleiding van de ouders, zoals gerepresenteerd in het formatiegewicht.

De achtergrondkenmerken waarvoor in deze modellen gecorrigeerd wordt in een hoofdeffectenmodel, zijn geslacht, leertijd, stratum en formatiegewicht. In aanvullende analyses zijn de overige variabelen, zoals regio en thuistaal toegevoegd (een extra variabele per analyse). Doordat de vaardigheden in een jaar op dezelfde schaal liggen als de voorgaande jaren zijn de resultaten hiermee direct vergelijkbaar over de jaren heen.

In de eerdere verslagen van de JMTR wordt dieper ingegaan op verschillen tussen de gecorrigeerde modellen en de niet gecorrigeerde modellen, en de verschillen in interpretatie. Ook de gevolgen voor de vergelijking over de jaren heen en binnen het analysejaar zijn daar uiteengezet (zie bijvoorbeeld Hemker & Van Weerden, 2009; Hemker, Kuhlemeier en Van Weerden, 2010).

2.3.3 Interpretatie van 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). Deze startwaarden vormen voor elk van de drie schalen het gemiddelde en de standaarddeviatie die we in 2013 hebben aangetroffen omdat dat het eerste peilingsjaar voor de WO-schalen is. Door het gebruik van een vaardigheidsschaal kunnen de prestaties voor verschillende vaardigheden en voor verschillende jaren naast elkaar gezet worden op een en dezelfde schaal. Merk wel op dat dit in het geval van de WO-schalen ook kan op scoreniveau omdat iedereen exact dezelfde toets gemaakt heeft.

Voor een inhoudelijke interpretatie zou dieper ingegaan moeten worden op de inhoud van de opgaven. Op basis daarvan kan bepaald worden wat de leerlingen nu precies wel en niet kunnen. Dat zal in dit verslag nog niet gedaan worden omdat deze opgaven nu nog niet publiek gemaakt kunnen worden. Dat kan pas plaatsvinden na afloop van de afname in 2015.

Wat wel mogelijk is, is om in de verdeling van de scores specifieke typische leerlingen te identificeren. We richten ons in de rapportage niet alleen op de gemiddelde leerling, maar ook op andere leerlingen in de vaardigheidsverdeling. Voor dit doel zijn vijf typische leerlingen gedefinieerd, elk gekoppeld aan kenmerkende percentielpunten (zie tabel 2.7). In de rapportage gaan we uit van het basisjaar 2013 en rapporteren we voor 2014 (en mogelijk de jaren daarop) de eventuele wijzigingen in percentielen die gevonden worden. Zo zal bij een eventuele verandering in vaardigheid blijken dat meer of minder leerlingen de schaalwaarde 250 bereiken, maar ook kunnen meer leerlingen de schaalwaarde 186 bereiken die staat voor een zeer laag vaardige leerling: zijn er nu meer of minder zeer laag vaardige leerlingen dan in het basisjaar? Op een vergelijkbare wijze kan dit uitgevoerd worden voor de andere vaardigheidsgroepen. Aan de schaalwaarde horende bij de laag vaardige leerling (P25) kan een speciale betekenis worden toegekend. Dit is het punt op de vaardigheidsschaal dat in het basisjaar door minstens 75% van de leerlingen wordt bereikt. De definitie van de overige punten is vergelijkbaar.

De keuze van de percentages P25, P50 en P75 is deels gebaseerd op de grenzen die initieel bedacht waren voor de referentieniveaus taal en rekenen. In het advies van de Expertgroep Doorlopende Leerlijnen Rekenen en Taal (SLO, 2008) wordt P25 regelmatig genoemd als inschatting van het referentieniveau voor 1F, zowel bij taal als bij rekenen. Het referentieniveau 1S zou dan voor Rekenen overeenkomen met P50. Bij Taal is daarvoor de P75 genoemd. Dit waren relatieve grenzen gebaseerd op eerdere PPON-rapportages die functioneerden als (grote) voorlopige inschattingen. Gekoppeld aan de specifieke inhoud van de schalen waren er echter nog geen uitspraken gedaan over het gewenste niveau. In de afgelopen jaren is er veel aan gedaan om de referentieniveaus inhoudelijk te definiëren. De voor het basisonderwijs relevante niveaus 1F en 1S zijn nu wel inhoudelijk vastgelegd, al is er in 2014 nog geen harde koppeling te maken tussen prestaties op de Eindtoets van 2013 enerzijds en de referentieniveaus anderzijds voor de taal- en rekenvaardigheden. Voor de WO-schalen is dat in het geheel niet mogelijk omdat er voor de WO-schalen (nog) geen referentieniveaus zijn geformuleerd. Daar waar een koppeling tussen de referentieniveaus en de Eindtoets voor taal- en rekenvaardigheden in 2015 mogelijk zal zijn, is dat vooralsnog niet het geval voor de WO-schalen.

Tabel 2.7 Definiëring typische leerlingen in de vaardigheidsverdeling

Aanduiding	Afkorting	Percentiel 2013	% kandidaten 2013 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

2.3.4 Significantie en effectgrootte

Naast de vaardigheidsschattingen worden ook de verschillen tussen groepen geëvalueerd. 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 aan te geven of de gevonden verschillen ook van betekenis kunnen worden genoemd is de term effectgrootte geïntroduceerd (Cohen, 1988). Om de resultaten zinvol met elkaar te kunnen vergelijken rapporteren wij behalve de statistische significantie ook deze effectgrootte. De effectgrootte wordt in ons geval berekend als het verschil tussen de gemiddelden van de twee jaren (of twee subgroepen) 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 2.8. Alles met een effectgrootte boven de 0,20 noemen we hier betekenisvol. Effectgroottes lager dan 0,20 worden doorgaans gekwalificeerd als verwaarloosbaar.

Tabel 2.8 Kwalificatie van effectgrootten

Effectgrootte (zowel plus als min)	Kwalificatie
0,0 tot 0,	verwaarloosbaar
0,2 tot 0,5	klein effect
0,5 tot 0,8	matig effect
0,8 of groter	groot effect

3 De resultaten van de klassieke analyses

Er is een drietal klassieke analyses dat voor de WO-schalen wordt uitgevoerd. In de eerste analyses worden de verdeling van de achtergrondvariabelen over de verschillende jaren met elkaar vergeleken. Ook wordt hier onderzocht of de verdeling van de achtergrondvariabelen binnen een jaar verschilt tussen leerlingen die wel en geen WO-toetsen gemaakt hebben. Voor de vergelijking tussen jaren kan ook de standaardscore gebruikt worden om te zien of de groepen sterk van elkaar verschillen bij de overige vaardigheden. Deze standaardscore geeft een goede samenvatting van de taal-, reken-, en studievaardigheden samen en is over alle toetsen en jaren heen vergelijkbaar.

In de tweede serie analyses worden de verschillende groepen, verdeeld op basis van hun achtergrondvariabelen, met elkaar vergeleken op de scores op de vaardigheden Natuuronderwijs, Aardrijkskunde en Geschiedenis. Waar nodig betreft dit gewogen scores. Tot slot wordt de interactie tussen jaar en de achtergrondvariabelen voor de drie WO-vaardigheden onderzocht.

3.1 Verdeling van de achtergrondvariabelen

Er kan onderzocht worden bij welke variabelen er een verschil in de verdeling is. Als er op basis van significantie gekeken wordt, dan is vrijwel ieder verschil significant. Alleen bij het verschil tussen jongens en meisjes die wel of geen WO-toetsen gemaakt hebben is er geen verschil. Om die reden is met name ook gekeken naar de wat grotere verschillen die ook betekenisvol lijken.

Tabel 3.1. Percentages voor achtergrondvariabelen per categorie voor 2013 en 2014

Percentages leerlingen	2013			2014			2013 & 2014	
	Nee	Ja	Totaal	Nee	Ja	Totaal	Nee	Ja
WO gedaan	20,1	79,9	100,0	23,2	76,8	100,0	21,6	78,4
Geslacht								
Jongen	49,7	49,4	49,5	49,4	49,8	49,7	49,5	49,6
Meisje	50,3	50,6	50,5	50,6	50,2	50,3	50,5	50,4
Formatiegewicht								
0.00 & ontbrekend	84,0	89,5	88,4	85,4	90,5	89,3	84,7	90,0
0.30	7,5	6,5	6,7	6,8	5,7	5,9	7,1	6,1
1.20	8,5	4,1	5,0	7,9	3,8	4,7	8,2	3,9
Advies doorstroom VO								
BB	7,9	5,7	6,1	7,4	5,9	6,2	7,6	5,8
BB/KB	5,1	4,9	5,0	4,7	4,5	4,5	4,9	4,7
KB	11,5	9,5	9,9	12,5	10,5	10,9	12,0	10,0
KB/GT	0,9	0,8	0,8	0,8	0,8	0,8	0,9	0,8
GT	20,6	20,3	20,3	20,5	20,3	20,3	20,5	20,3
GT/HA	13,0	13,9	13,7	12,3	12,7	12,6	12,6	13,3
HA	15,3	15,9	15,8	16,2	16,7	16,6	15,8	16,3
HA/VW	12,0	13,4	13,1	10,9	12,0	11,7	11,4	12,7
VW	13,8	15,6	15,2	14,8	16,6	16,2	14,3	16,1
Toets								
EB/EA	82,7	86,6	85,8	81,9	85,5	84,7	82,2	86,1
EN	17,3	13,4	14,2	18,1	14,5	15,3	17,8	13,9

Thuis taal Algemeen								
Alleen NLs	.	.	.	86,7	92,4	91,1	86,7	92,4
NL+Ander	.	.	.	2,1	1,2	1,4	2,1	1,2
Alleen Ander	.	.	.	11,1	6,4	7,5	11,1	6,4
Thuis taal detail								
Alleen NLs	.	.	.	86,7	92,4	91,1	86,7	92,4
Turks (+NL+evt ander)	.	.	.	3,9	2,0	2,4	3,9	2,0
Arabs (+NL+evt ander)	.	.	.	4,1	1,9	2,4	4,1	1,9
Sur/Ant (+NL+evt ander)	.	.	.	0,8	0,4	0,5	0,8	0,4
Anders (+NL+evt ander)	.	.	.	4,4	3,2	3,5	4,4	3,2
Leertijd (Leeftijd op 1 okt)								
Regulier (10-11jr)	81,7	83,5	83,1	83,1	84,7	84,3	82,5	84,1
Vertraagd (12-13jr)	18,3	16,5	16,9	16,9	15,3	15,7	17,5	15,9
Stratum								
S1: 0% t/m 10% gew. lln	53,4	64,9	62,6	55,6	68,5	65,5	54,6	66,7
S3: >10% tot 25% gew. lln	25,4	25,5	25,5	25,3	22,6	23,2	25,3	24,1
S3: >25% gew. lln	21,2	9,6	11,9	19,1	8,9	11,3	20,1	9,3
Regio								
Noord	12,1	9,8	10,3	12,7	9,2	10,0	12,4	9,5
Oost	23,3	24,9	24,6	24,3	24,5	24,5	23,9	24,7
West	52,3	39,6	42,1	50,2	39,9	42,3	51,2	39,7
Zuid	12,4	25,7	23,0	12,8	26,3	23,2	12,6	26,0
Mate van verstedelijking								
Zeer sterk stedelijk	23,4	12,9	15,0	23,4	12,8	15,3	23,4	12,8
Sterk stedelijk	28,8	23,5	24,6	27,8	23,3	24,4	28,3	23,4
Matig stedelijk	19,2	21,8	21,3	18,6	22,1	21,3	18,9	22,0
Weinig stedelijk	16,9	27,7	25,6	18,1	27,8	25,5	17,6	27,7
Niet stedelijk	11,7	14,1	13,6	12,0	14,0	13,5	11,8	14,0
Schoolgrootte								
Klein (1-200)	34,7	29,7	30,7	34,1	30,4	31,2	34,4	30,0
Groot (>200)	65,3	70,3	69,3	65,9	69,6	68,8	65,6	70,0
Denominatie code*								
Rooms-Katholiek	26,6	38,6	36,2	26,9	38,8	36,1	26,8	38,7
Openbaar	41,4	27,5	30,3	40,6	27,6	30,6	41,0	27,6
Protestants-Christelijk	19,8	24,4	23,5	21,0	24,1	23,4	20,4	24,3
Algemeen Bijzonder	5,0	4,5	4,6	4,2	4,6	4,6	4,6	4,6
Overig (<5%)*	7,3	5,0	5,5	7,3*	4,8	5,4	7,2*	4,9

*Incl. 35 lln waarbij onbekend *

Merk op dat bovenstaande percentages de leerlingen betreffen voor wie de achtergrondgegevens bekend zijn. Als ook de leerlingen meegenomen worden van wie de achtergrondvariabelen onbekend zijn, dan liggen de percentages lager.³ Onbekende waarden: zie hierna.

Onbekende waarden								
Geslacht onbekend	1,0	0,9	0,9	1,3	1,0	1,1	1,1	1,0
Advies onbekend	28,7	28,5	28,5	28,1	27,7	27,8	28,4	28,1
Thuis taal Algemeen	100,0	100,0	100,0	4,8	4,0	4,2	49,3	53,3
Thuis taal detail	100,0	100,0	100,0	4,9	4,0	4,2	49,3	53,3
Leeftijd =? of <10jr of >13jr	1,0	0,8	0,8	1,0	0,8	0,9	1,0	0,8
Leeftijd onbekend	0,9	0,7	0,7	0,9	0,7	0,8	0,9	0,7

³ Bijvoorbeeld bij doorstroomadvies zijn de percentages duidelijk lager in het geval leerlingen met ontbrekende waarden meegeteld worden. Als echter de verdeling van de adviezen van de groep waar die niet voor bekend is vergelijkbaar is met de groep waar de adviezen wel bekend voor zijn de gegevens in tabel 3.1 weer goed te interpreteren als landelijke gegevens. Het resultaat van leerlingen van wie de adviezen bekend zijn en van wie ze onbekend zijn, verschillen amper waardoor het niet waarschijnlijk geacht wordt dat de ontbrekende waarnemingen geen systematische vertekening in de percentages veroorzaken.

