

Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2013

Jaarlijks Peilingsonderzoek van het Onderwijsniveau

Jan van Weerden, Bas Hemker en Kees Mulder



Jaarlijks Peilingsonderzoek naar het Onderwijsniveau

Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2013

Jan van Weerden, Bas Hemker en Kees Mulder

© Stichting Cito Instituut voor Toetsontwikkeling Arnhem (2014)

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

Samenvatting 7

Inleiding 9

1 Vraagstelling en aanpak 11

- 1.1 Vraagstelling 11
- 1.2 Methode 11
 - 1.2.1 Kwaliteit van de meetinstrumenten 12
 - 1.2.2 Steekproeftrekking en analyse 13
 - 1.2.3. Achtergrondvariabelen 14
- 1.3 Hoe is er geanalyseerd? 15
 - 1.3.1 Vergelijking tussen en over de jaren 15
 - 1.3.2 Analyses voor de leerlingen in 2013 16
 - 1.3.3 Significantie en effectgrootte 16
 - 1.3.4 De gekozen rapportageschaal 17

2 De resultaten voor jaargroep 8 19

- 2.1 De vergelijking over de jaren 19
 - 2.1.1 Taal 20
 - 2.1.2 Rekenen-Wiskunde 21
- 2.2 Verschillen voor achtergrondvariabelen 22
 - 2.2.1 Effecten op leerlingniveau 23
 - 2.2.2 Effecten op schoolniveau 28

3 De resultaten voor jaargroep 4 31

- 3.1 De vergelijking over de jaren 31
- 3.2 Verschillen per achtergrondvariabele 33
 - 3.2.1 Effecten voor leerlingvariabelen 33
 - 3.2.2 Effecten van schoolvariabelen 37

4 Conclusies 39

- 4.1 Algemeen beeld 39
- 4.2 Invloed van achtergrondvariabelen 39
- 4.3 Discussie 40

Literatuur 43

Voorwoord

Dit rapport vormt de neerslag van de zesde meting van de taal- en rekenvaardigheden van leerlingen in het basisonderwijs sinds 2008. De meting is van start gegaan naar aanleiding van de kwaliteitsagenda voor het Primair Onderwijs, verschenen in 2007 onder de titel 'Scholen voor morgen' en wordt sindsdien gecontinueerd als landelijke monitor voor een aantal basisvaardigheden.

Bij de uitvoering van dit onderzoek wordt gebruikgemaakt van data uit toetssystemen die al bij de meeste scholen worden gebruikt voor het in kaart brengen van onderwijsopbrengsten bij rekenen en taal. Het gaat dan om de Eindtoets in jaargroep 8 en toetsen van het Cito Volgsysteem voor primair onderwijs in jaargroep 4. Door deze keuze wordt de extra toetslast voor leerlingen en leerkrachten beperkt.

De uitkomsten van deze jaarlijkse peiling geven een indicatie voor de stand van zaken met betrekking tot een belangrijk deel van de referentieniveaus behorende bij de doorlopende leerlijnen zoals geformuleerd in het advies van de Expertgroep Doorlopende leerlijnen "Over de drempels met taal en rekenen".

De resultaten voor jaargroep 8 worden meegenomen in het jaarlijkse verslag 'De staat van het onderwijs' van de Inspectie van het onderwijs.

Dit rapport zou niet mogelijk geweest zijn zonder de medewerking van de leerkrachten en schoolleiders van de basisscholen die bereid waren om hun gegevens ter beschikking te stellen van ons onderzoek. Wij danken hen hartelijk voor hun bijdrage aan het welslagen van deze peiling.

Bij de uitvoering van het project zijn diverse mensen betrokken geweest, waarvan er enkele met name genoemd moeten worden. De verspreiding van toetsmaterialen en de dataverzameling voor jaargroep 4 werd met zorg uitgevoerd door Gijs Marissink en Sylvia Praster. De analyses voor jaargroep 4 zijn uitgevoerd door Kees Mulder. De analyses voor jaargroep 8 werden door Bas Hemker uitgevoerd.

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

Drs. J.J. van Weerden
Projectleider PPON

Samenvatting

In dit rapport wordt verslag gedaan van de zesde jaarlijkse peiling van de reken- en taalvaardigheid in de jaargroepen 4 en 8. Daarvoor is in jaargroep 8 gebruikgemaakt van gegevens uit de Eindtoets van 2013 betreffende Woordenschat, Spelling, Begrijpend lezen en Rekenen-Wiskunde. Voor jaargroep 4 is een afzonderlijke steekproef van scholen getrokken en zijn scholen voorzien van LVS-toetsen uit het Cito Volgsysteem primair onderwijs voor Woordenschat, Spelling, Begrijpend lezen en Rekenen-Wiskunde.

Vergelijken we de prestaties van 2013 met die van 2012, dan is het beeld divers. De prestaties van de leerlingen is nog verder verbeterd bij Woordenschat, gestabiliseerd bij Begrijpend lezen en bij Spelling en rekenvaardigheid gedaald.

De daling die bij de rekenvaardigheid gevonden is van 2012 naar 2013, komt vrijwel overeen met de stijging van 2011 naar 2012. De verschillen tussen 2011 en 2013 zijn dan ook vrijwel nihil en ook niet significant. Bekijken we de trends over de gehele periode vanaf het begin van de meting, dan is de vaardigheid voor alle onderwerpen net als in 2012 nog steeds (significant) hoger dan in 2008. Het gaat om een verschil dat varieert van 2 tot 5 punten op de schaal met als gemiddelde 250 en als standaarddeviatie 50.

Wel moet geconstateerd worden dat de effectgroottes beneden de grens van 0.20 vallen en dus niet betekenisvol mogen worden genoemd.

In jaargroep 4 gaan alle geanalyseerde vaardigheden in meerdere of mindere mate achteruit tussen 2013 en 2012. Alleen bij Spelling is daarbij sprake van significantie. Ook in groep 4 zijn ondanks de gevonden dalingen de resultaten in 2013 beter dan die in 2008. Die stijging ten opzichte van 2008 is overal significant, behalve bij Begrijpend lezen. De effectgroottes zijn hier, op begrijpend lezen na, ook groter en variëren van 0,14 bij Woordenschat tot 0,20 bij Rekenen-wiskunde. Dat laatste ligt op de grens van betekenisvol.

Nemen we de twee jaargroepen bij elkaar dan zien we dat er ten opzichte van 2012 geen significante vooruitgang is geboekt, maar dat er bij enkele vaardigheden wel sprake is van significante achteruitgang, namelijk bij Spelling, zowel in groep 8 als groep 4 en bij Rekenen in groep 8. Bij de andere vaardigheden zijn er geen significante verschillen.

Alles bij elkaar genomen moeten we dus concluderen dat de trend ten opzichte van 2008 in alle gevallen nog steeds positief uitvalt, maar dat er sprake is van stabilisering of achteruitgang en in geen geval van verdere vooruitgang.

De resultaten zijn ook geanalyseerd met een aantal achtergrondvariabelen voor leerlingen en scholen. Jongens in jaargroep 8 behalen hogere scores in Rekenen-Wiskunde en Woordenschat dan meisjes. Meisjes zijn beter in Spelling en Begrijpend lezen. In jaargroep 4 zien we dezelfde verschillen, maar is het verschil bij Woordenschat niet significant. Vertraagde leerlingen blijken in alle gevallen een grote achterstand te hebben. Ook zien we verschillen tussen leerlingen met een verschillend formatiegewicht. Een opvallend contrast daarbij is in jaargroep 8 bij Spelling en Getallen/Bewerkingen, waarbij leerlingen met een hoog gewicht (1.20) het beter doen dan leerlingen met een laag gewicht (0.30). Leerlingen zonder formatiegewicht presteren bij alle onderwerpen beter dan 'gewogen' leerlingen.

Zeer grote effecten zien we bij de variabele doorstroomadvies, waarmee goed geïllustreerd wordt hoe groot de verschillen in vaardigheid tussen leerlingen zijn bij het verlaten van de basisschool.

Variabelen op schoolniveau zoals schoolpopulatie (stratum), schoolgrootte, regio en verstedelijking hebben wel vaak een significante invloed in jaargroep 8, maar de effectgroottes zijn doorgaans, als ze al betekenis hebben, klein.

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 de gesignaleerde vooruitgang en wijzigingen in onderwijsbeleid is nader onderzoek nodig naar tussenliggende factoren.

Inleiding

In het kader van de kwaliteitsagenda 'Scholen voor morgen' is in 2008 het Jaarlijks Peilingsonderzoek naar het Onderwijsniveau (JPON) van start gegaan. Dit jaarlijkse onderzoek is gericht op het monitoren van het onderwijsniveau op het gebied van taal- en rekenvaardigheid in het basisonderwijs. In 2008 is verslag gedaan van de eerste jaarlijkse niveaupeiling van taal en rekenen in jaargroep 8 en 4 (Hemker & Van Weerden, 2009). Het project is ondergebracht bij PPON, de periodieke peiling van het onderwijsniveau, maar kent een andere methodiek en gebruikt andere instrumenten. De overeenkomst is dat het ook hier gaat om periodiek onderzoek, zij het met een hogere frequentie, en 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.

De kwaliteitsagenda 'Scholen voor morgen' was gericht op duurzame verbetering van het primair onderwijs en legde de prioriteit bij de verhoging van de taal/lees- en rekenopbrengsten. Dit is gecontinueerd in het beleid van daaropvolgende kabinetten en heeft onder andere geleid tot nieuwe wetgeving en voorstellen daartoe, waarin de positie van rekenen en taal en het toezicht daarop verder is verstevigd.

In de aanbevelingen van de Expertgroep Doorlopende Leerlijnen Taal en Rekenen (EGDLL, ofwel de commissie Meijerink, SLO, 2008) werd vastgelegd wat leerlingen moeten kennen en kunnen op dit gebied bij het verlaten van de basisschool. Dit staat bekend als de referentieniveaus voor taal en rekenen en is bij wet geregeld in 2010. Om vast te stellen wat de beginsituatie was en hoe het niveau zich sindsdien heeft ontwikkeld in het basisonderwijs is het Jaarlijks Peilingsonderzoek (JPON) gestart.

Scholen beschikken doorgaans over voldoende toetsgegevens, gebaseerd op leerlingvolgsystemen en eind- of entreetoetsen, als input voor verbeterplannen op groeps-, school- en bestuursniveau. Deze gegevens kunnen echter ook, onder bepaalde condities, worden ingezet voor een landelijke niveau-bepaling, zoals dat in het project JPON wordt gedaan. Wel moet worden vastgesteld dat de bestaande toetssystemen inhoudelijk nog niet zijn afgestemd of ingericht op de referentieniveaus van de commissie Meijerink. De toetsen die nu ingezet kunnen worden zijn nog niet dekkend voor het hele terrein dat met de referentieniveaus wordt beschreven. Met name de productieve vaardigheden bij taal (schrijven en spreken) ontbreken nog, net als taalvaardigheden als begrijpend luisteren of grammaticaleiteit (of kunnen omgaan met elementen uit de 'begrippenlijst'). De toetsresultaten geven echter wel een duidelijke indicatie van de stand van zaken op een aantal belangrijke vaardigheden en maken het ook mogelijk daarin de ontwikkeling van jaar tot jaar te laten zien.

In 2008 is deze jaarlijkse peiling voor de eerste keer uitgevoerd. Daarmee is een start gemaakt met wat een reeks van jaarlijkse rapporten moet worden, waarin het niveau van taal en rekenen in het primair onderwijs in beeld wordt gebracht. Daarbij is 2008 steeds het uitgangspunt voor de vergelijking, behalve bij de toets Woordenschat in jaargroep 4. Deze toets werd pas in 2009 voor de eerste keer ingezet.

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 1). 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 (hoofdstuk 2). Daarna volgen de resultaten van jaargroep 4 (hoofdstuk 3). In beide hoofdstukken geven we een vergelijking over de jaren heen op de onderwerpen voor rekenen en taal, en de relatie van deze uitkomsten met enkele achtergrondvariabelen zoals geslacht en thuistaal.

Het laatste hoofdstuk bevat een samenvatting, conclusie en discussie.

1 Vraagstelling en aanpak

1.1 Vraagstelling

Het belangrijkste doel van JPON is het nauwkeurig vaststellen van veranderingen in de taal- en rekenvaardigheden van leerlingen in jaargroep 8 en 4 van het basisonderwijs. Daarnaast is het doel na te gaan in hoeverre de prestatieverschillen tussen bepaalde groepen leerlingen gelijk zijn gebleven dan wel groter of kleiner zijn geworden.

1.2 Methode

Voor deze peiling maken we gebruik van instrumenten die reeds in een ander kader worden ingezet. Het zijn dus geen nieuwe toetsen die afzonderlijk voor dit project zijn ontwikkeld. Voor jaargroep 8 maken we gebruik van de Eindtoets Basisonderwijs, een toets die een aantal onderdelen bevat voor rekenen en taal die we goed voor dit doel kunnen gebruiken, maar die elk jaar volledig wordt vernieuwd. Omdat deze toets elk jaar andere items bevat, kunnen de resultaten niet zomaar van jaar tot jaar met elkaar worden vergeleken. Door gebruik te maken van aanvullende databronnen is hier toch een goede oplossing voor gevonden. We beschrijven dit verderop in dit hoofdstuk (par. 1.2.1).

Voor jaargroep 4 zijn er geen data uit een landelijke toets beschikbaar. Wel wordt door veel scholen gebruikgemaakt van de LVS-toetsen van het Cito Volgstelsel voor primair onderwijs, het voormalige LOVS. Er is in dit geval een steekproef getrokken van scholen met de vraag of men de gegevens van een aantal LVS-toetsen aan ons wilde verstrekken. Om dat voor deze scholen mogelijk en ook aantrekkelijk te maken kregen ze de toetsmaterialen gratis toegezonden, met het verzoek de verkregen scores aan ons te retourneren. De LVS-toetsen worden echter niet elk jaar vernieuwd. Een voordeel hiervan is dat de meetschalen bekend zijn en niet meer veranderen. Doordat de inhoud van de toetsen onveranderd is, is het niet mogelijk dat wijzigingen in de inhoud van de toets een verklaring zijn voor de veranderingen over de tijd. Een nadeel is echter dat de bekendheid van het materiaal mogelijk van invloed kan zijn op de gevonden resultaten.

