

Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2009

Jaarlijks Peilingsonderzoek van het Onderwijsniveau - Technische rapportage
B.T. Hemker, J.B. Kuhlemeier en J.J. van Weerden



Jaarlijks Peilingsonderzoek van het Onderwijsniveau

**Peiling van de rekenvaardigheid en de taalvaardigheid
in jaargroep 8 en jaargroep 4 in 2009**

Technische rapportage

B.T. Hemker, J.B. Kuhlemeier en J.J. van Weerden

© Stichting Cito Instituut voor Toetsontwikkeling Arnhem (2010)

Alle rechten voorbehouden. 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

Verkorte samenvatting 5

1 Inleiding 7

2 De opzet van het onderzoek 9

2.1 Opzet van het onderzoek in jaargroep 8 9

2.2 Opzet van het onderzoek in jaargroep 4 13

3 Eigenschappen van de gebruikte toetsen 17

3.1 Eigenschappen van de gebruikte toetsen in jaargroep 8 17

3.2 Eigenschappen van de gebruikte toetsen in jaargroep 4 19

4 Analyse en rapportage: opgaven en groepeigenschappen op één schaal 21

4.1 Toepassing van een IRT-model 21

4.2 De vaardigheidsschaal: rapportage van de populatie 22

4.3 Analyse en rapportage van prestatieverschillen tussen 2008 en 2009 23

4.4 Het interpreteren van de figuren en tabellen 27

5 Resultaten voor taalvaardigheid in jaargroep 8 33

5.1 Nederlandse Taal: Woordenschat 34

5.2 Nederlandse Taal: Spelling 40

5.3 Nederlandse Taal: Begrijpend lezen 46

6 Resultaten voor Rekenen/Wiskunde in jaargroep 8 53

6.1 Rekenen/Wiskunde: Getallen en Bewerkingen 54

6.2 Rekenen/Wiskunde: Breuken, Procenten en Verhoudingen 61

6.3 Rekenen/Wiskunde: Meten, Meetkunde, Tijd en Geld 66

7 Resultaten voor taal- en rekenvaardigheid in jaargroep 4 71

7.1 Taalvaardigheid: Woordenschat 72

7.2 Taalvaardigheid: Spelling 78

7.3 Taalvaardigheid: Begrijpend lezen 81

7.4 Rekenvaardigheid 83

8 Discussie 91

8.1 Doel en context 91

8.2 Onderzoekopzet en de onderzoeksmethode 92

8.3 De resultaten voor jaargroep 8 94

8.4 De resultaten voor jaargroep 4 102

Literatuur 109

Bijlagen

1	Verdeling van de achtergrondvariabelen de steekproef in jaargroep 8	111
2	De opgaven in jaargroep 8 (inclusief tabellen en figuren)	113
3	Een alternatieve representatie van de vaardigheidsschalen	125

Verkorte samenvatting

In het kader van de kwaliteitsagenda 'Scholen voor morgen' is in 2008 het Jaarlijks Peilingsonderzoek van 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, 2008). In het onderhavige rapport wordt verslag gedaan van de tweede jaarlijkse peiling van de reken- en taalvaardigheid in jaargroep 4 en 8.

Vergelijken we de prestaties van 2009 met die van 2008, dan blijken de vierde- en achtstegroepers een kleine vooruitgang in respectievelijk Spelling en Begrijpend lezen te hebben geboekt. Daarnaast blijkt dat de kloof tussen de prestaties van autochtone en allochtone leerlingen iets kleiner is geworden. Dit geldt voor Woordenschat in jaargroep 8 en voor Rekenen en Spelling in jaargroep 4. Bij deze onderdelen hebben allochtone leerlingen dus een deel van de achterstand ingehaald.

De volgende jaarlijkse peilingen zullen uitwijzen in hoeverre deze ogenschijnlijke vooruitgang zich voortzet.

Deze tweede jaarlijkse peiling geeft voor het eerst de mogelijkheid een vergelijking te maken voor de vaardigheden taal en rekenen in jaargroep 4 en 8 voor twee opeenvolgende jaren. Het valt niet te verwachten dat er op een dergelijke korte termijn al op landelijk niveau relevante en eenduidig te interpreteren verschillen kunnen worden aangetroffen. Veranderingen in het onderwijs op landelijk niveau gaan daarvoor dikwijls te langzaam. Er is op zijn hoogst sprake van indicaties, waarbij op langere termijn wellicht een trend is te zien.

De taalvaardigheid en de rekenvaardigheid van de leerlingen in jaargroep 8 en jaargroep 4 zijn net als in 2008 weer gemeten met toetsen die al veel in het onderwijs worden gebruikt. Dat waren de Eindtoets Basisonderwijs voor jaargroep 8 en toetsen van het Leerlingvolgsysteem (LOVS) voor jaargroep 4. De gemeten taal vaardigheden waren voor beide leerjaren Begrijpend lezen, Spelling en Woordenschat. De gemeten rekenvaardigheden verschilden iets over de twee leerjaren. In jaargroep 4 waren dat *getallen en getalsrelaties, optellen en aftrekken, vermenigvuldigen en delen, en meten, tijd en geld*. In jaargroep 8 waren de drie rekenvaardigheden *getallen en bewerkingen, breuken, procenten en verhoudingen en meten, meetkunde, tijd en geld gemeten*.

Bij het meten van de leerlingen zijn ook achtergrondvariabelen van de leerlingen meegenomen. De belangrijkste was de variabele afnamejaar: het jaar waarin de leerling de toets had gemaakt. Door de groepen te vergelijken komen we er achter of er sprake is van een gewijzigd vaardigheidsniveau door de tijd. Andere basisvariabelen waren formatiegewicht (is de leerling een achterstandsleerling?), stratum (zitten er veel achterstandsleerlingen op school?), leertijd (is de leerling een zittenblijver?) en geslacht. Toegevoegde variabelen die meegenomen zijn in de analyses waren regio en urbanisatiegraad van de locatie van de school, gesproken thuistaal en voor de leerlingen in jaargroep 8 het advies van de leraar of lerares wat betreft het te volgen voortgezet onderwijs na jaargroep 8.

Bij de analyses kwam naar voren dat er nauwelijks betekenisvolle veranderingen in de tijd hebben plaats gevonden. In de meeste gevallen waren de resultaten niet significant. Het enige significante resultaat voor jaargroep 8 was een vooruitgang bij Begrijpend lezen, maar die vooruitgang was miniem. Wel was opvallend dat in jaargroep 4 bij de achterstandsscholen meer een vooruitgang geconstateerd werd bij de meeste vaardigheden terwijl bij scholen met juist weinig achterstandsleerlingen een lichte achteruitgang was gevonden. In jaargroep 8 is bij taal dit effect gevonden op leerlingniveau: Bij de achterstandsleerlingen was een lichte vooruitgang en bij de niet-achterstandskinderen een lichte achteruitgang. Het is nog wel zo dat achterstandsleerlingen en -scholen nog wel slechter presteren, maar de kloof tussen wel en niet achterstand is iets kleiner geworden. Een uitzondering is de vaardigheid Begrijpend lezen in jaargroep 8 waar juist een omgekeerd effect gevonden is.

1 Inleiding

In het kader van de kwaliteitsagenda 'Scholen voor morgen' is in 2008 het Jaarlijks Peilingsonderzoek van het Onderwijsniveau (JPON) van start gegaan. Dit jaarlijkse onderzoek is gericht op het monitoren van het onderwijsniveau op het gebied van Nederlandse taal en rekenen/wiskunde in het basisonderwijs. JPON heeft tot doel een antwoord te geven op de volgende onderzoeksvragen:

- Wat steken de leerlingen op van het onderwijs?
- Hoe groot zijn de prestatieverschillen tussen groepen van leerlingen (zoals jongens en meisjes of autochtone en allochtone leerlingen)?
- Welke wijzigingen voltrekken zich van jaar tot jaar in de leerresultaten en in de prestatieverschillen tussen groepen leerlingen?

In 2009 is verslag gedaan van de eerste jaarlijkse niveaupeiling van taal en rekenen in jaargroep 4 en 8 in 2008 (Hemker & Van Weerden, 2009). Voor u ligt het verslag en de technische verantwoording van de tweede peiling van de taal- en rekenvaardigheid die in 2009 is uitgevoerd in jaargroep 8 en 4. Het rapport bevat ook een vergelijking van de leerprestaties van de leerlingen van 2009 met die van 2008.

Dit verslag wijkt op enkele punten af van het vorige verslag. Het belangrijkste verschil is dat nu de leerresultaten van twee opeenvolgende jaren met elkaar zijn vergeleken. Zaken die niet veranderd zijn, krijgen in dit verslag minder aandacht. Dat betreft onder meer de relatie met ander onderzoek zoals het periodieke peilingonderzoek van het onderwijsniveau in Nederland (PPON) en het Cohort Onderzoek Onderwijs-Loopbanen onder leerlingen van 5 tot 18 jaar (COOL⁵⁻¹⁸). Ook is er minder aandacht voor de verschillen in leerresultaten tussen jaargroep 4 en jaargroep 8. Alleen als er verschuivingen zijn, hebben wij daar aandacht aan geschonken.

Na deze inleiding begint het verslag met een uiteenzetting van de opzet van het onderzoek (hoofdstuk 2). Daarna bespreken we de eigenschappen van de gebruikte toetsen (hoofdstuk 3) en de wijze waarop de gegevens geanalyseerd en gerapporteerd zijn (hoofdstuk 4). Vervolgens beschrijven we de taalvaardigheid van de leerlingen in jaargroep 8 (hoofdstuk 5), de rekenvaardigheid in jaargroep 8 (hoofdstuk 6) en de taal- en rekenvaardigheid in jaargroep 4 (hoofdstuk 7). De vaardigheden van de leerlingen illustreren we daarbij aan de hand van een beschrijving van de opgaven die zij wel of juist niet goed beheersen. Speciale aandacht wordt ook geschonken aan de relatie tussen leerresultaten en achtergrondvariabelen zoals geslacht en thuistaal. Het laatste hoofdstuk bevat een discussie waarin we onder meer de resultaten van 2009 vergelijken met die van 2008 (hoofdstuk 8). Tot slot zijn enkele bijlagen opgenomen.

2 De opzet van het onderzoek

In dit rapport wordt verslag gedaan van de tweede jaarlijkse niveaupeiling van het onderwijsniveau in de groepen 8 en 4 van het basisonderwijs. De onderzoeksmethoden zijn nagenoeg identiek aan die van de eerste peiling uit 2008. Dit geldt zowel voor de gebruikte toetsen, de achtergrondkenmerken als de statistische analyse. Vandaar dat wij de opzet van het onderzoek hier slechts summier beschrijven. Voor de details van de methoden en technieken verwijzen wij naar het eerste rapport dat is uitgegeven onder de titel *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2008* (Cito, 2009).

2.1 Opzet van het onderzoek in jaargroep 8

De vaardigheden en de gebruikte meetinstrumenten

In jaargroep 8 zijn in totaal zes vaardigheden gemeten. Voor Nederlandse taal waren dat Begrijpend lezen, Spelling en Woordenschat en bij Rekenen/Wiskunde ging het om Getallen en Bewerkingen, Breuken, Procenten en Verhoudingen en Meten, tijd en geld. Deze vaardigheden zijn gemeten met de Eindtoets Basisonderwijs, de Niveautoets, de Digitale Eindtoets en het Toetsboekje Extra. Hieronder lichten we de gebruikte toetsen en datasets kort toe.

Eindtoets Basisonderwijs

Voor het onderzoek in 2009 is net als in 2008 gebruikgemaakt van de Eindtoets Basisonderwijs (EB). De gegevens zijn verzameld in de periode van 3 tot en met 5 februari 2009. Merk op dat in de EB bij Taal ook het onderwerp 'Schrijven van teksten' is opgenomen. Dit onderwerp heeft weliswaar een hoge samenhang met het onderdeel Begrijpend lezen, maar moet toch inhoudelijk als een andere vaardigheid gezien worden. Ook analysetechnisch blijkt dat deze vaardigheid niet goed op dezelfde schaal geplaatst kan worden als Begrijpend lezen. Om die reden zijn de opgaven behorend bij 'Schrijven van teksten' buiten beschouwing gelaten. Een ander onderdeel van de EB dat buiten beschouwing gelaten is, is de vaardigheid Studievaardigheden.

Niveautoets

Het gebruik van de EB heeft als nadeel dat sommige scholen deze toets niet aan de minst vaardige leerlingen hebben voorgelegd. Om dat probleem te ondervangen hebben wij ook de resultaten op de Niveautoets (NT) gebruikt. De NT is bestemd voor leerlingen met een algehele leerachterstand van ten minste 1½ jaar. Het betreft overwegend leerlingen die in aanmerking komen voor praktijk- of leerwegondersteunend onderwijs. De NT wordt op de computer gemaakt, maar bevat dezelfde onderdelen en hetzelfde aantal opgaven als de reguliere, papieren Eindtoets. De opgaven van de NT zijn aangepast aan het niveau van leerlingen met een grote leerachterstand en dus overwegend gemakkelijker dan die van de EB. Desalniettemin konden de opgaven van de EB en de Niveautoets met behulp van Item Response Theorie op één schaal gebracht worden. Daarnaast hebben we een controlevariant van de EB gebruikt die tegelijkertijd en onder dezelfde omstandigheden als de reguliere EB is afgenomen. Deze variant die ook op papier wordt afgenomen wordt het Ankerboekje genoemd. Het vormt de verbinding tussen de toets van 2008 en die van 2009. We hebben overigens ook andere methoden gebruikt om de resultaten van de verschillende toetsen vergelijkbaar te maken, maar daarover later meer.

Digitale Eindtoets Basisonderwijs

Net als voorgaande jaren konden scholen in 2009 de Eindtoets Basisonderwijs ook digitaal afnemen, maar deze gegevens zijn niet in het onderzoek gebruikt. De digitale variant van de EB is ook gebruikt als inhaaltoets voor leerlingen die bijvoorbeeld vanwege ziekte niet aan de reguliere papieren eindtoets konden deelnemen. De Digitale Eindtoets bestaat uit dezelfde onderdelen en hetzelfde aantal opgaven als de reguliere, papieren Eindtoets, maar bevat andere opgaven. De moeilijkheid van de Digitale Eindtoets is echter gelijk aan die van de reguliere Eindtoets. De afname van de Digitale Eindtoets vindt plaats in de periode tussen 26 januari en 3 april 2009. Deze toets wordt bij een beperkt aantal leerlingen afgenomen

(1125 leerlingen), in totaal minder dan 1% van de leerlingen die de Eindtoets maakten. In tegenstelling tot de Niveautoets is er geen directe reden om aan te nemen dat deze leerlingen fundamenteel van de andere leerlingen verschillen, noch wat betreft de resultaten, noch wat betreft de achtergrondvariabelen. De periode van afname verschilt echter duidelijk van die van de EB

Toetsboekje Extra

Het Toetsboekje Extra bevat opgaven die niet meegerekend worden voor de score op de Eindtoets, maar die gebruikt worden om de gegevens van de EB op dezelfde schaal te plaatsen als de toetsen van het Leerlingvolgsysteem (LVS). Deze toets gebruiken we om de verschillende afnamejaren met elkaar te verbinden zodat een vergelijking mogelijk is.