Bij de meeste achtergrondvariabelen zien we verschillen tussen de groepen die de WO-toets wel of niet gemaakt hebben (tabel 3.1). Dat is het geval in beide schooljaren. Bij formatiegewicht zien we dat de groep die de WO-toets gemaakt heeft relatief meer leerlingen zonder gewicht bevat dan de groep die geen WO-toets gemaakt heeft. Over de jaren heen is dan 90,0% in de WO-groep en 84,7% bij de groep die geen WO-toets gemaakt heeft. Bij de leerlingen met het hoogste gewicht zien we juist het omgekeerde. Er zijn relatief veel van deze leerlingen die geen WO-toets maken (8,2% met gewicht 1.20) tegenover leerlingen die wel de WO-toets maken (3,9% met gewicht 1.20). Dat geldt ook, zij het in iets minder mate, voor gewicht 0.30. Er is dus een relatief grotere uitval op de gewichtsléerlingen bij de WO-toetsen dan bij de leerlingen die geen gewicht hebben.

Andere groepen waar we een relatief grote uitval op vinden, zijn de leerlingen met een advies op een lager niveau in het voortgezet onderwijs. We zien daar een oververtegenwoordiging van leerlingen met een havo/vwo- of vwo-advies. Daarmee samenhangend zien we ook dat de leerlingen die de Eindtoets Niveau maken ondervertegenwoordigd zijn in de WO-groep. Bij thuistaal zien we ook relatief veel uitval bij leerlingen die thuis een andere taal spreken dan Nederlands, waardoor de Nederlandstalige leerlingen wat oververtegenwoordigd zijn bij de WO-toetsen. Ook zijn er relatief minder leerlingen die ooit zijn blijven zitten die een WO-toets gemaakt hebben vergeleken met de gehele groep die een Eindtoets gemaakt heeft.

Voor wat betreft de schooleigenschappen zien we een ondervertegenwoordiging van de scholen in de sterk en zeer sterk stedelijke gebieden: daar worden de WO-toetsen relatief minder afgenomen. Dat levert dus een oververtegenwoordiging op bij de scholen in de meer landelijke gebieden, vooral de weinig stedelijke gebieden. Bij de variabele regio zien we een relatief grote uitval bij West terwijl er juist weinig uitval is in Zuid, waardoor die groep weer oververtegenwoordigd is. Daar enigszins mee samenhangend zien we dat openbare scholen relatief meer uitvallen bij de WO-toetsen, terwijl dat bij katholieke scholen (veel in Zuid) juist niet het geval is. Kijken we naar het stratum van de school dan zien we dat scholen met relatief veel "gewichtsléerlingen" (meer dan 25% van de leerlingen op school met gewicht 0.30 of 1.20) vaker uitvallen en geen WO-toetsen maken, terwijl relatief veel scholen met minder dan 10% "gewichtsléerlingen" juist wel de WO-toetsen maken.

Het is duidelijk dat de uitval op de WO-toetsen vooral plaatsvindt bij de groepen die juist relatief laag scoren op de reken- en taalvaardigheden, zoals gevonden bij de peilingen taal en rekenen. Ook de resultaten op de gehele Eindtoets zijn hoger bij de groepen die juist oververtegenwoordigd zijn bij de deelname aan het WO-deel. Wat de verschillen in resultaten zijn tussen de leerlingen die wel en geen WO-toetsen gemaakt hebben, valt uiteraard niet te onderzoeken voor wat betreft resultaten op de WO-toetsen – de resultaten op de WO-toetsen ontbreken immers juist voor de groep die deze toetsen niet gemaakt hebben –, maar wel voor de overige vaardigheden.

3.2 Vaardigheid per variabele bij leerlingen die wel en niet de WO-toetsen hebben gemaakt

Een goede indicatie om te zien of de groepen die wel of geen WO-toetsen gemaakt hebben van elkaar verschillen, is de standaardscore (zie tabel 3.2). Die is over beide jaren heen vergelijkbaar. Dat geldt niet voor de totaalscores voor de gehele set van 200 opgaven waar de standaardscore op gebaseerd is, omdat de toetsen voor de beide jaren inhoudelijk verschillen. Dat geldt uiteraard ook voor de vergelijkbaarheid van de scores van de verschillende subschalen, zoals Taal, Rekenen, of daarbinnen de subschalen, zoals Begrijpend lezen en Getallen en getalsbewerkingen. Binnen hetzelfde afnamejaar zijn de scores wel vergelijkbaar gemaakt.

Tabel 3.2 Gemiddelde standaardscores en betrouwbaarheidsintervallen voor de verschillende groepen leerlingen en het verschil in gemiddelde daartussen

Standaardscore	WO Gedaan		Verschil		
	Geen WO	Wel WO	Gemiddeld	95% betrouwbaarheidsinterval	
2013	534,16	535,31	1,15	1,02	1,28
2014	534,13	535,23	1,09	0,97	1,22
samen	534,14	535,27	1,12	1,03	1,21

Het verschil in standaardscore tussen de leerlingen die wel en geen toets Wereldoriëntatie hebben gemaakt is significant. Op een standaardscoreschaal lijkt een verschil van iets meer dan een punt niet veel, maar het komt ongeveer overeen met een effect van 0,10. Er is dus geen sprake van een willekeurige selectie van kandidaten voor de WO-toetsen. Hoewel de standaardscore weliswaar niet gelijk is aan de scores op de WO-schalen is er wel sprake van relatief hoge correlaties met de WO-score. De correlatie tussen de totaalscore voor de WO-toets en de standaardscore is 0,72. Voor de afzonderlijke kortere schalen liggen de correlaties tussen 0,60 en 0,65. Op de latente schaal, dus rekening houdend met de onnauwkeurigheid van de individuele toetsen, zien we dat deze voor de gehele WO-schaal en de afzonderlijke WO-schalen tussen de 0,70 en 0,80 ligt. Deze verschillen zijn ook voor de jaren afzonderlijk berekend, zowel voor de score op de 200 items waarop de standaardscore gebaseerd is, als voor de subschalen Taal en Rekenen. Het beeld blijkt daarbij identiek.

3.3 Weging

De vergelijking uitgevoerd in de vorige paragraaf laat zien dat de verdeling van de achtergrondvariabelen niet gelijk is voor de groep leerlingen die geen WO-toets gemaakt hebben en zij die dat wel gedaan hebben. We zien dat ook terug in de standaardscores.

Als geen rekening gehouden zou worden met het feit dat de leerlingen geen willekeurige steekproef vormen, zou dit dus zeer waarschijnlijk leiden tot een overschatting van de score op de vaardigheid voor WO. Er moet daarom een wegingsprocedure plaatsvinden. De vraag is nu welke variabelen daarvoor ingezet moeten worden.

Een belangrijke indicator voor de keuze van een variabele die goed kan functioneren als een wegingsfactor, is dat deze in de verdeling een verschil laat zien tussen leerlingen die wel of geen WO-toetsen hebben gemaakt. Hierboven hebben we gezien dat dit voor de meeste gebruikte achtergrondvariabelen het geval is. Een andere goede indicator is dat er sprake moet zijn van een effect op de vaardigheid: de verschillende niveaus moeten van elkaar verschillen in vaardigheid. Daarnaast is het een goede indicator als bij die variabele binnen een niveau juist een zeer gering verschil in vaardigheid gevonden wordt tussen leerlingen die wel of geen WO-toetsen hebben gemaakt. In dat geval zijn de leerlingen binnen dat niveau wel direct vergelijkbaar.

Tabel 3.3 Gemiddelde standaardscores per variabelecategorie voor leerlingen die niet en wel de WO-toets hebben gemaakt en het verschil daartussen.

	WO-gemaakt	niet	wel	Vershil
TOTAAL		534,1	535,3	1,1
Geslacht	Jongen	534,4	535,5	1,1
	Meisje	533,9	535,1	1,2
	<i>Onbekend</i>	532,1	532,9	0,8
Formatiegewicht	0.00 & Missing	535,3	536,0	0,7
	0.30	528,3	528,9	0,5
	1.20	527,0	527,7	0,7
Advies VO	BB	516,4	517,1	0,7
	BB/KB	521,9	522,3	0,4
	KB	525,7	525,8	0,1
	KB/GT	528,0	528,6	0,7
	GT	530,8	531,0	0,2
	GT/HA	535,3	535,5	0,1
	HA	539,0	539,3	0,3
	HA/VW	542,6	542,8	0,2
	VW	546,5	546,8	0,2
	<i>Onbekend</i>	534,3	535,3	1,1
Thuis taal Algemeen	Alleen NLs	534,9	535,6	0,7
	NL+Ander	530,9	531,4	0,5
	Alleen Ander	529,1	529,9	0,8
	Missing	534,1	535,3	1,2
Thuis taal Detail	Alleen Nederlands	534,9	535,6	0,7
	Turks (*)	527,3	527,9	0,6
	Arabisch (*)	529,3	529,1	-0,2
	Sur/Ant (*)	524,8	524,6	-0,2
	Anders (*)	532,3	533,0	0,7
	<i>Onbekend</i>	534,1	535,3	1,2
Toetsvariant	EB/EA	537,0	537,4	0,5
	EN	521,1	521,8	0,6
Leertijd	Regulier (10-11jr)	535,5	536,6	1,1
	Vertraagd (12-13jr)	527,8	528,5	0,7
	<i>LF=? of <10jr of >13jr</i>	533,5	534,3	0,8
Denominatie	RoomsKatholiek	534,6	535,5	0,8
	Openbaar	533,4	534,8	1,4
	ProtestantsChristelijk	534,2	535,2	1,0
	AlgemeenBijzonder	537,2	537,3	0,1
	Overig (<5%)	534,2	535,0	0,8
Regio	Noord	534,0	534,9	0,8
	Oost	534,1	534,9	0,8
	West	534,0	535,3	1,3
	Zuid	535,0	535,8	0,8
Urbanisatie	Zeer sterk stedelijk	532,7	534,3	1,7
	Sterk stedelijk	534,3	535,3	1,0
	Matig stedelijk	534,8	535,6	0,9
	Weinig stedelijk	534,9	535,3	0,4
	Niet stedelijk	534,5	535,4	0,9
Schoolgrootte	Klein (1-200)	533,1	534,5	1,3
	Groot (>200)	534,7	535,6	0,9
Stratum	S1: 0% t/m 10% gew. IIn	536,2	536,3	0,2
	S3: >10% tot 25% gew. IIn	533,5	534,1	0,6
	S3: >25% gew. IIn	529,5	530,5	1,0

In tabel 3.3. is dit gegeven voor de standaardscores van 2013 en 2014 samen weergegeven. Als per variabele gekeken wordt zien we weinig verschillen over de jaren heen zodat deze groep zonder bezwaar samengenomen kan worden. In deze tabel is te zien dat van de gegeven variabelen de variabele “Advies Voortgezet Onderwijs”, een advies dat al gegeven wordt voordat de toetsen gemaakt worden, sterk samenhangt met de gevonden scores. De score voor de leerlingen met een vmbo-bb advies is aanzienlijk lager dan voor de leerlingen met een vwo-advies. Dat is ook gebruikelijk en dergelijke verschillen worden ook bij de afzonderlijke vaardigheden gevonden. Een ander kenmerk van deze variabele is dat het gevonden verschil tussen de leerlingen die wel en niet de WO-toetsen gemaakt hebben per categorie relatief klein is, getuige de cijfers in de laatste kolom.

De grootste waardeverschillen worden gevonden bij de categorie “Advies onbekend” (ontbrekende waarde). Daar is het verschil tussen de leerlingen die wel en niet de WO-toets gemaakt hebben ongeveer even groot als bij de gehele toets. Dat is ook volgens de verwachting, omdat deze groep niet opgedeeld is. De groep “onbekend” zou dan ook sterk vergelijkbaar moeten zijn met de gehele groep. Wanneer ook gekeken wordt naar het verschil tussen de gemiddelde score van de leerlingen van wie het advies niet bekend is met de groep van wie het advies wel bekend is, dan zien we dat dit bij de 234.396 leerlingen die in 2013 en 2014 de WO-toets gemaakt hebben geen significant verschil oplevert ($\alpha = 0,01$).

De leerlingen die de WO-toets hebben gemaakt en van wie het advies bekend is, zijn dus vergelijkbaar met de leerlingen van wie het advies niet bekend is. Kijken we nu naar het gewogen⁴ gemiddelde verschil per adviescategorie tussen leerlingen die wel en niet de WO-toets hebben gemaakt, dan zien we dat dit 0,26 is, terwijl dat voor de gehele toets 1,12 is.

Een nadeel van het gebruik van deze variabele voor de weging is het relatief grote aantal leerlingen van wie dat advies niet bekend is. Een andere variabele, waarbij het verschil op de standaardscore tussen leerlingen die wel of niet de WO-toets hebben gemaakt kan worden berekend, is de variabele Stratum. Het gaat hier om een gewogen gemiddeld verschil van 0,42 standaardscorepunt. Het voordeel van het gebruik van deze variabele is dat alle leerlingen tot een van de drie betreffende categorieën behoren.

Deze analyse op de mogelijke inzet van achtergrondvariabelen als controlevariabele is ook gedaan per jaar afzonderlijk, zowel voor de scores op de 200 opgaven die de standaardscore bepalen in hun geheel, als op de scores op de twee subschalen Taal en Rekenen afzonderlijk. Dit leverde dezelfde uitkomst op.

Het resultaat geeft aan dat een weging zinvol is en dat een goede variabele om mee te wegen de variabele Advies Voortgezet onderwijs kan zijn. Wat we willen bereiken met deze weging is dat we met de resultaten zoals gevonden bij leerlingen die de WO-toets hebben gemaakt een goede schatting krijgen van het resultaat voor de gehele populatie. Dat kunnen we niet controleren voor de WO-schalen zelf, maar wel door de scores te vergelijken die wel bekend zijn van de leerlingen die al of niet de WO-toets hebben gemaakt. Het gaat dan om het resultaat op de Eindtoets, zonder het WO-deel.

Deze weging hebben we voor de twee jaren afzonderlijk uitgevoerd. Dat houdt in dat dit ook gedaan kan worden op het niveau van de scores op de 200 opgaven die de standaardscore bepalen en de scores op de twee subschalen Taal en Rekenen afzonderlijk. Gecontroleerd is of de score per jaar voor gehele populatie gelijk is aan de gewogen score op de groep leerlingen die de WO-toetsen hebben gemaakt. In tabel 3.4 worden de resultaten gegeven voor de verschillende mogelijke wegingen.