De hier genoemde toetsen zijn in aantal items en mogelijk te onderscheiden eenheden lang niet zo gedifferentieerd als in het reguliere onderzoek van PPON gebruikelijk is. Het aantal rapportage-eenheden is daardoor beperkt tot een zevental onderwerpen in jaargroep 8 en een viertal onderwerpen in jaargroep 4. Ook worden bij PPON meer en andere itemvormen ingezet dan alleen de vierkeuzevragen van de Eindtoets of de meerkeuzevragen en kort-antwoordmodellen van de LVS-toetsen.

Een belangrijk verschil met PPON is verder de afnameconditie. Bij PPON worden alle toetsen en taken afgenomen door een getrainde toetsleider. In dit geval was de eigen leerkracht de toetsleider. Ook dat vormt een mogelijke aantasting van de objectiviteit van de afname.

Tot slot moet gewezen worden op het verschil in impact van de verschillende toetssituaties. De Eindtoets kan worden beschouwd als een 'high-stake' toets, een toets waar leerlingen hun uiterste best op zullen doen, niet alleen vanwege de gevolgen van de uitslag voor hun verdere toekomst, maar ook vanwege de grote mate van aandacht en publiciteit er om heen. Voor LVS-toetsen geldt dat in mindere mate en voor afnames voor PPON lijkt dat nog minder het geval. Deze verschillen in conditie en daardoor in motivatie van de leerlingen hebben hun weerslag op de hoogte van de toetsscores, zo is inmiddels uitgezocht (zie verder Hemker, 2013).

Voor een vergelijking over de jaren heen zijn deze risico's van minder groot belang, omdat de condities immers in de meeste gevallen niet over de jaren heen zullen variëren. Het is dus geen probleem om een trend in beeld te brengen. Voor een vergelijking van deze uitkomsten met die van andere onderzoeken zijn ze dat echter wel. Zo zullen resultaten op de Eindtoets een ander beeld kunnen geven van de mate waarin

de landelijk vastgestelde referentieniveaus zijn gehaald dan de uitkomsten van een LVS-meting of een PPON-meting.

Uit overwegingen van efficiëntie en kosten zijn voor de jaarlijkse niveaupeilingen dus overwegend gegevens gebruikt die toch al door scholen worden verzameld. Voor de meting van de taal- en rekenvaardigheid in jaargroep 8 zijn dat de verzamelde gegevens uit campagne van de Eindtoets Basisonderwijs 2012. Daarnaast is gebruikgemaakt van gegevens verzameld met de bijbehorende Niveautoets, de ankertoets en de nu voor de tweede keer op een aantal scholen afgenomen toets Basisvaardigheden. Voor taal betreft het de onderdelen Begrijpend lezen (BL), Spelling (Sp) en Woordenschat (Wo). Bij rekenen gaat het om de onderdelen Getallen en bewerkingen (GB), Breuken, procenten en verhoudingen (BPV) en Meten, meetkunde, tijd en geld (MMTG).

In jaargroep 4 is gebruikgemaakt van toetsen van het Cito Volgsysteem primair onderwijs (LOVS). Voor taal is gekozen voor dezelfde onderdelen als in jaargroep 8. Bij rekenen zijn in eerdere peilingen vier onderdelen onderscheiden in plaats van drie, namelijk Getallen en getalsrelaties (G/G), Optellen en aftrekken (O/A), Vermenigvuldigen en delen (V/D) en Meten, tijd en geld (MTG). Vanaf de peiling 2011 worden deze echter ook als samengestelde rekenschaal onderzocht. Dat is mogelijk aangezien de verschillende subvaardigheden (zeer) hoog met elkaar correleren. Bij de peiling van 2013 is deze samengestelde rekenschaal de enige schaal waarop de resultaten worden gegeven. Een steekproef van scholen is gevraagd om gegevens te leveren over deze toetsen voor jaargroep 4. Deze toetsen worden doorgaans afgenomen aan het eind van het schooljaar in de maanden mei-juni. In beide gevallen, dus zowel in jaargroep 4 als in jaargroep 8, is sprake van een steekproefprocedure, zodat de resultaten als representatief mogen worden gezien voor het niveau in Nederland op de beide meetmomenten.

1.2.1 Kwaliteit van de meetinstrumenten

Er is gebruikgemaakt van de resultaten op de Eindtoets Basisonderwijs (EB) van 2013. Deze toets wordt elk jaar opnieuw samengesteld en bestaat uit opgaven die al in proeftoetsen in voorgaande jaren zijn uitgetoetst. De psychometrische kenmerken zijn van tevoren goed in te schatten en blijken telkens van hoog niveau te zijn. De eigenschappen van de Eindtoets worden elk jaar gepubliceerd in een afzonderlijke publicatie en dat is ook voor 2012 het geval (Cito, 2013). We mogen hier stellen dat de psychometrische kwaliteit bij de Eindtoets 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).

We moeten wel wijzen op het feit dat de wijze waarop deze toets en de versies daarvan in dit onderzoek worden ingezet wel een andere is dan waarvoor zij is bedoeld. Het primaire doel van de Eindtoets is namelijk het opleveren van een goede prognose voor de kans van slagen in verschillende vormen van voortgezet onderwijs voor individuele leerlingen. In deze studie gaat het niet om de totaalscore voor alle onderdelen, de standardscore, maar om de uitkomsten op delen van de toets. In JPON worden verschillende onderwerpen van de toets afzonderlijk gebruikt om een beeld te schetsen van de kwaliteit van het onderwijs op systeemniveau. Voor dat doel zijn de psychometrische eisen die aan een toets moeten worden gesteld anders. Met dat fenomeen is rekening gehouden in de analyses, op een vergelijkbare wijze als in voorgaande jaren (zie verder Hemker & Van Weerden, 2009; Hemker, Kuhlemeier & Van Weerden, 2010; Hemker, Kordes en Van Weerden, 2011).

Daar waar het de toetsen voor jaargroep 4 betreft, kunnen we opmerken dat die niet veranderd zijn ten opzichte van vorige jaren en dat daarom ook de meeteigenschappen van de toetsen niet veranderd zijn. De toetsen voor jaargroep 4 maken deel uit van het Cito Volgsysteem primair onderwijs. Deze toetsen worden pas in productie genomen na een uitvoerige pretest en kunnen bogen op een goede psychometrische kwaliteit. De in dit onderzoek gehanteerde toetsen zijn regulier aangeboden aan de COTAN (Commissie Test Aangelegenheden Nederland) en zijn op de relevante criteria als voldoende en goed beoordeeld. De psychometrische kwaliteit is weergegeven in de respectievelijke wetenschappelijke

verantwoordingen van deze toetsen (Van Berkel e.a., 2010; Feenstra e.a., 2010; Janssen e.a., 2010; De Wijs e.a., 2010). De eigenschappen van de toetsen in de context van JPON zijn in eerder verslagen beschreven. Zie voor verdere informatie: www.toetswijzer.nl.

1.2.2 Steekproeftrekking en analyse

Jaargroep 8

In dit onderzoek zijn twee databestanden gebruikt om tot steekproeven te komen. Die steekproeven verschilden wel in grootte. De steekproef van jaargroep 8 is veel omvangrijker dan die van jaargroep 4. De steekproef in jaargroep 8 is gebaseerd op de data van het totaal aantal deelnemers aan de Eindtoets in 2013. Na een eerste schifting van leerlingen die buiten de onderzoekspopulatie vallen, resteerden 150.756 deelnemers. Achterwege zijn gelaten vooral leerlingen van scholen in het buitenland, speciaal onderwijs en voortgezet onderwijs en leerlingen die de digitale versie van de Eindtoets hadden gemaakt. Vervolgens zijn uit het databestand twee steekproeven getrokken: één ten behoeve van de schaalconstructie en één voor de analyse met achtergrondvariabelen.

Voor de schaalanalyses (OPLM) is gebruikgemaakt van de resultaten van alle leerlingen die beschikbaar zijn in de koppeling tussen jaargangen. Bij het schalen van de opgaven is de representativiteit van de steekproef van minder groot belang, vanwege populatieonafhankelijkheid van de itemparameterschattingen bij de gebruikte analysemethode (item response theorie; IRT). In deze steekproef hebben we alle leerlingen ingesloten die de Niveautoets (NT) of de toets Basisvaardigheden (TBV) of de ankertoets hebben gemaakt. Vervolgens is die dataverzameling aangevuld met een aselechte steekproef uit alle overige leerlingen van de Eindtoets totdat er uiteindelijk 30.000 leerlingen in het bestand zaten. De NT-leerlingen werden ingesloten omdat zij een andere Eindtoets hebben gemaakt. De ankertoets en de TBV werden ingesloten om de toetsen op onderdelen te kunnen koppelen. Daarbij is het gebruik van data van de TBV in de analyse nieuw dit jaar. De kwaliteit van de gebruikte schalen is niet anders dan die in voorgaande jaren. Merk op dat met de gebruikte aantallen iedere kleine afwijking in veel gevallen als significant gekwalificeerd wordt maar dat deze afwijking geen effect heeft op de resultaten (zie voor meer hierover in de vorige JPON-rapportages). Na de schaalanalyse is met de geconstrueerde schalen, die in principe dezelfde psychometrische kenmerken bezaten als in voorgaande jaren, de analyse gedaan op achtergrondvariabelen, met in de eerste plaats de jaarvergelijking. Daartoe is een nieuwe steekproefprocedure uitgevoerd waarbij 5 aselechte steekproeven zijn getrokken van ongeveer 30.000 leerlingen zonder teruglegging, waarbij representativiteit wel van belang is. De resultaten in dit rapport zijn gebaseerd op de meest representatieve steekproef van de vijf. Om de robuustheid van deze steekproef te bepalen zijn de SAUL-analyses op drie verschillende steekproeven uitgevoerd. 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 de terugblik bij de Eindtoets (Cito, 2013). We mogen dus stellen dat deze steekproeven een representatief beeld geven van de Nederlandse populatie in jaargroep 8.

Jaargroep 4

Voor jaargroep 4 is een afzonderlijke steekproef getrokken uit het scholenbestand van Cito. Dit betrof een gestratificeerde steekproef met strata gebaseerd op de formatiegewichten van de scholen, conform de procedure bij PPON, waarbij werd gemikt op ca. 100 scholen. Scholen die al participeerden in andere onderzoeken van Cito werden buiten de steekproef gehouden. De scholen werden aangeschreven in februari 2013 en kregen bij deelname gratis toezending van de benodigde toetsmaterialen. Deze waren ook bruikbaar en beschikbaar voor administratie in de rapportagemodule van het Cito Volgstelsel. Als tegenprestatie leverden de scholen een ingevuld digitaal scoringsblad in na afloop van de toetsafname in de maanden mei-juni. De werving leverde uiteindelijk 100 scholen op met in totaal 2299 leerlingen waarvan achtergrondgegevens bekend waren.

Doordat de gebruikte toetsen onveranderd waren en in de eerdere peilingen de schaling van de opgaven al was uitgevoerd is er bij de data-verzameling gebruik gemaakt van de totaalscores op de verschillende toetsonderdelen. Ook waren de toets- en itemeigenschappen van de toetsen al bekend. De itemscores zijn zodoende niet meer opgevraagd, maar zijn rechtsreeks uit de totaalscores verkregen. Deze zijn door de leerkracht opgestuurd, samen met informatie over welke (versie) van de toetsen de leerlingen gemaakt

hadden en de achtergrondgegevens van de leerlingen. Op basis van de informatie uit de eerder uitgevoerde peilingen kunnen de scores van de verschillende toetsen op de vaardigheidsschalen geplaatst worden. Bij de toetsen Begrijpend lezen en Spelling wordt met aparte – in moeilijkheid van elkaar verschillende – versies gewerkt. Bij deze toetsen moet naast de somscore ook bekend zijn welke versie de leerling gemaakt heeft voordat een latente trek kan worden toegekend op basis van de somscore. Met dit fenomeen is rekening gehouden.

Zoals in de beschrijving van de methode in paragraaf 1.2 al is aangegeven, is de analyse sinds 2012 anders uitgevoerd dan in vorige jaren. Er is nu direct gewerkt met de totaalscores op de toets, zonder dat de gegevens op itemniveau beschikbaar waren. Deze totaalscores zijn door de leerkrachten ingevuld op een scoreblad, samen met de variabelen voor de achtergrondkenmerken en de gegevens over de versie die leerlingen gemaakt hebben.

Bij de rekenschalen is nu gebruikgemaakt van de samengestelde rekenschaal. Dit kan zonder bezwaar, aangezien de verschillende mogelijke subschalen voor rekenen zeer hoog met elkaar correleerden (correlatie was gemiddeld hoger dan 0,90). De resultaten op de verschillende subschalen lijken daardoor ook zeer sterk op elkaar en konden zodoende goed samengevat worden in de samengestelde rekenvaardigheidsschaal. De correlaties voor de verschillen schalen bij taal waren aanzienlijk lager (gemiddeld 0,66) waardoor informatie verloren zou gaan als alleen die taalschaal gerapporteerd zou worden. De resultaten op een algemene schaal voor taal zou ook erg afhankelijk zijn van de specifieke samenstelling van deze schaal.

De analyses van de effecten van de achtergrondvariabelen zijn gedaan op de vaardigheidsscores, zodat de gerapporteerde resultaten vergelijkbaar zijn met die van de eerdere rapportages. Naast de geschatte gemiddelden (en andere verdelingseigenschappen) van latente variabelen is er bij de analyse gebruikgemaakt van GLM (Generalized Linear Model). Met deze methode kunnen, vergelijkbaar met de eerdere uitgevoerde analyses met SAUL, de gecorrigeerde effecten berekend worden.

Zowel voor jaargroep 8 als jaargroep 4 zijn de steekproeven een goede afspiegeling van de populaties van leerlingen.

1.2.3 Achtergrondvariabelen

Behalve de genoemde taal- en rekenvaardigheden zijn ook een aantal achtergrondkenmerken in het onderzoek betrokken. Deze zijn te onderscheiden in 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);
- thuistaal: Nederlands gesproken, een andere taal of een combinatie;
- advies VO: ingeschat niveau vervolgonderwijs door leerkracht (alleen jaargroep 8);
- wel of niet een IJK-code (alleen jaargroep 8), wat staat voor respectievelijk: (Allochtone) leerlingen die aan het begin van jaargroep 8 vier jaar of korter in Nederland zijn en die het Nederlands onvoldoende beheersen om de opgaven in de Eindtoets goed te kunnen lezen (Code I), of leerlingen die naar verwachting naar het (voortgezet) speciaal onderwijs of naar het praktijkonderwijs (pro) gaan (Code J) of naar verwachting in aanmerking komen voor het leerwegondersteunend onderwijs (lwoo) (Code K).
- gemaakte toets: Eindtoets of Niveautoets.