Het vergelijken van jaren

De reguliere EB's voor de verschillende jaren hebben geen opgaven gemeenschappelijk. Daardoor is vergelijking in de tijd niet zonder meer mogelijk. Om de resultaten van 2009 toch met die van 2008 te kunnen vergelijken is gebruikgemaakt van Item Response Theorie (IRT). Om die techniek toe te kunnen passen moeten de te construeren vaardigheidsschalen wel een aantal opgaven gemeenschappelijk hebben. Deze overlap is op drie manieren verkregen. Ten eerste is er een proefonderzoek gehouden waarin leerlingen voor de EB bestemde conceptopgaven voorgelegd kregen. Ten tweede hebben we het eerder genoemde Ankerboekje gebruikt dat in 2009 is gemaakt door 3847 leerlingen. Dat bevat opgaven die ook in 2008 bij een klein aantal (1211) leerlingen zijn afgenomen. Ten derde is het Toetsboekje Extra gebruikt.

Om de jaren met elkaar te verbinden met IRT moet de verbinding tussen de jaren zo sterk mogelijk zijn. Hiermee kunnen IRT-schalen gemaakt worden: meetlatten waarmee de jaren vergeleken kunnen worden. Voor het maken van die schalen (meetlatten) hoeft je geen willekeurige steekproef te gebruiken. Je moet er alleen wel voor zorgen dat zoveel mogelijk leerlingen in de analyses meegenomen worden die opgaven gemaakt hebben die zowel in 2008 als in 2009 zijn afgenomen. Dat is de kalibratieset: de dataset waarmee de meetlat gemaakt wordt. Dat deze dataset geen willekeurige steekproef van alle leerlingen in de twee afnamejaren is, is voor de toepassing van de techniek van IRT geen probleem. Binnen deze theorie zijn de schattingen immers onafhankelijk van de populatie.

Voor het meten van de vaardigheden in de verschillende jaren is echter in het kader van de gewenste representativiteit, wel een willekeurige steekproef noodzakelijk. Daarom zal, na het maken van de vaardigheidsschalen, een andere dataset die wel een willekeurige steekproef is, gebruikt worden voor de analyse van o.a. de jaarverschillen. Die dataset wordt de onderzoekssteekproef genoemd.

Kalibratieset

Voor de schaling van de verschillende toetsen is een speciale dataset aangemaakt. Bij de samenstelling van deze zogeheten kalibratieset is ervoor gezorgd dat het aantal leerlingen met meerdere toetsen in het bestand zo groot mogelijk is. De kalibratieset bevat derhalve alle leerlingen die zowel de Niveautoets, het Ankerboekje als het Toetsboekje Extra gemaakt hebben. De dataset is aangevuld met een (wel willekeurige) steekproef van leerlingen die alleen de reguliere EB maakten, maar niet al in de dataset zaten. Het aantal leerlingen in de kalibratieset per toets per afnamejaar is weergegeven in tabel 2.1.

Tabel 2.1 Aantal leerlingen in de kalibratieset per toets per afnamejaar

Aantal leerlingen	2008	2009	Totaal
Alleen Eindtoets	6608	3941	10549
Ankerboekje	1164	3847	5011
Niveautoets	2315	2347	4662
Eindtoets en Toetsboekje Extra	5840	5966	11806
Totaal	15927	16101	32028

Bij deze dataset zijn nog twee opmerkingen te maken. Ten eerste zijn niet alle beschikbare leerlingen in de statistische analyse betrokken. Het analyseprogramma kan namelijk niet meer dan ongeveer 32.000 observaties per opgave aan. Dit is alleen “ten koste” gegaan van de leerlingen die alleen de EB gemaakt hebben. Omdat de itemparameterschattingen in IRT-analyses populatieonafhankelijk zijn, is dit voor de schaling geen probleem.

Ten tweede hebben we in deze analyses ook de parameters van de opgaven uit het Toetsboekje Extra geschat. De desbetreffende opgaven zijn echter alleen gebruikt voor de vergelijking in de tijd (dit wil zeggen: het schalen van de Eindtoetsen uit 2008 en 2009).

De schattingen van de itemparameters zijn op twee manieren gecontroleerd. Ten eerste zijn de parameters vergeleken met de parameters die gevonden zijn bij het eerder genoemde proefonderzoek met de EB. Ten tweede is gecontroleerd of de gevonden parameters ook stand hielden bij de dataset die gebruikt is voor de vergelijking van de leerresultaten in 2008 en 2009. Deze controles leverden bevredigende resultaten op.

De onderzoekspopulatie en de onderzoekssteekproef

De analyse van de taal- en rekenvaardigheden in Nederland en de ontwikkeling daarvan van 2008 naar 2009 is uitgevoerd op een andere dataset dan de hiervoor besproken kalibratieset. De onderzoekspopulatie van JPON betreft de reguliere leerlingen in 2008 en 2009 in het basisonderwijs in Nederland. Tabel 2.2 vermeldt het aantal leerlingen en het aantal scholen waarvan gegevens beschikbaar zijn.

Tabel 2.2 Aantal leerlingen en scholen in de onderzoekspopulatie

	2008	2009	Totaal
Aantal leerlingen			
Totaal	146608	148396	295004
In onderzoekspopulatie	140337	141931	282268
Percentage van totaal	95,72	95,64	95,68
Aantal scholen			
Totaal	6252	6194	12446
In onderzoekspopulatie	5850	5789	11639
Percentage van totaal	93,57	93,46	93,52

Leerlingen kunnen om verschillende redenen buiten de onderzoekspopulatie vallen. Daartoe behoren de leerlingen die de digitale variant van de EB gemaakt hebben (957 leerlingen in 2008; 1125 in 2009) en leerlingen die de EB in het buitenland gemaakt hebben (218 leerlingen in 2008; 270 in 2009). Net als in 2008 is er ook in 2009 een groep instellingen die geen basisonderwijs aanbod of waarbij het niet duidelijk was of het een reguliere basisschool betrof. In deze groep zitten bijvoorbeeld ook leerlingen uit het speciaal (basis-)onderwijs of het voortgezet onderwijs. Omdat de leerlingen van deze scholen niet tot de onderzoekspopulatie behoren, zijn deze niet meegenomen in de analyses.

Merk op dat het bij de uitgesloten leerlingen vaak gaat om individuele, in plaats van klassikale afnamen. Men is dan specifiek in die ene leerling geïnteresseerd, die niet willekeurig gekozen is. Dit verklaart ook waarom het aantal leerlingen per school bij de uitgesloten leerlingen in vergelijking met de onderzoekspopulatie relatief klein is.

Er zijn geen gegevens van CBS of CFI beschikbaar per leerjaar. Maar op basis van andere gegevens schatten we er dat er in 2008 en 2009 tussen de 180.000 en de 185.000 leerlingen in jaargroep 8 van het reguliere basisonderwijs zaten. Dat betekent dat ten minste tussen de 75% en 80% van de populatie in ons onderzoek betrokken is. Zouden we de leerlingen toevoegen die de Digitale Eindtoets maakten, dan ligt dit percentage eerder richting de 80%, waarbij het percentage in 2009 iets hoger ligt dan in 2008.

Achtergrondkenmerken

Naast de toetsresultaten zijn van de leerlingen dezelfde achtergrondkenmerken verzameld als in 2009. Ook de definitie van deze variabelen is gelijk gebleven. Het gaat daarbij om de volgende achtergrondkenmerken: geslacht, leertijd, formatiegewicht, thuistaal en advies VO. In vergelijking met de eerdere peiling is de belangrijkste toevoeging de variabele afnamejaar: 2008 tegenover 2009.

Bij de variabele leertijd onderscheiden we twee categorieën:

- regulier: de leerlingen in jaargroep 8 die op 1 oktober van het jaar voor de afname niet ouder zijn dan 11 jaar (voor de afname van de Eindtoets Basisonderwijs van 2008 is deze datum 1 oktober 2007 en voor 2009 is dat 1 oktober 2008).
- vertraagd: de leerlingen die op dat moment 12 jaar of ouder zijn.

Als toelichting op het kenmerk formatiegewicht het volgende. De leerlinggewichten voor het basisonderwijs zijn opnieuw gedefinieerd en op het ogenblik worden deze gefaseerd ingevoerd. In het schooljaar 2007-2008 vielen de leerlingen in jaargroep 8 nog onder de oude gewichtenregeling. Daarom zijn de volgende gewichten gehanteerd :

- 1.25 (nu factor 0.25) voor Nederlandse arbeiderskinderen (in termen van opleidings- en/of beroepsniveau van de ouders);
- 1.40 (nu factor 0.40) voor schipperskinderen in een internaat of pleeggezin;
- 1.70 (nu factor 0.70) voor kinderen uit de reizende en trekkende bevolking;
- 1.90 (nu factor 0.90) voor kinderen van wie ten minste één van de ouders van niet-Nederlandse herkomst is (en beperkingen kent in opleidings- en/of beroepsniveau);
- 1.00 (nu factor 0.00) voor alle andere kinderen.

De gewichten tussen haakjes zijn de gewichten die tegenwoordig gebruikt worden *naast* de nieuwe gewichten. Ondanks de grote aantallen leerlingen in de analyses komen de gewichten 1.40 en 1.70 zeer weinig voor. Slechts 207 leerlingen in 2009 (en 191 in 2008) hadden een leerlinggewicht van 1.4 en 188 hadden een leerlinggewicht van 1.7 (was 202 in 2008). Om die reden zijn in de analyses de leerlingen met gewichten 1.25, 1.40 en 1.70 samengenomen als een groep.

Met de variabele advies VO wordt bedoeld het (ingeschatte) niveau van de opleiding in het voortgezet onderwijs die de leerling na jaargroep 8 gaat volgen. Dit betreft het advies van de leerkracht voor de leerling voordat de Eindtoets is afgenomen. Hier zijn negen categorieën onderscheiden. De vijf hoofdcategorieën zijn:

- BB: basisberoepsgerichte leerweg binnen het vmbo;
- KB: kaderberoepsgerichte leerweg binnen het vmbo;
- GT: gemengde of theoretisch leerweg binnen het vmbo;
- havo;
- vwo.

De vier overige categorieën zijn de tussenliggende categorieën: BB/KB, KB/GT, GT/havo en havo/vwo. In de rapportage van de resultaten worden alleen de resultaten van deze hoofdcategorieën gegeven. Merk op dat de gemengde categorie KB/GT in 2009 als afzonderlijke categorie geschrapd is. Toch vinden we deze gemengde categorie nog terug in onze gegevens. De reden is dat leraren zowel KB als GT aankruisten. Het aantal waarnemingen in die categorie nam echter wel relatief sterk af. De leerresultaten voor de tussenliggende categorieën zijn overigens conform de verwachting: telkens liggen de resultaten van een tussencategorie tussen de resultaten van de twee aangrenzende hoofdcategorieën.

Bij de variabele thuistaal zijn er drie hoofdcategorieën onderscheiden, te weten alleen Nederlands, Nederlands plus een andere taal en alleen een andere taal. Er is bij thuistaal ook een onderscheid te maken naar specifieke andere talen. Deze talen zijn Turks, Arabisch, Surinaams/Antilliaans en een West-Europese taal. Om voldoende leerlingen in de verschillende groepen te hebben zijn de leerlingen die alleen deze andere taal spreken samengenomen met de leerlingen die deze andere taal samen met Nederlands thuis spreken.

Naast deze leerlingkenmerken zijn drie schoolkenmerken in het onderzoek betrokken, te weten stratum, regio en urbanisatiegraad.

De stratumindeling is een indeling van scholen op basis van schoolgewicht. De schoolscore is gebaseerd op de formatiegewichten van de leerlingen en bestaat uit de ratio van het gewogen aantal leerlingen en het nominale aantal leerlingen, met aftrek van een correctieterm van het gewogen aantal leerlingen.

Deze correctieterm bedraagt 9% van het nominale aantal leerlingen. Hierdoor heeft de schoolscore (uitgaande van voorheen geldende formatiegetallen) een bereik van 0.91 tot en met 1.81.

Deze stratumindeling weerspiegelt in globale termen een indeling van de schoolpopulatie op basis van sociaal-economische achtergrond van de schoolpopulatie. Voor de variabele stratum is een driedeling van de schoolgewichten gebruikelijk:

Stratum	Schoolscore	Omschrijving
Stratum 1	0.91 - 1.00	overwegend kinderen van ouders met afgeronde opleiding, weinig allochtone leerlingen
Stratum 2	1.01 - 1.20	relatief meer autochtone 1.25 leerlingen, weinig allochtone 1.90 leerlingen
Stratum 3	1.21 - 1.81	vooral autochtone 1.25 leerlingen en allochtone 1.90 leerlingen

De indeling naar regio en urbanisatiegraad is gebaseerd op de indeling van het CBS. Dat betekent dat het kenmerk regio uit vier categorieën bestaat op basis van de provincie waarin de school zich bevindt:

- Noord: Groningen, Friesland, Drenthe
- Oost: Overijssel, Gelderland, Flevoland
- West: Utrecht, Noord-Holland, Zuid-Holland, Zeeland
- Zuid: Noord-Brabant, Limburg

De indeling naar urbanisatiegraad gebeurt op postcode. Het betreft een vijfdeling die ook binnen CBS-publicaties gebruikt wordt:

- zeer sterk stedelijk;
- sterk stedelijk;
- matig stedelijk;
- weinig stedelijk;
- niet stedelijk.

Voor de verdeling van de achtergrondvariabelen van de leerlingen in 2009, zie Bijlage 1.

2.2 Opzet van het onderzoek in jaargroep 4

De vaardigheden, meetinstrumenten en de vergelijking van de afnamejaren

In jaargroep 4 zijn niet helemaal dezelfde vaardigheden gemeten als in jaargroep 8. De vaardigheden verschillen niet alleen qua niveau, maar ook inhoudelijk. Dat kan natuurlijk ook niet anders omdat de leerlingen beduidend jonger zijn en het onderwijs duidelijk anders is. Van de ene kant worden sommige basisvaardigheden in jaargroep 8 niet meer direct getoetst omdat vrijwel alle leerlingen deze dan beheersen. Van de andere kant worden er in jaargroep 8 ook vaardigheden gemeten die in jaargroep 4 nog niet onderwezen worden. Al met al zijn in jaargroep 4 voor Nederlands drie vaardigheden gemeten, te weten Begrijpend lezen, Spelling en Woordenschat. Bij Rekenen/Wiskunde ging het om vier vaardigheden, te weten Getallen en getalrelaties, Optellen en aftrekken, Vermenigvuldigen en delen en Meten, tijd en geld.

Voor de vaardigheden Begrijpend lezen en Spelling zijn gegevens van bestaande startmodules en vervolgmodes van het leerlingvolgsysteem (LOVS) gebruikt. Hiermee zijn de vaardigheden zowel in 2008 als in 2009 gemeten. Voor de vaardigheid Woordenschat zijn de gegevens in 2008 afkomstig van de proeftoetsen voor nieuw te ontwikkelen toetsen binnen het LOVS. Voor de meting in 2009 is de nieuwe woordenschattoets gebruikt. Voor Rekenen/Wiskunde is slechts één toets gebruikt waarbij de twee delen door alle leerlingen aan het einde van jaargroep 4 gemaakt werden.

De onderzoekspopulatie en de achtergrondvariabelen

Als in jaargroep 4 en 8 dezelfde toets is afgenomen (behalve bij Woordenschat) is vergelijking van de leerprestaties tussen deze twee groepen in principe mogelijk. In jaargroep 8 is met de Eindtoets zonder extra toetsdruk en extra werk voor de scholen een enorme steekproef van meer dan 80% van de leerlingen gerealiseerd. De verdeling van de achtergrondgegevens van de gemeten leerlingen verschilt niet van de landelijke verdeling. Er is dan ook geen reden om aan te nemen dat het vaardigheidsniveau van de gemeten en niet gemeten leerlingen in jaargroep 8 substantieel verschilt. Sommige scholen nemen de Eindtoets niet bij de minst vaardige leerlingen af. Van deze leerlingen hebben we echter wel gegevens van de Niveautoets. Daardoor beschikken we toch over gegevens over het onderste gedeelte van vaardigheidsverdeling waar zich de minder vaardige leerlingen bevinden.