Ook is in deze tabel te zien dat de beste weging met de variabele Advies Voortgezet Onderwijs gevonden wordt. Als de resultaten van de leerlingen die de WO-toets gemaakt hebben gewogen worden met die variabele, dan komen daar vrijwel de populatiewaarden uit. De beste alternatieven daarna zijn het stratum van de school (3 niveaus gebaseerd op het percentage gewichtsléerlingen) en de toets die de leerlingen gemaakt hebben (Eindtoets Niveau tegenover de Eindtoets Basis). Geen enkel effect heeft het wegen op

⁴ Weging vind plaats door het verschil per variabele per niveau te wegen met het percentage dat dit niveau voorkomt in 2013 en 2014. Voor die percentages zie Bijlage 3.

geslacht. Merk ook op dat in 2013 weging op basis van Thuistaal niet mogelijk is. Denominatie is enige variabele waarbij enig serieus verschil in effect tussen 2013 en 2014 gevonden wordt bij de weging. In 2014 is de weging op basis van denominatie redelijk succesvol, maar in 2013 niet zo.

Tabel 3.4 *Verskil in standaardscores voor populatie en WO-toets gewogen en ongewogen*

	Gemiddelde			Verskil met populatie			Verskil gewogen - ongewogen		
	2013	2014	samen	2013	2014	samen	2013	2014	samen
Populatie	535,1	535,0	535,0						
WO-toets ongewogen	535,3	535,2	535,3	,23	,25	,24	,23	,25	,24
Gewogen met leerlingkenmerken									
Geslacht	535,3	535,2	535,3	,23	,25	,24	,00	,00	,00
Formatiegewicht	535,2	535,1	535,2	,14	,16	,15	-,09	-,10	-,09
Advies VO	535,1	535,0	535,0	,00	,05	,02	-,23	-,20	-,22
Thuistaal algemeen	.	535,2	.	.	,19	,21	.	-,07	.
Thuistaal detail	.	535,1	.	.	,17	,20	.	-,08	.
Toetsvariant	535,2	535,1	535,1	,11	,12	,11	-,12	-,13	-,13
Leertijd	535,3	535,2	535,2	,20	,22	,21	-,03	-,03	-,03
Leeftijd)	535,3	535,2	535,2	,20	,22	,21	-,03	-,03	-,03
Gewogen met schoolkenmerken									
Stratum	535,2	535,1	535,1	,09	,10	,10	-,14	-,15	-,14
Regio	535,3	535,2	535,2	,22	,22	,22	-,01	-,03	-,02
Verstedelijking	535,3	535,2	535,2	,20	,22	,21	-,03	-,03	-,03
Schoolgrootte	535,3	535,2	535,3	,22	,24	,23	-,01	-,01	-,01
Denominatie	535,3	535,1	535,2	,22	,10	,16	-,01	-,15	-,08

Dit is ook onderzocht per jaar afzonderlijk en ook voor de scores op de 200 opgaven die de standaardscore bepalen en de scores op de twee subschalen Taal en Rekenen. Dit leverde hetzelfde beeld op, zij het dat daar de resultaten voor 2013 en 2014 niet samen kunnen worden geanalyseerd.

Naast de weging en het effect op de schattingen op het niveau van het gehele jaar is er ook gekeken wat het verschil was tussen de twee meest belovende wegingsmethodes, namelijk op basis van het VO-advies en op basis van stratum, als we kijken naar de resultaten op het niveau van de achtergrondvariabelen. Zo kan er bijvoorbeeld gekeken worden welke weging het beste resultaat oplevert bij jongens in 2014: dat wil zeggen, welke weging op de gegevens van de leerlingen die de WO-toets wel gemaakt hebben komt het dichtst bij de werkelijke waarde van de standaardscore van de gehele groep, dus inclusief de leerlingen die de WO-toets niet gemaakt hebben? In dat geval zien we (niet gegeven in tabel 3.4) dat de weging met de variabele Stratum net iets betere resultaten oplevert. Het voordeel van de weging met stratum is dan ook dat voor alle leerlingen bekend is tot welk stratum ze behoren, terwijl het advies voor het voortgezet onderwijs bij ruim 28% van de leerlingen onbekend is.

Op basis van deze resultaten is besloten de weging van de resultaten plaats te laten vinden door de resultaten op de drie WO-schalen zowel te wegen met de variabele advies VO als de variabele stratum. De gewichten zijn bepaald met de percentages in de gehele populatie. De percentages worden voor de twee jaren afzonderlijk bepaald. De leerlingen van wie het VO-advies niet bekend was, zijn niet meegenomen bij de eerste correctie. Deze groep verschilde overigens niet van de groep waar het advies wel van bekend was, maar er kon niet mee gewogen worden. De weging op basis van de variabele stratum bevat wel alle leerlingen. Bij de presentatie hierna zullen de gegevens zowel gewogen als ongewogen weergegeven worden.

3.4 Resultaten op de drie toetsen

3.4.1 Jaarverschillen

Er worden resultaten voor drie analyses gerapporteerd. Ten eerste wordt de ongewogen score gegeven voor de groep leerlingen die deze toets gemaakt heeft. Vervolgens geven we twee gewogen scores: die met als wegingsvariabele Advies VO en die met de wegingsvariabele Stratum. Deze uitkomsten geven aan welke scores te verwachten zijn als de verdeling van voor de variabele Advies VO, dan wel de variabele Stratum in de groep leerlingen die de WO-toets heeft gemaakt gelijk zou zijn aan die in de gehele EB-populatie. In tabel 3.4 geven we de uitkomsten voor beide jaren en voor alle drie de onderdelen van de WO-toets.

Tabel 3.4 Gemiddelde score per toets, ongewogen en gewogen

Toetsonderdeel	Jaar	Ongewogen			Gewogen: Advies VO			Gewogen: Stratum		
		2013	2014	'14-'13	2013	2014	'14-'13	2013	2014	'14-'13
Natuuronderwijs		19,84	19,86	0,03	19,72	19,76	0,05	19,76	19,78	0,02
Aardrijkskunde		20,03	20,01	-0,01	19,91	19,91	0,00	19,95	19,94	-0,01
Geschiedenis		21,09	21,23	0,14	20,97	21,13	0,16	21,02	21,16	0,14

Zoals te verwachten zijn de gewogen scores lager dan de ongewogen scores. Het verschil tussen de ongewogen score en de score na de weging met het VO-advies is gemiddeld 0,11 punten (in 2013 gemiddeld 0,12 en in 2014 0,10). Het verschil tussen ongewogen en de weging met stratum varieert van 0,07 tot 0,08. Doordat de verschillen vergelijkbaar zijn over de jaren heeft weging relatief weinig effect op de verschillen tussen de twee jaren. De grootste verschillen worden gevonden bij de weging met het VO-advies, maar de verschillen zijn maximaal 0,02⁵. Voor het verschil in jaren heeft de weging een zeer gering effect.

Bij de toetsing van de significantie van de verschillen, gedaan op de ongewogen scores, zijn de gevonden verschillen tussen de jaren voor Natuuronderwijs en Aardrijkskunde niet significant bij een alpha van 0,01. Bij Geschiedenis zijn de verschillen wel significant ($p < 0,00001$). Ook bij weging met Stratum of VO-Advies is dat het geval. Als een significantieniveau van 0,05 gekozen wordt, dan is het verschil bij Natuuronderwijs met weging met Advies-VO wel significant. Vanwege het grote aantal waarnemingen, dus meer dan 100.000 observaties per groep, heeft significantie nog maar weinig betekenis.

Een maat die beter inzicht geeft in de betekenis van de gevonden verschillen is de effectgrootte. In deze maat wordt het verschil tussen twee gemiddelden gerelateerd aan de gedeelde standaarddeviatie van de betrokken toetsen. In dit geval is er per toets weinig verschil in standaarddeviatie tussen de twee jaren (tabel 3.5). De standaarddeviaties per jaar zijn vrijwel gelijk aan de standaarddeviaties over de jaren heen zoals gegeven in tabel 2.1.

Bij al deze vergelijkingen zijn de verschillen zo klein dat ze volgens de gebruikelijke vuistregels voor interpretatie (zie hoofdstuk 2, tabel 2.1) als verwaarloosbaar, of niet betekenisvol, kunnen worden geïnterpreteerd. We mogen dus concluderen dat de resultaten op de toetsen WO van 2013 en 2014 niet van elkaar verschillen.

⁵ Uit tabel 3.5 zou geconcludeerd kunnen worden dat het verschil maximaal 0,03 is (verschil tussen '14 en '13 bij Natuuronderwijs) tussen gewogen met Advies VO (0,05) en gewogen met Stratum (0,02), maar deze waarden zijn afgeronde waarden. De onderliggende waarde van 0,05 is 0,046 en van 0,02 is dat 0,024. Het verschil tussen die twee waarden is 0,022.

Tabel 3.5 Effecten en standaarddeviaties: ongewogen en gewogen

Toets	SD	Effecten 2014-2013		
		Ongewogen	Weeg: Advies VO	Weeg: Stratum
WO1: Natuuronderwijs	4,60	0,01	0,01	0,01
WO2: Aardrijkskunde	5,00	-0,00	-0,00	-0,00
WO3: Geschiedenis	4,83	0,03	0,03	0,03

3.4.2 Leerlingvariabelen

Hieronder worden de resultaten gegeven op de scores voor de verschillende subgroepen. Hierbij zijn de wegingen uitgevoerd op basis van de variabele Stratum aangezien deze weging bij de Eindtoetsresultaten de beste resultaten opleverde (zie paragraaf 3.3). Ook de ongewogen resultaten en de resultaten voor de weging met Advies-VO voor de drie vaardigheden zijn berekend en opvraagbaar, maar zijn in deze publicatie niet opgenomen. In tegenstelling tot hierboven bij de jaarvergelijking, worden de effectgroottes en significantie verder niet aangegeven. Over het algemeen zijn verschillen groter dan 0,10 scorepunt (effectgroottes groter dan ,02) significant. De effectgroottes en significantie worden wel gegeven bij de IRT-resultaten in hoofdstuk 4.

Als we kijken naar de jaarverschillen per variabele categorie (tabel 3.6) dan zien we dat er bij de verschillende categorieën weinig grote verschillen zijn met wat er over de gehele populatie gevonden wordt. Het grootste negatieve verschil wordt gevonden bij leerlingen met een hoog formatiegewicht (categorie 1.20). Daar zien we een kleine achteruitgang van 2013 naar 2014 bij Natuuronderwijs. Bij vertraagde leerlingen zien we ook nog een kleinere achteruitgang bij Aardrijkskunde. Bij een aantal andere toetsen en groepen zien we ook een lichte achteruitgang, bijvoorbeeld bij jongens bij Natuuronderwijs en Aardrijkskunde, maar deze zijn altijd (in absolute waarde) kleiner dan 0,10 punt en dus weinig betekenisvol. De vooruitgang is juist iets groter bij de lagere vmbo-groepen voor alle vaardigheden en bij de leerlingen die de Eindtoets Niveau gemaakt hebben.

De jaarverschillen bij de verschillende subgroepen zijn verwaarloosbaar qua omvang. De verschillen tussen de categorieën binnen een achtergrondvariabele zijn wel aanzienlijk. Zoals ook bij de onderdelen taal en rekenen in de Eindtoets het geval is (zie Hemker & Van Weerden, 2015) worden de grootste verschillen gevonden tussen de leerlingen die een advies vmbo-bb krijgen en de leerlingen met een vwo-advies. Bij de aardrijkskunde- en de geschiedenisstoets is er een verschil van meer dan 10 punten (van de maximaal 30). Bij de toets Natuuronderwijs ligt het verschil rond de 9 punten. De verschillen tussen de jaren 2013 en 2014 zijn hier klein.

Tabel 3.6 *Uitkomsten voor de drie vaardigheden WO per jaar per categorie van de achtergrondvariabelen op leerlingniveau (weging met stratum)*

		Natuuronderwijs		Aardrijkskunde		Geschiedenis	
		2013	2014	2013	2014	2013	2014
Totaal	Jaar	19,76	19,78	19,95	19,94	21,02	21,16
Geslacht	Jongen	20,30	20,25	21,08	21,02	21,80	21,90
	Meisje	19,23	19,33	18,86	18,88	20,27	20,44
Formatiegewicht	0.00 & Missing	20,16	20,15	20,37	20,31	21,40	21,50
	0.30	17,50	17,58	16,95	17,06	18,40	18,61
	1.20	15,46	15,28	16,38	16,35	17,63	17,69
Leertijd	Regulier	20,07	20,08	20,34	20,31	21,37	21,50
	Vertraagd	18,23	18,19	18,07	17,95	19,27	19,31
Toets	EB/EA	20,40	20,47	20,72	20,76	21,77	21,96
	EN	15,74	15,83	15,11	15,22	16,27	16,59
Advies VO	vmbo-bb	14,34	14,53	13,66	13,78	14,70	14,99
	vmbo-kb	16,77	16,95	16,39	16,50	17,66	17,87
	vmbo-tl/gt	18,47	18,62	18,46	18,45	19,70	19,93
	havo	20,83	20,81	21,24	21,26	22,32	22,46
	vwo	23,39	23,43	24,37	24,28	24,99	25,12
Thuis taal (alg)	Alleen NLs	.	20,11	.	20,15	.	21,39
	NL+Ander	.	16,99	.	18,28	.	19,41
	Alleen Ander	.	16,11	.	17,45	.	18,53
Thuis taal (det)	Alleen Nederlands	.	20,11	.	20,15	.	21,39
	Turks(+)	.	15,06	.	16,44	.	17,42
	Arabisch(+)	.	15,50	.	16,99	.	18,25
	Sur/Ant(+)	.	15,58	.	15,83	.	17,02
	Anders(+)	.	17,62	.	18,98	.	20,01

Dat de verschillen tussen 2013 en 2014 klein zijn, zeker ten opzichte van de verschillen tussen de categorieën, wordt ook binnen andere achtergrondvariabelen gevonden. Leerlingen die de Niveautoets maken, scoren aanzienlijk lager dan de leerlingen die de Eindtoets Basis maken. Het verschil is het kleinst bij Natuuronderwijs (verschil ruim 4,5) en het grootst bij Aardrijkskunde (verschil ruim 5,5).