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 (noord, oost, west, zuid);
- urbanisatiegraad van de locatie van de school;
- schoolgrootte (alleen jaargroep 8).

De achtergrondvariabelen zijn niet in beide gepeilde jaargroepen hetzelfde. Voor jaargroep 4 was uiteraard het advies VO niet beschikbaar, evenmin als de IJK-codering, en was ook de schoolgrootte niet bekend. Ook zijn in de loop van de tijd variabelen veranderd, zoals de definitie van het formatiegewicht en van stratum. Hoe hier mee om is gegaan is in eerdere verslagen van de peilingen beschreven. Dit jaar is ook de variabele IJK meegenomen in de analyses als alternatief voor de variabele thuistaal die niet opgevraagd is de in de Eindtoets-campagne van 2013.

De achtergrondvariabelen worden bij de Eindtoets verzameld via het antwoordblad dat bij deze toets hoort en waarop leerlingen en leerkrachten een aantal gegevens invoeren. Bij de steekproef in jaargroep 4 werden de achtergrondgegevens per leerling op een afzonderlijke leerlingenlijst ingevuld door de betrokken leerkracht.

Opgemerkt moet worden dat het feit dat de wijziging in definitie van formatiegewichten in de loop van de in dit onderzoek bestreken periode ook een rol speelt bij de vergelijking over de jaren. Tevens heeft deze wijziging gevolgen gehad voor de indeling in de strata van de scholen. Hoe daarmee is omgegaan is uitvoerig toegelicht in de rapportage van het peilingsjaar 2010 (Hemker, Kordes & Van Weerden, 2011)

1.3 Hoe is er geanalyseerd?

Hieronder geven we beknopt weer hoe we de resultaten van 2013 hebben vergeleken met die van 2012 en de jaren daarvoor. Voor een uitgebreide technische verantwoording van de steekproef, de gebruikte toetsen, de statistische analyse en de rapportagemethodiek wordt verwezen naar de technische rapportages van eerdere jaren (Hemker & Van Weerden, 2009; Hemker, Kuhlemeier en Van Weerden, 2010; Hemker, Kordes & Van Weerden, 2011; Van Weerden, Hemker, Straat & Mulder, 2012). Wel gaan we in op de afwijkingen in methodiek die er dit peilingsjaar zijn toegepast.

1.3.1 Vergelijking tussen en over de jaren

Een voorwaarde om vaardigheden van de verschillende jaren direct met elkaar te kunnen vergelijken, is dat de vaardigheden op dezelfde schaal gemeten zijn. Aan deze voorwaarde is voldaan als de leerlingen in de steekproef van 2012 precies dezelfde toetsen hebben gemaakt als de leerlingen in eerdere steekproeven. Voor jaargroep 4 is aan deze voorwaarde voldaan, aangezien de leerlingen in beide gevallen ongewijzigde toetsen uit het reguliere Cito Volgsysteem hebben gemaakt. De voor jaargroep 8 gebruikte Eindtoets Basisonderwijs wordt echter ieder jaar volledig vernieuwd. Wij hebben de vergelijkbaarheid echter kunnen waarborgen doordat de achtstegroepers telkens een aantal extra opgaven zijn voorgelegd (namelijk via ankertoetsen), die wel over de jaren heen ongewijzigd zijn gebleven. Ter aanvulling op die gegevens kon ook gebruik worden gemaakt van een overlap in items met de Niveautoets en de toets Basisvaardigheden. Deze laatste toets werd voor het eerst in 2011 op een groot aantal scholen in jaargroep 8 afgenomen en is in 2012 en 2013 opnieuw gebruikt.

Met behulp van een speciale analysetechniek: het One Parameter Logistic Model (OPLM) (Verhelst, 1993; Verhelst en Glas, 1995), een variant van het item response model, zijn schalen geconstrueerd voor alle afzonderlijke onderwerpen. Dit is reeds in 2008 en 2009 uitgevoerd, waarbij de toetsen allen op dezelfde schaal werden gezet, met een gemiddelde van 250 punten en een standaardafwijking van 50. Op deze wijze kunnen de prestaties op de verschillende toetsen over de jaren heen met elkaar worden vergeleken (vgl. Hemker, Kordes & Van Weerden, 2011).

Voor het vergelijken van prestaties over de jaren heen moet ook rekening gehouden worden met de samenstelling van de responsgroep. Als de samenstelling gelijk is dan is een directe vergelijking 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. Stel dat alle basisscholen bijvoorbeeld minder leerlingen zijn gaan verwijzen naar het speciaal onderwijs. De responsgroep in 2013 zou dan meer 'zorgleerlingen' bevatten dan die in 2012. We zouden dan ten onrechte kunnen concluderen dat de vaardigheid van de leerlingen achteruit is gegaan. Gelukkig zijn er statistische technieken beschikbaar die ons voor dit soort verkeerde conclusies kunnen behoeden. Vandaar dat wij voor jaargroep 8 zowel ongecorrigeerde als gecorrigeerde gegevens verstrekken (d.w.z. gecorrigeerd voor veranderingen in de samenstelling van de populatie). In dit geval wordt gebruikgemaakt van een speciale programma voor regressieanalyse van vaardigheidsscores geschat met behulp van OPLM, te weten SAUL: Structural Analysis of Univariate Latent variabels (Verhelst & Verstralen, 2002).

In jaargroep 4 is de steekproef veel minder groot dan in jaargroep 8 (namelijk 'slechts' ongeveer 2500 leerlingen van ongeveer 100 scholen). De omvang van de te vergelijken steekproeven in jaargroep 4 is te klein om steekproeffluctuaties met voldoende zekerheid te kunnen vaststellen en hiervoor vervolgens statistisch te kunnen corrigeren. Wel is het mogelijk te corrigeren voor veranderingen in de samenstelling van de populatie van vierdegroepers. Hierbij zijn zowel gecorrigeerde als ongecorrigeerde analyses uitgevoerd. De achtergrondkenmerken waarvoor gecorrigeerd wordt, zijn in beide modellen geslacht, leertijd, stratum en formatiegewicht.

1.3.2 Analyses voor de leerlingen in 2013

De gegevens van 2013 zijn als afzonderlijke groep geanalyseerd. In dit rapport zijn de gegevens van de gemiddelden van de verschillende te onderscheiden groepen vergeleken op de vaardigheidsschalen. Daarnaast zijn ook hier gecorrigeerde effecten geschat. 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 hoofdeffecten-model, 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 2013 op dezelfde schaal liggen als de voorgaande jaren zijn de resultaten hiermee direct vergelijkbaar over de jaren heen.

In de eerdere verslagen 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.

1.3.3 Significantie en effectgrootte

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. Voor jaargroep 8 is de steekproef zeer veel groter dan voor jaargroep 4. Om de resultaten toch zinvol met elkaar te kunnen vergelijken rapporteren wij behalve de statistische significantie ook de zogeheten effectgrootte. De effectgrootte wordt in ons geval berekend als het verschil tussen de gemiddelden van 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 1.1. Alles met een effectgrootte boven de 0.20 noemen we hier betekenisvol.

Tabel 1.1 *Kwalificatie van effectgrootten*

Effectgrootte (zowel plus als min) Kwalificatie

0,0 tot 0,2	geen effect
0,2 tot 0,5	klein effect
0,5 tot 0,8	matig effect
0,8 of groter	groot effect

1.3.4 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 PPO (zie bijv. Janssen, Van der Schoot & Hemker, 2005). De startwaarde is voor elke schaal het gemiddelde dat we in 2008 hebben aangetroffen. Dat gemiddelde is op 250 gesteld (zie verder Hemker & Van Weerden, 2009). De transformatie heeft als voordeel dat we de prestaties voor verschillende vaardigheden en voor verschillende jaren naast elkaar kunnen zetten op een en dezelfde schaal. Om de resultaten goed te kunnen beschrijven richten we ons in de rapportage niet alleen op de gemiddelde leerling, maar ook op andere groepen leerlingen in de vaardigheidsverdeling. Voor dit doel zijn vijf typische leerlingen gedefinieerd gekoppeld aan kenmerkende percentielpunten (zie tabel 1.2). In de rapportage gaan we uit van het basisjaar en rapporteren in de jaren daarop de eventuele wijziging in percentielwaarde. Zo zal bij een eventuele vooruitgang in vaardigheid blijken dat meer leerlingen de schaalwaarde 250 bereiken, maar ook kunnen meer leerlingen de schaalwaarde 186 bereiken. Teruggerekend betekent dat een wijziging in het percentage leerlingen dat tot een groep typische leerlingen behoort ten opzichte van 2008.

In dit verband moet worden opgemerkt dat aan de schaalwaarde horende bij de laag vaardige leerling een speciale betekenis kan worden toegekend. Dit is het punt op de vaardigheidsschaal dat door minstens 75% van de leerlingen wordt bereikt. In het advies van de Expertgroep Doorlopende Leerlijnen Rekenen en Taal (SLO, 2008) wordt dit punt regelmatig genoemd als referentieniveau voor 1F. Het referentieniveau 1S zou dan voor Rekenen overeenkomen met de P50. Bij Taal is daarvoor de P75 genoemd. Gekoppeld aan de specifieke inhoud van de schalen die we hier rapporteren zijn er echter nog geen uitspraken gedaan over het gewenste niveau.

Tabel 1.2 *Definiëring typische leerlingen in de vaardigheidsverdeling*

Aanduiding	Afkorting	Percentiel 2008	Schaalwaarde
Zeer laag vaardig	ZLV	P10	186
Laag vaardig	LV	P25	216
Gemiddelde/standaard	G/S	P50	250
Hoog vaardig	HV	P75	284
Zeer hoog vaardig	ZHV	P90	314

2 De resultaten voor jaargroep 8

2.1 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 2013 met voorgaande jaren tot en met 2008. We presenteren hier dus de ongecorrigeerde verschillen.

In tabel 2.1 zijn de uitkomsten voor de zes opeenvolgende jaren weergegeven als gemiddelde op de vaste schaal. Het gemiddelde voor 2008 is per definitie 250 en de standaarddeviatie is 50. De schaalwaarde van 250 is vastgelegd in 2008 en vormt het startpunt waarmee we de uitkomsten van 2009 en later vergelijken (zie verder Hemker & Van Weerden, 2009).

In tabel 2.2 geven we weer hoe groot de verschillen van jaar tot jaar zijn in termen van effectgrootte. Bovendien geven we de effectgrootte voor het contrast 2013-2008, dat wil zeggen het verschil in uitkomst van de meting in 2013 met die van de eerste meting in 2008, dus over de afgelopen 6 jaar. Als het verschil tussen 2013 en 2012 statistisch significant is, is het effect vetgedrukt. Evenzo geldt dat voor het verschil tussen 2013 en 2008.

Vergelijken we de uitkomsten voor 2013 met 2012 dan zien we dat bij Woordenschat het gemiddelde is gestegen, maar dat bij alle andere vaardigheden er sprake is van een daling ten opzichte van 2012, dan wel een stabilisering. Begrijpend lezen komt op dezelfde schaalwaarde uit, spelling en alle rekenonderwerpen vertonen een daling en, zoals weergegeven in tabel 2.2, deze negatieve verschillen zijn alle significant. Deze daling komt vrijwel overeen met de eerder gevonden stijging van 2011 naar 2012. De grootte van de verschillen varieert van nul tot drie punten op de schaal. Dat laatste zien we alleen bij de rekenonderwerpen.

Vergelijken we 2013 met 2008 dan zien we dat voor alle vaardigheden nog steeds geldt dat het in 2013 bereikte gemiddelde hoger is dan in 2008, het uitgangsjaar. In tabel 2.2 zien we tevens dat die positieve verschillen allen significant zijn. De effectgrootte varieert van 0,05 tot 0,08 en is het grootst bij de schaal woordenschat.

Tabel 2.1 Gemiddelde schaalwaarden jaargroep 8*

Vaardigheden	2008	2009	2010	2011	2012	2013
T: Woordenschat	250	249	257	251	252	254
T: Spelling	250	250	252	252	256	255
T: Begrijpend lezen (anker)	250	252	254	257	253	253
RW: Rekenen: over-all	250	250	253	253	256	253
RW: Getallen en bewerkingen	250	250	252	252	255	252
RW: Breuken, procenten en verhoudingen	250	250	254	253	257	254
RW: Meten, tijd en geld	250	249	254	254	256	255

*standaarddeviaties variëren van 49 tot 53; in 2008 per definitie 50

Tabel 2.2 Effectgroottes van de jaarverschillen in jaargroep 8*

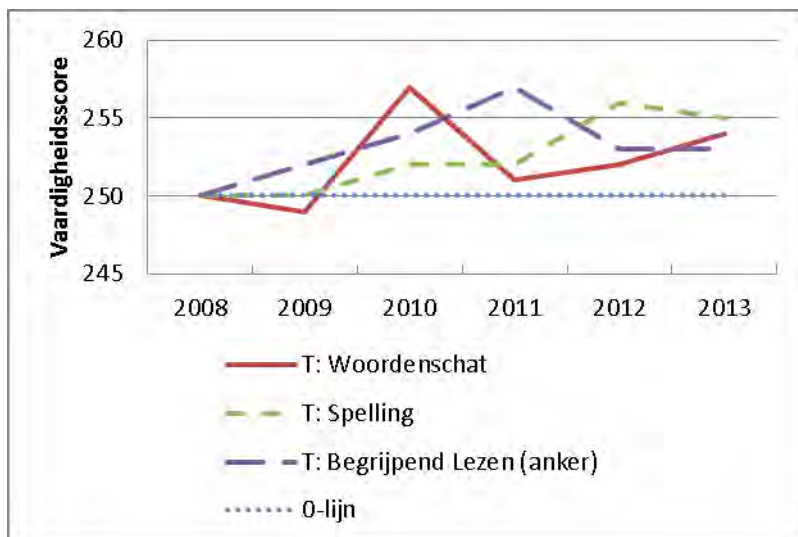
Vaardigheden	2009-2008	2010-2009	2011-2010	2012-2011	2013-2012	2013-2008
T: Woordenschat	-0,02	0,15	-0,11	0,02	0,04	0,08
T: Spelling	0,00	0,03	0,00	0,07	-0,03	0,07
T: Begrijpend lezen (anker)	0,04	0,04	0,07	-0,09	0,00	0,06
RW: Rekenen: over-all	-0,01	0,07	0,00	0,06	-0,06	0,06
RW: Getallen en bewerkingen	-0,01	0,05	0,00	0,07	-0,06	0,05
RW: Breuken, procenten en verhoudingen	-0,01	0,08	-0,01	0,07	-0,07	0,06
RW: Meten, tijd en geld	0,01	0,07	0,00	0,05	-0,04	0,07

*Vet = significant verschil met voorgaande jaar of, in laatste kolom, met 2008 ($\alpha < 0,01$)

2.1.1 Taal

De ontwikkeling in schaalscores voor de verschillende onderwerpen is afzonderlijk in beeld gebracht voor taal en rekenen. We zien bij taal dat de drie trendlijnen duidelijk verschillen van elkaar, maar wel alle drie bij elkaar komen op een wat hoger niveau in 2013. Het niet-lineaire verloop van de trend in de resultaten bij Woordenschat is al nader besproken in de rapportage van 2011, waarbij gewezen is op het verschil tussen jongens en meisjes bij de uitkomsten. Vanaf 2011 is er sprake van een redelijk stabiel stijgende lijn. De opgaande lijn van spelling lijkt echter tot stilstand te zijn gekomen en bij begrijpend lezen is de situatie ten opzichte van 2012 ook nauwelijks gewijzigd.