De metingen in jaargroep 4 leverden wel extra werk op voor de scholen. De antwoorden per opgave moesten bij de scholen worden opgevraagd, evenals de achtergrondgegevens per leerling. Aangezien we de scholen niet te veel wilden belasten hebben we in jaargroep 4 slechts een steekproef van zo'n 2500 leerlingen van ongeveer 100 scholen in het onderzoek betrokken.

In 2009 zijn in jaargroep 4 dezelfde achtergrondkenmerken gebruikt als in 2008 (met dien verstande dat in 2009 ook de variabele afnamejaar is meegenomen). De vergelijking van de beide onderzoeksjaren zal niet direct op basis van de totaalscores gedaan worden. Een belangrijke reden is dat de samenstelling van de steekproef kan fluctueren. Dat is bij een relatief kleine steekproef van 2500 leerlingen in jaargroep 4 eerder het geval dan bij de omvangrijke steekproef in jaargroep 8. Aangezien de verwachte veranderingen van de leerprestaties in een tijdbestek van één jaar relatief klein zijn, kunnen kleine steekproeffluctuaties relatief grote effecten hebben. Voor deze steekproeffluctuaties wordt gecorrigeerd. Dit doen we met behulp van wegingen en het berekenen van gecorrigeerde effecten. De basis voor de analyses is het item response model (zie ook hoofdstuk 4).

De onderzoekspopulatie betreft jaargroep 4 in het reguliere basisonderwijs. Net zoals bij het onderzoek in jaargroep 8 worden hier de leerlingen uit het Speciaal (Basis-)Onderwijs buiten beschouwing gelaten. De gebruikte achtergrondvariabelen in jaargroep 4 komen grotendeels overeen met de achtergrondvariabelen die gebruikt zijn voor het onderzoek binnen jaargroep 8. De basisvariabelen zijn gelijk. Dat zijn de variabelen afnamejaar (2008 of 2009), stratum, formatiegewicht, geslacht (jongen of meisje) en leertijd. Extra variabelen zijn regio, urbanisatiegraad en thuistaal. De definitie van de meeste van deze variabelen is vergelijkbaar met die in jaargroep 8. Er zijn wel een paar verschillen die we hieronder toelichten. Bij de variabele leertijd zijn de leeftijden uiteraard anders dan bij jaargroep 8: daar waar 11 jaar de referentieleeftijd op 1 oktober bij jaargroep 8 is, is dat 7 jaar bij jaargroep 4. Bij de variabele thuistaal is er een niveau toegevoegd, nl. streektaal. Het grootste verschil zit in de definitie van de variabele formatiegewicht. De definitie van de gewichten van de leerlingen voor de formatietelling is gewijzigd en deze verandering wordt gefaseerd ingevoerd. In jaargroep 8 worden de oude definities nog gebruikt, terwijl in jaargroep 4 (vrijwel) alle leerlingen de nieuwe gewichten hebben. Er worden voor de nieuwe gewichten drie typen ouders onderscheiden (zie 'Brochure nieuwe gewichtenregeling basisonderwijs', april 2008, van het CFI –zie ook www.cfi.nl) Vereenvoudigd weergegeven ziet dat er als volgt uit:

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

De nieuwe gewichten zijn dan:

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

Merk op dat in vergelijking met de oude gewichten bij de nieuwe definities van de leerlinggewichten de herkomst van de ouders geen rol meer speelt. Het aantal leerlingen met oude gewichten was in 2009 vrij klein. Leerlingen met een oud gewicht worden in de groep geplaatst met de vergelijkbare rangorde bij de nieuwe gewichten. Hierdoor ontstaan drie groepen formatiegewichten:

- F1: Leerlingen met nieuw gewicht 0.00 (of oud gewicht 1.00);
- F2: Leerlingen met nieuw gewicht 0.30 (of oud gewicht 1.25, 1.40 of 1.70);
- F3: Leerlingen met nieuw gewicht 1.20 (of oud gewicht 1.90).

Deze samenvoegingen zijn gebaseerd op de gegevens uit het PRIMA-onderzoek waarmee een zogeheten was-woordt-tabel is gemaakt. Daarin bleek de grootste overlap te zitten tussen 0.0 en 1.00, tussen 0.3 en 1.25 en tussen 1.2 en 1.90. Ook zijn in 2008 de effecten onderzocht tussen 0.0 en 1.00, tussen 0.3 en 1.25 en tussen 1.2 en 1.90. Deze effecten waren bij de meeste vaardigheden afwezig of niet significant of beide.

3 Eigenschappen van de gebruikte toetsen

Zoals uiteengezet in hoofdstuk 2 zijn in deze studie (onderdelen van) verschillende meetinstrumenten gebruikt die via overlap in opgaven op één schaal gebracht zijn. In dit hoofdstuk worden de eigenschappen van de gebruikte subtoetsen beschreven voor de jaargroepen 8 en 4 in 2008 en 2009.

3.1 Eigenschappen van de gebruikte toetsen in jaargroep 8

In deze paragraaf beschrijven we achtereenvolgens de eigenschappen van de Eindtoets, het Ankerboekje en de Niveautoetsen. De eigenschappen van het Toetsboekje Extra zijn hier niet gegeven: die toetsen zijn alleen gebruikt voor het maken van de meetschalen (het koppelen van de jaren) maar niet voor het meten van de leerlingen.

Eindtoets

De eigenschappen van de zes subtoetsen van de Eindtoets in 2008 en 2009 zijn weergegeven in tabel 3.1.

Tabel 3.1 *Toetseigenschappen van de subtoetsen Nederlands en Rekenen/Wiskunde in de Eindtoets in 2008 en 2009*

	Aantal items	Gem. score	SD	GemP	Alpha	GLB
Nederlandse Taal in 2008 (N = 136.858)						
Nederlandse taal						
Spelling	20	14,8	3,2	74,2	0,69	0,73
Woordenschat	20	13,5	3,0	67,7	0,61	0,66
Begrijpend lezen	30	23,1	4,5	77,0	0,78	0,80
Nederlandse Taal in 2009 (N = 135.737)						
Nederlandse taal						
Spelling	20	14,9	3,4	74,4	0,73	0,76
Woordenschat	20	14,2	3,2	70,9	0,66	0,69
Begrijpend lezen	30	23,1	4,2	77,1	0,76	0,78
Rekenen/Wiskunde in 2008 (N = 136.858)						
Rekenen/Wiskunde						
Getallen en Bewerkingen	25	18,1	4,6	72,49	0,82	0,83
Breuken, Procenten en Verhoudingen	20	14,6	4,0	72,86	0,80	0,82
Meten, Meetkunde, Tijd en Geld	15	10,9	2,8	72,77	0,69	0,72
Rekenen/Wiskunde in 2009 (N = 135.737)						
Rekenen/Wiskunde						
Getallen en Bewerkingen	25	18,3	4,4	73,1	0,79	0,80
Breuken Procenten en Verhoudingen	20	14,0	4,1	70,2	0,80	0,82
Meten, Meetkunde, Tijd en Geld	15	10,6	3,1	70,5	0,73	0,75

Naast de gemiddelde score (Gem.score) en de standaarddeviatie (SD) van de scores is ook de gemiddelde P-waarde gegeven (GemP). Deze gemiddelde P-waarde is de gemiddelde score gedeeld door de maximale score (in dit geval is dat gelijk het aantal items in de schaal) vermenigvuldigd met 100. De gemiddelde P-waarde valt goed te interpreteren als de moeilijkheid van de subtoets voor deze groep leerlingen. Hij kan

ook gezien worden als het gemiddelde percentage correcte antwoorden voor de groep leerlingen die deze toets gemaakt heeft.

De alpha die hier genoemd is, is Cronbach's alpha. Deze veel gebruikte maat geeft een indicatie van de betrouwbaarheid van een toets. Deze alpha is echter een ondergrens van de feitelijke ofwel 'ware' betrouwbaarheid. Als er genoeg observaties zijn, kan ook de GLB (Greatest Lowerbound; Verhelst, 1998) bepaald worden. Ook dat is een ondergrens van de betrouwbaarheid, maar een die dichterbij de feitelijke betrouwbaarheid ligt dan Cronbach's alpha. Volgens het beoordelingsstelsel voor toetsen van de Commissie Toetsaangelegenheden (COTAN) van het Nederlands Instituut van Psychologen (NIP) zijn de betrouwbaarheden van de meest subschalen goed te noemen voor onderzoek op groepsniveau (Evers, Lucassen, Meijer & Sijtsma, 2009). Alleen van de subtoets Woordenschat is de betrouwbaarheid met name in 2008 wat lager, maar volgens de criteria van de COTAN nog steeds voldoende.

Bij de interpretatie van de gegevens in tabel 3.1 is een waarschuwing op zijn plaats. De leerprestaties van 2008 en 2009 zijn niet direct vergelijkbaar. De EB van 2009 bevatte immers volledig andere opgaven dan die van 2008. De resultaten zijn echter vergelijkbaar gemaakt door ze met behulp van Item Response Theorie op één schaal te brengen (zie ook hoofdstuk 4).

Eindtoets-Anker (toetsen in het Ankerboekje)

In 2008 hebben meer leerlingen de standaard Eindtoets gemaakt dan in 2009, terwijl het totaal aantal leerlingen in 2009 hoger was dan in 2008. Dat komt omdat in 2009 meer leerlingen de Ankertoets gemaakt hebben. Van deze Ankertoets zijn 42 opgaven gebruikt die niet in de standaard Eindtoets zitten: 20 rekenopgaven en 22 taalopgaven. De overige 88 opgaven zijn gelijk aan de Eindtoets (40 rekenopgaven, en 48 taalopgaven). Merk op dat er voor Spelling geen enkele overlap is met de opgaven van de Ankertoets. Bij Spelling is de overlap geheel verkregen met behulp van het Toetsboekje Extra (zie tabel 3.2).

Tabel 3.2 Toetseigenschappen van de subtoetsen in de Eindtoets-Anker in 2008 en 2009 (N = 1164)

	Aantal items	Gem. score	SD	GemP	Alpha
Eindtoets-Anker in 2008 (N = 1164)					
Spelling	20	15,0	3,1	74,9	0,66
Woordenschat	19	13,4	2,9	70,7	0,61
Begrijpend lezen	29	21,7	4,5	74,7	0,77
Getallen en Bewerkingen	21	16,2	3,7	77,1	0,78
Breuken, Procenten en Verhoudingen	20	14,1	4,1	70,7	0,81
Meten, Meetkunde, Tijd en Geld	15	11,0	2,6	73,1	0,64
Eindtoets-Anker in 2009 (N = 3847)					
Spelling	20	14,8	3,4	74,2	0,72
Woordenschat	20	14,4	3,2	72,0	0,66
Begrijpend lezen	28	21,3	4,2	76,1	0,76
Getallen en Bewerkingen	24	17,3	4,2	72,3	0,78
Breuken, Procenten en Verhoudingen	20	13,7	4,2	68,4	0,80
Meten, Meetkunde, Tijd en Geld	15	10,9	2,7	72,6	0,66

Voor de schalen bij de Eindtoets-Anker is geen GLB geleverd omdat het aantal leerlingen te klein was om deze maat te kunnen berekenen. Gezien de alpha's zijn de subtoetsen in de EB-Anker echter voldoende betrouwbaar voor analyses op groepsniveau. De waarden van alpha liggen voor 2009 over het algemeen ook iets hoger dan in 2008.

Niveautoetsen

De eigenschappen van de subtoetsen van de Niveautoets in 2008 en 2009 zijn weergegeven in tabel 3.3. Vanwege het kleine aantal leerlingen is voor de schalen van de Niveautoets geen GLB beschikbaar.

Tabel 3.3 Eigenschappen van de subtoetsen in de Niveautoets in 2008 en 2009

	Aantal items	Gem. score	SD	GemP	Alpha
Niveautoets in 2008 (N = 2315; bk 3)					
Spelling	20	12,8	3,7	64,0	0,72
Woordenschat	20	15,2	3,1	76,1	0,67
Begrijpend lezen	29	19,8	4,8	68,4	0,78
Getallen en Bewerkingen	25	17,8	4,3	71,1	0,78
Breuken, Procenten en Verhoudingen	20	13,4	3,9	67,0	0,76
Meten, Meetkunde, Tijd en Geld	15	9,9	2,8	66,0	0,63
Niveautoets in 2009 (N = 2347; bk 7)					
Spelling	20	16,0	3,1	80,1	0,72
Woordenschat	20	16,1	2,8	80,7	0,68
Begrijpend lezen	29	20,1	4,7	69,2	0,77
Getallen en Bewerkingen	25	17,9	4,2	71,5	0,77
Breuken, Procenten en Verhoudingen	20	13,6	3,8	67,8	0,75
Meten, Meetkunde, Tijd en Geld	15	9,8	2,7	65,2	0,60

De taaltoetsen van de Niveautoets zijn in 2009 gedeeltelijk ververs. Bij Spelling zijn 9 van de 20 opgaven gelijk gebleven, bij Woordenschat zijn dat er 15 en bij Begrijpend lezen zijn 23 van de 30 gelijk gebleven. Bij Spelling is het gemiddelde prestatieniveau behoorlijk gestegen. Bij de opgaven die gelijk gebleven zijn, zien we echter geen verschil tussen de gemiddelde prestaties in 2008 en 2009. De verklaring voor de stijging van de prestaties is dat de alleen in 2008 gebruikte opgaven aanzienlijk moeilijker waren (lagere P-waarden) dan de opgaven die daar in 2009 voor in de plaats zijn gekomen.

De Niveautoetsen van 2009 waren identiek aan die van 2009. Hierdoor zijn de gemiddelde prestaties direct vergelijkbaar. Vanwege het kleine aantal leerlingen zijn vergaande conclusies echter niet mogelijk. Daarnaast – en wellicht nog belangrijker – is dat de groep leerlingen die de Niveautoets gemaakt heeft niet representatief is. De Niveautoets is immers vooral gemaakt door leerlingen die niet vaardig genoeg geacht werden om de reguliere Eindtoets te maken.

3.2 Eigenschappen van de gebruikte toetsen in jaargroep 4

Voor het vierde leerjaar zijn ook in 2009 bestaande toetsen uit het LVS gebruikt. Voor de vaardigheden Spelling en Begrijpend lezen hebben de leerlingen eerst een zogeheten startmodule gemaakt. Naar aanleiding van de score op die begintoeets kregen de leerlingen een vervolgmodule toegevoerd. Leerlingen met een lagere score op de startmodule maakten een relatief gemakkelijke vervolgmodule (Vervolg 1) en degenen met een hogere score een relatief moeilijke vervolgmodule (Vervolg 2). De vaardigheid in Spelling en Begrijpend lezen is met behulp van Item Response Theorie bepaald op basis van de geschaalde scores op de start- en vervolgmodule.

Zowel de toets Woordenschat als de toets Rekenen/Wiskunde bestaat uit twee delen die door alle leerlingen gemaakt zijn. Voor Woordenschat is de volledige toets gebruikt. Van de 54 opgaven van de gehele rekentoets hebben er 13 betrekking op Getallen en getalrelaties, 12 op Optellen en aftrekken, 14 op Vermenigvuldigen en delen en 15 op Meten, tijd en geld.