Er worden ook verschillen gevonden tussen jongens en meisjes: jongens scoren ongeveer 1 punt hoger bij de toets Natuuronderwijs, 1,5 punt hoger bij Geschiedenis en iets meer dan 2 punten hoger bij Aardrijkskunde. Bij alle toetsen wordt een verschil van rond de 2 punten gevonden tussen de vertraagde leerlingen en de leerlingen die niet vertraagd zijn.

Bij Natuuronderwijs zien we dat het verschil in score tussen een leerling zonder gewicht (0.00) en een leerling met een laag gewicht (0.30) maar iets groter is dan dat tussen een leerling zonder en met een hoog gewicht (1.20). De eerste twee groepen schelen iets meer dan 2,5 punt, terwijl de twee gewichtsgroepen een verschil van 2 punten laten zien. Bij de andere twee toetsen zien we dat het verschil tussen een 0.00- en een 0.30-leerling juist groter is (ruim 3 punten bij Aardrijkskunde, rond de 3 punten bij Geschiedenis), terwijl het verschil tussen de leerling met een hoog en een laag gewicht juist kleiner is (minder dan 1 punt bij beide toetsen).

Bij Thuis taal waren er geen gegevens beschikbaar voor 2013. Gezien de geringe verschillen tussen de jaren bij de overige variabelen gaan we er vanuit dat dit ook het geval is bij de verschillen tussen de categorieën leerlingen die onderscheiden worden op basis van thuis taal. Net zoals bij taal en rekenen zien we dat de groep die thuis alleen Nederlands spreekt de hoogste score behaalt.

3.7 *Uitkomsten voor de drie vaardigheden WO per jaar per categorie van de achtergrondvariabelen op schoolniveau (weging met stratum)*

		Natuuronderwijs		Aardrijkskunde		Geschiedenis	
		2013	2014	2013	2014	2013	2014
Totaal	Jaar	19,76	19,78	19,95	19,94	21,02	21,16
Stratum	Stratum 1	20,38	20,37	20,60	20,53	21,60	21,70
	Stratum 2	19,54	19,42	19,54	19,39	20,71	20,71
	Stratum 3	16,96	17,09	17,43	17,60	18,61	18,91
Regio	Noord	19,81	19,89	19,85	19,91	21,05	21,27
	Oost	19,93	19,84	19,97	19,87	21,06	21,06
	West	19,61	19,64	20,03	19,99	21,05	21,20
	Zuid	19,81	19,92	19,85	19,92	20,92	21,14
Verstedelijking	Zeer sterk	18,32	18,53	19,12	19,23	20,07	20,35
	Sterk	19,69	19,61	20,00	19,90	21,10	21,11
	Matig	20,03	20,07	20,19	20,26	21,30	21,46
	Weinig	20,08	20,07	20,09	20,02	21,11	21,28
	Niet	20,34	20,36	20,14	20,10	21,26	21,42
Schoolgrootte	Klein (1-200)	19,61	19,59	19,56	19,53	20,67	20,79
	Groot (>200)	19,83	19,87	20,12	20,12	21,17	21,32
Denominatie code	RK	19,85	19,86	20,01	20,04	21,01	21,20
	Openbaar	19,50	19,53	19,74	19,77	20,75	20,93
	PC	19,81	19,83	20,07	19,97	21,17	21,27
	AB	20,57	20,54	21,05	20,94	22,10	22,02
	Overig (<5%)	19,61	19,81	19,29	19,26	20,94	20,95

4 De resultaten van de itemresponsetheorie-analyses

In dit hoofdstuk worden de effecten van de achtergrondvariabelen op de schalen zoals die bepaald zijn met de IRT-modellen beschreven. Deze analyses zijn uitgevoerd met behulp van de programma's OPLM en SAUL. Er worden steekproeven getrokken om analyses op uit te voeren omdat de basisbestanden hiervoor te groot zijn. Het betreft een tiental steekproeven, ieder van 10% van de gehele populatie, zodat de gehele populatie in de analyses gebruikt is. De resultaten van de steekproeven zijn met elkaar vergeleken. In dit hoofdstuk worden de resultaten gegeven van de steekproeven die het meest representatief zijn voor de tien steekproeven.

4.1 De vergelijking over de jaren

Ervan uitgaande dat steekproeffluctuaties geen rol spelen (zie paragraaf 1.2), vergelijken we in deze paragraaf de resultaten van jaargroep 8 in 2014 met die in het voorgaande jaar 2013. In de modellering wordt zo voor de effecten geen correctie gebruikt om de jaren te vergelijken. Er wordt echter wel rekening gehouden met de afwijking van de verdeling van de achtergrondvariabelen om tot een weging te komen, zodat de resultaten representatief voor de populatie zijn. Bij de klassieke resultaten bleek een weging op basis van de variabele Stratum de beste resultaten op te leveren om de resultaten van de steekproef op het niveau van de populatie te krijgen. Ook hier is gebruikgemaakt van een weging met behulp van de variabele Stratum, waarbij de verdeling van stratum bij de gehele Eindtoets als referentie is genomen. Er zijn ook controles uitgevoerd met een weging op basis van formatiegewicht, met de DUO-gegevens als referentie.

Net zoals bij de analyses bij taal en rekenen (zie Hemker & Van Weerden, 2015) wordt de eerste meting, in dit geval in 2013, als startpunt genomen. Dat betekent dat de effecten vergeleken worden met dat eerste jaar, waarbij een positief effect een stijging in vaardigheid over de tijd betekent en een negatief effect een daling. Ook wordt het resultaat van het eerste jaar zodanig geschaald dat het gemiddelde van dat jaar op 250 ligt, met een standaarddeviatie van 50. Dit geldt voor alle drie de gemeten vaardigheden, waardoor de gemiddelde waarde in 2013 voor zowel Natuuronderwijs, Aardrijkskunde als Geschiedenis ligt op 250.

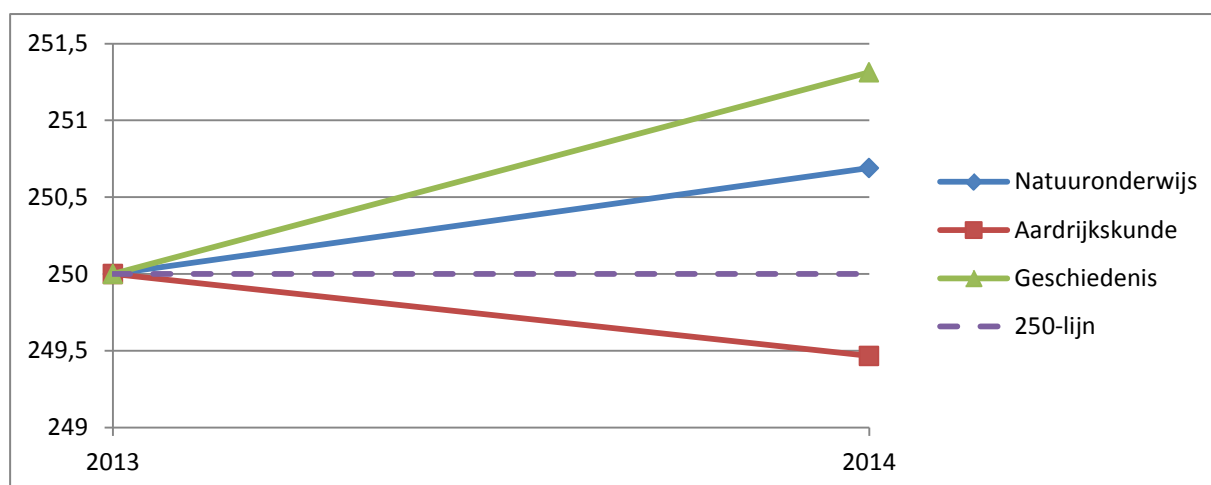
In tabel 4.1 zijn de uitkomsten voor de twee opeenvolgende jaren weergegeven als gemiddelde op de vaste schaal. In deze tabel zijn de waarden afgerond op gehele getallen. De niet afgeronde gemiddelden zijn 250,69 (natuuronderwijs), 249,47 (aardrijkskunde) en 251,31 (geschiedenis).

Tabel 4.1 Gemiddelde vaardigheid en standaarddeviatie bij de 3 WO-schalen voor 2013 en 2014

Vaardigheden	2013		2014	
	Gem.	SD	Gem.	SD
Natuuronderwijs	250	50	251	50
Aardrijkskunde	250	50	249	50
Geschiedenis	250	50	251	50

In figuur 4.1 zijn de resultaten afgebeeld in een grafiek. Te zien is dat de vaardigheden voor Natuuronderwijs en Geschiedenis iets zijn gestegen en die van Aardrijkskunde iets is gedaald. De verschillen met 2013 zijn echter zeer klein te noemen. Dat blijkt ook uit de effectschattingen (tabel 4.2). De effectschattingen met waarden tussen -0,01 en 0,03 geven aan dat de verschillen als betekenisloos beschouwd kunnen worden. Deze effecten zijn ook in de steekproeven met een grootte van meer dan 20.000 personen

niet significant. In praktische zin kan dus gesteld worden dat de vaardigheid bij de drie schalen in 2014 niet verschilt van die van 2013.



Figuur 4.1 Trend voor de onderwerpen wereldoriëntatie in jaargroep 8

Tabel 4.2 Effectgroottes voor de jaarverschillen bij de drie WO-schalen

Vaardigheden	2014-2013
Natuuronderwijs	0,01
Aardrijkskunde	-0,01
Geschiedenis	0,03

Het is mogelijk dat de effectschattingen wel betekenis krijgen bij het volgen over meerdere jaren. Ook bij de taal- en rekenvaardigheden wordt gemiddeld een jaarlijkse stijging met een effectgrootte van 0,01 gevonden (Hemker en Van Weerden, 2015). Deze stijging is van jaar op jaar als verwaarloosbaar te beschouwen, maar als we een grotere periode overzien en deze stijging over meer jaren berekend wordt, dan tellen deze stijgingen over de jaren heen bij elkaar op en begint het verschil wel meer betekenis te krijgen. Om te zien of dat ook hier het geval is zullen deze vaardigheden vanaf nu ook over de tijd gevolgd moeten worden.

Een andere manier om de resultaten met elkaar te vergelijken is te kijken naar de percentielen typische leerlingen. Een typisch zeer laag vaardige leerling is gedefinieerd als een leerling op het niveau waar 10% van de leerlingen op of onder scoort (P10). Dat betekent dat 90% van de leerlingen een vaardigheid heeft die hoger is dan de vaardigheid van een zeer laag vaardige leerling. Dit geldt voor de vaardigheidsverdeling in 2013, het referentiejaar, die immers per definitie een gemiddelde heeft van 250 en een standaarddeviatie van 50 en het geldt voor alle drie de subschalen. Zo scoort per definitie 50% van de leerlingen in 2013 boven het gemiddelde van het referentiejaar (2013) en 10% boven het niveau van de zeer hoog vaardige leerling (gedefinieerd als P90 in het referentiejaar).

De resultaten van 2014 kunnen met die van 2013 vergeleken worden, omdat deze immers op dezelfde meetschaal liggen. De resultaten zijn weergegeven in tabel 4.3. In deze tabel is te zien dat afgerond 51% van de leerlingen in 2014 bij de vaardigheden Natuuronderwijs en Geschiedenis boven het gemiddelde van 2013 liggen. De verschillen tussen 2013 en 2014 zijn ook hier, zoals te verwachten zeer gering.

De analyses waarbij in plaats van een weging op basis van de variabele stratum gebruikgemaakt is van een weging op basis van de variabele formatiegewicht laat zeer vergelijkbare resultaten zien. De verschillen tussen 2014 en 2013 zijn daarbij zelfs iets kleiner.

Tabel 4.3 Percentage leerlingen per vaardigheid vergeleken met de score van de typische leerling

Leerlingengroep*	Percentiel-grens	2013	2014		
		Alle vaardigheden	Natuuronderwijs	Aardrijkskunde	Geschiedenis
> dan ZLV	P10	90	90	90	90
> LV	P25	75	75	75	76
> Standaard 2008	P50	50	51	50	51
> HV	P75	25	25	25	26
> ZHV	P90	10	10	10	10

4.2 Verschillen voor achtergrondvariabelen

In deze paragraaf rapporteren we de verschillen tussen groepen van leerlingen gebaseerd op een categorisering naar achtergrondvariabele. We presenteren hier de resultaten van het referentiejaar 2013 en de meest recente meting in 2014. Tevens geven we het resultaat van de combinatie van jaren aangezien de effecten van de achtergrondvariabelen in 2013 en 2014 zeer vergelijkbaar waren, op een enkele uitzondering na: jaar-interactie was in veruit de meest gevallen niet significant.

In de Jaarlijkse meting Taal en Rekenen (Hemker en Van Weerden, 2015) ging er speciale aandacht uit naar de resultaten bij de leertijd. Die aandacht was ingegeven door de geobserveerde trend dat het percentage leerlingen dat een jaar (of meer) vertraging opgelopen had in de afgelopen jaren aanzienlijk gedaald was. In 2008 was het percentage vertraagde leerlingen in de peilingspopulatie nog 20,2% terwijl dat in 2014 nog maar 15,7% was. Het percentage vertraagde leerlingen in 2013 was 16,9%. De daling in het percentage van het eerste jaar met resultaten voor de Wereldoriëntatievakken naar de meest recente peiling was uiteraard kleiner dan de daling van 2008 naar 2014, maar relatief groot als we ons bedenken dat het een verandering in slechts één jaar betreft.

Er is gebruikgemaakt van drie verschillende basismodellen: een model voor de analyses van de jaren gezamenlijk, een voor de analyses van 2013 afzonderlijk en een voor 2014 afzonderlijk. Die basismodellen verklaren de variabelen geslacht, leertijd, formatiegewicht en stratum. Deze laatste variabele definiëren we als het percentage gewichtslernen op een school, onderscheiden naar drie categorieën. Voor de gezamenlijke analyses van de 2013 en 2014 bevatte het model ook de variabele jaar.

Ieder van de geschatte effecten per variabele wordt gecorrigeerd voor effecten van de overige variabelen. Hiermee kan het unieke effect van de variabele bepaald worden. Aan dit basismodel zijn in de verdere analyses variabelen toegevoegd. Dit betreft de variabelen toetsvariant (Eindtoets Basis of Eindtoets Niveau), advies VO, schoolgrootte, regio en verstedelijking. Dit levert een vijftal extra modellen op de basisvariabelen, telkens aangevuld met één extra variabele. De gerapporteerde effecten zijn telkens gecorrigeerde effecten. Voor een weergave van de gemiddelden, standaarddeviaties, overschrijdingskansen en effectgroottes per achtergrondvariabele verwijzen we naar de tabellen in de bijlagen.