Figuur 2.1 Trends over de jaren voor de taalonderwerpen in jaargroep 8



Bij de ontwikkeling van begrijpend lezen moet nog de kanttekening gemaakt worden dat de trendlijn is gebaseerd op de resultaten zoals gevonden met de ankertoets. Dat is ook de wijze waarop de eerdere jaarvergelijkingen gedaan zijn (zie Van Weerden e.a. 2013). De vergelijking met de toets Basisvaardigheden, zoals in de vorige rapportage verkend, was ook mogelijk geweest, maar is lastiger te interpreteren.

Tabel 2.3 Vergelijking over de jaren heen in percentage per leerlingengroep voor de taalonderwerpen in jaargroep 8

Leerlingengroep*	Start- waarde	Woordenschat					Spelling					Begrijpend lezen				
	2008	'09	'10	'11	'12	'13	'09	'10	'11	'12	'13	'09	'10	'11	'12	'13
> dan ZLV (<i>p</i> 10)	90	90	92	90	91	91	90	91	91	92	92	91	91	92	91	91
> LV (<i>p</i> 25)	75	74	79	76	76	78	75	76	76	79	78	76	77	79	77	77
> Standaard 2008 (<i>p</i> 50)	50	49	55	51	52	53	50	51	51	55	54	52	53	56	52	52
> HV (<i>p</i> 75)	25	24	29	26	26	28	25	26	26	29	28	26	27	30	27	27
> ZHV (<i>p</i> 90)	10	10	13	10	11	12	10	11	11	12	12	11	11	13	11	11

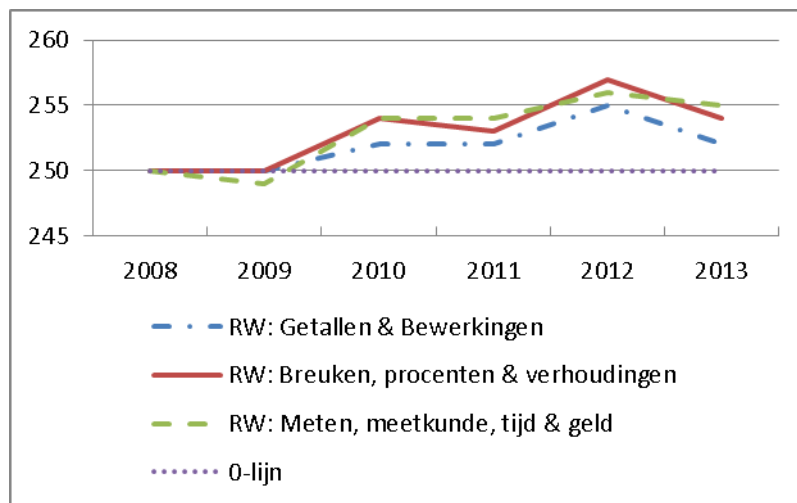
*Zie tabel 1.2

In tabel 2.3 is weergegeven welk percentage van de leerlingen een bepaalde percentielscore behaalt, waarbij de startwaarde gebaseerd is op de situatie van 2008. Zo was het percentage leerlingen dat percentielwaarde 10 behaald, de zeer laag vaardige leerling, per definitie 90%. Het gaat dan concreet om het percentage leerlingen dat een schaalwaarde van 186 of hoger heeft op de schaal van 250 (sd=50), zie ook tabel 1.2. In 2013 blijkt dat nu op 91% uit te komen, hetgeen betekent dat iets meer leerlingen dat niveau in 2013 bereiken dan in 2008. Uitgaande van de 75% die behoort bij percentiel 25 zien we dat bij begrijpend lezen de toename een percentiel punt lager is dan bij de andere twee, namelijk 77 naast 78. Bij de gemiddelde leerling, *p*50, is het verschil soms nog wat groter: van 50% naar 54% bij Spelling bijvoorbeeld.

2.1.2 Rekenen-Wiskunde

Voor Rekenen-Wiskunde is het mogelijk om zowel een totaalschaal te presenteren als drie afzonderlijke schalen: voor elk onderscheiden onderwerp een. De onderlinge samenhang is dermate groot dat hier statistisch geen enkel bezwaar tegen is. We presenteren hier de afzonderlijke schalen. De sterke onderlinge samenhang van de drie onderwerpen zien we in de grafische weergave terug. Over de hele linie zien we ongeveer hetzelfde beeld: in 2010 gaan alle schalen enkele punten omhoog, in 2011 is er nauwelijks verschil, maar in 2012 is er weer een duidelijke vooruitgang. In 2013 gaan alle schalen echter weer naar beneden. Over de hele linie bekeken is de uitkomst in 2013 evident lager dan in 2012. Maar ten opzichte van 2008 blijft er nog steeds sprake van een positief verschil, waarbij Getallen en bewerkingen relatief het laagst scoort van de drie onderwerpen en Meten, meetkunde, tijd & geld het hoogst.

Figuur 2.3 Trends over de jaren voor de rekenonderwerpen in jaargroep 8



Tabel 2.5 Vergelijking over de jaren heen in percentage per leerlingengroep voor de rekenonderwerpen in jaargroep 8

Leerlingengroep*	Start- waarde	Getallen en bewerkingen					Breuken, procenten en verhoudingen					Meten, meetkunde, tijd en geld				
	2008	'09	'10	'11	'12	'13	'09	'10	'11	'12	'13	'09	'10	'11	'12	'13
> dan ZLV (p_{10})	90	90	91	91	92	91	90	91	91	92	91	90	91	91	92	92
> LV (p_{25})	75	75	76	76	78	76	75	77	77	79	77	74	77	77	79	78
> Standaard 2008 (p_{50})	50	50	52	51	54	52	50	53	53	56	53	49	53	53	55	54
> HV (p_{75})	25	25	26	26	28	26	25	27	27	30	27	24	28	27	29	28
> ZHV (p_{90})	10	10	11	11	12	11	10	11	11	13	11	10	11	11	12	12

*Zie tabel 1.2

Ook voor Rekenen zijn de percentages leerlingen in kaart gebracht voor de verschillende percentielwaarden, p_{10} , p_{25} , p_{50} , p_{75} en p_{90} . Dit is weergegeven in tabel 2.5. De toename in percentage ten opzichte van 2008 varieert van 1 tot 4 procent. Het percentage laagvaardige leerlingen (LV) is met 1 tot 3 procent toegenomen, het meeste bij Meten, meetkunde, tijd en geld. Dat is nog meer waarneembaar bij de gemiddelde leerling, daar is een toename van 50% naar 54%. Bij Getallen en bewerkingen is dat 2% en de schaal Breuken, procenten en verhoudingen zit daar met 3% tussen in.

2.2 Verschillen voor achtergrondvariabelen

In deze paragraaf rapporteren we de verschillen tussen groepen van leerlingen gebaseerd op een categorisering naar achtergrondvariabele. We presenteren hier in de eerste plaats de gevonden effecten in 2013 en gaan bij enkele variabelen in op de trend over zes jaar.

Het gebruikte basismodel voor de analyses van 2013 bevatte de verklarende variabelen geslacht, leertijd, formatiegewicht en stratum. Ieder van de geschatte effecten per variabelen 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 of Niveautoets), 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. In tabel 2.6 zijn de effectgroottes voor de variabelen op leerlingniveau weergegeven.

Tabel 2.6 Effecten op leerlingniveau in jaargroep 8 voor het contrast 2013-2012*

Variabele	Contrast	Taal			Rekenen		
		WS	Sp	BL	G/B	BPV	MMTG
Geslacht	Meisjes - Jongens	-0,24	0,27	0,15	-0,30	-0,46	-0,44
Leertijd	Regulier - Vertraagd	0,49	0,72	0,58	0,67	0,69	0,65
Stratum*	1 – 2	0,11	0,10	0,19	0,10	0,13	0,12
	2 – 3	0,37	0,02	0,37	0,05	0,14	0,16
	1 – 3	0,47	0,12	0,56	0,14	0,27	0,28
Formatiegewicht	0.0 - 0.3	0,39	0,36	0,52	0,44	0,47	0,46
	0.0 - 1.2	0,81	0,10	0,69	0,20	0,33	0,34
	0.3 - 1.2	0,42	-0,27	0,17	-0,24	-0,15	-0,12
Advies VO	vmbo-kb - vmbo-bb	0,63	0,67	1,09	1,02	1,13	1,00
	vmbo-gt - vmbo-kb	0,53	0,49	0,74	0,70	0,75	0,71
	havo - vmbo-gt	0,85	0,94	1,15	1,20	1,24	1,25
	vwo - havo	1,32	1,49	1,75	1,56	1,66	1,78
	vwo - vmbo-bb	3,31	3,58	4,73	4,48	4,77	4,73
Toets	Eindoets-Niveautoets	1,13	1,14	1,41	1,39	1,45	1,39

*vet is significant ($p=0.01$); stratum is gebaseerd op het aantal gewichtsl leerlingen: 1 = <10%; 2 = 10-25%, 3 = >25%

2.2.1 Effecten op leerlingniveau

Geslacht

Voor het verschil in resultaten tussen jongens en meisjes zien we in alle gevallen significante en betekenisvolle effecten. Deze zijn bij de rekenonderwerpen het grootst en in het voordeel van de jongens. Bij de taalonderwerpen zien we meer variatie van richting. Jongens behalen een hogere vaardigheid in woordenschat, maar een lagere vaardigheid bij spelling en begrijpend lezen.

In tabel 2.7 hebben we deze resultaten onderscheiden naar geslacht bij elkaar gezet, verder in beeld gebracht met een grafische weergave per vaardigheid in figuur 2.3. Doorgaans zien we parallelle ontwikkelingen, maar Woordenschat is daarop een uitzondering. De resultaten van zowel de jongens als de meisjes zijn grillig te noemen. De hogere resultaten van de jongens in 2013 worden bijna teniet gedaan door de negatieve resultaten bij de meisjes, zodat de algemene richting wel positief is, maar niet veel. Bij de overige vaardigheden zien we wel steeds een min of meer parallelle trendlijn. Opvallend is het verschil in ontwikkeling bij Getallen en bewerkingen. Daar zien we een terugval bij de jongens naar de situatie van 2008, terwijl de resultaten bij de meisjes zich in positieve richting blijven ontwikkelen. Bij de andere rekenvaardigheden is echter ook bij de meisjes een terugslag in 2013 ten opzichte van 2012. Bij Meten, meetkunde, tijd en geld zien we dat de jongens op hetzelfde niveau blijven.

Als we de ontwikkeling over de gehele periode van 6 jaar overzien dan blijkt dat het verschil tussen jongens en meisjes in zijn geheel, voor alle onderwerpen bij elkaar, iets minder is geworden. De verschillen in 2008 en 2013 zijn per onderwerp weergegeven in tabel 2.8. In de laatste rij zien we dat het per saldo betekent dat bij de taalonderwerpen het verschil positief is en bij de rekenonderwerpen negatief. De grootste verandering in verschil tussen jongens en meisjes is opgetreden bij de rekenschaal Getallen/bewerkingen.