Met uitzondering van Woordenschat zijn in 2009 in jaargroep 4 dezelfde toetsen gebruikt als in 2008. Volgens de richtlijnen van de COTAN (Evers, Lucassen, Meijer & Sijtsma, 2009) zijn al deze toetsen voldoende betrouwbaar voor meting op zowel individueel als groepsniveau. Voor gedetailleerde gegevens over de eigenschappen van deze toetsen verwijzen we naar het verslag van het eerste jaarlijkse peilingsonderzoek (Hemker & Van Weerden, 2009). De enige toets die daarin niet beschreven is, is de nieuwe toets Woordenschat (Van Berkel en Alberts, 2009). In hoofdstuk 7 volgt een nadere beschrijving van deze toets. Voor meer informatie over deze toetsen en de wijze van afname wordt verwezen naar de handleidingen van de toetsen uit het LOVS voor jaargroep 4.

4 Analyse en rapportage: opgaven en groepseigenschappen op één schaal

In dit hoofdstuk worden de statistische analyse en de wijze van rapporteren beknopt beschreven. Eerst geven we een inleiding op het gebruikte IRT-model (paragraaf 4.1). Daarna leggen we uit hoe we de leerlingen in de jaargroepen 8 en 4 in vijf verschillende vaardigheidsniveaus hebben ingedeeld (paragraaf 4.2). Vervolgens beschrijven we hoe we het jaareffect gemodelleerd hebben, dit wil zeggen het verschil tussen de leerprestaties in 2008 en 2009 (paragraaf 4.3). Tot slot geven we aan hoe de lezer de gegevens in de tabellen en figuren kan interpreteren (paragraaf 4.4).

4.1 Toepassing van een IRT-model

Voor het schalen van de opgaven maken we in deze studie gebruik van een zogeheten conditional maximum likelihood (CML)-schattingsprocedure (Verhelst & Eggen, 1989) onder het *One-Parameter Logistic Model* (OPLM). Dit is uitgevoerd met het bijbehorende computerprogramma OPLM (Verhelst, Glas & Verstralen, 1995). Deze schattingsprocedure vereist dat de toetsen uit deels dezelfde opgaven bestaan. Hoe deze overlap gecreëerd is, is eerder besproken in hoofdstuk 2. Het aantal overlappende opgaven uit de kalibratieset van jaargroep 8 is weergegeven in tabel 4.1.

Merk op dat in jaargroep 4 alle opgaven overeenkomen aangezien in beide jaren dezelfde toetsen zijn gebruikt. Die worden hier niet apart weergegeven. Overigens zijn de toetsen Woordenschat wel een uitzondering op die regel, maar daarover meer in hoofdstuk 7.

Tabel 4.1 Aantallen opgaven met overlap

	<u>Overlap tussen 2008 en 2009</u>				<u>Overlap EA – EB</u>
	<u>EA</u>	<u>TE</u>	<u>Tot EB</u>	<u>NT</u>	<u>2009 en 2008</u>
Spelling	0	30	30	9	20
Woordenschat	8	25	32	15	12
Begrijpend lezen	14	15	29	23	16
Getallen en Bewerkingen	5	5	10	25	20
Breuken, Procenten en Verhoudingen	7	5	12	20	13
Meten, Meetkunde, Tijd en Geld	8	10	18	15	7

EA: overlap in de ankerboekjes tussen 2008 en 2009

TE: aantal opgaven op die schaal in Toetsboekje Extra (gelijk in 2008 en 2009)

Tot EB: totaal aantal opgaven dat gebruikt wordt om de overlap te verkrijgen van de twee eindtoetsen (=EA+TE)

NT: overlap in de ankerboekjes tussen 2008 en 2009

Overlap EA – EB: aantal opgaven dat de ankertoetsen en de Eindtoets gemeen hebben

Merk op dat de opgaven van het Toetsboekje Extra in tabel 4.1 ontbreken. De reden is dat de groep leerlingen die deze opgaven maakt waarschijnlijk niet representatief is voor gehele populatie. Voor de schatting van de populatiegegevens is representativiteit echter wel noodzakelijk. Zoals in hoofdstuk 2 is uitgelegd is voor het plaatsen van de opgaven op de schaal zelf een gebrek aan representativiteit echter geen probleem aangezien de parameterschattingen binnen het IRT-model populatieonafhankelijk zijn (Hambleton, Swaminathan & Rogers, 1991).

De overlap bij de afzonderlijke taaltoetsen is met circa 30 opgaven redelijk groot te noemen. Bij de reken-toetsen is de overlap bij de Eindtoets wat lager. Doordat de correlaties tussen de rekentoetsen alle zeer hoog zijn (latente correlaties tussen de .92 en .96) is het mogelijk om de rekenschalen gezamenlijk te

schalen, zodat de overlap veel groter wordt (40 opgaven). Na deze kalibratie van de gehele rekenschaal zijn de drie afzonderlijke schalen gemaakt op basis waarvan de leerprestaties in 2008 en 2009 beschreven zijn.

Voor de schattingen van eigenschappen van de populatie en het controleren of de itemparameters inderdaad populatie-onafhankelijk geschat zijn, is een steekproef van 56.433 leerlingen getrokken. Bij dit aantal werd het aantal van 32.000 observaties per opgave niet overschreden. Dit betrof 28.063 leerlingen uit 2008 en 28.380 leerlingen uit 2009. Dit was voor beide jaren 20,0% van de leerlingen uit de onderzoekspopulatie. De steekproef is getrokken uit de gehele verzameling van leerlingen, en niet op basis van eerst een trekking op schoolniveau. Het aantal scholen dat hierdoor in de steekproef terechtgekomen is was 5783 voor 2008 en 5710 voor 2009. Voor beide leerjaren was dat iets minder dan 99% van de beschikbare scholen. Deze steekproef is vergelijkbaar met de totale onderzoekspopulatie voor wat betreft de verdeling van de achtergrondvariabelen en voor wat betreft de verdeling van de leerlingen over de drie toetsen. Deze dataset is afzonderlijk geanalyseerd voor de zes verschillende vaardigheden, waarbij gekeken is of de eerder gevonden parameters een goede beschrijving van de data gaven.

Door het grote aantal observaties per opgave – meer dan 27.000 voor de EB-opgaven – was de statistische *power* zeer groot. Een statistische passing is in een dergelijk geval moeilijk te verkrijgen. We hebben daarom ook gekeken naar de grootte van de verschillen van de geschatte en de geobserveerde beschrijving van de opgaven. Dat is gedaan door de geschatte item response functie te vergelijken met punten op de geobserveerde item response functie (Verhelst, Glas & Verstralen, 1995). Dan valt op dat de gevonden verschillen mogelijk wel significant zijn, maar dusdanig klein dat zij vrijwel geen effect hebben in de beschrijving van de opgaven en de schattingen van de vaardigheden. In een enkel geval is een opgave uit de analyse verwijderd als de afwijking net iets te groot was. In de beschrijving van de vaardigheden wordt hier verder op ingegaan.

4.2 De vaardigheidsschaal: rapportage van de populatie

Met behulp van IRT is het mogelijk de gegevens van 2008 en 2009 op één vaardigheidsschaal te plaatsen. Per vaardigheid is er een afzonderlijke schaal. Dit houdt in dat er voor jaargroep 8 zes vaardigheidsschalen zijn (drie taal- en drie rekenschalen) en zeven voor jaargroep 4 (drie taal- en vier rekenschalen). Op deze schalen worden de populaties en de opgaven van 2008 en 2009 afgebeeld.

Voor bijna iedere vaardigheid zijn de opgaven in 2009 op dezelfde schaal gepresenteerd als in 2008. De uitzondering is Woordenschat in jaargroep 4, waarover later meer bij de presentatie van de resultaten voor jaargroep 4. Deze schaalwaarden zijn per vaardigheid zo gedefinieerd dat het gemiddelde van de populatie in 2008 de waarde 250 kreeg, en de standaarddeviatie gelijk was aan 50. Deze schaaltransformatie is gebruikelijk in PPON-onderzoek. Hiermee zijn de populatiewaarden van 2008 de referentiewaarden geworden. De populatie van 2009 is ook op deze schaal geplaatst. Als nu het gemiddelde van 2009 onder de 250 ligt, dan is de populatie als geheel minder vaardig geworden op deze vaardigheid in vergelijking met 2008, onder de aanname van een gelijke standaarddeviatie. Ligt het gemiddelde van 2009 daarboven dan zijn de leerlingen vaardiger geworden.

Het is niet alleen interessant om te vergelijken hoe de gemiddelden zich tot elkaar verhouden, maar het is ook nuttig om de vaardigheid van “typische leerlingen” te beschrijven: wat kan een zeer laag vaardige leerling nog net wel en wat is precies het beheersingsniveau van een zeer hoog vaardige leerling? In het verslag van 2008 is hierbij gebruikgemaakt van een relatieve schaal. De minst vaardige leerling was een leerling die zich op percentiepunt 10 van de vaardigheidsverdeling bevond. Dit werd in het verslag van 2008 een percentiel-10 leerling (ZLV-leerling, zeer laag vaardig) genoemd, omdat dit het niveau was waarbij 10% van de leerlingen *in 2008* een gelijke of lage vaardigheid had en 90% van de leerlingen een hogere vaardigheid. Zo zijn ook de LV (laag vaardige)-, ST (standaard)-, HV (hoog vaardige)- en de ZHV (zeer hoog vaardige)-leerlingen gedefinieerd, waarbij de ZHV-leerling de meest vaardige leerling was. Voor

deze ZHV-leerling geldt dat 90% van de leerlingen *in 2008* een vaardigheid gelijk aan of lager dan deze leerling had (dus had slechts 10% van de leerlingen een hogere vaardigheid dan de ZHV-leerling).

De beheersingsniveaus van deze vijf ideaaltypische leerlingen zijn afgebeeld op de vaardigheidsschaal. Een ZLV-leerling heeft op deze schaal een vaardigheid van 186, een LV-leerling een vaardigheid van 216, een ST-leerling een vaardigheid van 250, een HV-leerling een vaardigheid van 284 en een ZHV-leerling een vaardigheid van 314. Hiermee liggen de vaardigheden van de typische leerlingen vast. Echter, het gebruik van de termen ZLV tot en met ZHV om deze typische leerlingen te beschrijven is lastig als meer populaties op de vaardigheidsschaal afgebeeld moeten worden (dit wil zeggen: de populaties van 2008 en 2009). Als de vaardigheid toegenomen is over de jaren dan zullen er minder dan 10% van de leerlingen tot de groep ZLV-leerlingen behoren, wat gezien de gebruikelijke definitie van een ZLV-leerling niet goed kan. Een ZLV-leerling is dan ook de ZLV-leerling *van 2008*. Om nu een dergelijke verwarring te voorkomen en niet telkens het referentiejaar 2008 toe te hoeven voegen, worden nu de volgende definities van ideaaltypische leerlingen gebruikt:

- **ZLV-leerling** **Zeër laag vaardige leerling**: Een leerling met een vaardigheid van 186 op de vaardigheidsschaal – komt overeen met een percentiel-10 (P10) leerling in 2008.
- **LV-leerling** **Laag vaardige leerling**: Een leerling met een vaardigheid van 216 op de vaardigheidsschaal – komt overeen met een percentiel-25 (P25) leerling in 2008.
- **ST-Leerling** **Standaard 50 leerling**: Een leerling met een vaardigheid van 250 op de vaardigheidsschaal – komt overeen met een percentiel-50 (P50) leerling in 2008.
- **HV-leerling** **Hoog vaardige leerling**: Een leerling met een vaardigheid van 284 op de vaardigheidsschaal – komt overeen met een percentiel-75 (P75) leerling in 2008.
- **ZHV-leerling** **Zeër hoog vaardige leerling**: Een leerling met een vaardigheid van 314 op de vaardigheidsschaal – komt overeen met een percentiel-90 (P90) leerling in 2008.

Bij de typering en naamgeving van de niveaus van de vijf typische leerlingen zijn drie opmerkingen op zijn plaats. Ten eerste merken we op dat we geen gebruikmaken van de termen “minimumniveau”, “fundamenteel niveau”, “basisniveau”, “streefniveau” en “expertniveau”. Deze termen hebben een specifieke betekenis in andere vormen van onderzoek, bijvoorbeeld naar de referentieniveaus (vergelijk de publicaties van Expertgroep Doorgaande Leerlijnen Rekenen en Taal, 2008-2009). Om verwarring te voorkomen zijn deze termen hier niet gebruikt.

Ten tweede merken we op dat de definitie van de vijf ideaaltypische leerlingen gerelateerd is aan de geobserveerde vaardigheidsverdeling in 2008 en niet aan een inhoudelijke of didactische interpretatie van de vaardigheidsschaal. Een inhoudelijke interpretatie kan beter plaatsvinden aan de hand van de opgaven die op die schaal geplaatst zijn. In de rapportage van de resultaten worden de vijf beheersingsniveaus daarom geïllustreerd aan de hand van een inhoudelijke beschrijving van de opgaven die de leerling op het desbetreffende niveau al dan niet beheerst.

Ten derde valt op te merken dat het vaardigheidsniveau van een leerling van schaal tot schaal kan verschillen. Zo kan een leerling voor de ene vaardigheid tot de groep ZLV-leerlingen behoren en voor de andere vaardigheid tot de groep ZHV-leerlingen. Dat zelfde geldt uiteraard ook voor de overige vier omschrijvingen van “typische leerlingen”.

4.3 Analyse en rapportage van prestatieverschillen tussen 2008 en 2009

Om de vaardigheden van 2008 en 2009 direct met elkaar te kunnen vergelijken, moet er aan ten minste twee voorwaarden zijn voldaan:

- A De vaardigheden moeten in 2009 op dezelfde schaal gemeten zijn als in 2009.
- B De samenstelling van de responsgroep mag niet gewijzigd zijn (bijvoorbeeld meer of minder zittenblijvers bevatten).

Aan voorwaarde A is voldaan als de steekproef van 2009 precies dezelfde toetsen heeft gemaakt als de steekproef van 2008. Voor jaargroep 4 is aan deze voorwaarde voldaan, aangezien de leerlingen in beide gevallen ongewijzigde toetsen uit het reguliere LOVS hebben gemaakt. De voor jaargroep 8 gebruikte Eindtoets Basisonderwijs wordt echter omwille van de geheimhouding ieder jaar volledig vernieuwd. Zoals eerder beschreven, hebben wij de vergelijkbaarheid gewaarborgd door de achtstegroepers in 2008 en 2009 een aantal extra opgaven voor te leggen. Met behulp van een item response model konden de prestaties van 2009 toch met die van 2008 op dezelfde meetschaal met elkaar worden vergeleken.

Aan voorwaarde B is voldaan als de samenstelling van responsgroep in 2009 vergelijkbaar is met die in 2008. Veranderingen in de samenstelling van een responsgroep kunnen een gevolg zijn van zogenaamde steekproeffluctuaties en van 'echte' verandering in de samenstelling van de populatie. Het probleem van steekproeffluctuaties zal zich voor jaargroep 8 niet zo gauw voordoen. Zowel in 2008 als in 2009 is namelijk een zeer grote aselechte steekproef van vele honderdduizenden leerlingen getrokken (telkens ongeveer 80% van de populatie). Wel kan er zich een wijziging in de samenstelling van de populatie hebben voorgedaan. Stel dat scholen bijvoorbeeld overgaan op een strenger doubleerbeleid. De responsgroep in 2009 zou dan minder zittenblijvers bevatten dan die in 2008. We zouden dan ten onrechte kunnen concluderen dat de vaardigheid van de leerlingen vooruit is gegaan. Gelukkig zijn er tegenwoordig statistische technieken beschikbaar die ons voor dit soort verkeerde conclusies kunnen behoeden. Vandaar dat wij zowel ongecorrigeerde als gecorrigeerde gegevens rapporteren (d.w.z. gecorrigeerd voor veranderingen in de samenstelling van de populatie).