De modellen waren hoofdeffectenmodellen. Ter controle zijn in de analyses waarin beide jaren tegelijkertijd zijn meegenomen ook interactiemodellen uitgevoerd, waarbij de interactie tussen jaar en de overige variabelen zijn geanalyseerd. Dat interactie-effect was slechts bij één vaardigheid (Natuuronderwijs) bij één variabele (Formatiegewicht) significant. Hierover meer bij de bespreking van de resultaten van deze variabele.

Alle gecorrigeerde effecten voor 2013 en 2014 afzonderlijk zijn gegeven in Bijlage 1. De belangrijkste resultaten worden ook in dit hoofdstuk gegeven. In Bijlage 2 staan de geschatte gemiddelden en standaarddeviaties van de onderscheiden groepen. De belangrijkste resultaten worden ook in de loop van de volgende paragraaf gegeven.

4.2.1 Effecten voor de achtergrondvariabelen op leerlingniveau

In tabel 4.4 zijn de effectgroottes voor de variabelen op leerlingniveau weergegeven. Het betreft hier de resultaten van de hoofdeffectenmodellen. De vetgedrukte resultaten zijn de meest relevante resultaten. Door het ontbreken van significante interacties zijn dat in de meest gevallen de resultaten voor beide jaren samen (kolom Totaal). De resultaten voor de jaren afzonderlijk wijken daar in de meeste gevallen met kleine afwijkingen tussen -0,05 en 0,05 van af. Een uitzondering hierop zijn de resultaten bij de variabele en de schaal waarbij het enige interactie-effect is gevonden: formatiegewicht bij Natuuronderwijs.

Tabel 4.4 Effecten op leerlingniveau in jaargroep 8 in 2013 en 2014 (na correctie)

Variabele (Contrast)	Natuuronderwijs			Aardrijkskunde			Geschiedenis		
	Totaal	2013	2014	Totaal	2013	2014	Totaal	2013	2014
Geslacht									
Meisjes - Jongens	-0,35	-0,35	-0,34	-0,61	-0,61	-0,60	-0,39	-0,36	-0,43
Leertijd									
Regulier - Vertraagd	0,38	0,38	0,39	0,51	0,51	0,51	0,44	0,45	0,43
Formatiegewicht									
0.0 - 0.3	0,49	0,47	0,50	0,59	0,60	0,58	0,51	0,48	0,54
0.3 - 1.2	0,25	0,16	0,35	0,00	0,01	-0,01	0,09	0,07	0,12
0.0 - 1.2	0,73	0,63	0,85	0,59	0,61	0,57	0,60	0,55	0,66
Advies VO									
vmbo-KB - vmbo-BB	0,50	0,47	0,53	0,60	0,62	0,58	0,55	0,52	0,58
vmbo-GT - vmbo-KB	0,43	0,46	0,40	0,44	0,41	0,47	0,55	0,52	0,58
havo - vmbo-GT	0,63	0,64	0,62	0,74	0,72	0,77	0,67	0,72	0,62
vwo - havo	0,93	0,89	0,97	1,04	1,02	1,05	1,00	1,00	1,01
vwo - vmbo BB	2,49	2,46	2,52	2,81	2,76	2,86	2,77	2,76	2,78
Toets									
Eindtoets-Niveautoets	1,05	1,05	1,05	1,15	1,11	1,19	1,16	1,17	1,15
Taal									
Alleen NLS - NLS+Ander			0,50			0,20			0,02
NLS+Ander - Alleen Ander			0,10			0,11			0,27
Alleen NLS - Alleen Ander			0,61			0,32			0,29
Alleen NLS - Turks(+NLS)			0,83			0,62			0,46
Alleen NLS - Arabisch(+NLS)			0,79			0,36			0,26
Alleen NLS - Sur./Ant.(+NLS)			0,91			0,83			0,59
Alleen NLS - Ov. ander taal(+NLS)			0,33			0,06			0,05

Legenda

klein effect en
p < 0,01
p < 0,05
niet significant

p < 0,001 en
klein effect
matig effect
groot effect

minder dan klein effect, niet significant, of beiden

Geslacht

Het verschil in resultaten tussen jongens en meisjes laat in alle gevallen significante en betekenisvolle effecten zien. Deze zijn bij Aardrijkskunde met een matige effectgrootte van rond de -0,60 het grootst. Bij de vaardigheden Natuuronderwijs (effecten rond -0,35) en Geschiedenis (effecten tussen -0,36 en -0,43) moeten ze als klein worden gekwalificeerd. Bij de vaardigheid Geschiedenis lijkt het verschil tussen jongens en meisjes van 2013 naar 2014 groter geworden te zijn, maar deze verandering is niet significant.

In tabel 4.5 geven we de resultaten op de vaardigheidsschaal, onderscheiden naar geslacht en jaar.

We zien bij de vaardigheden voor wereldoriëntatie steeds min of meer vergelijkbare resultaten voor de jaarvergelijking bij jongens en meisjes, zoals te verwachten valt als er geen sprake is van een significante interactie.

Tabel 4.5 Gemiddelde schaalscores per onderwerp per categorie voor 2013 en 2014 en het jaarverschil

	Natuuronderwijs			Aardrijkskunde			Geschiedenis		
	2013	2014	verschil	2013	2014	verschil	2013	2014	verschil
Geslacht									
Samen	250	251	1	250	249	-1	250	251	1
Jongens	258	258	0	263	263	0	259	261	3
Meisjes	242	243	1	237	236	-1	242	242	0
Leertijd									
Samen	250	251	1	250	249	-1	250	251	1
Regulier	254	254	1	254	254	-1	254	255	1
Vertraagd	231	232	0	229	227	-2	229	231	2
Formatiegewicht									
Samen	250	251	1	250	249	-1	250	251	1
0.00	254	255	1	254	254	-1	254	255	1
0.30	224	223	-1	219	217	-2	222	222	-1
1.20	204	197	-6	212	211	-2	212	211	-1

Leertijd

De variabele leertijd wordt hier, net als in eerdere rapporten, geanalyseerd voor twee categorieën: vertraagde leerlingen en reguliere leerlingen. Het betreft respectievelijk de groep vertraagde leerlingen die gezien hun leeftijd (op 1 oktober) het primair onderwijs al verlaten hadden kunnen hebben, en de groep leerlingen die gezien hun leeftijd nog thuis hoort in het primair onderwijs. De vertraagde leerlingen hebben een verlengde leertijd in het primair onderwijs gehad ten opzichte van de reguliere leerlingen. In welk leerjaar deze leerlingen die verlengde leertijd hebben gehad is daarbij niet bekend. Een relatief groot deel zal die mogelijk al in de kleuterperiode (groep 1 / 2) hebben opgelopen. Het betreft ook een relatief grote groep leerlingen geboren in de maanden augustus en september. Dit zouden de jonge leerlingen in de klas zijn als ze het reguliere traject zouden volgen.

Over de jaren heen wordt steeds gevonden dat de vertraagde leerlingen op de tot nu toe in dit onderzoek gemeten vaardigheden, rekenen en taal, gemiddeld minder goed presteren dan de reguliere leerlingen. Dat wordt ook bij de vaardigheden voor wereldoriëntatie gevonden. Bij Aardrijkskunde zien we de grootste effecten, die met waarden rond 0,51 als matig gekwalificeerd moeten worden. De effecten van Natuuronderwijs (effecten rond 0,38) en Geschiedenis (effecten rond 0,44) zijn klein te noemen, al wijken ze niet veel af van die bij Aardrijkskunde.

Over de twee jaren heen is er relatief weinig variatie binnen de vaardigheden ten gevolge van het effect van de variabele leertijd. In tabel 4.5 is ook te zien dat de waarden binnen een schaal en binnen een leertijd niet echt van elkaar verschillen. Dat is een positief resultaat als we ons realiseren dat er in 2014 relatief minder vertraagde leerlingen zijn dan in 2013. Daar waar bij de gehele groep die de eindtoets gemaakt heeft het percentage vertraagde leerlingen daalde van 16,9% naar 15,7%, is dat bij de groep die de toets

Wereldoriëntatie gemaakt heeft een daling van 16,5% naar 15,3%: een vergelijkbare daling. De gevonden daling in het percentage vertraagde leerlingen komt overeen met andere bronnen zoals die van de inspectie en DUO (zie verder Hemker en Van Weerden, 2015).

Doordat het percentage vertraagde, minder goed presterende leerlingen afneemt, kan de groep als geheel dus een sterker stijgende trend vertonen, dan de subgroepen. Het is een interessante vraag hoe dit geïnterpreteerd moet worden. Is het zo dat de leerlingen in de jaren ervoor beter opgeleid worden, zodat er minder noodzaak is voor leerlingen om te blijven zitten? Of is het beleid geworden om leerlingen minder vaak te laten doubleren en minder vaak langer te laten 'kleuteren'?

Als hetzelfde percentage leerlingen vertraagd zou zijn in 2014 als in 2013, dan zouden 1,2% van de leerlingen – met een vaardigheid waarschijnlijk tussen die van de reguliere en vertraagde leerlingen in – een jaar extra leertijd hebben.

Aardrijkskunde levert voor de variabele Leertijd een effectgrootte op van rond de 0,50, wat gekwalificeerd kan worden als een matig effect. Voor Natuuronderwijs en Geschiedenis liggen de waarden lager en kunnen we het effect als klein betitelen. Bij Geschiedenis is er een klein verschil tussen 2013 en 2014, maar dat verschil is niet significant (geen interactie met jaar).

Formatiegewicht

Het verschil tussen de leerlingen met een laag gewicht en leerlingen zonder gewicht kan bij alle drie de onderdelen als een matig effect worden gekwalificeerd. De waarden variëren van 0,50 tot 0,58.

Bij het verschil tussen de leerlingen met een laag gewicht en een hoog gewicht zien we wisselende resultaten. Bij de vaardigheden Aardrijkskunde en Geschiedenis zien we dat de gecorrigeerde effecten niet groter zijn dan 0,20 of kleiner zijn dan -0,20. Deze effecten zijn daarmee geclassificeerd als verwaarloosbaar. Bij Aardrijkskunde ligt het gevonden gecorrigeerde effect zelfs op 0,00. Dit hoeft overigens niet te betekenen dat de leerlingen met een laag gewicht (0,30) ook exact even goed presteren als leerlingen met een hoog gewicht (1,20). Dat blijkt ook uit de gevonden gemiddelden die gegeven zijn in tabel 4.5. Daarin is te zien dat bij Aardrijkskunde en Geschiedenis het gemiddelde van de leerlingen met een hoog gewicht lager ligt dan het gemiddelde van leerlingen met een laag gewicht. Dat dit verschil niet in de effectschatting teruggevonden wordt, heeft te maken met het feit dat de effectschatting gecorrigeerde effecten bevat, waarbij gecorrigeerd wordt voor het effect van stratum, leertijd en geslacht. Het betekent dat het additionele effect van een hoog ten opzichte van een laag effect vrijwel nihil is. Ook kan opgemerkt worden dat het ongecorrigeerde effect bij Aardrijkskunde kleiner dan 0,20 is en daarmee als betekenisloos geclassificeerd wordt. Bij Geschiedenis ligt het ongecorrigeerde effect rond 0,20, te classificeren als een klein effect.

Omdat het gecorrigeerde effect bij Aardrijkskunde en Geschiedenis nihil is bij de leerlingen met een laag (0,30) en een hoog (1,20) gewicht, lijken de effecten die gevonden worden bij de vergelijking tussen leerlingen zonder gewicht (0,00) en leerlingen met een hoog gewicht sterk op de effecten zoals gevonden bij leerlingen met een laag gewicht en leerlingen zonder een gewicht. De effectgrootte voor dit verschil ligt bij beide vaardigheden rond de 0,60 en is daarmee matig te noemen. Bij de vaardigheid Geschiedenis ziet het er naar uit dat het verschil tussen leerlingen zonder gewicht en met een hoog gewicht in 2014 groter is geworden dan in 2013. Dit verschil tussen de jaren is echter niet significant.

Bij Natuuronderwijs worden wel significante interactie-effecten gevonden. De leerlingen met een hoog formatiegewicht presteren in 2014 slechter dan in 2013. Dat is te zien aan de gemiddelden in tabel 4.5 maar ook aan het verschil in de effecten in tabel 4.4. Waar in 2013 het effect tussen leerlingen met een laag tegenover een hoog gewicht 0,16 is, is het in 2014 met 0,25 een klein effect geworden. Het verschil tussen 0,00-leerlingen en 1,20-leerlingen verandert van 0,63 in 2013 (matig effect) naar 0,85 in 2014 (groot effect). De achterstand die de leerlingen met een hoog formatiegewicht al hadden in 2014 bij Natuuronderwijs is dus vergroot.

VO-advies

Bij deze variabele komen we zeer grote verschillen tegen. We zien dat er sprake is van contrasten die overall significant zijn en overall betekenisvolle effectgroottes opleveren⁶. Die effectgroottes variëren van klein tot (zeer) groot. Het kleinste effect (0,40) wordt gevonden bij Natuuronderwijs in 2014 bij het verschil tussen vmbo-gt en vmbo-kb. Het grootste verschil wordt gevonden bij Aardrijkskunde in 2014 tussen vmbo-bb in vergelijking met vwo. Dat effect is 2,86 wat betekent dat de gemiddelde leerling in vmbo-bb bijna drie standaarddeviaties onder het niveau van de gemiddelde vwo-leerling zit. Er worden bij deze achtergrondvariabele wel wat kleine verschillen gevonden tussen de jaren, maar geen van die verschillen is significant. De verschillen tussen de adviesniveaus zijn bij Aardrijkskunde en Geschiedenis vergelijkbaar, en iets groter dan bij Natuuronderwijs. De gevonden verschillen zijn echter kleiner dan gevonden wordt bij de taal- en rekenvaardigheden, waar het grootste contrast, dat tussen vwo en vmbo-bb, zelfs een effectgrootte van 4,70 laat zien (Hemker & Van Weerden, 2015).