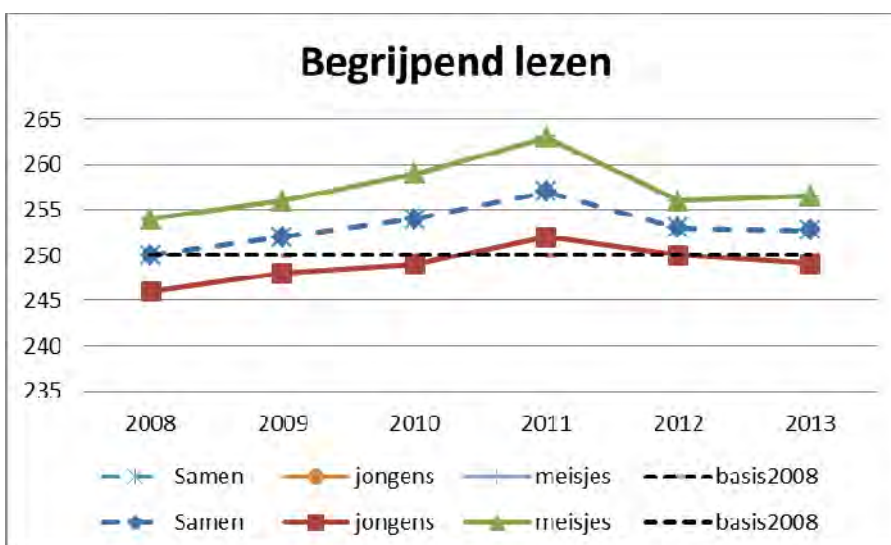
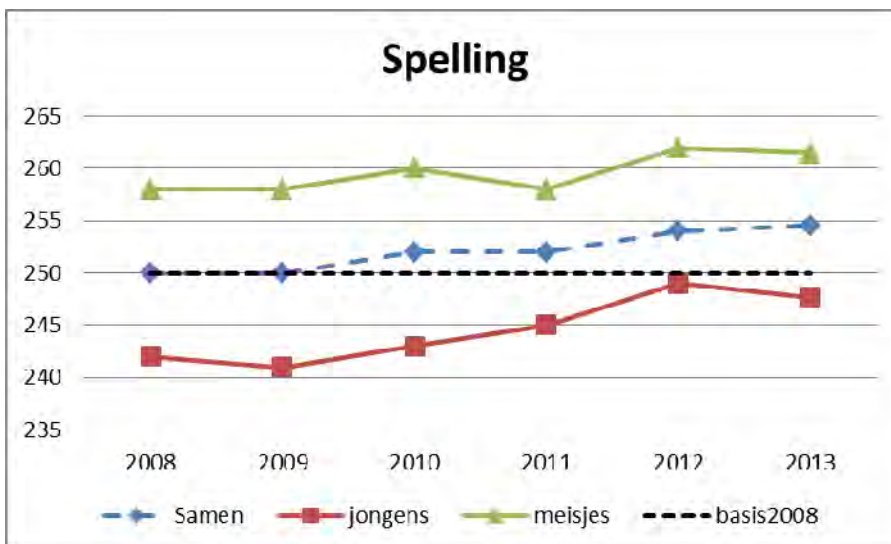
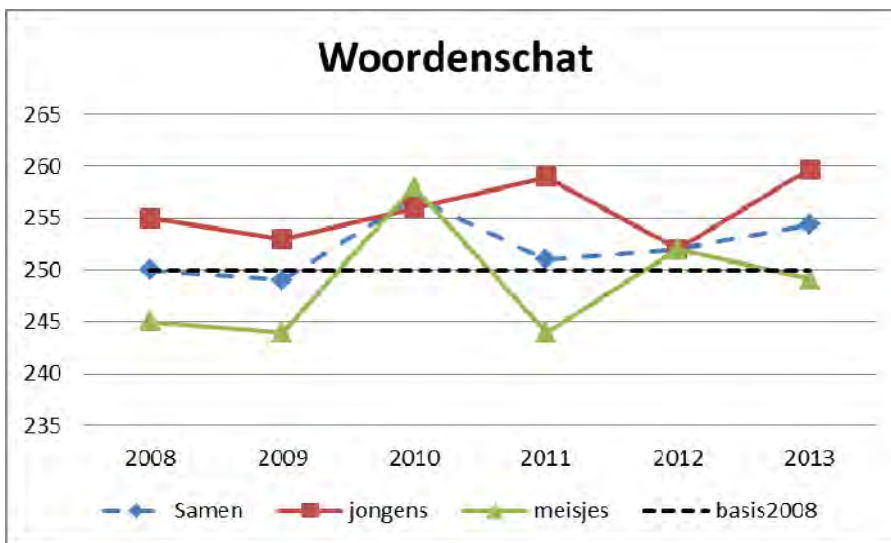
Tabel 2.7 Vaardigheidsscores per onderwerp naar geslacht van 2008 tot en met 2013

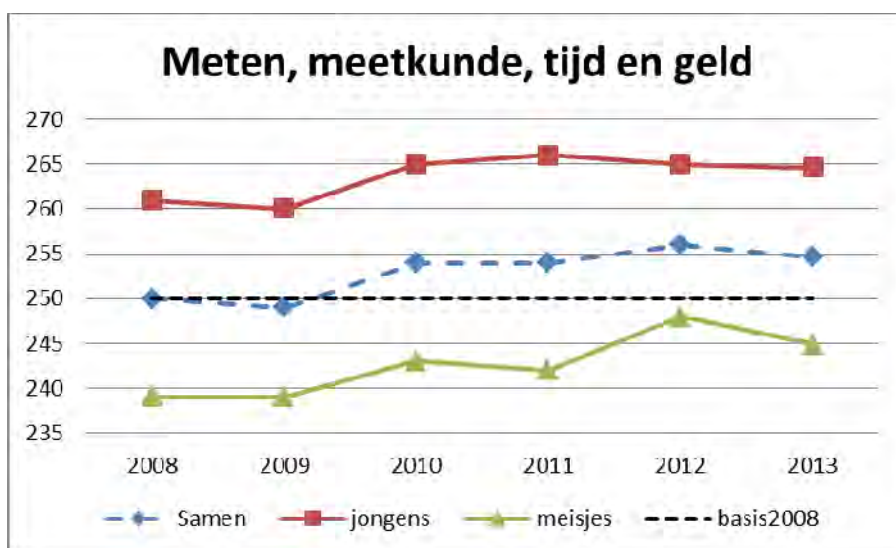
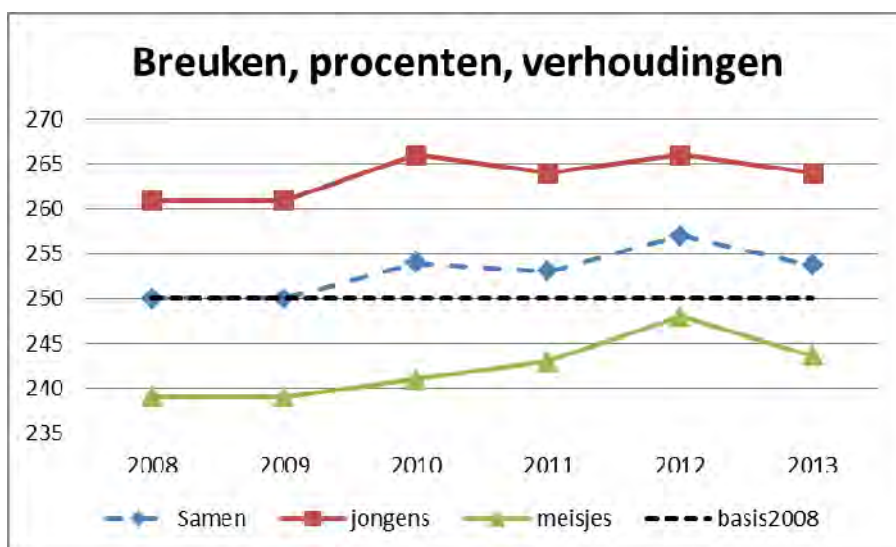
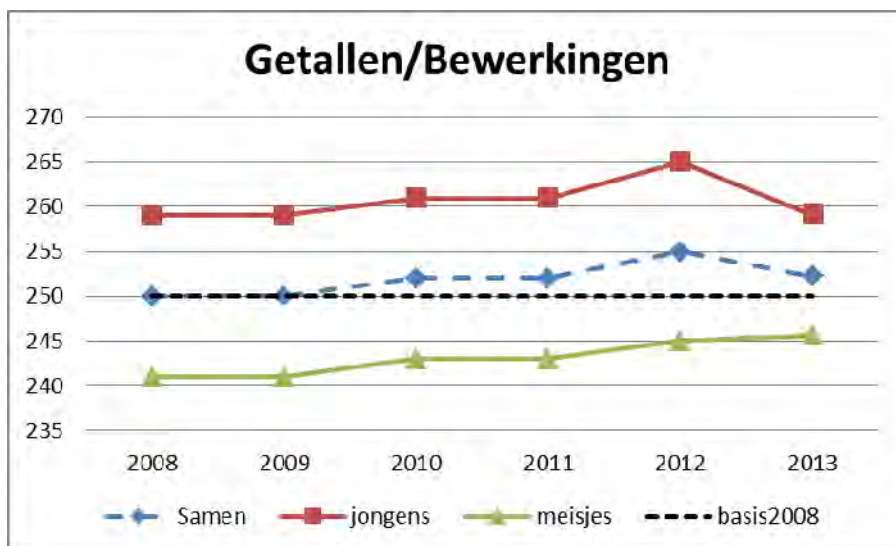
Taal	Geslacht	Gemiddelde vaardigheidsscore						Vershil
		2008	2009	2010	2011	2012	2013	2013-2008
Woordenschat	samen	250	249	257	251	252	254	4
	jongens	255	253	256	259	252	260	5
	meisjes	245	244	258	244	252	249	4
Spelling	samen	250	250	252	252	256	255	5
	jongens	242	241	243	245	249	248	6
	meisjes	258	258	260	258	262	261	3
Begrijpend Lezen	samen	250	252	254	257	253	253	3
	jongens	246	248	249	252	250	249	3
	meisjes	254	256	259	263	256	257	3
Rekenen-Wiskunde								
Getallen/Bewerkingen	samen	250	250	252	252	255	252	2
	jongens	259	259	261	261	265	259	0
	meisjes	241	241	243	243	245	246	5
Breuken, procenten, verhoudingen	samen	250	250	254	253	257	254	4
	jongens	261	261	266	264	266	264	3
	meisjes	239	239	241	243	248	244	5
Meetkunde, meten, tijd en geld	samen	250	249	254	254	256	255	5
	jongens	261	260	265	266	265	265	4
	meisjes	239	239	243	242	248	245	6

Tabel 2.8 Wijziging in het verschil tussen jongens en meisjes van 2008 tot en met 2013 weergegeven in punten op de schaal.

	Taal			Rekenen		
	WS	Sp	BL	G/B	BPV	MMTG
Vershil J-M 2008	10	-16	-8	18	22	22
Vershil J-M 2013	11	-14	-7	14	20	20
Verandering 2008-2013	1	2	1	-4	-2	-2

Figuur 2.3 Trends in vaardigheden naar geslacht





Leertijd

Het verschil in resultaat ten gevolge van de categorisering naar leertijd is overal significant en betekenisvol. De effectgrootte is meestal matig, dat wil zeggen, de waarde ligt boven de 0.50, maar nog onder de 0.80. Alleen bij woordenschat is het effect klein. Leerlingen die zijn vertraagd hebben bij alle onderwerpen een lagere vaardigheid.

Formatiegewicht

Voor de variabele formatiegewicht vinden we overal significante effecten.

Het contrast geen gewicht versus een laag gewicht levert overal een significant en betekenisvol effect op. Leerlingen zonder gewicht hebben een hogere vaardigheid. De effectgrootte is klein en in één geval matig te noemen. Het contrast 0.0 versus 1.2 laat ook overal positieve effecten zien, maar deze zijn nogal wisselend in grootte. We vinden hier het grootste verschil bij deze variabele een effectgrootte van 0,81, dat als groot kan worden gekwalificeerd, namelijk bij Woordenschat. Voor Begrijpend lezen is het effect matig en voor Spelling zelfs betekenisloos. Bij Rekenen leveren alle drie contrasten kleine effecten op.

Het contrast tussen leerlingen met een gewicht onderling, dus formatiegewicht 0.3 versus leerlingen met formatiegewicht 1.2 levert vaak het kleinste effect op van de drie. Alleen Woordenschat wijkt daar van af. Daar is het effect nagenoeg even groot als bij het contrast van leerlingen zonder gewicht met een klein gewicht. Opvallend is hier dat de effecten verschillende kanten op wijzen: leerlingen met een hoog gewicht scoren beter in Spelling, maar slechter in Woordenschat en Begrijpend lezen. Ook bij de rekenonderwerpen scoren leerlingen met een hoog gewicht beter dan die met een laag gewicht. De effectgrootte is echter alleen betekenisvol bij Woordenschat, Spelling en Getallen en bewerkingen.

VO-advies

Bij deze variabele komen we zeer grote verschillen tegen. We zien dat er sprake is van contrasten die overal significant zijn en overal betekenisvolle effectgroottes opleveren. Die effectgroottes variëren van klein tot (zeer) groot. Bij het grootste contrast, dat tussen vwo en vmbo-BB, zien we effectgroottes tot 4,77, hetgeen betekent dat het verschil tussen deze groepen van leerlingen, gecorrigeerd voor geslacht, leertijd, en formatiegewicht, bijna vijf standaarddeviaties bedraagt. Het illustreert de grote spreiding in vaardigheid die we in jaargroep 8 al aantreffen. Opvallend in dit verband is ook de aanzienlijke effectgrootte bij het contrast vwo-havo. Deze leerlingen zitten in het VO soms nog bij elkaar in een klas, maar ook hier is met name bij de onderdelen Begrijpend lezen en Meten, meetkunde, tijd & geld het verschil aanzienlijk. De kleinste verschillen vinden we bij het contrast vmbo-GT en vmbo-KB. Daar variëren de effectgroottes van 0,53 bij Woordenschat tot 0,75 bij Breuken, procenten en verhoudingen.

Toets

Ook de toetsvorm die is gemaakt kan als variabele worden gezien. Leerlingen die een achterstand hebben van anderhalf jaar of meer krijgen van de leerkracht de niveautoets toegewezen. Dit is uiteraard een naar verhouding kleinere groep leerlingen. Wel moet hier opgemerkt worden dat dit aantal flink is toegenomen met de introductie van de papieren Niveautoets. In voorgaande jaren werd deze toets alleen digitaal aangeboden en was het aantal deelnemers kleiner. Het te verwachten verschil in prestatie zien we terug in de effectgroottes. Die variëren van 1,13 bij Spelling tot 1,45 bij de schaal Breuken, procenten en verhoudingen bij rekenen en zijn dus allen als groot te kwalificeren.

De variabele IJK

Net als in de vorige rapportage gebruiken we hier de categorisering voor speciale leerlingen, de IJK-code, bij gebrek aan een code voor de thuistaal van de leerlingen. Het gaat om de volgende aanduiding¹:

Code I: (Allochtone) leerlingen die aan het begin van jaargroep 8 vier jaar of korter in Nederland zijn en die het Nederlands onvoldoende beheersen om de opgaven in de Eindtoets goed te kunnen lezen.

¹ Of de leerkracht deze codes op de juiste wijze heeft gebruikt is maar beperkt te controleren.

Code J: Leerlingen die naar verwachting naar het (voortgezet) speciaal onderwijs of naar het praktijk-onderwijs (pro) gaan.

Code K: Leerlingen die naar verwachting in aanmerking komen voor het leerwegondersteunend onderwijs (lwoo).