Bij jaargroep 8 is het aantal leerlingen in de steekproef zo groot dat verschillen in de verdeling van de achtergrondvariabelen tussen 2008 en 2009 waarschijnlijk 'echte' verschuivingen in de populaties vertegenwoordigen. In jaargroep 4 zijn de steekproeven echter te klein om een onderscheid tussen steekproeffluctuaties en 'echte' verschuivingen te kunnen maken. Als daar verschuivingen in de verhoudingen van achtergrondvariabelen gevonden worden, dan zijn dit veel eerder steekproeffluctuaties, en die zullen dan ook zo geïnterpreteerd moeten worden. Dat houdt in dat correctie van de steekproef op basis van de verdeling van de achtergrondvariabelen geen optie is maar een noodzaak.

Voor de correctie voor eventuele verschillen in de samenstelling van de beide responsgroepen zijn twee typen statistische correctiemodellen beschikbaar: het hoofdeffectenmodel en het interactiemodel. De belangrijkste achtergrondkenmerken waarvoor gecorrigeerd wordt, zijn in beide modellen geslacht, leertijd, stratum en formatiegewicht. De keuze van het model is afhankelijk van de vraag of er behalve van een hoofdeffect ook sprake is van een interactie tussen achtergrondkenmerken en het jaar van afname. We lichten dit toe aan de hand van een voorbeeld. We spreken van een hoofdeffect van geslacht als het vaardigheidsverschil tussen jongens en meisjes in beide jaren gelijk is. Een interactie tussen geslacht en jaareffect wil zeggen dat prestatieverschil tussen jongens en meisjes het ene jaar groter of kleiner is dan het andere jaar. In het hoofdeffectenmodel wordt alleen gecorrigeerd voor eventuele hoofdeffecten van achtergrondkenmerken zoals geslacht, leertijd, stratum en formatiegewicht. Als er sprake is van een interactie van een of meer van deze achtergrondvariabelen met het afnamejaar, is gekozen voor correctie volgens het interactiemodel en rapporteren we de verandering in het prestatieniveau per categorie van de achtergrondvariabele.

Significantie en effectgrootte

Of een gemiddeld vaardigheidsverschil tussen 2009 en 2008 statistisch significant is, hangt in belangrijke mate af van de steekproefgrootte. Hoe groter de steekproef, hoe eerder een gevonden verschil statistisch significant is. Voor jaargroep 8 is de steekproef zoals gezegd zeer veel groter dan voor jaargroep 4. Om de resultaten in de jaargroepen 4 en 8 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 2008 en 2009 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 4.2.

Tabel 4.2 Kwalificatie van effectgrootten

Effectgrootte	Kwalificatie
0,0	geen effect
0,2 of -0,2	klein effect
0,5 of -0,5	matig effect
0,8 of -0,8	groot effect

In de rapportage van de resultaten wordt naast het gevonden effect ook direct de kwalificatie van het effect gegeven. Hierbij worden kwalificaties zoals gegeven in tabel 4.2 als ondergrenzen genomen. Dat betekent dat een effect van 0,4 als klein wordt gekwalificeerd omdat het de grens van 0,5 nog niet heeft gepasseerd. De effecten tussen -0,2 en 0,2 kunnen als dus als zodanig klein beschouwd worden dat zij geen relevantie hebben. Er worden geen aparte significantieniveaus aangegeven. Voor alle kleine effecten geldt dat zij bij overschrijdingskans $p < .01$ significant zijn. Voor de effecten die als matig of groot gekwalificeerd kunnen worden, geldt dat zij alle bij een overschrijdingskans van $p < .000001$ significant zijn.

De gerapporteerde effecten betreffen telkens het gecorrigeerde verschil tussen twee niveaus van een variabele. De richting van het effect wordt bepaald door de volgorde van het contrast. Zo zijn de volgende twee effecten exact aan elkaar gelijk:

Contrast	Effectgrootte
Meisjes – jongens	0,2
Jongens – meisjes	-0,2

In beide gevallen betekent het dat er een klein effect is waarbij meisjes beter presteren dan jongens. De gerapporteerde effecten van de hoofdvariabelen betreffen telkens de resultaten van het eerste model met alleen die variabelen.

Het modelleren van de jaarvergelijkingen voor jaargroep 8

Na in het voorgaande globaal in te zijn gegaan op de vergelijking tussen jaargroep 8 en 4, geven we hier een meer gedetailleerde beschrijving. In de statistische analyse is nagegaan welke invloed de achtergrondkenmerken hebben op de leerprestaties van de leerlingen in 2008 en 2009. Met behulp van multiple regressie analyse is voor elk achtergrondkenmerk nagegaan wat het gezuiverde, of gecorrigeerde effect is. Dit wil zeggen de unieke bijdrage van het achtergrondkenmerk aan de leerprestaties waarbij de invloed van de andere achtergrondkenmerken, voor zover wij die gemeten hebben, constant zijn gehouden. Voor jaargroep 8 is een aantal verschillende statistische modellen geanalyseerd. Omdat het afnamejaar de belangrijkste variabele is, speelt deze variabele in alle modellen een belangrijke rol. De drie belangrijkste typen modellen zijn:

- 1 het ongecorrigeerd jaareffect;
- 2 het jaareffect gecorrigeerd voor de hoofdvariabelen geslacht, leertijd, stratum en formatiegewicht;
- 3 het jaareffect gecorrigeerd voor de hoofdvariabelen en daarnaast telkens een van de aanvullende achtergrondkenmerken type toets, advies VO, regio, urbanisatiegraad en thuistaal.

Binnen model 2 en 3 zijn nog twee submodellen geanalyseerd, te weten het eerder besproken hoofd-effectenmodel en het interactiemodel (respectievelijk model 2A, 2B, 3A en 3B). Hieronder lichten we de hoofd- en submodellen kort toe.

Model 1

In het eerste model is afnamejaar de enige onafhankelijke variabele. Het regressiegewicht van afnamejaar representeert het ongecorrigeerde jaareffect, dit wil zeggen het verschil tussen de gemiddelde prestaties in 2009 en 2008. Dit effect is het feitelijke jaareffect als we ervan uitgaan dat de samenstelling van de populatie gelijk is gebleven. Dit ongezuiverde jaareffect is ook belangrijk omdat een verandering in de

samenstelling van de populatie ook een echte verbetering in vaardigheid kan representeren. Een voorbeeld hiervan wordt hieronder gegeven.

Leerlingen die zijn blijven zitten hebben over het algemeen lagere schoolse vaardigheden dan leerlingen in dezelfde klas die nooit zijn blijven zitten. Stel dat binnen de groep zittenblijvers de vaardigheid over de verschillende jaren niet wijzigt, en dat dit ook het geval is bij de niet-zittenblijvers. Als er gecorrigeerd wordt voor de variabele leertijd zal dat leiden tot de conclusie dat er geen vooruitgang is in vaardigheid. Als echter blijkt dat er in het tweede peilingsjaar minder zittenblijvers zitten, dan zal de groep als geheel wel beter zijn geworden. Juist doordat de leerlingen vaardiger worden, zijn er dan minder zittenblijvers. Als je daar dan voor corrigeert, merk je niet op dat er wel degelijk vooruitgang in vaardigheid is.

Het is bijvoorbeeld zelfs mogelijk dat men gecorrigeerd een daling rapporteert terwijl een sprake is van een vooruitgang. Als rekenvoorbeeld is de vaardigheid voor vijf verschillende even grote groepen gegeven in tabel 4.3. In het eerste jaar (jaar 1), stromen de leerlingen tot en met percentiel 60 door naar het vmbo, terwijl de 40% daarboven naar havo/vwo gaat. De gemiddelde vaardigheid binnen vmbo is dan 230 en binnen havo/vwo 280. Als in het jaar daarna bij alle vijf percentielgroepen de vaardigheid 5 punten omhoog zou gaan, en daarmee de middelste percentielgroep nu wel het naar havo/vwo wordt verwezen, dan zal het gemiddelde van zowel vmbo als van vwo dalen. Als er nu gecorrigeerd zou worden voor doorstroom, dan zou er een achteruitgang van 5 vaardigheidspunten gerapporteerd worden.

Tabel 4.3 Voorbeeld van verschil in uitkomst door correctie met achtergrondvariabele

	Percentiel	1-20.	21-40.	41-60.	61-80.	81-100.	Gemiddelde
Vaardigheid in 2008		210	230	250	270	290	250
Vaardigheid in 2009		215	235	255	275	295	255
Vmbo in jaar 1	60%	230		40%	280		vwo in jaar 1
Vmbo in jaar 2	40%	225		60%	275		vwo in jaar 2

Een ander voorbeeld is dat men besluit om leerlingen pas bij een grotere achterstand te laten doubleren. Dat betekent dat er relatief meer minder vaardige leerlingen gewoon doorstromen, en de niet doorstromende leerlingen nog minder vaardig zijn. In dat geval daalt de gemiddelde vaardigheid dus zowel bij de zittenblijvers als bij de niet-zittenblijvers. Gecorrigeerd voor leertijd zou de vaardigheid dan dalen in de tijd. Kijkt men naar het ongecorrigeerde resultaat dan is het heel goed mogelijk dat de vaardigheid juist wel vooruit gaat.

De belangrijke vraag is natuurlijk of er nu vooruitgang is of juist niet. Als de verhouding van de verschillende groepen over de tijd verandert, hoeft het ongecorrigeerde effect niet hetzelfde te zijn als het gecorrigeerde effect. Wat dan het juiste effect zal zijn, kan onderwerp van discussie zijn, waarbij verschillende argumenten aangevoerd kunnen worden. In dit rapport zullen voor jaargroep 8 daarom zowel de gecorrigeerde als de ongecorrigeerde jaareffecten gerapporteerd worden.

Model 2A en 2B

In het tweede model zijn naast het afnamejaar ook de achtergrondkenmerken geslacht, leertijd, stratum en formatiegewicht als onafhankelijke variabelen opgenomen. Het effect van ieder van deze variabelen is geschat gecorrigeerd voor de overige variabelen. Voor iedere schaal zijn van het tweede model twee submodellen geanalyseerd: het hoofdeffectenmodel (model 2A) en het interactiemodel (model 2B). De keuze van het submodel is afhankelijk van de vraag of er behalve van een hoofdeffect ook sprake is van een interactie tussen het jaar van afname en een of meer van de achtergrondkenmerken. We lichten dit toe aan de hand van een voorbeeld. We spreken van een hoofdeffect van geslacht als het vaardigheidsverschil tussen jongens en meisjes in beide jaren gelijk is. Een interactie tussen afnamejaar en geslacht wil zeggen dat prestatieverschil tussen jongens en meisjes het ene jaar groter of kleiner is dan het andere jaar. In het

hoofdeffectenmodel wordt alleen gecorrigeerd voor eventuele hoofdeffecten van geslacht, leertijd, stratum en formatiegewicht. Als er sprake is van een interactie van een of meer van deze achtergrondvariabelen met het afnamejaar, is gekozen voor correctie volgens het interactiemodel. We rapporteren de verandering in het prestatieniveau dan per categorie van de achtergrondvariabele. Bij de taalvaardigheden was er sprake van een interactie tussen afnamejaar en formatiegewicht. Hoe deze interactie inhoudelijk te interpreteren is, wordt verder besproken bij de resultaten voor taalvaardigheid in hoofdstuk 5. Bij de rekenschalen was er geen sprake van interactie en is de correctie volgens het hoofdeffectenmodel uitgevoerd.

Model 3A en 3B

Het derde model bouwt voort op het tweede model. Aan de vijf onafhankelijke variabelen in het tweede model – afnamejaar, geslacht, leertijd, stratum en formatiegewicht – is vijf keer een zesde achtergrondkenmerk toegevoegd (met teruglegging). De vijf aanvullende achtergrondkenmerken zijn in jaargroep 8: gemaakte toets, advies VO, regio, urbanisatiegraad en thuistaal. Bij taalvaardigheid bouwen de vijf derde modellen voort op het tweede model met een interactie tussen jaar en formatiegewicht. Bij rekenvaardigheid kon zoals gezegd bij elk van de drie vaardigheden worden voortgebouwd op het model met alleen hoofdeffecten. Voor iedere schaal zijn van het derde model twee submodellen geanalyseerd: het hoofdeffectenmodel (model 3A) en het interactiemodel (model 3B) met interactietermen tussen afnamejaar en het desbetreffende aanvullende achtergrondkenmerk.

4.4 Het interpreteren van de figuren en tabellen

In de statistische analyse zijn de opgaven en de onderscheiden subgroepen van leerlingen op één schaal geplaatst. Zoals hiervoor is uiteengezet, zijn de opgaven van 2009 nu aan deze schaal toegevoegd. Omdat de schaal gebaseerd is op de verdeling van 2008 is de positie van de opgaven in 2009 gelijk aan die in 2008 (onder de aanname van een passend model). De relatieve moeilijkheid van de opgaven ten opzichte van elkaar is, met andere woorden, gelijk gebleven.

De positie van een opgave op de vaardigheidsschaal wordt gegeven door twee waarden, namelijk de positie op de vaardigheidsschaal die een kans geeft van 0,50 om de opgave goed te maken (het P50-punt van de opgave) en de positie op de vaardigheidsschaal die een kans geeft van 0,80 om de opgave goed te maken (het P80-punt van de opgave). Uiteraard ligt bij een enkele opgave het P80-punt hoger op de vaardigheidsschaal dan het P50-punt van die opgave. De leerling moet immers een hogere vaardigheid hebben om een grotere kans te hebben de opgave goed te maken.

Bij opgaven die een sterk onderscheid maken tussen vaardige en minder vaardige leerlingen liggen het P50- en het P80-punt dicht bij elkaar. Bij die opgaven levert een klein verschil in vaardigheid al een groot verschil in kansen op succes op. Bij minder goed onderscheidende opgaven is de afstand juist relatief groot. De leerling moet sterk in vaardigheid stijgen wil deze een duidelijk grotere kans hebben om die opgave goed te maken.

Het is hiermee ook duidelijk dat de kans om een opgave goed te maken voor leerlingen met verschillende vaardigheidsniveaus verschillend is. Naarmate een leerling een onderwerp beter beheerst, heeft hij of zij een grotere kans om een opgave over dat onderwerp goed te beantwoorden. Dit betekent ook dat de vijf verschillende typische leerlingen verschillende kansen hebben om de opgaven goed te maken. Hiermee kan de relatie tussen de opgaven en de typische leerlingen gegeven worden in termen van “moeilijk voor de leerling”, “uitdagend voor een leerling” of “gemakkelijk voor een leerling”:

- Moeilijke opgave. Een opgave is moeilijk voor een leerling als deze leerling met zijn of haar vaardigheidsniveau een kans lager dan 0,50 heeft om de opgave goed te maken ($< 0,50$)
- Uitdagende opgave. Een opgave is uitdagend voor een leerling als deze leerling met zijn of haar vaardigheidsniveau een kans van minstens 0,50 maar ten hoogste 0,80 heeft om de opgave goed te maken.

- Makkelijke opgave. Een opgave is gemakkelijk voor een leerling als deze leerling met zijn of haar vaardigheidsniveau een kans hoger dan 0,80 heeft om de opgave goed te maken.