Toets

Bij deze variabele wordt onderscheid gemaakt tussen leerlingen die de reguliere Eindtoets Basis – of de even moeilijke ankertoets – maken en de leerlingen die de gemakkelijkere Eindtoets Niveau (EN) maken. De EN, die in deze vorm bestaat sinds 2013, is een papieren alternatief voor de reguliere eindtoets, waardoor leerlingen gewoon in de klas samen met de andere leerlingen de toets kunnen maken (voor 2013 kon deze toets alleen digitaal worden gemaakt). Deze versie van de toets is bestemd voor leerlingen waarvan de leerkracht verwacht dat ze doorstromen naar vmbo-bb of vmbo-kb.

De twee onderscheiden groepen leerlingen maken bij de eindtoets weliswaar verschillende toetsen (regulier tegenover EN), maar dat is niet het geval bij de toetsen Wereldoriëntatie. Daar maken beide groepen dezelfde toetsen. Zoals is te verwachten presteren de leerlingen die de reguliere eindtoets maken beter op de toets Wereldoriëntatie. De effectgrootte moet als groot worden gekwalificeerd met een waarde van 1,05 bij Natuuronderwijs, en rond 1,15 bij Aardrijkskunde en Geschiedenis. Deze effecten zijn vergelijkbaar met de effecten die voor deze achtergrondvariabele gevonden worden bij de taalvaardigheden.

Thuis taal

Deze achtergrondvariabele is in 2013 niet opgevraagd en er zijn dus geen resultaten voor 2013. Er zijn alleen resultaten voor 2014 en de groep “totaal” is hier dus niet aan de orde. Deze resultaten betreffen een tweetal verschillende opdelingen voor de variabele thuis taal. De eerste opdeling is een driedeling met de volgende categorieën: (1) de leerling spreekt thuis alleen Nederlands, (2) de leerling spreekt naast Nederlands ook een andere taal, of (3) de leerling spreekt thuis geen Nederlands. De tweede opdeling maakt ook onderscheid naar de verschillende thuistalen: Turks, Arabisch, Surinaams & Antilliaans, en een groep overige talen. Om de groepen voldoende groot te houden wordt geen onderscheid meer gemaakt tussen leerlingen die thuis alleen die andere taal spreken of naast die taal thuis ook Nederlands spreken.

Bij de gecorrigeerde effecten, waarbij is gecorrigeerd met de variabelen formatiegewicht, stratum, geslacht en leertijd, valt op dat de taal die thuis gesproken wordt alleen bij de vaardigheid Natuuronderwijs betekenisvolle, dan wel als matig te kwalificeren effecten oplevert. Dat betreft de contrasten van de leerlingen die thuis alleen Nederlands spreken vergeleken met de groep die naast Nederlands ook een andere taal spreekt (effect: 0,50) of de groep die alleen een andere taal spreekt (effect: 0,60). Het contrast tussen de groepen die alleen een buitenlandse taal spreken of thuis een gemengd taalgebruik kennen is bij Natuuronderwijs te verwaarlozen (effect: 0,10). Voor Aardrijkskunde en Geschiedenis worden geen significante en betekenisvolle effecten gevonden voor het al dan niet een andere taal spreken naast het Nederlands. Dat geldt ook voor het verschil tussen leerlingen die alleen een buitenlandse taal spreken of daarnaast ook nog Nederlands spreken. Dat verschil is niet significant bij een alpha van 0,01, ondanks dat de effectgrootte 0,27 is, zoals bij Geschiedenis). Het verschil tussen de leerlingen die alleen een andere taal dan Nederlands spreken en zij die alleen Nederlands spreken is wel significant en betekenisvol. De effectgrootte is met een waarde rond 0,30 als klein te kwalificeren.

⁶ De effecten zijn berekend op een wijze waarbij de ontbrekende waarden niet zijn meegenomen aangezien die een hogere standaard hebben dan de homogenere groepen van leerlingen waarvan het doorstroomadvies wel bekend is.

Kijken we naar de contrasten met de specifieke andere talen die al dan niet naast het Nederlands gesproken worden dan valt op dat de groep die thuis (ook) Surinaams dan wel Antilliaans spreekt het minst goed presteert. Het effect is groot bij Natuuronderwijs en Aardrijkskunde en matig bij Geschiedenis. Een ander groot effect wordt gevonden bij Natuuronderwijs voor leerlingen die Turks spreken (al dan niet naast het Nederlands). Bij deze groep wordt een matig effect van thuistaal gevonden bij Aardrijkskunde en een klein effect bij Geschiedenis. Bij de groep die Arabisch spreekt is het effect met 0,79 bij Natuuronderwijs bijna als groot te kwalificeren,. Bij de andere twee vaardigheden is het effect juist klein, en bij Geschiedenis weliswaar significant bij alpha 0,01 maar niet meer bij alpha⁷ is 0,001. Deze groep onderscheidt zich daarmee van de categorieën die thuis Turks of Surinaams/Antilliaans spreken. Voor leerlingen die een andere buitenlandse taal spreken dan de eerder genoemde talen is alleen bij Natuuronderwijs nog een klein betekenisvol effect gevonden. Bij de andere twee vaardigheden is het effect niet meer als betekenisvol te kwalificeren.

De gegeven gemiddelde scores in tabel 4.6 zijn gewogen voor stratum, maar zijn niet gezuiverd voor het effect van stratum, formatiegewicht, geslacht en leertijd. Daardoor zijn de verschillen tussen de groep die alleen Nederlands spreekt en de andere groepen wat groter. Bij Natuuronderwijs is bijvoorbeeld te zien dat het verschil tussen die groep leerlingen en de groepen die thuis Turks, Arabisch of Surinaams, dan wel Antilliaans spreken, meer dan een hele standaarddeviatie (= 50 punten) is. De verschillen zijn wat minder groot bij Aardrijkskunde en Geschiedenis.

Tabel 4.6 Gemiddelde schaalscores per onderwerp per categorie van de variabele thuistaal en het verschil ten opzichte van het gemiddelde in 2014

	Natuuronderwijs		Aardrijkskunde		Geschiedenis	
	Gem	Vershil	Gem	Vershil	Gem	Vershil
	251		249		250	
Thuistaal -algemeen						
Alleen Nederlands	255	4	252	2	254	4
Nederlands + Anders	218	-32	226	-24	242	-8
Alleen een andere taal	207	-44	218	-31	221	-29
Thuistaal - detail						
Alleen Nederlands	255	4	252	2	254	4
Turks (+)	194	-57	202	-47	213	-37
Arabisch (+)	197	-54	215	-34	220	-30
Surinaams/Antilliaans(+)	195	-55	198	-51	206	-44
Overige andere taal (+)	226	-25	234	-16	237	-13

Gem=gemiddelde voor de categorie; Verschil ten opzichte van gemiddelde voor het onderwerp; (+) = en Nederlands

4.2.2 Effecten voor de achtergrondvariabelen op schoolniveau

Bij de analyse van de invloed van de achtergrondvariabelen stratum, schoolgrootte, regio en verstedelijking is ook een correctie toegepast. Bij die variabelen is namelijk eveneens duidelijk een relatie met andere variabelen die maken dat voor een goede vergelijking een gezuiverd effect moet worden berekend. Daarin zijn de variabelen geslacht, leertijd en formatiegewicht als correctiefactoren meegenomen. De gecorrigeerde effecten zijn te vinden in tabel 4.7.

⁷ Bij alle andere betekenisvolle effecten in dit rapport is het resultaat wel significant bij een alpha van 0,001.

Tabel 4.7 Effecten van de variabelen stratum, schoolgrootte, regio en verstedelijking (na correctie)

Variabele	Contrast	Natuuronderwijs			Aardrijkskunde			Geschiedenis		
		Totaal	2013	2014	Totaal	2013	2014	Totaal	2013	2014
Stratum*										
Stratum 1-Stratum 2		0,15	0,16	0,14	0,19	0,17	0,20	0,19	0,19	0,18
Stratum 2-Stratum 3		0,46	0,49	0,42	0,26	0,24	0,27	0,22	0,24	0,21
Stratum 1-Stratum 3		0,61	0,65	0,56	0,44	0,41	0,48	0,41	0,44	0,38
Regio										
Oost - Noord		0,02	0,05	0,00	0,08	0,07	0,09	-0,02	0,02	-0,06
West - Noord		0,05	0,03	0,06	0,15	0,17	0,13	0,09	0,13	0,06
Zuid - Noord		0,06	0,03	0,09	0,08	0,08	0,08	-0,02	-0,02	-0,02
West - Oost		0,02	-0,02	0,07	0,07	0,10	0,04	0,11	0,11	0,12
Zuid - Oost		0,04	-0,02	0,10	0,00	0,01	-0,01	0,00	-0,04	0,04
Zuid - West		0,02	0,00	0,03	-0,07	-0,09	-0,05	-0,11	-0,15	-0,07
Verstedelijking										
Sterk - Zeer sterk		-0,11	0,10	0,12	0,02	0,05	-0,01	0,04	0,07	0,01
Matig - Sterk		-0,03	0,04	0,02	-0,03	-0,06	0,01	0,01	-0,01	0,03
Weinig - Matig		0,01	0,00	-0,02	-0,02	0,00	-0,04	-0,09	-0,09	-0,10
Niet - Weinig		-0,06	0,06	0,07	-0,02	-0,01	-0,03	0,02	0,02	0,02
Schoolgrootte										
Groot (>200 lln) - Klein (1-200 lln)		0,03	0,06	0,00	0,08	0,07	0,09	0,08	0,07	0,10

*Stratum 1= 0 tot 10%; stratum 2= 10 tot 25%; stratum 3= 25 tot en met 100% gewichtsleerlingen.

Stratum

De indeling van scholen naar stratum is gebaseerd op dezelfde variabele als formatiegewicht en we zien dat in de effecten bij de contrasten terug. Het beeld is vergelijkbaar, maar de verschillen zijn wat kleiner, niet altijd significant en het aantal betekenisvolle effecten is minder. Het verschil tussen de Stratum 1-scholen die minder dan 10% leerlingen met een formatiegewicht hebben en scholen uit Stratum 2 (10 tot 25% leerlingen met een formatiegewicht) is in de meeste gevallen net iets kleiner dan 0,20 en daarmee net niet betekenisvol (bijna een klein effect). Het verschil tussen Stratum 2-scholen en Stratum 3-scholen, die 25% of meer leerlingen met een formatiegewicht hebben, is bij alle vaardigheden wel betekenisvol. Dat is bij Natuuronderwijs het grootst, met waarden tussen 0,40 en 0,50. Bij de ander twee vaardigheden liggen deze effecten tussen de 0,20 en de 0,30. Al deze effecten zijn als klein te kwalificeren. Een zelfde tweedeling in effecten bij de vaardigheden wordt ook gevonden bij de vergelijking van Stratum 1- en Stratum 3-scholen: de gevonden effecten bij Geschiedenis en Aardrijkskunde zijn vergelijkbaar (tussen 0,40 en 0,50, en dus klein), terwijl deze bij Natuuronderwijs hoger liggen. Met waarden rond 0,60 is dat effect als matig te kwalificeren. Bij Natuuronderwijs worden wel wat verschillen gevonden tussen 2013 en 2014, maar die interactie is niet significant.

Als we naar de niet gezuiverde gemiddelden kijken bij Stratum dan zijn de verschillen (uiteraard) groter, omdat bij de hier gepresenteerde resultaten gecorrigeerd is voor de verdeling van formatiegewicht van de individuele leerling, hetgeen per definitie samenhangt met de variabele Stratum. De hier gepresenteerde effecten zijn dan ook eerder te zien als het effect van relatief meer of minder schoolgenoten met (ook) een gewicht. Het is wel duidelijk dat op de Stratum 3-scholen de prestaties het laagst zijn.⁸

⁸ Wellicht dat op de scholen met veel gewichtsleerlingen relatief veel aandacht uit gaat naar taal en rekenen, en dat daarmee het onderwijs in wereldoriëntatie wat achterblijft. Het kan ook zo zijn dat veel van de wereldoriëntatievaardigheden thuis bijgebracht worden en dat dit de oorzaak van de gevonden verschillen is.

Tabel 4.8 Gemiddelde schaalscores per onderwerp per categorie voor 2013 en 2014 en het jaarverschil

	Natuuronderwijs			Aardrijkskunde			Geschiedenis		
	2013	2014	verschil	2013	2014	verschil	2013	2014	verschil
Stratum									
Samen	250	251	1	250	249	-1	250	251	1
Stratum 1	258	257	0	257	256	0	257	258	1
Stratum 2	247	248	1	245	243	-2	245	245	0
Stratum 3	215	218	2	226	223	-3	225	227	3
Regio									
Samen	250	251	1	250	249	-1	250	251	1
Noord	252	250	-1	248	248	0	250	252	2
Oost	252	249	-3	249	251	1	250	249	-1
West	248	250	2	252	250	-2	252	253	1
Zuid	251	254	3	249	249	0	247	251	4
Urbanisatiegraad									
Samen	250	251	1	250	249	-1	250	251	1
Zeer Sterk	234	236	3	242	242	1	240	244	4
Sterk	248	250	2	251	250	-1	252	251	0
Matig	253	253	0	251	252	1	253	255	2
Weinig	254	253	0	251	251	-1	250	251	1
Niet	257	257	0	252	250	-2	251	253	2

Schoollocatie: Regio en verstedelijking

De locatie van de school en de mate van verstedelijking hebben geen toegevoegd effect bij de variabele formatiegewicht, stratum, leertijd en geslacht: geen van de gecorrigeerde effecten is betekenisvol. Als we naar de gewogen, maar niet gezuiverde gemiddelden kijken dan zien we dat leerlingen op scholen in zeer sterk stedelijke gebieden iets minder goed presteren, maar daar zijn ook meer leerlingen die thuis een andere taal spreken en een lager formatiegewicht hebben. Over het algemeen geldt: hoe minder stedelijk het gebied, hoe groter het percentage leerlingen dat thuis alleen Nederlands spreekt en hoe groter het percentage leerlingen zonder formatiegewicht.

Schoolgrootte en denominatie

Bij de variabele schoolgrootte is een school tot 200 leerlingen gedefinieerd als een kleine school en een school met 200 of meer leerlingen als een grote school. De richting van het effect is in alle gevallen, net als bij de taal- en rekvaardigheden, in het voordeel van de grotere scholen, maar in geen van de jaren is voor geen van de vaardigheden sprake van een betekenisvol effect. Dat is vergelijkbaar met de uitkomst bij de taal- en rekvaardigheden (Hemker en Van Weerden, 2015).