Tabel 2.9 Effecten van de variabelen IJK en de variabele I, gemiddelden en standaarddeviaties in 2013

		IJK			Alleen I			
		Effect	M	SD	Effect	M	SD	
Spelling	Geen IJK	0	258	50	Geen I	0	255	50
	Wel IJK	-0,93	205	49	Wel I	-0,19	237	52
Begrijpend Lezen (anker)	Geen IJK	0	258	50	Geen I	0	253	50
	Wel IJK	-1,14	192	50	Wel I	-0,26	226	54
Woordenschat	Geen IJK	0	258	50	Geen I	0	255	50
	Wel IJK	-0,91	204	50	Wel I	-0,43	221	54
Getallen en bewerkingen	Geen IJK	0	257	47	Geen I	0	252	50
	Wel IJK	-1,14	195	48	Wel I	0,04	245	52
Breuken, procenten en verhoudingen	Geen IJK	0	258	47	Geen I	0	254	50
	Wel IJK	-1,16	195	48	Wel I	-0,01	242	52
Meten, tijd en geld	Geen IJK	0	259	47	Geen I	0	255	50
	Wel IJK	-1,12	197	49	Wel I	-0,02	243	52

* Alle verschillen zijn significant (p=0.01)

Leerlingen in de categorieën I en J doen doorgaans niet mee aan de Eindtoets Basis, maar kunnen de Niveautoets voorgelegd krijgen. Voor leerlingen in de categorie K is dat zeker het geval. Daarvan vinden we er ook het meest terug in de data. Er is een grote samenhang tussen formatiegewicht en deze speciale codes. Leerlingen met de code K komen procentueel het meest voor bij successievelijk 1.20- en 0.30-leerlingen. Dat geldt ook voor de codes I en J, maar dan met veel lagere percentages.

In tabel 2.8 zien we dat bij alle onderwerpen het contrast tussen leerlingen die wel een IJK-code hebben en leerlingen die dat niet hebben, significant en betekenisvol is. De effectgroottes zijn zonder uitzondering groot en hebben in de meeste gevallen de omvang van meer dan een hele standaarddeviatie. Alleen bij Spelling en Woordenschat is het effect iets kleiner, maar nog steeds groot te noemen (>0.80).

Voor de volledigheid zijn ook de effecten van alleen de variabele I in kaart gebracht. Deze variabele focust op de leerlingen die te kort in Nederland zijn om de taal goed te beheersen. Bij de rekenschalen en bij Spelling zijn de verschillen niet significant. De effectgroottes kunnen in twee gevallen als betekenisvol gekwalificeerd worden, namelijk bij Begrijpend lezen en Woordenschat. Het gaat dan om kleine effecten.

2.2.2 Effecten 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 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 kleiner, niet altijd significant en het aantal betekenisvolle effecten is minder.

Bij Spelling en bij Rekenen: Getallen en bewerkingen zijn de verschillen voor het contrast tussen stratum 2 (matig) en 3 (veel) niet significant. Bij de overige twee rekenonderwerpen is alleen het contrast 'weinig vs

veel'-gewichtleerlingen betekenisvol.

Bij de taalonderwerpen Woordenschat en Begrijpend lezen levert ook het contrast 'matig versus veel gewichtenleerlingen nog een betekenisvol effect op evenals het contrast 'weinig' versus 'veel'. Daar wordt bij Begrijpend lezen een effectgrootte bereikt die als matig kan worden gekwalificeerd.

Voor de overige variabelen blijkt na correctie weinig verschil over te blijven, ook al zijn de contrasten hier en daar wel significant. Ze hebben echter nergens een effectgrootte die we betekenisvol kunnen noemen.

De schoolvariabelen grootte, regio en verstedelijking doen er zo bezien weinig toe.

Tabel 2.10 Effecten van de variabelen stratum, schoolgrootte, regio en verstedelijking (na correctie)

Variabele	Contrast	Taal			Rekenen		
		WS	Sp	BL	G/B	BPV	MMTG
Stratum (% gew lln.)*	Weinig - Matig	0,11	0,10	0,19	0,10	0,13	0,12
	Matig - Veel	0,37	0,02	0,37	0,05	0,14	0,16
	Weinig - Veel	0,47	0,12	0,56	0,14	0,27	0,28
Schoolgrootte	Groot - Klein	0,01	0,11	0,07	0,08	0,09	0,04
Regio	Oost - Noord	-0,10	-0,01	0,04	0,00	0,03	0,00
	West - Noord	-0,06	0,12	0,09	0,08	0,11	0,06
	Zuid - Noord	-0,13	0,18	0,09	0,14	0,13	0,08
	West - Oost	0,05	0,13	0,05	0,08	0,08	0,06
	Zuid - Oost	-0,02	0,19	0,05	0,14	0,10	0,08
	Zuid - West	-0,07	0,06	0,00	0,06	0,02	0,02
Verstedelijking	Sterk - Zeer sterk	-0,04	-0,06	-0,02	-0,05	-0,06	-0,02
	Matig - Sterk	0,02	-0,05	-0,02	-0,02	-0,03	-0,02
	Weinig - Matig	-0,04	-0,04	-0,06	-0,04	-0,03	-0,02
	Niet - Weinig	0,02	-0,05	-0,06	0,00	0,00	0,02

*vet is significant ($p=0.01$); Stratumindeling naar % gewichtsleerlingen: <10% = weinig, 10 – 25%= matig, > 25% = veel.

3 De resultaten voor jaargroep 4

Vanaf de peiling van 2011 wordt gerapporteerd op een algemene rekenschaal. Dat kan omdat de subschalen voor Rekenen zeer hoog met elkaar correleren. Daarnaast rapporteren we voor de drie taalvaardigheden Spelling, Begrijpend lezen en Woordenschat.

3.1 De vergelijking over de jaren

Bij jaargroep 4 zien we een beeld dat afwijkt van de lijn van de eerdere jaren: de gemiddelde scores liggen in 2013 over het algemeen iets lager dan in 2012. Bij alle onderwerpen is de score in schaalwaarde wel hoger dan in 2008 (zie tabel 3.1). Bij Woordenschat zijn de metingen met het betreffende instrument pas in 2009 gestart en ook daar is de score gestegen.

Tabel 3.1 Effectgroottes en gemiddelde schaalwaarden voor taal en rekenen jaargroep 4

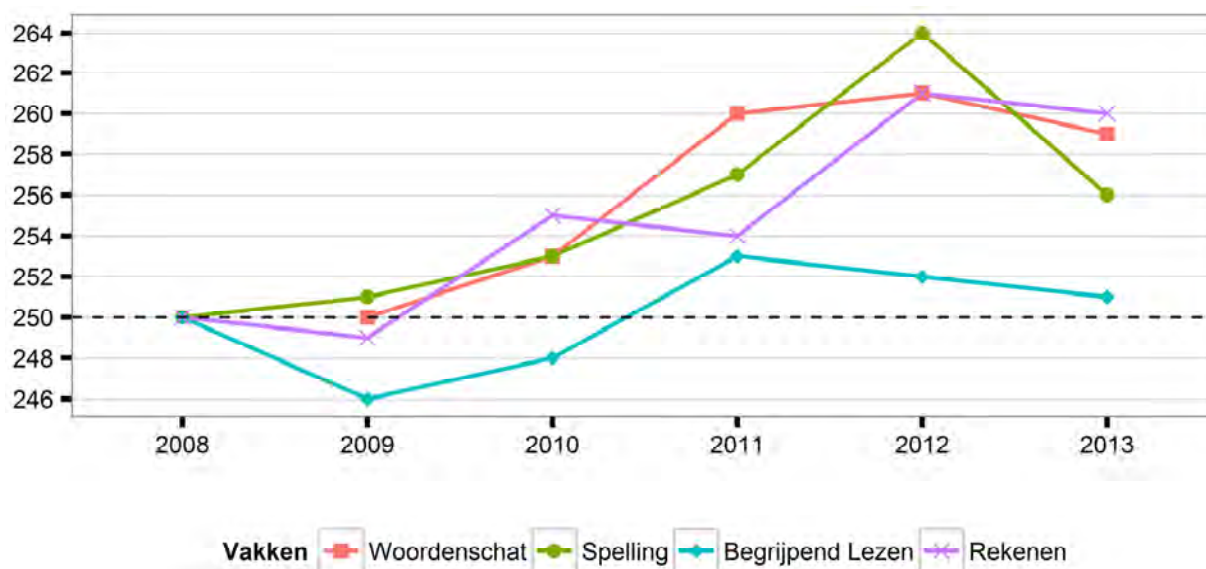
	Effecten*		Jaren					
	2013-2012	2013-2008	2008	2009	2010	2011	2012	2013
Woordenschat*	-0,05	0.19*	-	250	253	260	261	259
Spelling	-0,13*	0.14*	250	251	253	257	264	256
Begrijpend Lezen	-0,02	0.01	250	246	248	253	252	251
Rekenen-Wiskunde	-0,07	0.15*	250	249	253	254	261	260

* Significantie: * : $a < ,01$; ** : $a < ,001$; *Woordenschat is vanaf 2009 gemeten, contrast is hier 2013-2009

Onder "effecten", weergegeven in tabel 3.3, is te zien dat alleen voor Spelling de daling significant is. De daling van 8 punten is opvallend, maar lijkt deels verklaard te kunnen worden door de relatief hoge score in 2012. Immers, in vergelijking met 2011 is de score voor spelling gemiddeld slechts 1 punt gezakt. Over een periode van vijf jaar, 2008-2013, lijkt de gemiddelde schaalwaarde voor Begrijpend Lezen behoorlijk stabiel, terwijl de score voor de andere drie onderwerpen significant toegenomen is.

De trends voor de vier vaardigheden zijn in figuur 3.1 grafisch weergegeven.

Figuur 3.1 Trends voor jaargroep 4



Met behulp van de percentages in tabel 3.2 kunnen we een inschatting maken van de omvang van de verschillen. Per vak en per jaar wordt bekeken hoeveel procent van de leerlingen een score behaalt die in 2008 bij een specifiek percentiel hoort. Zo zien we dat in 2013 bij Woordenschat 57% van de leerlingen de gemiddelde score (dat is, de score die hoort bij p_{50} , het 50^{ste} percentiel) van 2008 het basisjaar behaalt. Er is dus sprake van een toename van 7%. Voor Spelling is dat 5%, voor Begrijpend lezen is dat 1%, en voor de rekenvaardigheid is er een toename te zien van 8%. Bij rekenvaardigheid lagen de scores in 2012 en 2013 op een gemeenschappelijke schaal, terwijl de resultaten van de voorliggende jaren (2008 tot en met 2011) zijn verkregen door de gemiddelden over de drie eerder gerapporteerde rekenschalen te nemen. We zien dus bij alle onderwerpen dat meer dan de helft van de leerlingen in 2013 boven het gemiddelde van 2008 (de standaard) presteert. Ook de percentages voor de laag vaardige leerling LV (= percentiel 25) zijn gestegen ten opzichte van 2008 variërend van 1% bij Begrijpend lezen tot 7% bij Rekenen-wiskunde.

Tabel 3.2 Vergelijking over de jaren heen in percentage per leerlingengroep voor jaargroep 4

Leerlingengroep*	'08	Woordenschat					Spelling					Begrijpend lezen					Rekenen-wiskunde				
		'09	'10	'11	'12	'13	'09	'10	'11	'12	'13	'09	'10	'11	'12	'13	'09	'10	'11	'12	'13
> ZLV (p_{10})	90	90	91	93	93	93	90	92	92	94	92	89	89	91	91	91	90	91	91	94	94
> LV (p_{25})	75	75	77	81	81	81	76	77	79	82	79	72	74	77	76	76	74	77	78	83	82
> Standaard 2008 (p_{50})	50	50	53	58	59	57	51	53	56	59	55	47	48	52	52	51	49	52	54	61	58
> HV (p_{75})	25	25	27	32	32	31	26	27	30	33	29	23	23	27	26	25	24	27	28	34	31
> ZHV (p_{90})	10	10	11	14	14	13	10	11	13	15	12	9	9	11	11	10	10	11	12	15	13

*zie tabel 1.2

Behalve als verschillen op de schaalwaarden kunnen de jaarverschillen ook als effecten weergegeven worden, waarbij per jaar wordt bekeken of er een (significante) stijging of daling is sinds het vorige jaar. Hiervoor zijn gecorrigeerde effectgroottes gebruikt, welke te zien zijn in tabel 3.3. Er is gecorrigeerd voor geslacht, formatiegewicht, stratum en leeftijd. Wederom zien we voor alle onderwerpen een lichte daling, die in het geval van Spelling significant is.

Ook is te zien dat over de periode van vijf jaar de gemiddelde schaalwaarden Woordenschat, Spelling en Rekenen-Wiskunde significant zijn toegenomen. Effectgroottes worden betekenisvol genoemd bij een waarde boven de .20. Dit komt bij deze vergelijkingen echter niet voor. Begrijpend Lezen lijkt ook hier stabiel gebleven over de periode van zes jaar.

Tabel 3.3 Effectgroottes voor de jaarvergelijkingen per onderwerp (gecorrigeerd)

	Woorden- schat	Spelling	Begrijpend lezen	Rekenen
2009 – 2008	n.v.t.	0,07	-0,05	-0,01
2010 – 2009	0,04	0	0	0,06
2011 – 2010	0,14**	0,08*	0,11**	0,04
2012 – 2011	0,03	0,15**	-0,03	0,12**
2013 – 2012	-0,04	-0,12**	-0,01	-0,03
2013 – 2008*	0,14**	0,17**	0,03	0,20**

Significantie: * : $a < ,01$; ** : $a < ,001$; Rekenen-Wiskunde betreft effectgrootte van 4 onderdelen gezamenlijk; Woordenschat 2013-2008 is 2013-2009.

3.2 Verschillen per achtergrondvariabele

In deze paragraaf gaan we in meer detail in op de prestatieverschillen tussen vierdegroepers met een verschillende achtergrond. We presenteren voor 2013 de verschillen voor de achtergrondvariabelen van de leerlingen, nl. geslacht, leertijd, formatiegewicht en thuistaal. Op schoolniveau rapporteren we over de achtergrondvariabelen stratum, regio en verstedelijking (urbanisatiegraad). Ook hier moet bedacht worden dat niet alle in statistisch opzicht significante contrasten ook betekenisvol zijn (zie paragraaf 3.1). De hier gerapporteerde effecten betreffen, net als in eerdere rapportages, gecorrigeerde effecten. Er wordt gecorrigeerd op de (basis-)variabelen geslacht, leertijd, formatiegewicht en stratum. Voor de overige extra variabelen zijn de effecten gerapporteerd zoals gevonden in het model met de basisvariabelen waarbij alleen die enkele extra variabele is toegevoegd, wederom net als in eerdere rapportages. We rapporteren deze effecten gescheiden naar variabelen op leerlingniveau en variabelen op schoolniveau.