Deze definities zijn gegeven om opgaven te classificeren. Merk op dat de classificatie van een opgave afhangt van het niveau van een leerling. Een opgave kan voor een ZLV- leerling immers moeilijk zijn, terwijl diezelfde opgave voor een ZHV-leerling juist gemakkelijk is.

Het interpreteren van de figuren

In dit rapport zijn de prestaties van de typische leerlingen en de beide afnamejaren in een figuur afgebeeld. De vaardigheidsniveaus van de verschillende groepen leerlingen zijn daarbij geïllustreerd via voorbeelden van opgaven die deze leerlingen al dan niet beheersen. Een fictief voorbeeld is gegeven in figuur 4.1.

Het interpreteren van de informatie in deze figuren behoeft enige toelichting. In figuur 4.1 is de vaardigheidsschaal gegeven door de verticale as. Deze vaardigheidsschaal loopt hier van 100 tot 400 (zie de rechter verticale as). In de afbeelding zijn een aantal horizontale lijnen afgebeeld. De lichte horizontale lijnen geven vaardigheidsscores aan, oplopend met een waarde van 50. Zie hiervoor ook de rechterzijde van het diagram (vaardigheidsscore). De vaardigheidsschaal wordt steeds afgebeeld tussen de vaardigheidsscores 100 en 400. De bredere (paarse) horizontale lijnen geven de vijf onderscheiden typische leerlingen: zeer laag vaardig (ZLV), laag vaardig (LV), standaard (ST), hoog vaardig (HV) en zeer hoog vaardig (ZHV).

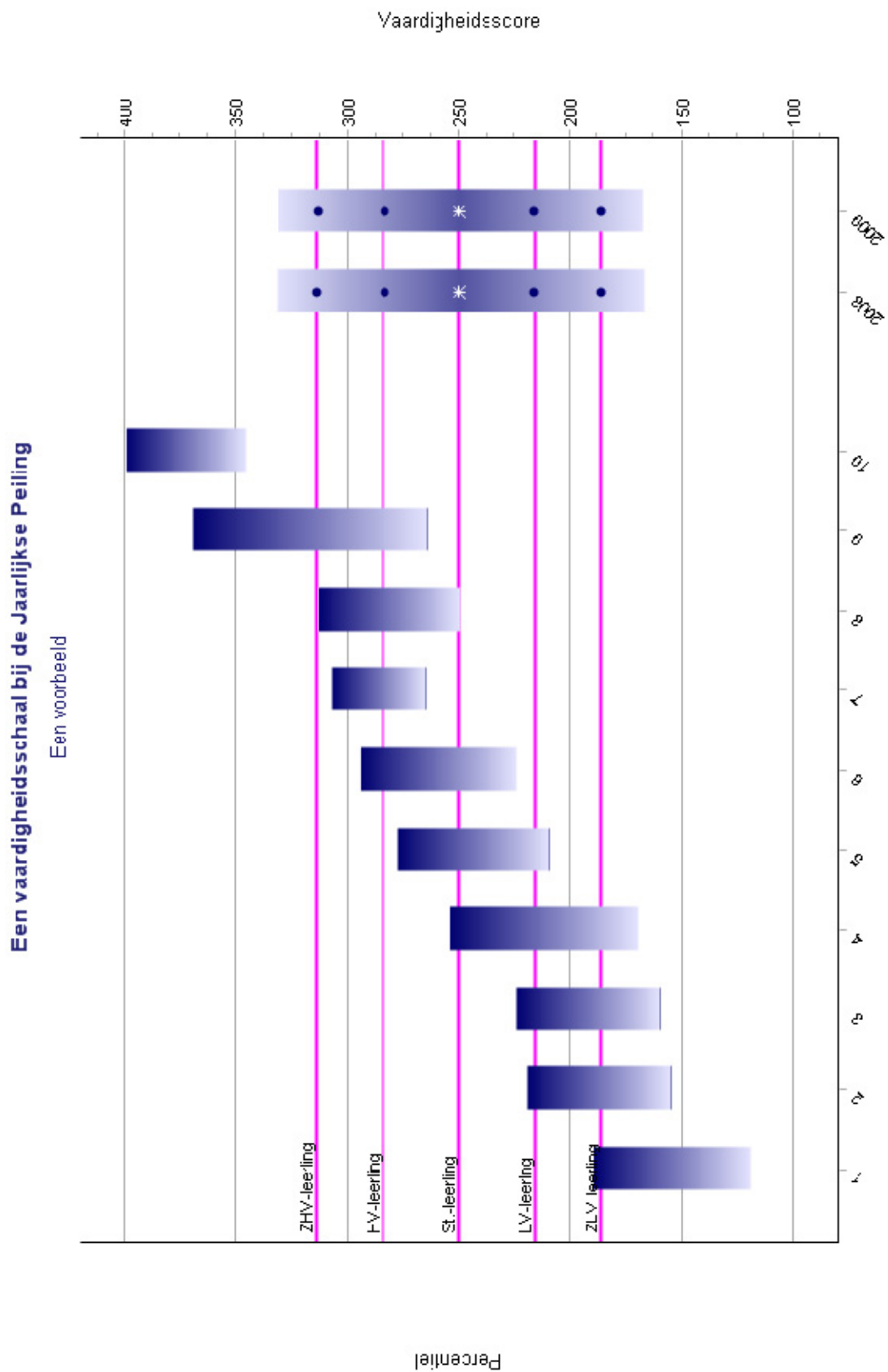
De opgaven en de populaties zijn weergegeven met verticale balken. De balken zonder stippen die beginnen aan de linkerkant van de figuur verbeelden de opgaven. Deze kolommen zijn onder aan het diagram genummerd en verkleuren van boven naar beneden van licht naar donker. Het verticale balkje begint op het punt op de vaardigheidsverdeling waarbij de kans om die opgave goed te maken een half is, oftewel 50%. Dat is het onderste (lichtste) punt van bij iedere opgavenkolom. Leerlingen op dit vaardigheidsniveau zullen gemiddeld vijf van de tien van precies dit type opgaven goed maken. Naarmate een opgave moeilijker is, zal dat beginpunt steeds hoger op de schaal komen te liggen. Het balkje eindigt op het punt dat de kans op het correcte antwoord 0,8 bedraagt. Dat wil dus zeggen dat leerlingen op dit vaardigheidsniveau gemiddeld acht van de tien opgaven van precies dit type goed zullen maken. Het kleurverloop in het balkje, van lichter naar donkerder, symboliseert de toename in de kans om de opgave goed te maken. Uiteraard ligt het vaardigheidsniveau om een kans van 0,8 te hebben om de opgave correct te maken altijd boven dat om een kans van 0,5 te hebben. Aan de hand van het balkje onderscheiden we drie niveaus in de beheersing van een opgave:

- We spreken van **goede** beheersing wanneer de kans op een goed antwoord groter is dan 0,8. De leerling heeft dan een vaardigheidsscore die hoger ligt dan het balkje aangeeft.
- Wanneer de kans op een goed antwoord tussen 0,5 en 0,8 ligt, spreken we van een **matige** beheersing. Dit gebied op de vaardigheidsschaal komt dus overeen met wat het balkje weergeeft.
- We spreken van **onvoldoende** beheersing van een opgave wanneer de kans op een goed antwoord kleiner is dan 0,5. De vaardigheidsscore van de leerling ligt dan onder het beginpunt van het balkje.

Merk op dat definities erg lijken op de classificaties van de opgaven. Nu wordt echter niet de opgaven geclassificeerd, maar het beheersingsniveau. In principe komt het neer op hetzelfde, maar verschilt de focus. In de classificatie naar moeilijkheid ligt de focus op de opgave, en bij de classificatie van de beheersing ligt de focus op de leerling. Het zijn twee verschillende wijzen om naar dezelfde relatie tussen opgave en leerling te kijken. Uiteraard is het ook mogelijk de moeilijkheid van de opgaven uit de figuur af te leiden.

Figuur 4.1

Een voorbeeld van een vaardigheidsschaal bij JPON



Ter verdere illustratie nemen we opgave 8 uit figuur 4.1. Leerlingen met vaardigheidsscore 250 hebben een kans van 0,5 om die opgave goed te maken. Leerlingen met een lagere vaardigheidsscore beheersen opgave 8 dus onvoldoende. Die opgave is voor hen moeilijk te noemen. Als we nu naar de lijnen van de typische leerlingen kijken, dan zien we dat 50% van de leerlingen een vaardigheidsscore heeft die lager is dan 250. Daaruit kunnen we concluderen dat 50% van de leerlingen deze opgave onvoldoende beheerst. Pas vanaf de lijn voor de ZHV-leerling wordt deze opgave goed beheerst: vanaf het vaardigheidsniveau 314 wordt de kans om de opgave goed te beantwoorden groter dan 0,80. Opgemerkt wordt dat een matige beheersing van een opgaven niet automatisch op een probleem wijst. Dat is een inhoudelijke vraag die alleen beantwoord kan worden door naar de inhoud van de opgaven te bestuderen.

Een voorbeeld van een makkelijker opgave is opgave 2. Die wordt goed beheerst door LV-leerlingen. Leerlingen op ZLV-niveau hebben een matige beheersing van deze opgave. Alleen leerlingen met een extreem lage vaardigheid van onder de 155 beheersen deze opgave onvoldoende

Daar waar in 2008 gekozen is voor het weergeven van voorbeeldopgaven, worden in dit rapport alle opgaven van de Eindtoets die op de schaal passen weergegeven. Als een opgave niet op de schaal past, wordt deze niet weergegeven. Ook de opgaven uit de Ankertoetsen en de Niveautoetsen zijn niet afgebeeld. De opgaven uit de Ankertoets zijn niet gepresenteerd omdat deze opgaven niet openbaar gemaakt mogen worden. De opgaven uit de NT zijn niet afgebeeld omdat computeropgaven moeilijk op papier zijn weer te geven. De opgaven van de Anker- en Niveautoetsen zijn alle aan de makkelijke kant en zijn in de rapportage gerepresenteerd door gemakkelijke opgaven uit de EB.

Naast de opgaven zijn ook de populaties van 2008 en 2009 in de figuur weergegeven. De kolommen aan de rechterzijde representeren de beide onderzoeksjaren. Deze kolommen hebben onder aan het diagram een label voor het onderzoeksjaar. Met de asterisk in het midden van de kolom en de beide puntjes daarboven en daaronder zijn de gemiddelde prestaties van de vijf typische leerlingen aangeduid. De gemiddelde vaardigheidsscore van een groep (percentiel 50 van die groep) is met de witte asterisk aangeduid. De punten geven, van boven naar beneden, de percentiepunten 10, 25, 75 en 90 aan binnen het jaar. De onderkant van de kolom wordt gegeven bij percentiel 5, en de bovenkant bij percentiel 95 van het jaar. De verdeling van 2008 lag al vast, zoals eerder in dit hoofdstuk is aangegeven. De verdeling van 2009 is nu ook op deze schaal gelegd. Deze vaardigheidsverdelingen zijn overigens niet gecorrigeerd voor andere factoren die mogelijkkerwijs van invloed zijn op de resultaten.

Het interpreteren van de tabellen

In de rapportage van 2008 konden de marginale populatieverdelingen van de verschillende afnamejaren niet worden weergegeven (er was toen ook maar een jaar). Wel zijn de populatieverdelingen van een viertal achtergrondvariabelen gerapporteerd, te weten stratum (drie categorieën), geslacht (twee categorieën), leertijd (twee categorieën) en formatiegewicht (drie categorieën). Die ontbreken nu in de figuren. In plaats daarvan zijn alle opgaven die gebruikt zijn in 2009 afgebeeld in plaats van slechts een deel van de opgaven. De verdelingen van de verschillende categorieën van de achtergrondvariabelen zijn nog wel terug te vinden in de tabellen.

In de tabellen zijn ook het gemiddelde en de standaarddeviatie per categorie van de achtergrondkenmerken weergegeven. Als voorbeeld van een dergelijke tabel bespreken we tabel 5.1.3. Het gaat daarbij om de volgende aanvullende achtergrondkenmerken en categorieën: gemaakte toets (2 categorieën: EB en NT), advies VO (de vijf hoofdcategorieën: BB, KB, GT, havo en vwo), regio (vier categorieën: Noord, Oost, West en Zuid), urbanisatiegraad (vijf categorieën: zeer sterk, sterk, matig, weinig en niet stedelijk) en thuistaal (drie hoofdcategorieën: alleen Nederlands, Nederlands plus een andere taal en alleen een andere taal, en de vier extra categorieën).

Op basis van de tabellen kunnen vergelijkbare conclusies getrokken worden als op basis van de figuren. De tabellen hebben als voordeel dat de verschillen in vaardigheidsniveaus tussen groepen leerlingen inhoudelijke betekenis kunnen krijgen aan de hand van de voorbeeldopgaven. Zo kan bijvoorbeeld nagegaan worden welk type opgaven zeer gemakkelijk is voor de gemiddelde leerling op het vwo. Als de

positie van de P80-waarde van een opgave op de schaal ver onder het gemiddelde niveau van een leerling op het vwo zit, dan is die opgave voor die groep zeer gemakkelijk. Dus als de P80-positie van een opgave op 200 ligt en de gemiddelde positie van een groep ligt op 300, dan is die opgave voor die groep zeer gemakkelijk. Omgekeerd is een opgave waarbij de P50-waarde ver boven het gemiddelde van een groep ligt zeer moeilijk voor die groep.

In het verslag van het onderzoek in 2008 werden de vaardigheden ook gerelateerd aan de percentielschaal, maar aangezien we nu niet één populatie (2008) hebben maar twee (ook 2009) is een dergelijke representatie eerder verwarrend dan verhelderend. Deze is dan nu ook weggelaten. Ter referentie worden wel bij iedere tabel de waarden weergegeven die horen bij de vijf onderscheiden typische leerlingen.

Ten slotte merken we op dat niet alle effecten in de tabellen zijn weergegeven. Zo is het effect van EB-Anker nergens gegeven. Dat effect ontbreekt omdat de resultaten van de EB en het EB-Anker nooit significant verschilden en het effect van het EB-Anker en de NT vergelijkbaar was met dat van de EB en de NT. Bij de variabele advies VO zijn alleen de effecten van de opvolgende hoofdcategorieën weergegeven. Wil men bijvoorbeeld het effect kennen van een grotere stap, bijvoorbeeld tussen BB en vwo, dan is dat effect de som van alle tussenliggende effecten. Ook bij urbanisatiegraad zijn alleen de effecten gegeven van de opvolgende categorieën van deze variabele. Wanneer bij de variabele thuistaal de effecten gegeven worden per specifieke thuistaal, dan wordt die alleen gegeven in vergelijking met de leerlingen die alleen Nederlands spreken. De effecten tussen de andere talen onderling zijn daaruit af te leiden.

5 Resultaten voor taalvaardigheid in jaargroep 8

In dit hoofdstuk worden de prestaties van de leerlingen in jaargroep 8 beschreven voor de drie taalvaardigheden Woordenschat, Spelling en Begrijpend lezen. We rapporteren de uitkomsten alleen per vaardigheid. In de discussie (hoofdstuk 8) wordt een samenvatting van de resultaten gegeven. Deze samenvatting vindt voor jaargroep 8 en jaargroep 4 plaats per achtergrondvariabele in plaats van per vaardigheid.

De resultaten bij de taalvaardigheden in jaargroep 8 worden telkens op dezelfde wijze weergegeven. Eerst wordt de gemeten vaardigheid in algemene termen beschreven. Daarna laten we zien met welke opgaven de vaardigheid gemeten is. De posities van de opgaven op de vaardigheidsschaal zijn gepresenteerd in Bijlage 2.