Er is ook een klein onderzoek uitgevoerd naar het effect van denominatie. Daarbij is de opdeling in rooms-katholiek (38,7% van de leerlingen die in 2013 en 2014 mee hebben gedaan aan de toetsen Wereld-oriëntatie), openbaar (27,6%) en protestants-christelijk (24,3%) gebruikt. Bijna 90% van de leerlingen zit op een van dit soort scholen. De gecorrigeerde effecten blijken geen van alle betekenisvol (tabel 4.9).

Kijken we naar de marginalen of de ongecorrigeerde effecten dan levert dat weinig op voor de verschillen tussen de leerlingen op de rooms-katholieke scholen en die op de protestants-christelijke scholen. Bij de openbare scholen pakken de ongecorrigeerde effecten ten opzichte van de andere twee denominaties in lichte mate negatief uit: Openbare scholen hebben relatief veel leerlingen met een formatiegewicht.

Tabel 4.9 Effecten van de variabele denominatie (na correctie)

Variabele	Contrast	Natuuronderwijs			Aardrijkskunde			Geschiedenis		
		Totaal	2013	2014	Totaal	2013	2014	Totaal	2013	2014
Denominatie										
	Openbaar - RK	0,00	-0,02	0,02	-0,03	-0,03	-0,03	0,00	0,03	-0,02
	PC - RK	0,00	0,00	0,00	0,00	0,01	-0,02	0,07	0,09	0,05
	PC - Openbaar	0,00	0,02	-0,02	0,03	0,03	0,02	0,07	0,06	0,07

5 Conclusies

In deze rapportage worden de scores op de toets Wereldoriëntatie, een facultatief toetsonderdeel bij de Eindtoets, voor het eerst in kaart gebracht als basis voor een jaarlijkse peiling. Daarbij worden dezelfde analysetechnieken ingezet als die reeds sinds 2008 gehanteerd worden voor de onderdelen taal en rekenen in het kader van de jaarlijkse meting taal en rekenen (Hemker en Van Weerden, 2015). Deze zijn aangevuld met extra analyses, die uitgevoerd konden worden omdat de toetsen in 2013 en 2014 identiek waren. Startjaar voor de vergelijking over de tijd is nu 2013. In deze eerste analyse kunnen we reeds 2014 met 2013 vergelijken. We onderscheiden daarbij de onderdelen Aardrijkskunde, Geschiedenis en Natuuronderwijs.

5.1 Algemeen beeld

Het verschil in resultaten tussen 2013 en 2014 blijkt minimaal te zijn. Het maakt daarbij niet uit of dit op klassieke wijze of met behulp van IRT wordt geanalyseerd. Wel is er enig verschil tussen de onderscheiden onderdelen van wereldoriëntatie.

Geschiedenis wordt in 2014 beter gemaakt dan in 2013, en ook bij Natuuronderwijs is er sprake van een zeer geringe stijging van de prestaties. Bij Aardrijkskunde zien we een zeer geringe daling. De aangetroffen verschillen zijn echter vooralsnog zo klein dat ze niet als betekenisvol kunnen worden gekwalificeerd. Pas als deze verschillen zich in de loop van de tijd als trends zouden doorzetten, kan dat leiden tot geaccumuleerde effecten die, net zoals bij de jaarlijkse metingen rekenen en taal (Hemker en Van Weerden, 2015), betekenisvol kunnen worden genoemd.

5.2 Invloed van achtergrondvariabelen

Geslacht

Jongens in jaargroep 8 behalen hogere scores op alle drie de vaardigheden voor wereldoriëntatie. De verschillen zijn het grootst bij Aardrijkskunde.

Leertijd

Vertraagde leerlingen blijken in alle gevallen een behoorlijke achterstand te hebben. Die achterstand is het grootst bij Aardrijkskunde en het kleinst bij Natuuronderwijs. Opvallend bij deze variabele is de verandering in het percentage vertraagde leerlingen op school. We zien dat in vergelijking met 2013 in plaats van iets minder dan 17% nu ruim 15% van de leerlingen vertraagd is⁹. Het feit dat minder leerlingen een verlengde leertijd krijgen, heeft niet geleid tot een daling in de vaardigheid.

Formatiegewicht

We zien dat leerlingen zonder formatiegewicht (0.00) op alle drie de vaardigheden beter presteren dan leerlingen mét een formatiegewicht, laag (0.30) of hoog (1.20). Het gaat dan om kleine tot matige effectgroottes. Bij Aardrijkskunde en Geschiedenis blijkt er nauwelijks verschil te zijn tussen leerlingen met een laag of een hoog gewicht. Bij Natuuronderwijs is dat er wel, maar is de effectgrootte klein. Het contrast tussen leerlingen met een zwaar gewicht en leerlingen zonder gewicht is hier dan ook matig (2013) tot groot (2014).

⁹ Merk op dat anders dan de percentages 17% en 15% doen vermoeden het verschil tussen de twee jaren 1% is (1,2%); 17% en 15% zijn beiden afgeronde waarden.

Advies Voortgezet onderwijs

Het verschil in vaardigheid verschilt aanzienlijk tussen de groepen van leerlingen met verschillende doorstroomadviezen. Gemiddeld is het verschil tussen de vaardigheid van een vmbo-bb-leerling en een vwo-leerling 2,5 tot 3 standaarddeviaties. Het geeft aan hoe groot de vaardigheidsverschillen aan het eind van de basisschool al zijn, niet alleen voor taal en rekenen maar ook voor wereldoriëntatie. De verschillen bij taal en rekenen zijn wel groter dan bij wereldoriëntatie. Het contrast vmbo bb-vwo is bij Aardrijkskunde het grootst en bij Natuuronderwijs het kleinst. Ook de contrasten bij alle tussenliggende doorstroomniveaus leveren significante en betekenisvolle effectgroottes op. Tussen havo en vwo is die bijvoorbeeld al minstens een hele standaarddeviatie, wat staat voor een groot effect, maar ook tussen de onderscheiden leerwegen in het vmbo zijn de contrasten steeds betekenisvol en wat effectgrootte betreft minstens klein te noemen en vaak matig.

Thuis taal

We zien dat de variabele thuis taal er bij wereldoriëntatie ook toe doet. Over het algemeen geldt voor deze onderwerpen: hoe meer Nederlands de leerling thuis spreekt hoe hoger de vaardigheid is, maar de contrasten zijn lang niet overal significant of betekenisvol. Het grootste effect vinden we bij Natuuronderwijs en wel voor het contrast van leerlingen die alleen Nederlands spreken met leerlingen die thuis ook Surinaams/Antilliaans spreken, gevolgd door leerlingen die Turks of Arabisch spreken. Bij Aardrijkskunde en Geschiedenis leveren de contrasten met Arabisch een klein effect op. Geschiedenis levert, vergeleken met de andere twee vaardigheden, de kleinste effecten op.

Stratum

Deze variabele is geconstrueerd op basis van de verdeling in formatiegewichten op school. De verschillen en effecten zijn dan ook grotendeels vergelijkbaar met die bij de variabele formatiegewicht. De verschillen zijn nagenoeg overal significant, maar niet in alle gevallen met betekenisvolle effectgroottes. Het contrast tussen scholen met veel en weinig gewichtsleerlingen is het grootst bij Natuuronderwijs.

Regio en verstedelijking

De effecten voor deze twee variabelen blijken, gecorrigeerd voor stratum, formatiegewicht, leertijd en geslacht, niet betekenisvol. We vinden dat ook bij de analyse van de taal- en rekenvaardigheden (zie Hemker en Van Weerden, 2015). De gevonden verschillen tussen de gemiddelde scores van de leerlingen in de zeer sterk stedelijke gebieden en de minder stedelijke gebieden bij de ongecorrigeerde vergelijking komen voort uit de situatie dat er relatief veel achterstandsleerlingen in die stedelijke gebieden wonen. Door de correctie valt dat verschil weg.

Schoolgrootte

Deze achtergrondvariabele levert geen betekenisvolle effectgroottes op.

5.3 Discussie

Dit is de eerste rapportage over het niveau van de leerprestaties in Aardrijkskunde, Geschiedenis en Natuuronderwijs, zoals dat tot uiting komt in de prestaties op het onderdeel wereldoriëntatie in de Eindtoets Basisonderwijs. Deze rapportage kan als verbreding gezien worden van het jaarlijkse peilingsonderzoek taal en rekenen, dat sinds 2008 wordt uitgevoerd, en waarin op basis van een vaste schaal de prestaties over de jaren heen gevolgd kunnen worden.

Hoewel het de eerste analyse en rapportage betreft voor deze onderwerpen kan al wel een jaarvergelijking worden gemaakt, namelijk voor de jaren 2013 en 2014. In beide jaren is exact dezelfde toets afgenomen.

We zien in de vergelijking dat er enige beweging is in de hoogte van de schaalscores, maar de verschillen zijn vooralsnog te klein om betekenisvol te kunnen worden genoemd. Zo zien we een lichte vooruitgang bij Geschiedenis en in mindere mate bij Natuuronderwijs. Aardrijkskunde lijkt iets achteruit te gaan.

Als we deze meting doorzetten en dit een vast onderdeel van het jaarlijkse peilingsonderzoek laten zijn, dan is de kans aanwezig dat we op termijn wel tot betekenisvolle verschillen in resultaat komen. Effecten zouden zich dan over meerdere jaren moeten doorzetten.

Een belangrijk verschil met de meting bij taal en rekenen is wel de beleidsachtergrond. Voor taal en rekenen zijn duidelijk impulsen vanuit de overheid gegeven om in 2008 te beginnen met de Kwaliteitsagenda Scholen voor morgen. Daarna is dit gecontinueerd en heeft dit tot nieuwe wetgeving geleid (referentieniveaus, verplichte eindtoets voor taal en rekenen en gebruik leerlingvolgsysteem). Gelijktijdig werden scholen uitgenodigd tot deelname aan verbeterprojecten voor taal en rekenen. Dit soort impulsen zijn er niet geweest voor de vakken vallend onder wereldoriëntatie. De vraag is dan ook of het reëel is te verwachten dat er zich wijzigingen zullen voordoen in positieve zin, indien het onderwijsaanbod niet wezenlijk beïnvloed wordt. Wel zijn er mogelijk invloeden vanuit de samenleving die het onderwijsaanbod en de leerlingprestaties ten goede of ten kwade kunnen beïnvloeden. Alleen al om dat te kunnen vaststellen is verdere continuering van deze meting toch wenselijk.

Zoals al hiervoor geïmpliceerd: het is lastig op grond van de uitkomsten van dit onderzoek een adequate beleidsevaluatie te plegen. Voor het vinden van verklaringen ontbreken ook gegevens over wat scholen precies hebben gedaan op het gebied van de gemeten vaardigheden. We kunnen dus niet vaststellen in hoeverre hierin wijzigingen zijn opgetreden in de loop der jaren. Dat geldt ook voor de eventuele wijzigingen in achtergrondvariabelen, zoals bijvoorbeeld het al of niet vertraagd zijn van leerlingen

Het verdient daarom aanbeveling om, net als bij het regulier peilingsonderzoek (PPON) gebruikelijk was, het onderwijsaanbod of veranderingen daarin te inventariseren zodanig dat gegevens betreffende de leerlingprestaties daaraan gekoppeld kunnen worden. Met behulp van informatie over bijvoorbeeld het gebruik van andere methoden of tijdsbesteding aan bepaalde onderwerpen kan wellicht een plausibele interpretatie worden verbonden aan de verandering in resultaten bij leerlingen. Een goede inventarisatie van het onderwijsaanbod bij een steekproef van scholen kan hierin al volstaan. Het verdient aanbeveling een dergelijke inventarisatie bij een volgende meting te organiseren.

Literatuur

Boxtel, H. van, R. Engelen, A. de Wijs (2012). *Wetenschappelijke verantwoording van de Eindtoets Basisonderwijs 2010*. Cito, Arnhem.

Cito (2014). *Terugblik en resultaten 2014. Eindtoets basisonderwijs jaargroep 8*. Cito, Arnhem.

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*. SLO, Enschede.

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.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*. Cito, Arnhem. (zie www.cito.nl, Onderzoek en wetenschap, PPON)

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*. Cito, Arnhem. (<http://www.minocw.nl/documenten/133682d.pdf>)

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 - Technische rapportage*. Cito, Arnhem. (zie www.cito.nl, Onderzoek en wetenschap, PPON)

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

Inspectie van het onderwijs (2012). *Monitor verbetertrajecten taal en rekenen 2008/2009, 2009/2010 en 2010/2011*, Inspectie van het Onderwijs/OCW, Utrecht.

Kneepkens, B., Van der Schoot, F., Hemker, B.T. (2011). *Balans van het natuurkunde- en techniekonderwijs aan het einde van de basisschool 4. Uitkomst van de vierde peiling in 2010*. PPON-reeks nummer 43. Cito, Arnhem.

Notté, H., Van der Schoot, F., Hemker, B.T. (2011). *Balans van het aardrijkskundeonderwijs aan het einde van de basisschool 4. Uitkomst van de vierde peiling in 2008*. PPON-reeks nummer 41. Cito, Arnhem.

Thijssen, J., Van der Schoot, F., Hemker, B.T. (2011). *Balans van het biologieonderwijs aan het einde van de basisschool 4. Uitkomst van de vierde peiling in 2010*. PPON-reeks nummer 44. Cito, Arnhem.

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 - Jaarlijks Peilingsonderzoek van het Onderwijsniveau*. Cito, Arnhem. (zie www.cito.nl, Onderzoek en wetenschap, PPON)

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.

Wagenaar, H., Van der Schoot, F., Hemker, B.T. (2010). *Balans van het geschiedenisonderwijs aan het einde van de basisschool 4. Uitkomst van de vierde peiling in 2008*. PPON-reeks nummer 42. Cito, Arnhem.