Voordat we de verschillen in effecten bespreken kunnen we al constateren dat er meer significante verschillen en betekenisvolle effecten zijn bij de leerlingvariabelen dan bij de schoolvariabelen. Variabelen als regio en verstedelijking blijken weinig verschil in resultaten op te leveren als gecorrigeerd wordt voor de basisvariabelen. Bij de achtergrondvariabelen voor de leerlingen ligt dat anders. Daar vinden we veel kleine en enkele matige effecten en één keer een groot effect. Bij de schoolvariabelen zien we alleen kleine effecten.

3.2.1 Effecten voor leerlingvariabelen

De contrasten bij de leerlingvariabelen leveren in veel gevallen significantie op en de effectgrootte is vaak betekenisvol. We bespreken de uitkomst van deze analyses per variabele.

Tabel 3.4 Effectgrootten* van achtergrondvariabelen op leerlingenniveau voor 2013 in jaargroep 4 (gecorrigeerd)

Variabele	Contrast	Woordenschat	Begrijpend lezen	Spelling	Rekenen-Wiskunde
Geslacht	Jongens-meisjes	0,06	-0,17**	-0,18**	0,36**
Leertijd	Regulier-Vertraagd	0,23**	0,36**	0,45**	0,44**
Formatiegewicht	0.0 - 0.3	0,09	0,30*	0,2	0,2
	0.0 - 1.2	0,37**	0,43**	0,19	0,27*
	0.3 - 1.2	0,28	0,13	-0,01	0,07
Herkomst	Nederland-buitenlands	0,54**	0,29*	-0,12	0,40**
Thuis taal	Alleen NLs - NLs & Bui	0,86**	0,59**	0,22	0,48**
	Alleen NLs - Alleen Bui	0,51**	0,33*	0,07	0,32*
	NLs & Bui - Alleen Bui	-0,35	0,26	0,29	-0,16

* Significantie: * a < ,01 ; ** a < ,001; Rekenen-Wiskunde betreft effectgrootte van 4 onderdelen gezamenlijk.

Geslacht

Jongens behalen over het algemeen lagere scores voor de taalvaardigheden dan meisjes en in tabel 3.4 is te zien dat dat verschil ook in 2013 gevonden werd. Voor Woordenschat, waarvoor het resultaat eerdere jaren vaak wel significant, maar verwaarloosbaar klein was, is het effect van geslacht nu ook niet significant. Hierdoor is het aannemelijk dat op dit onderwerp amper of zeer kleine geslachtsverschillen bestaan. De effecten van begrijpend lezen en spelling zijn significant, maar klein. Bij Rekenen-Wiskunde is de situatie anders: jongens scoren hierop beter, en dit contrast is significant en betekenisvol. Ook hier is de effectgrootte als klein te kwalificeren. Dit effect is overeenkomstig dat gevonden in 2012.

In tabel 3.4 zijn deze verschillen van geslacht op de vaardigheidsschaal² weergegeven, en in figuur 3.2 zijn deze grafisch weergegeven per vak. Als we de trend voor jongens en meisjes bekijken in de trendlijnen over de afgelopen vijf jaren, dan valt op dat die lijnen meestal parallel lopen: dat wil zeggen, er lijkt weinig verandering te zijn over de jaren heen in het verschil tussen jongens en meisjes. Wat verschillen op geslacht betreft is duidelijk dat bij de taalonderdelen de meisjes gemiddeld hoger scoren dan de jongens, maar bij Rekenen-Wiskunde is deze situatie andersom.

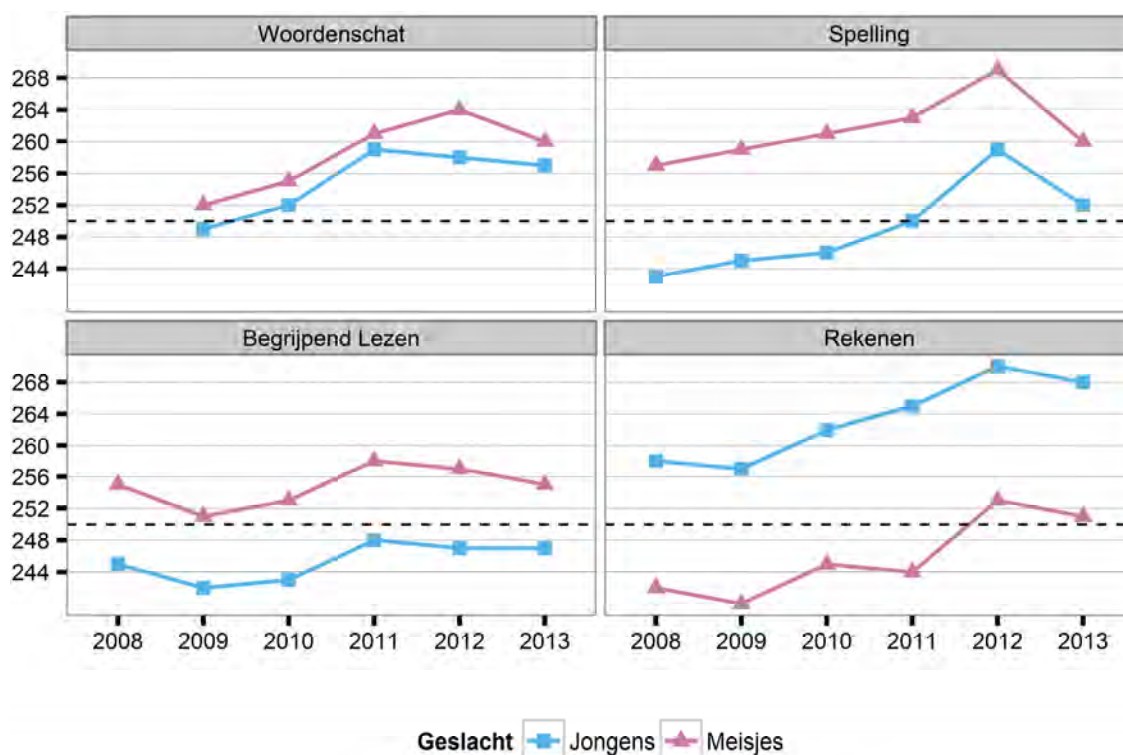
² Resultaten op de vaardigheidsschaal zijn de resultaten waarbij niet gecorrigeerd wordt voor de effecten van de basisvariabelen.

Tabel 3.5 Trend in schaalwaarden van 2008 tot en met 2013 naar geslacht

	Gemiddelde vaardigheidsscores						toename
	2008	2009	2010	2011	2012	2013	2013-2008*
Woordenschat*	-	250	253	260	261	259	9
Jongens	-	249	252	259	258	257	8
Meisjes	-	252	255	261	264	260	8
Spelling	250	251	253	257	264	256	6
Jongens	243	245	246	250	259	252	11
Meisjes	257	259	261	263	269	260	3
Begrijpend Lezen	250	246	248	253	252	251	1
Jongens	245	242	243	248	247	247	2
Meisjes	255	251	253	258	257	255	0
Rekenen-Wiskunde	250	249	253	254	261	260	10
Jongens	258	257	262	265	270	268	10
Meisjes	242	240	245	244	253	251	9

* Start van de meting Woordenschat in 2009, dus voor Woordenschat is dit 2013-2009.

Figuur 3.2 Trends in vaardigheid naar geslacht



Leertijd

In tabel 3.4 zien we dat vertraagde leerlingen, dat wil zeggen de leerlingen die ten minste een keer doubleerden, aanzienlijk lagere scores behaalden dan de niet-vertraagde leerlingen. Dit contrast levert over de hele linie een klein effect op. Net als in 2012 was het effect het kleinst bij Woordenschat en het grootst bij Spelling en Rekenen-Wiskunde.

Formatiegewicht

Bij de start van het project (2008) rapporteerden scholen (ook) nog oude formatiegewichten, maar de laatste jaren hebben we alleen te maken met de nieuwe formatiegewichten. Bij Woordenschat en Rekenen-Wiskunde scoren de leerlingen zonder formatiegewicht (0.0) significant beter dan leerlingen met een hoog formatiegewicht (1.2), waarbij sprake is van een kleine tot matige effectgrootte. Bij Begrijpend Lezen is ook een significant effect te zien tussen leerlingen zonder formatiegewicht en leerlingen met een laag formatiegewicht (0.3). Bij Spelling zijn geen significante effecten te zien. Voor geen van de vakken is er een sterk verschil gevonden tussen leerlingen met een laag en een hoog formatiegewicht.

Er zijn kleine verschillen waar te nemen met 2012, maar in grote lijnen is het resultaat vergelijkbaar. In 2012 was met name het verschil tussen leerlingen met formatiegewicht 0.0 en 0.3 groter dan in 2013 voor de onderwerpen Woordenschat en Rekenen-Wiskunde.

Herkomst

Deze variabele laat significante verschillen zien bij Woordenschat, Begrijpend Lezen en Rekenen-Wiskunde. Hierbij zijn de effectgroottes matig. Voor Spelling is geen significant effect te vinden. In 2012 leek hier nog een omgekeerd effect te zijn, waarbij leerlingen met een niet-Nederlandse herkomst hogere scores behaalden. Dit vinden we in 2013 niet meer terug.

Thuis taal³

We kunnen zien in tabel 3.4 bij de onderwerpen Woordenschat, Begrijpend Lezen en Rekenen-Wiskunde dat leerlingen, die thuis alleen Nederlands spreken, significant hoger scoren dan leerlingen die thuis zowel Nederlands als een buitenlandse taal spreken (hierbij zijn de effectgroottes matig) en als leerlingen die thuis alleen een buitenlandse taal spreken (hierbij zijn de effectgroottes klein). Voor Spelling zijn wederom geen verschillen te vinden.

Voor deze variabele hebben we de resultaten van de meting ook over de jaren heen per categorie in kaart gebracht. We zien in tabel 3.6 en figuur 3.3 duidelijke verschillen in de ontwikkeling per groep.

In het algemeen zien we vooral een opvallende daling van de scores van de groep leerlingen die naast Nederlands een andere taal spreken. Van 2011 op 2012 was dit al het geval, maar deze daling lijkt nu nog sterker aanwezig. Bij de alle vakken scoort deze groep inmiddels minder goed dan de groep leerlingen die thuis slechts één buitenlandse taal spreekt. Dit is opvallend, omdat dit in de eerdere jaren van het peilingsonderzoek nog niet het geval was. Aangezien de groep relatief op alle vakken erg slecht scoort, is het echter ook mogelijk dat de steekproef dit jaar relatief zwakke leerlingen bevat.

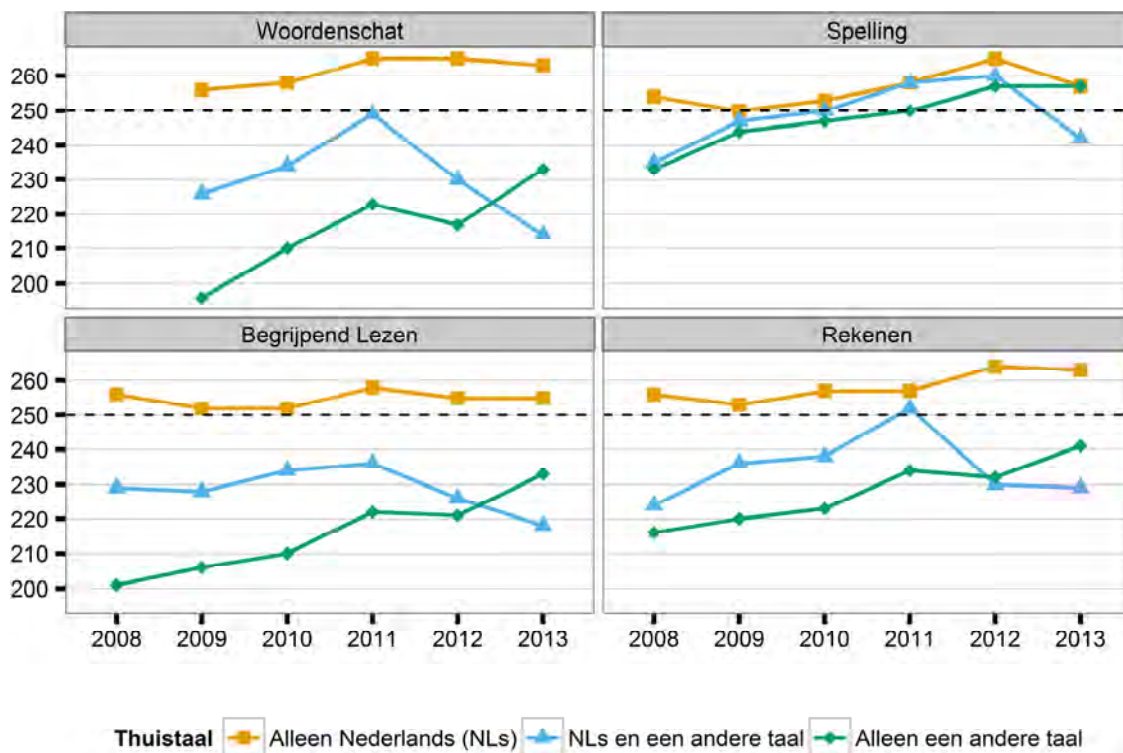
³ In deze variabele is geen onderscheid gemaakt naar buitenlandse taal. In de meeste gevallen gaat het om Marokkaans, Turks en Surinaams/Antilliaans.