De posities van de opgaven van de Eindtoets 2009 op de punten P50 en P80 van de vaardigheidsschaal worden in dit hoofdstuk gepresenteerd in de volgorde waarin de opgaven aan de leerlingen zijn voorgelegd. Voor de vijf typische leerlingen wordt per opgave aangegeven of een opgave moeilijk, uitdagend of gemakkelijk is. In de kaders hieronder zijn de definities van de typische leerlingen en de definities van de moeilijkheidsgraad gegeven. Eerder is in hoofdstuk 4 een nadere toelichting op deze beschrijving gegeven.

Code	Label typische leerling	Definitie (positie op vaardigheidsschaal)
ZLV	Zeer laag vaardige leerling	186
LV	Laag vaardige leerling	214
ST	Standaard leerling	250
HV	Hoog vaardige leerling	286
ZHV	Zeer hoog vaardige leerling	314

De interpretatie van schaalwaarden is soms lastig, maar in het geval van IQ-scores zijn we dit eigenlijk al gewend. Meer hierover in Bijlage 4 waar de resultaten voor jaargroep 8 gegeven worden alsof het IQ-scores zouden zijn.

Code	Label moeilijkheidsgraad	Definitie (P = kans op een goed antwoord)
0	Gemakkelijk	$P > 0,80$
1	Uitdagend	$0,50 \leq P \leq 0,80$
2	Moeilijk	$P < 0,50$

In een tabel wordt ook het aantal gemakkelijke, uitdagende en moeilijke opgaven voor de typische leerlingen gegeven.

Net als vorig jaar gaan we niet verder in op de opgaven in de EB-Anker en de Niveautoetsen. De opgaven van de EB-Anker zijn al beschreven in de rapportage van de peiling in 2008 en zijn vergelijkbaar met de overige EB-opgaven. Het merendeel van de Niveautoets-opgaven is ook beschreven in deze rapportage 2008. Deze opgaven waren, zoals eerder vermeld, duidelijk gemakkelijker dan de opgaven uit de reguliere eindtoets. In dit hoofdstuk rapporteren we alleen over de opgaven van de Eindtoets 2009. Voor een beschrijving van de gebruikte opgaven van de Eindtoets 2008 verwijzen wij ook naar de rapportage van de peiling in 2008 (Hemker & Van Weerden, 2009). Hierbij zij opgemerkt dat de relatieve moeilijkheid van de opgaven ten opzichte van elkaar niet is veranderd.

Behalve over de moeilijkheidsgraad van de opgaven rapporteren we ook over de mate waarin de leerlingen deze opgaven beheersen. Hoewel deze twee direct aan elkaar gerelateerd zijn, geeft de mate van beheersing ook net een andere blik op wat leerlingen nu werkelijk kennen en kunnen.

Na de omschrijving van de vaardigheidsschaal worden de populatieverdelingen op deze vaardigheidsschaal geplaatst. Merk op dat deze populatieverdelingen niet gecorrigeerd zijn voor de eventuele invloed van de achtergrondkenmerken. Voor de interpretatie van deze verschillen kan beter gekeken worden naar de gecorrigeerde effectschattingen (zie hoofdstuk 4). Merk overigens op dat het gebruik van de term effect niet veronderstelt dat de achtergrondkenmerken de *oorzaak* van het verschil zijn. Er worden geen veronderstellingen gedaan over de causaliteit.

In de rapportage besteden we speciale aandacht aan het jaareffect, dit wil zeggen het verschil tussen de prestaties in 2009 en 2008 (2009 minus 2008). Dit effect is onder verschillende statistische modellen geschat. Bij elke vaardigheid beginnen we met model met het afnamejaar als enige onafhankelijke variabele en verkrijgen dan het zogeheten ongecorrigeerde jaareffect. Daarna rapporteren we het gecorrigeerde jaareffect volgens een model waarin ook de achtergrondkenmerken geslacht, leertijd, stratum en formatiegewicht zijn opgenomen. Tot slot is een model geanalyseerd waarbij we enkele aanvullende achtergrondkenmerken hebben toegevoegd, te weten type toets, advies VO, regio, urbanisatiegraad en thuistaal. Voor een beschrijving van de modellen en de interpretatie van de gecorrigeerde en ongecorrigeerde jaareffecten verwijzen we naar paragraaf 4.3.

Bij de taalvaardigheden was er sprake van een interactie tussen Afnamejaar en formatiegewicht en is de correctie uitgevoerd volgens het interactiemodel. Vandaar dat we het jaareffect bij taal per formatiegewicht rapporteren. Bij de rekenschalen was er geen sprake van interactie en is de correctie volgens het hoofdeffectenmodel uitgevoerd.

5.1 Nederlandse Taal: Woordenschat

De vaardigheid

Woorden vormen de bouwstenen van onze taal en maken allerlei soorten van kennis toegankelijk. Zo vervult het beschikken over een uitgebreide woordenschat een sleutelfunctie bij het verwerven van kennis via teksten. Goed begrijpend kunnen lezen hangt voor een groot deel af van de breedte en de diepte van de woordkennis van een leerling. Bij de breedte van de woordkennis gaat het om het beheersen van veel verschillende woorden, bij een diepe woordkennis staat de vraag centraal: 'hoe goed kent de leerling woorden of begrippen in relatie tot andere woorden en begrippen?' Dit hoogste niveau van woordkennis is belangrijk in verband met de schoolse taalvaardigheid die leerlingen vooral in de bovenbouw van de basisschool nodig hebben. Zij moeten dan over een uitgebreid netwerk van begrippen beschikken met woordkennis die snel kan worden ingezet, zodat ze, onder meer tijdens het leesproces, verbanden en principes begrijpen en problemen kunnen oplossen. Hoe meer woorden én woordbetekenissen leerlingen kennen, hoe beter en preciezer ze in staat zijn om de betekenis van nieuwe woorden in een tekst af te leiden en om teksten te begrijpen. Omgekeerd geldt dat naarmate leerlingen meer en vlotter gaan lezen, hun woordenschatuitbreiding in steeds grotere mate bepaald wordt door de geschreven taal waarmee ze in aanraking komen: het is een zichzelf versterkend proces. Het netwerk van begrippen zal zich steeds verder uitbreiden aan de hand van nieuwe begrippen én door het leggen van steeds meer relaties tussen al aanwezige begrippen.

Wat leerlingen kunnen

De schaal voor Woordenschat is gemaakt met behulp van 98 opgaven. Hiervan kwamen er 25 uit het Toetsboekje Extra. Deze opgaven zijn alleen gebruikt voor de koppeling tussen Eindtoetsen uit 2008 en 2009. In de schaal voor de analyses van de achtergrondvariabelen zitten zodoende hooguit 73 opgaven. Van deze opgaven zijn er 25 afkomstig uit de Niveautoetsen (15 zowel 2008 als 2009, 5 alleen in 2008 en 5 alleen in 2009), en 48 uit eindtoetsen (20 in 2008, 20 in 2009 en 8 ankeropgaven). Bij de analyses bleken alle opgaven op de schaal te passen. Van de 20 opgaven uit de Eindtoets 2009 is de positie op de vaardigheidsschaal in tabel 5.1.1 beschreven.

Tabel 5.1.1 Posities van de opgaven Woordenschat op de vaardigheidsschaal geordend naar de volgorde van afname

Taak	Item (tk)	Label	JPON_p50	JPON_p80	ZLV	LV	ST	HV	ZHV
Taal 3 ('09)	01	T3 - i01	161	272	1	1	1	0	0
Taal 3 ('09)	02	T3 - i02	212	323	2	1	1	1	1
Taal 3 ('09)	03	T3 - i03	188	299	2	1	1	1	0
Taal 3 ('09)	04	T3 - i04	164	219	1	1	0	0	0
Taal 3 ('09)	05	T3 - i05	159	269	1	1	1	0	0
Taal 3 ('09)	06	T3 - i06	187	298	2	1	1	1	0
Taal 3 ('09)	07	T3 - i07	200	255	2	1	1	0	0
Taal 3 ('09)	08	T3 - i08	165	275	1	1	1	0	0
Taal 3 ('09)	09	T3 - i09	159	270	1	1	1	0	0
Taal 3 ('09)	10	T3 - i10	213	323	2	1	1	1	1
Taal 3 ('09)	11	T3 - i11	236	319	2	2	1	1	1
Taal 3 ('09)	12	T3 - i12	177	259	1	1	1	0	0
Taal 3 ('09)	13	T3 - i13	103	186	0	0	0	0	0
Taal 3 ('09)	14	T3 - i14	188	244	2	1	0	0	0
Taal 3 ('09)	15	T3 - i15	133	299	1	1	1	1	0
Taal 3 ('09)	16	T3 - i16	184	267	1	1	1	0	0
Taal 3 ('09)	17	T3 - i17	169	224	1	1	0	0	0
Taal 3 ('09)	18	T3 - i18	187	228	2	1	0	0	0
Taal 3 ('09)	19	T3 - i19	217	549	2	2	1	1	1
Taal 3 ('09)	20	T3 - i20	248	359	2	2	1	1	1

De eerste kolom van tabel 5.1.1 vermeldt de taak. Het betreft hier de taak Taal 3, 2009 (de eerste taak voor taal op de derde toetsdag van de Eindtoets). De tweede en derde kolom bevat ten respectievelijk de volgnummers en labels van de opgaven binnen de taak. In Bijlage 2 zijn de bijbehorende opgaven nogmaals weergegeven, maar nu geordend op basis van de waarde van P80 (zowel als tabel als in de vorm van een figuur). De verdeling van de opgaven naar moeilijkheidsgraad voor de vijf verschillende typische leerlingen is gegeven in tabel 5.1.2.

Tabel 5.1.2 Aantal opgaven bij Woordenschat in de Eindtoets 2009 per type leerling en moeilijkheidsgraad

Moeilijkheidsgraad / Typische leerling	ZLV	LV	ST	HV	ZHV
0 Gemakkelijk	1	1	5	12	15
1 Uitdagend	9	16	15	8	5
2 Moeilijk	10	3	0	0	0

Uit deze tabel is onder ander te achterhalen dat voor de standaard leerling en leerlingen vaardiger dan de standaardleerling geen enkele opgave als moeilijk te classificeren is: op alle opgaven is de kans groter dan 0,50 dat dergelijke leerlingen de opgave goed hebben.

Op basis van tabel 5.1.1 kan de vaardigheid ook beschreven worden in termen van de typische leerlingen. Deze wordt hieronder gegeven.

De zeer laag vaardige (ZLV) leerling beheerst 10 van de 20 EB-opgaven slecht: T3-i2 /i3 /i6 /i7 /i10 /i11 /i14 /i14 /i18 /i19 en /i20. Er is 1 opgave die een dergelijk leerling goed beheerst: T3-i13. De overige 9 opgaven zijn uitdagend.

De laag vaardige (LV) leerling beheerst nog maar 3 opgaven slecht: T3- i11 /i19 en /i20.. Opgave T3-i13 is echter ook voor een dergelijke leerling nog wel de enige opgave die goed beheerst wordt. De overige 16 opgaven zijn uitdagend.

De standaard (ST) leerling beheerst geen enkele opgave slecht. Dat geldt dus ook voor alle hogere percentiel leerlingen. Een ST-leerling beheerst 5 opgaven goed: T3-i4 /i13 /i14 /i17 en /i18. De overige 15 opgaven zijn uitdagend.

De hoog vaardige (HV) leerling beheerst 12 opgaven goed. De overige 8 opgaven zijn uitdagend: T3-i2 /i3 /i6 /i10 /i11 /i15 /i19 /i20.

De zeer hoog vaardige (ZHV) leerling beheerst 15 van de EB opgaven goed. De enige 5 uitdagende opgaven zijn T3-i2 /i10 /i11 / i19 en /i20.

Na omschrijving van de vaardigheidsschaal en wat Typische leerlingen kunnen volgt nu de omschrijving van de populatieverdelingen.

Wat kunnen groepen van leerlingen

In tabel 5.1.3 is de verdeling van de vaardigheid over de gehele groep leerlingen in 2008 en 2009 weergegeven. Daarnaast is de verdeling gegeven per subgroep binnen de twee afnamejaren. De posities zijn aangeduid op dezelfde schaal als de opgaven. Merk op dat dit marginale schattingen zijn: er wordt hierbij niet gecorrigeerd voor eventuele verschillen in de verdeling van de achtergrondvariabelen. Voor de gecorrigeerde resultaten, waar gecorrigeerd wordt voor de basis variabelen stratum, formatiegewicht, leertijd en geslacht, moeten de effectschattingen vergeleken worden. Bij de taalschalen is rekening gehouden met de interactie tussen formatiegewicht en jaar. De marginalen van de genoemde aanvullende achtergrondkenmerken zijn ieder voor zich aan dit basismodel toegevoegd.

Het gemiddelde van de leerlingen in 2009 ligt op deze schaal iets lager dan in 2008. Bij de effect-schattingen is nagegaan is of dit verschil significant is en dat blijkt niet het geval te zijn. De iets lagere vaardigheid vinden we terug bij alle achtergrondkenmerken, maar de verschillen zijn overal zeer klein. De enige uitzondering is de variabele formatiegewicht waar bij leerlingen met formatiegewicht 1.00 en 1.25 iets meer omlaag gegaan zijn dan gemiddeld. Daar staat echter tegenover dat de leerlingen met formatiegewicht 1.90 juist gestegen zijn in vaardigheid. Hierover later meer bij de beschrijving van de effecten.

De verschillen tussen de groepen onderling is vergelijkbaar met de resultaten van 2008. Als we de resultaten van 2009 vergelijken met de resultaten van 2008 (Cito, 2008) dan zien we wat hele kleine verschillen die veroorzaakt worden doordat het model iets aangepast is. In de meeste gevallen gaat het om afrondingsverschillen. Alleen bij het achtergrondkenmerk thuistaal zijn wat grotere verschillen te vinden. Die zijn echter te herleiden tot het kleinere aantal leerlingen.

De opmerkingen die in 2008 geplaatst zijn over de verschillen tussen de groepen gelden in hoge mate ook voor 2009 (Cito, 2008; p. 39-41). De enige uitzondering is de variabele formatiegewicht waarbij de vaardigheid van de 1.90 leerlingen dichterbij die van de 1.25 en 1.00 leerlingen is komen te liggen. Dit had ook een gevolg voor het gemiddelde van de populatie leerlingen die thuis alleen een andere taal spreken dan Nederlands: hun vaardigheid is iets dichterbij de vaardigheid gaan liggen van leerlingen die thuis Nederlands spreken. Het verschil in vaardigheid tussen deze groepen is echter nog steeds relatief groot.