Relevante websites:

www.cito.nl

www.toetswijzer.nl

<http://ppon.cito.nl>

www.minocw.nl

Bijlagen

Bijlage 1 Effectschattingen voor alle vaardigheden gecategoriseerd naar achtergrondvariabele in jaargroep 8 in 2013 en 2014

Effecten		Natuuronderwijs		Aardrijkskunde		Geschiedenis	
		2013	2014	2013	2014	2013	2014
Contrast							
Formatie- gewicht	F 0.00 – F 0.30	0,47 ***	0,50 ***	0,60 ***	0,58 ***	0,48 ***	0,54 ***
	F 0.30 – F 1.20	0,16 *	0,35 ***	0,01	-0,01	0,07	0,12
	F 0.00 – F 1.20	0,63 ***	0,85 ***	0,61 ***	0,57 ***	0,55 ***	0,66 ***
Geslacht	Meisjes - Jongens	-0,35 ***	-0,34 ***	-0,61 ***	-0,60 ***	-0,36 ***	-0,43 ***
Leertijd	Regulier - Vertraagd	0,38 ***	0,39 ***	0,51 ***	0,51 ***	0,45 ***	0,43 ***
Stratum	Stratum 1 – Stratum 2	0,16 ***	0,14 ***	0,17 ***	0,20 ***	0,19 ***	0,18 ***
	Stratum 2 – Stratum 3	0,49 ***	0,42 ***	0,24 ***	0,27 ***	0,24 ***	0,21 ***
	Stratum 1 – Stratum 3	0,65 ***	0,56 ***	0,41 ***	0,48 ***	0,44 ***	0,38 ***
Soort toets	Eindtoets Basis - Niveau	1,05 ***	1,05 ***	1,11 ***	1,19 ***	1,17 ***	1,15 ***
Advies VO	vmbo-KB - vmbo-BB	0,47 ***	0,53 ***	0,62 ***	0,58 ***	0,52 ***	0,58 ***
	vmbo-GT - vmbo-KB	0,46 ***	0,40 ***	0,41 ***	0,47 ***	0,52 ***	0,58 ***
	havo - vmbo-GT	0,64 ***	0,62 ***	0,72 ***	0,77 ***	0,72 ***	0,62 ***
	vwo - havo	0,89 ***	0,97 ***	1,02 ***	1,05 ***	1,00 ***	1,01 ***
	vwo - vmbo BB	2,46 ***	2,52 ***	2,76 ***	2,86 ***	2,76 ***	2,78 ***
Regio	Oost - Noord	0,05	0,00	0,07	0,09 *	0,02	-0,06
	West - Noord	0,03	0,06	0,17 ***	0,13 **	0,13 **	0,06
	Zuid - Noord	0,03	0,09 *	0,08	0,08	-0,02	-0,02
	West - Oost	-0,02	0,07 *	0,10 ***	0,04	0,11 ***	0,12 ***
	Zuid - Oost	-0,02	0,10 **	0,01	-0,01	-0,04	0,04
	Zuid - West	0,00	0,03	-0,09 **	-0,05	-0,15 ***	-0,07 **
Urbanisatiegraad	Sterk - Zeer sterk	0,10 **	0,12 **	0,05	-0,01	0,07	0,01
	Matig - Sterk	0,04	0,02	-0,06	0,01	-0,01	0,03
	Weinig - Matig	0,00	-0,02	0,00	-0,04	-0,09 **	-0,10 **
	Niet - Weinig	0,06	0,07	-0,01	-0,03	0,02	0,02
Thuis taal - algemeen	Alleen NLS - NLS+Ander		0,50 ***		0,20		0,02
	NLS+Ander - Alleen Ander		0,10		0,11		0,27 *
	Alleen NLS - Alleen Ander		0,61 ***		0,32 ***		0,29 ***
Thuis taal - detail: Alleen Nederlands tegenover –	Turks (+)		0,83 ***		0,62 ***		0,46 ***
	Arabisch (+)		0,79 ***		0,36 ***		0,26 **
	Surinaams/Antilliaans (+)		0,91 ***		0,83 ***		0,59 ***
	Overige andere taal (+)		0,33 ***		0,06		0,05
Schoolgrootte	Groot - Klein	0,06	0,00	0,07 **	0,09 ***	0,07 **	0,10 ***

- geclassificeerd als “geen effect”, maar wel significant: *; 0,01 < α < 0,001; **; 0,001 < α < 0,0001; ***: α < 0,0001
- Sp = Spelling; BL-A = Begrijpend lezen (anker); BL-B = Begrijpend lezen (toets BV); WS = Woordenschat; G/B = Getallen en bewerkingen; BPV = Breuken, procenten, verhoudingen; MMTG = Meetkunde, meten, tijd en geld.
- (+) bij Thuis taal details betekent: de genoemde andere taal, al dan niet in combinatie met Nederlands

Bijlage 2 Gemiddelden en standaarddeviaties per vaardigheid gecategoriseerd naar achtergrondvariabele in jaargroep 8 in 2013 en 2014

	Vaardigheid	Natuuronderwijs				Aardrijkskunde				Geschiedenis			
	Jaar	2013		2014		2013		2014		2013		2014	
Variabele	Categorie	Gem	SD	Gem	SD	Gem	SD	Gem	SD	Gem	SD	Gem	SD
Jaar	2013 en 2014	250	50	251	50	250	50	249	50	250	50	251	50
Formatie- gewicht	F 0.00	254	48	255	48	254	48	254	48	254	49	255	49
	F 0.30	224	49	223	49	219	49	217	49	222	49	222	49
	F 1.20	204	49	197	49	212	49	211	49	212	49	211	49
Geslacht	Jongens	258	49	258	49	263	48	263	48	259	49	261	49
	Meisjes	242	49	243	49	237	48	236	48	242	49	242	49
Leertijd	Regulier	254	49	254	49	254	49	254	49	254	49	255	49
	Vertraagd	231	51	232	51	229	50	227	50	229	50	231	50
Stratum	Stratum (lovs*) 1	258	48	257	48	257	48	256	48	257	48	258	49
	Stratum (lovs*) 2	247	48	248	49	245	49	243	49	245	49	245	49
	Stratum (lovs*) 3	215	50	218	51	226	51	223	51	225	50	227	51
Soort toets	Eindtoets Basis	257	46	259	46	258	46	258	45	258	46	260	45
	Eindtoets Niveau	203	48	205	48	201	47	200	47	199	47	202	46
Advies Voortgezet Onderwijs	vmbo-BB	188	41	191	41	186	39	185	37	185	38	188	38
	vmbo-KB	212	40	216	41	213	38	210	37	209	38	213	38
	vmbo-GT	233	40	235	40	232	38	231	37	231	38	238	38
	havo	261	39	262	39	262	38	261	37	261	37	263	38
	vwo	298	39	303	39	303	37	303	36	302	37	304	37
Thuis taal – algemeen	Alleen Nederlands	.	.	255	48	.	.	252	49	.	.	254	49
	Nederlands+Anders	.	.	218	51	.	.	226	51	.	.	242	51
	Alleen een andere taal	.	.	207	51	.	.	218	51	.	.	221	51
Thuis taal – detail	Alleen Nederlands	.	.	255	48	.	.	252	49	.	.	254	49
	Turks (+)	.	.	194	50	.	.	202	51	.	.	213	51
	Arabisch (+)	.	.	197	50	.	.	215	51	.	.	220	51
	Surinaams/Antilliaans(+)	.	.	195	49	.	.	198	51	.	.	206	51
	Overige andere taal (+)	.	.	226	51	.	.	234	50	.	.	237	51
Regio	Noord	252	49	250	49	248	49	248	49	250	49	252	49
	Oost	252	49	249	50	249	50	251	50	250	50	249	50
	West	248	51	250	51	252	50	250	50	252	50	253	50
	Zuid	251	50	254	50	249	50	249	50	247	50	251	50
Urbanisatie- graad	Zeer Sterk	234	52	236	52	242	52	242	52	240	52	244	52
	Sterk	248	50	250	50	251	50	250	50	252	50	251	50
	Matig	253	49	253	49	251	50	252	50	253	50	255	50
	Weinig	254	49	253	49	251	49	251	49	250	49	251	49
	Niet	257	48	257	48	252	49	250	49	251	49	253	49
Schoolgrootte	Klein (1-200 ln)	246	50	248	51	246	50	245	50	246	50	246	50
	Groot (>200 ln)	252	50	252	50	252	50	252	50	252	50	254	50

Bijlage 3 Verdeling van de achtergrondvariabelen (AV) in de onderzoekspopulatie in 2013 2014 die een eindtoets gemaakt heeft, al dan niet met de toetsen Wereldoriëntatie

Aantallen WO gedaan	2013			2014			2013 & 2014	
	Nee	Ja	Totaal	Nee	Ja	Totaal	Nee	Ja
TOTAAL	30244	120512	150756	34430	113884	148314	64674	234396
AV: GESLACHT								
Jongen	14879	59016	73895	16782	56163	72945	31661	115179
Meisje	15054	60405	75459	17216	56547	73763	32270	116952
onbekend	311	1091	1402	432	1174	1606	743	2265
AV: Formatiegewicht 1								
0.00 & onbekend	25396	107802	133198	29398	103082	132480	54794	210884
0.30	2273	7794	10067	2329	6493	8822	4602	14287
1.20	2575	4916	7491	2703	4309	7012	5278	9225
AV: Formatiegewicht 2								
0.00	20615	88761	109376	24229	85317	109546	44844	174078
0.30	2273	7794	10067	2329	6493	8822	4602	14287
1.20	2575	4916	7491	2703	4309	7012	5278	9225
Onbekend	4781	19041	23822	5169	17765	22934	9950	36806
AV: Doorstroomadvies								
BB	1698	4901	6599	1824	4857	6681	3522	9758
BB/KB	1105	4234	5339	1170	3700	4870	2275	7934
KB	2486	8200	10686	3083	8642	11725	5569	16842
KB/GT	189	709	898	210	665	875	399	1374
GT	4440	17455	21895	5070	16729	21799	9510	34184
GT/HA	2810	11973	14783	3032	10450	13482	5842	22423
HA	3300	13736	17036	4004	13792	17796	7304	27528
HA/VW	2582	11547	14129	2696	9877	12573	5278	21424
VW	2969	13421	16390	3654	13668	17322	6623	27089
Onbekend	8665	34336	43001	9687	31504	41191	18352	65840
AV: Thuis taal 1								
Alleen NLs	0	0	0	28413	101108	129521	28413	101108
NL+Ander	0	0	0	701	1259	1960	701	1259
Alleen Ander	0	0	0	3648	7013	10661	3648	7013
Onbekend	30244	120512	150756	1668	4504	6172	31912	125016
AV: Thuis taal 2								
Alleen NLs	0	0	0	28413	101108	129521	28413	101108
Turks (*)	0	0	0	1294	2150	3444	1294	2150
Arabisch (*)	0	0	0	1338	2104	3442	1338	2104
Sur/Ant (*)	0	0	0	277	491	768	277	491
Anders (*)	0	0	0	1438	3522	4960	1438	3522
Onbekend	30244	120512	150756	1670	4509	6179	31914	125021
AV: Type toets								
EB/EA	25000	104404	129404	28191	97408	125599	53191	201812
EN	5244	16108	21352	6239	16476	22715	11483	32584

AV: Leertijd (LF op 1 okt)								
Regulier (10-11jr)	24457	99834	124291	28334	95649	123983	52791	195483
Vertraagd (12-13jr)	5484	19715	25199	5748	17279	23027	11232	36994
Onbekend **	303	963	1266	348	956	1304	651	1919
*** of LF=? of <10jr of >13jr								
AV: Leertijd extended (LF op 1 okt)								
Regulier (t/m 11jr)	24483	99940	124423	28368	95754	124122	52851	195694
Vertraagd (vanaf 12jr)	5485	19721	25206	5748	17284	23032	11233	37005
Onbekend 276	851	1127	314	846	1160	590	1697	
AV: Stratum								
S1: 0% t/m 10% gew. Iln	16150	78247	94397	19148	78010	97158	35298	156257
S3: >10% tot 25% gew. Iln	7676	30707	38383	8717	25744	34461	16393	56451
S3: >25% gew. Iln	6418	11558	17976	6565	10130	16695	12983	21688
AV: Regio								
Noord	3656	11809	15465	4380	10513	14893	8036	22322
Oost7042		30039	37081	8383	27956	36339	15425	57995
West	15810	47684	63494	17273	45419	62692	33083	93103
Zuid 3736		30980	34716	4392	29996	34388	8128	60976
Onbekend	0	0	0	2	0	2	2	0
AV: Urbanisatie 1								
Zeer sterk stedelijk	7077	15497	22574	8066	14602	22668	15143	30099
Sterk stedelijk	8720	28322	37042	9587	26570	36157	18307	54892
Matig stedelijk	5808	26314	32122	6395	25187	31582	12203	51501
Weinig stedelijk	5112	33427	38539	6247	31603	37850	11359	65030
Niet stedelijk	3527	16952	20479	4133	15922	20055	7660	32874
Onbekend	0	0	0	2	0	2	2	0
AV: Urbanisatie 2								
(Zeer) Sterk stedelijk	15797	43819	59616	17653	41172	58825	33450	84991
Matig tot niet stedelijk	14447	76693	91140	16775	72712	89487	31222	149405
Onbekend	0	0	0	2	0	2	2	0
AV: Schoolgrootte								
Klein (1-200)	10491	35772	46263	11742	34584	46326	22233	70356
Groot (>200)	19753	84740	104493	22688	79300	101988	42441	164040
AV: Denominatie								
Rooms-Katholiek	8032	46474	54506	9268	44215	53483	17300	90689
Openbaar	12509	33166	45675	13979	31426	45405	26488	64592
Protestants-Christelijk	5983	29440	35423	7229	27439	34668	13212	56879
Algemeen Bijzonder	1503	5419	6922	1458	5295	6753	2961	10714
Overig (<5%)	2217	6013	8230	2461	5509	7970	4678	11522
Onbekend	0	0	0	35	0	35	35	0

Cito helpt je inzicht te krijgen in je ontwikkeling en mogelijkheden. Door kennis, vaardigheden en competenties objectief meetbaar te maken en de ontwikkeling er van te volgen, kun je het beste uit jezelf halen, verantwoorde keuzes maken en beter richting geven aan je toekomst. Cito draagt daaraan bij door wereldwijd werk te maken van goed en eerlijk toetsen, vanuit de kernwaarden kundig, toonaangevend, integer, innovatief en betrokken.

Cito

Amsterdamseweg 13
Postbus 1034
6801 MG Arnhem
T (026) 352 11 11
www.cito.nl

Fotografie: Ron Steemers