Tabel 3.6 Trend in schaalwaarden van 2008 tot en met 2013 voor de categorieën van thuistaal

		Gemiddelde vaardigheidsscores						Toename 2013-2008*
		2008	2009	2010	2011	2012	2013	
Woordenschat*	Alleen Nederlands (NLs)	-	256	258	265	265	263	7
	NLs en een andere taal	-	226	234	249	230	214	-12
	Alleen een andere taal	-	196	210	223	217	233	37
Spelling	Alleen Nederlands (NLs)	254	250	253	258	265	257	3
	NLs en een andere taal	235	247	250	258	260	242	7
	Alleen een andere taal	233	244	247	250	257	257	24
Begrijpend lezen	Alleen Nederlands (NLs)	256	252	252	258	255	255	-1
	NLs en een andere taal	229	228	234	236	226	218	-11
	Alleen een andere taal	201	206	210	222	221	233	32
Rekenen-Wiskunde	Alleen Nederlands (NLs)	256	253	257	257	264	263	7
	NLs en een andere taal	224	236	238	252	230	229	5
	Alleen een andere taal	216	220	223	234	232	241	25

* Start van de meting in 2009

Figuur 3.3 Trends in vaardigheid naar thuistaal



3.2.2 Effecten van schoolvariabelen

Voor de verschillen van variabelen op schoolniveau is gekeken naar drie eigenschappen: *Stratum*, *Regio* en *Verstedelijking*. De resultaten zijn te zien in tabel 3.7. Het is goed om bij de interpretatie in het achterhoofd te houden dat dit wederom gecorrigeerde modellen zijn. Dit wil zeggen dat als bijvoorbeeld een school A met veel achterstandsl leerlingen een lagere score heeft dan een school B met weinig achterstandsl leerlingen, dat dit komt omdat de reguliere leerlingen op school A slechter scoren dan de

reguliere leerlingen op school B en/of de achterstandsleerlingen op school A slechter scoren dan die op school B. Het komt dus niet omdat school A meer achterstandsleerlingen bevat die het gemiddelde omlaag halen, omdat daarvoor gecorrigeerd is. De basisvariabelen waarvoor gecontroleerd is zijn wederom geslacht, formatiegewicht, leertijd en stratum.

Stratum

Bij Woordenschat scoren scholen met weinig tot geen achterstandsleerlingen (stratum 1) significant beter dan zowel scholen met wat meer achterstandsleerlingen (stratum 2) als scholen met relatief veel achterstandsleerlingen (stratum 3). Dit effect lijkt wat groter te zijn dan in 2012. De effectgrootte is als matig te kwalificeren. Voor Spelling is er wederom geen significant effect te zien. Voor Begrijpend Lezen en Rekenen-Wiskunde geldt dat stratum 1-scholen significant beter scoren dan stratum 3-scholen, met een kleine effectgrootte. Er is in 2013 geen significant effect te zien tussen stratum 2- en stratum 3-scholen, waar dit in 2012 nog wel het geval was.

Regio

Het contrast tussen de onderscheiden regio's levert geen significante effecten op.

Verstedelijking

In 2012 werd er geen enkel significant effect gevonden tussen de leerlingen uit een stedelijke en een landelijke omgeving. In 2013 zien we echter voor alle vakken behalve Spelling een klein effect dat wel significant is.

Tabel 3.7 Effectgrootten* van achtergrondvariabelen op schoolniveau voor 2013 (gecorrigeerd)

Variabele	Contrast	Woordenschat	Spelling	Begrijpend lezen	Rekenen-Wiskunde
Stratum	Stratum 1-Stratum 2	0,09	0,06	0,10	0,13
	Stratum 1-Stratum 3	0,55**	0,02	0,25*	0,31**
	Stratum 2-Stratum 3	0,46**	-0,03	0,15	0,18
Regio	Noord – Oost	0,06	0,14	0,18	0,12
	Noord – West	0,02	0,07	0,12	0,15
	Noord – Zuid	0,14	0,02	0,19	0,07
	Oost – West	-0,04	-0,07	-0,06	0,02
	Oost – Zuid	0,09	-0,11	0,01	-0,06
	West – Zuid	0,13	-0,04	0,07	-0,08
Verstedelijking	Stad-Land	-0,17*	-0,09	-0,17**	-0,20**

* Significantie: * a < ,01 ; **: a < ,001;

4 Conclusies

4.1 Algemeen beeld

Vergelijken we de prestaties van 2013 met die van 2012, dan is het beeld divers. De prestaties van de leerlingen is nog verder verbeterd bij Woordenschat, gestabiliseerd bij Begrijpend lezen en bij Spelling en rekenvaardigheid gedaald.

De daling die bij de rekenvaardigheid gevonden is van 2012 naar 2013, komt vrijwel overeen met de stijging van 2011 naar 2012. De verschillen tussen 2011 en 2013 zijn dan ook vrijwel nihil en ook niet significant. Bekijken we de trends over de gehele periode vanaf het begin van de meting, dan is de vaardigheid voor alle onderwerpen net als in 2012 nog steeds (significant) hoger dan in 2008. Het gaat om een verschil dat varieert van 2 tot 5 punten op de schaal met als gemiddelde 250 en als standaarddeviatie 50.

Wel moet geconstateerd worden dat de effectgroottes beneden de grens van 0.20 vallen en dus niet betekenisvol mogen worden genoemd.

In jaargroep 4 gaan alle geanalyseerde vaardigheden in meerdere of mindere mate achteruit tussen 2013 en 2012. Alleen bij Spelling is daarbij sprake van significantie. Ook in groep 4 zijn ondanks de gevonden dalingen de resultaten in 2013 beter dan die in 2008. Die stijging ten opzichte van 2008 is overal significant, behalve bij Begrijpend lezen. De effectgroottes zijn hier, op begrijpend lezen na, ook groter en variëren van 0,14 bij Woordenschat tot 0,20 bij Rekenen-Wiskunde. Dat laatste ligt op de grens van betekenisvol.

Nemen we de twee jaargroepen bij elkaar dan zien we dat er ten opzichte van 2012 geen significante vooruitgang is geboekt, maar dat er bij enkele vaardigheden wel sprake is van significante achteruitgang, namelijk bij Spelling, zowel in groep 8 als groep 4 en bij Rekenen in groep 8. Bij de andere vaardigheden zijn er geen significante verschillen.

Alles bij elkaar genomen moeten we dus concluderen dat de trend ten opzichte van 2008 in alle gevallen nog steeds positief uitvalt, maar dat er sprake is van stabilisering of achteruitgang en in geen geval van verdere vooruitgang.

4.2 Invloed van achtergrondvariabelen

Geslacht

Jongens in jaargroep 8 behalen hogere scores in Rekenen-Wiskunde en Woordenschat dan meisjes. Meisjes zijn beter in Spelling en Begrijpend lezen. In jaargroep 4 zien we dezelfde verschillen, maar is het verschil bij Woordenschat niet significant.

Leertijd

Vertraagde leerlingen blijken in alle gevallen een grote achterstand te hebben. De effectgrootte is bij jaargroep 8 forsler dan in jaargroep 4 en over het algemeen spreken we in jaargroep 8 van een matig effect en in jaargroep 4 van een klein effect. In vergelijking met de andere onderwerpen zien we bij Woordenschat het kleinste verschil.

Formatiegewicht

In jaargroep 8 zien we dat leerlingen zonder formatiegewicht (0.00) het op alle vaardigheden beter doen dan leerlingen met een formatiegewicht, laag (0.30) of hoog (1.20). Het gaat dan om kleine en matige effectgroottes. Woordenschat is de enige vaardigheid waarbij leerlingen met een laag formatiegewicht significant hogere scores dan leerlingen met een hoog formatiegewicht. Er is dan sprake van een klein effect. Bij Spelling en Getallen/Bewerkingen is het effect andersom: leerlingen met een hoog gewicht presteren beter dan leerlingen met een laag gewicht.

In jaargroep 4 is het beeld wel vergelijkbaar, maar zijn veel contrasten niet significant. Het contrast tussen leerlingen zonder gewicht en een laag formatiegewicht is alleen bij Begrijpend lezen significant. Het contrast van leerlingen zonder gewicht met leerlingen met een hoog gewicht is naast begrijpend lezen ook nog significant bij Woordenschat en Rekenen-Wiskunde. Het contrast laag gewicht (0.30) versus hoog gewicht (1.20) is nergens significant. De gevonden effectgroottes zijn klein.

Stratum

Deze variabele is geconstrueerd op basis de verdeling in formatiegewichten op school. De verschillen en effecten zijn dan ook grotendeels vergelijkbaar met die bij de variabele formatiegewicht. In jaargroep 8 zijn de verschillen nagenoeg overal significant, maar niet in alle gevallen met betekenisvolle effectgroottes. Die zien we vooral bij taalonderwerpen, waar leerlingen in stratum 1 hogere prestaties behalen dan leerlingen in stratum 3. Het aantal significante verschillen en betekenisvolle effectgroottes is in groep 4 beperkter. Alleen bij Woordenschat bereikt de effectgrootte de kwalificatie matig.

Regio en verstedelijking

In jaargroep 8 zien we significante verschillen voor de variabele regio en verstedelijking. Ze komen voor verschillen tussen regio's het meest voor bij Spelling en Rekenen, minder bij Begrijpend lezen en het minst bij Woordenschat. Vooral het contrast Zuid-Noord blijkt significante verschillen op te leveren, doorgaans ten gunste van de regio Zuid, maar bij Woordenschat juist andersom. De effectgroottes zijn allen echter betekenisloos. Bij verstedelijking zijn er enkele significante verschillen, maar de effectgroottes zijn nog kleiner. In jaargroep 4 levert verstedelijking voor Woordenschat en Begrijpend lezen en Rekenen-Wiskunde ook significante verschillen op, steeds in het voordeel van de leerlingen in stedelijke gebieden. Bij Rekenen-Wiskunde is de effectgrootte betekenisvol, maar klein. Regio levert als variabele in groep 4 geen significante verschillen op.

Doorstroom VO

Een variabele in jaargroep 8 die wel veel significante verschillen oplevert en ook grote effecten laat zien is het doorstroomadvies voor het VO. De grootste waarden zien we bij Begrijpend lezen en Rekenen-Wiskunde, waar het contrast tussen leerlingen met vwo-advies en leerlingen met vmbo BB-advies bijna vijf standaarddeviaties beslaat. Dat zijn in de context van de sociale wetenschappen en met name die van onderwijsonderzoek enorme verschillen. Het geeft aan hoe groot de vaardigheidsverschillen aan het eind van de basisschool al zijn. Voor Woordenschat en Spelling zijn de verschillen dan nog het kleinst, nl. circa 3.5 standaarddeviaties. Vooral het verschil tussen de leerlingen met adviezen vwo en havo leveren een grote bijdrage aan de grootte van het totaaleffect.

4.3 Discussie

Alles bij elkaar genomen mogen we concluderen dat de opwaartse trend voor alle reken- en taalvaardigheden die vanaf 2008 waarneembaar was, met ups en downs, in 2013 niet wordt doorgezet. Alleen bij Woordenschat in groep 8 zien we nog een stijgende lijn. Bij de andere vaardigheden is er sprake van stabilisatie of teruggang.

De resultaten zijn verschillend voor de beide jaargroepen in omvang. De vooruitgang in jaargroep 4 ten opzichte van 2008 is groter en heeft een in het algemeen een minder grillig verloop dan in jaargroep 8. Uitzondering daarbij is Begrijpend lezen dat in jaargroep 8 nu meer vooruitgang ten opzichte van 2008 vertoont dan in jaargroep 4. Bij Woordenschat en Rekenen-Wiskunde zijn de verschillen echter duidelijk groter in jaargroep 4 dan in jaargroep 8.

Opvallend is verder dat de trend in jaargroep 4 redelijk consistent is over alle vaardigheden heen.

De negatieve ontwikkeling van 2012 naar 2013 is uniform voor alle vaardigheden en verschilt alleen in omvang. In jaargroep 8 zijn de trendlijnen grilliger. Dit is vooral voor de taalonderdelen het geval, waar we zowel neerwaartse als opwaartse bewegingen zien in het laatste jaar.

Hoewel deze evaluatie gestart is in het kader van de Kwaliteitsagenda (2008), waarin o.a. het opbrengstgericht werken wordt gestimuleerd, met name bij taal en rekenen, kan op grond van de uitkomsten van dit onderzoek slechts ten dele een adequate beleidsevaluatie worden gepleegd. Voor het vinden van verklaringen ontbreken 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.

Het verdient daarom aanbeveling om, net als bij het regulier peilingsonderzoek (PPON) gebruikelijk is, 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 het besteden van meer tijd aan bepaalde onderwerpen kan wellicht een plausibele interpretatie worden verbonden aan de verandering in resultaten bij leerlingen.

Literatuur

Berkel, S. van, M. Hilde, R. Engelen, F. Kamphuis, F. Kleintjes, R. Krom (2010). *Woordenschat Groep 3 t/m 5. Wetenschappelijke verantwoording*. Cito, Arnhem.

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

Cito (2012). *Terugblik en resultaten 2012. 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.

Feenstra, H., F. Kamphuis, F. Kleintjes, R. Krom (2010). *Wetenschappelijke verantwoording Begrijpend lezen voor groep 3 tot en met 6*. Cito, Arnhem.

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

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)

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

Janssen, Jan, Frank van der Schoot, Bas Hemker (2005). *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 4. Uitkomst van de vierde peiling in 2004*. PPON-reeks nummer 32. Cito, Arnhem. (http://www.cito.nl/po/ppon/rekwisk/eind_fr.htm)

Janssen, J., N. Verhelst, R. Engelen en F. Scheltens (2010). *Wetenschappelijke verantwoording van de toetsen LOVS Rekenen-Wiskunde voor groep 3 tot en met 8*. 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.

Wijs, A. de, F. Kamphuis, F. Kleintjes, M. Tomesen (2010). *Wetenschappelijke verantwoording Spelling voor groep 3 tot en met 6*. Cito, Arnhem.

Relevante websites:

www.cito.nl

www.toetswijzer.nl

<http://ppon.cito.nl>

www.minocw.nl

Cito maakt wereldwijd werk van goed en eerlijk toetsen en beoordelen. Met de meet- en volgmethoden van Cito krijgen mensen een objectief beeld van kennis, vaardigheden en competenties. Hierdoor zijn verantwoorde keuzes op het gebied van persoonlijke en professionele ontwikkeling mogelijk. Onze expertise zetten we niet alleen in voor ons eigen werk maar ook om advies, ondersteuning en onderzoek te bieden aan anderen.

Cito

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

Klantenservice

T (026) 352 11 11
klantenservice@cito.nl

Fotografie: Ron Steemers