Tabel 5.1.3 Verdelingen van de leerlingen op de vaardigheid Nederlandse taal - Woordenschat

Jaar Variabele	Niveau	2008		2009		
		Gem.	SD	Gem.	SD	
Jaar		250	50	249	49	
Geslacht	Jongens	255	50	253	49	
	Meisjes	245	50	244	49	
Leertijd	Regulier	256	48	254	47	
	Vertraagd	225	50	225	49	
Stratum	Stratum 1	257	47	255	47	
	Stratum 2	247	50	245	49	
	Stratum 3	218	52	222	50	
Formatiegewicht	F100	259	46	256	46	
	F125	229	47	225	47	
	F190	199	47	210	47	
Toets	EB	251	49	250	48	
	NT	172	51	174	49	
Advies VO	vmbo-BB	182	41	183	39	
	vmbo-KB	211	40	210	39	
	vmbo-GT	233	39	230	39	
	Havo	265	39	262	38	
	VWO	313	38	310	38	
Regio	Noord	250	48	248	48	
	Oost	249	49	248	48	
	West	250	51	249	50	
	Zuid	251	49	249	48	
Urbanisatiegraad	Zeer Sterk	238	54	240	52	
	Sterk	250	51	249	49	
	Matig	254	49	252	48	
	Weinig	252	48	250	48	
	Niet	252	48	249	47	
Thuis taal	Alleen NLS	255	48	253	47	
	NLS + bu	220	51	223	49	
	Alleen bu	201	50	208	49	
	Turks	184	49	190	48	
	Arabisch	201	49	208	47	
	Surinaams	212	50	216	49	
	WestEuro	247	50	245	49	
	Ter referentie:	Typische leerling	Schaalwaarde (geen verschil tussen de jaren)			
		Zeer laag vaardige leerling (ZLV)	186		186	
	Laag vaardige leerling (LV)	216		216		
	Standaard leerling (ST)	250		250		
	Hoog vaardige leerling (HV)	284		284		
	Zeer hoog vaardige leerling (ZHV)	314		314		

De effecten van de achtergrondvariabelen

Behalve de waarden van de gemiddelden van de onderscheiden groepen willen we ook weten of de gevonden verschillen relevant zijn: wat is de effectgrootte? Ook willen we de verschillen tussen de groepen corrigeren voor de andere achtergrondkenmerken. De effectschattingen in tabel 5.1.4 geven deze informatie. In hoofdstuk 4 is ook te vinden hoe deze getallen zijn te interpreteren.

In de presentatie van de effectschattingen maken we zoals gezegd een verschil tussen de jaareffecten en die van de overige achtergrondkenmerken. Een van de jaareffecten is geschat volgens het model met als onafhankelijke variabele alleen het afnamejaar. Dit model levert het ongecorrigeerde, direct geobserveerde verschil tussen de afnamejaren. Bij Woordenschat is te zien dat het effect iets negatief is (-0,02), maar dit effect verschilt ondanks het zeer grote aantal waarnemingen niet significant van 0. Het effect is dan ook extreem klein.

Vanwege de interactie tussen afnamejaar en formatiegewicht is het jaareffect in tabel 5.1.4 per formatiegewicht gerapporteerd. Het jaareffect in het hoofdeffecten model verschilde overigens weinig van het ongecorrigeerde jaareffect – een resultaat wat we ook bij rekenen zullen zien waar geen interactie tussen jaar en formatiegewicht gevonden was. Bij de verschillende formatiegewichten zien we echter wel een aantal verschillen. Zo is de Woordenschat van de leerlingen met een formatiegewicht 1.00 iets achteruitgegaan. Het effect is klein en wordt geclassificeerd als “geen effect”, maar is wel significant ($\alpha = ,0001$). Het effect bij de F1.25 leerlingen is ook zwak negatief en ondanks dat het in absolute waarde groter is dan bij F1.00 niet significant (aangezien de groep met leerlingen met gewicht 1.25 veel kleiner is dan die met gewicht 1.00). Een positief effect wordt echter gevonden bij de leerlingen met formatiegewicht 1.90. Dit zijn de allochtone leerlingen bij wie thuis vaak geen Nederlands gesproken wordt en waarvan de ouders vaak een laag opleidingsniveau hebben (zie hoofdstuk 2 voor de volledige definities van de formatiegewichten).

Tabel 5.1.4 Jaareffect Taal – Woordenschat

Variabele: Contrast	Model	Effectgrootte	Effectkwalificatie
Jaar: 2009 – 2008	Alleen jaareffect	-0,02	geen effect
Jaar: 2009 – 2008	Basismodel taal (G L S F J JF)		
	Per formatiegewicht: F1.00	-0,07	geen effect **
	F1.25	-0,09	geen effect
	F1.90	0,22	klein

** Is geclassificeerd als “geen effect” (niet groter dan 0,2 en niet kleiner dan -,2), maar wel significant ($\alpha = ,0001$)

In de vorige niveaupeiling presteerden de leerlingen met formatiegewicht 1,90 slechter dan de autochtone leerlingen, en dat is nog steeds het geval, maar het verschil is wel kleiner geworden. Dat is ook te zien als de effecten per formatiegewicht in tabel 5.1.5, waar de effecten voor formatiegewicht per jaar zijn weergegeven. Ook daarin is te zien dat het verschil tussen de F1.90 en de F1.00 leerlingen, en de F1.90 en de F1.25 leerlingen in 2009 kleiner geworden is dan in 2008.

Wat betreft de effecten van de andere basisvariabelen (geslacht, leertijd en stratum) zijn in 2008 en 2009 vrijwel dezelfde resultaten gevonden. Dat was ook te verwachten aangezien voor deze variabelen geen interactie met jaar gevonden was. Bij de aanvullend toegevoegde variabelen zijn wat kleine verschillen gevonden. Alleen bij thuistaal was dit groter dan 0,10.

Tabel 5.1.5 Effectgrootte op de vaardigheid Nederlandse taal – Woordenschat

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	-0,23	klein
Leertijd	Vertraagd – Regulier	-0,54	matig
Stratum	Stratum2 – Stratum 1	-0,08	geen effect
	Stratum 3 – Stratum 2	-0,15	geen effect
	Stratum 3 – Stratum 1	-0,23	klein
Formatiegewicht	F1.25 – F1.00 (2008)	-0,54	matig
	F1.25 – F1.00 (2009)	-0,56	matig
	F1.90 – F1.25 (2008)	-0,51	matig
	F1.90 – F1.25 (2009)	-0,20	klein
	F1.90 – F1.00 (2008)	-1,05	groot
	F1.90 – F1.00 (2009)	-0,76	matig
Toets	NT – EB	-1,42	groot
Advies VO	vmbo-KB – vmbo-BB	0,64	matig
	vmbo-GT – vmbo-KB	0,46	klein
	Havo – vmbo-GT	0,77	matig
	VWO – Havo	1,24	groot
Regio	Oost – Noord	-0,02	geen effect
	West – Noord	0,07	geen effect
	Zuid – Noord	0,03	geen effect
	West – Oost	0,09	geen effect
	Zuid – Oost	0,05	geen effect
	Zuid – West	-0,04	geen effect
Urbanisatiegraad	Sterk – Zeer sterk	-0,04	geen effect
	Matig – Sterk	-0,01	geen effect
	Weinig – Matig	-0,06	geen effect
	Niet – Weinig	-0,04	geen effect
Thuis taal	NLs+Ander – Alleen NLs	-0,28	klein
	Alleen ander – Alleen NLs	-0,47	klein
	Alleen ander – NLs+Ander	-0,19	geen effect
	Turks – Alleen NLs	-0,84	groot
	Arabisch – Alleen NLs	-0,48	klein
	Surinaams/Antilliaans – Alleen NLs	-0,44	klein
	West-Europees – Alleen NLs	0,01	geen effect

De leerlingen die de NT gemaakt hebben, waren in 2009 relatief iets minder vaardig dan in 2008. Bij het achtergrondkenmerk advies VO is voor Woordenschat de afstand tussen vmbo-BB en vmbo-KB iets grotere geworden en tussen vmbo-KB en vmbo-GT iets kleiner. De vmbo-KB zijn relatief iets vooruitgegaan ten opzichte van 2008, maar de verschillen zijn zeer klein (zie ook de marginalen). De effecten van regio waren

over de hele linie ook wat kleiner dan vorig jaar. De effecten waren toen al klein, maar die zijn nu nog kleiner geworden. Er is geen verschil tussen de verschillende regio's gevonden. De verschillen tussen de verschillende gradaties van verstedelijking zijn nog steeds afwezig als gecorrigeerd wordt voor de basisvariabelen. Bij thuistaal waren zoals gezegd wel enige verschillen gevonden. Het bijkomende negatieve effect van een andere thuistaal, bovenop de basisvariabelen, is kleiner geworden. Deze verkleining van het gat tussen leerlingen die thuis een andere taal dan Nederlands spreken, bovenop het verkleinde effect van formatiegewicht 1.90, lijkt te wijzen op een verbeterde vaardigheid van de allochtone leerlingen bij woordenschat.

5.2 Nederlandse Taal: Spelling

De vaardigheid

Net als Woordenschat wordt Spelling als ondersteunende activiteit voor schrijfvaardigheid beschouwd. Het betreft in de vorm zoals die hier gevraagd wordt een kennisaspect, waarbij de vragen in meerkeuzevorm met vier antwoordalternatieven gesteld zijn. Bij de hier bevraagde Spelling gaat het om het herkennen van de (in)correcte schrijfwijze van woorden volgens de vastgelegde regels. Bij het spellen wordt onderscheid gemaakt tussen het spellen van werkwoorden en van niet-werkwoorden. Dat komt neer op een verdeling in de spelling van het lexicon (woorden met een vast woordbeeld) en de spelling van woorden met verbuigingen (morfologie). Het aantal werkwoorden en niet-werkwoorden is gelijk verdeeld. Er is voor gekozen om deze samen te nemen als één vaardigheid aangezien de samenhang hoog is.

Wat leerlingen kunnen

De schaal voor Spelling bestaat uit 101 opgaven. Hiervan waren er 30 uit het Toetsboekje Extra. Deze opgaven zijn alleen gebruikt voor de koppeling tussen eindtoetsen uit 2008 en 2009. In de schaal die is gebruikt bij de analyses van de achtergrondvariabelen zitten zodoende hooguit 71 opgaven. Van deze opgaven zijn er 31 afkomstig uit de niveautoetsen (9 zowel 2008 als 2009, 11 alleen in 2008 en 11 alleen in 2009) en 40 uit eindtoetsen (20 in 2008, 20 in 2009; er is echter geen enkele ankeropgaven- de koppeling gaat met behulp van het Toetsboekje Extra). Uit de analyses is geen enkele opgaven verwijderd uit de schaal. Van de 20 opgaven uit de eindtoets 2009 is in tabel 5.2.1 de positie op de vaardigheidsschaal van Spelling weergegeven.

De opgaven voor Spelling zijn afgenomen in de eerste taak voor Taal (Taal 1), wat ook de eerste taak van de eindtoets is op de eerste toetsdag. De eerste tien opgaven (T1-i01 t/m i10) waren opgaven met betrekking tot niet-werkwoordspelling. De tweede set van tien opgaven (T1-i21 t/m i30) hadden betrekking op spelling van werkwoorden. Merk op dat tussen deze twee sets van opgaven ook nog een aantal andere taalopgaven zaten in deze eerste taak voor taal (T1 – i11 t/m i20).

De verdeling van de moeilijkheidsgraad bij de vijf verschillende typische leerlingen is weergegeven in tabel 5.2.2. De eerste drie kolommen vermelden achtereenvolgens de naam, het volgnummer en het label van de taak. In Bijlage 2 zijn de opgaven nogmaals weergegeven, maar nu geordend op basis van de waarde van P80 (zowel als tabel als in de vorm van een figuur).

Tabel 5.2.1 Posities van de opgaven Spelling op de vaardigheidsschaal geordend naar de volgorde van afname

Taak	Item (tk)	Label	JPON_p50	JPON_p80	ZLV	LV	ST.	HV	ZHV
Taal 1 ('09)	01	T1 - i01	79	142	0	0	0	0	0
Taal 1 ('09)	02	T1 - i02	142	206	1	0	0	0	0
Taal 1 ('09)	03	T1 - i03	222	286	2	2	1	1	0
Taal 1 ('09)	04	T1 - i04	142	226	1	1	0	0	0
Taal 1 ('09)	05	T1 - i05	158	221	1	1	0	0	0
Taal 1 ('09)	06	T1 - i06	215	278	2	1	1	0	0
Taal 1 ('09)	07	T1 - i07	200	285	2	1	1	1	0
Taal 1 ('09)	08	T1 - i08	136	187	1	0	0	0	0
Taal 1 ('09)	09	T1 - i09	201	285	2	1	1	1	0
Taal 1 ('09)	10	T1 - i10	164	206	1	0	0	0	0
Taal 1 ('09)	11	T1 - i21	177	262	1	1	1	0	0
Taal 1 ('09)	12	T1 - i22	170	255	1	1	1	0	0
Taal 1 ('09)	13	T1 - i23	231	295	2	2	1	1	0
Taal 1 ('09)	14	T1 - i24	183	267	1	1	1	0	0
Taal 1 ('09)	15	T1 - i25	190	253	2	1	1	0	0
Taal 1 ('09)	16	T1 - i26	238	322	2	2	1	1	1
Taal 1 ('09)	17	T1 - i27	233	360	2	2	1	1	1
Taal 1 ('09)	18	T1 - i28	187	250	2	1	1	0	0
Taal 1 ('09)	19	T1 - i29	176	219	1	1	0	0	0
Taal 1 ('09)	20	T1 - i30	206	269	2	1	1	0	0

Tabel 5.2.2 Aantal opgaven bij Spelling in de Eindtoets 2009 per type leerling en moeilijkheidsgraad

Moeilijkheidsgraad \Typische leerling	ZLV	LV	ST	HV	ZHV
0 Gemakkelijk	1	4	7	14	18
1 Uitdagend	9	12	13	6	2
2 Moeilijk	10	4	0	0	0

Uit deze tabel is onder ander te achterhalen dat net als bij Woordenschat voor de standaard leerling geen enkele opgave als moeilijk te classificeren is. Op basis van tabel 5.2.1 kan ook de vaardigheid van de vijf typische leerlingen worden beschreven:

De zeer laag vaardige (ZLV) leerling beheerst 1 van de 20 EB opgaven goed: T1-i1. Er worden 10 opgaven slecht beheerst: T1-i3 /i6 /i7 /i9 /i23 /i25 /i26 /i27 /i28 en /i30. Er is 1 opgave die een dergelijk leerling goed beheerst: T1-i1. De overige 9 opgaven zijn uitdagend.

De laag vaardige (LV) leerling beheerst 4 opgaven slecht (T1- i3 /i23 /i26 en /i27) en 4 opgaven goed (T1-i1 /i2 i8 en /i10). De overige 12 opgaven zijn uitdagend.

De standaard (ST) leerling beheerst geen enkele opgave slecht. Dat geldt dus ook voor alle hogere percentiel leerlingen. Er zijn voor een ST-leerling 7 opgaven die goed beheerst worden: T1-i1 /i2 /i4 /i5 /i8 /i10 en /i29. De overige 13 opgaven zijn uitdagend.

De hoog vaardige (HV) leerling beheerst 14 opgaven goed. De overige 6 opgaven zijn uitdagend: T1-i3 /i7 /i9 /i23 /i26 en /i27.

De zeer hoog vaardige (ZHV) leerling beheerst 18 van de 20 EB opgaven goed. De enige 5 uitdagende opgaven zijn T1-i26 en /i27.

Wat groepen van leerlingen kunnen

In tabel 5.2.3 is de verdeling van de vaardigheid over de gehele groep leerlingen in 2008 en 2009 weergegeven. Daarnaast is de verdeling gegeven per subgroep binnen de twee afnamejaren.

De posities van de subgroepen zijn weergegeven op dezelfde schaal als de opgaven. Merk op dat dit marginale schattingen zijn: er wordt hierbij niet gecorrigeerd voor eventuele verschillen in de verdeling van de overige achtergrondvariabelen. De opmerkingen die gemaakt zijn bij Woordenschat over mogelijke verschillen met de eerdere rapportage van 2008 (Cito) zijn ook hier van toepassing. De verschillen zijn echter zeer klein.

De algemeen gemiddelden van 2008 en 2009 zijn gelijk aan elkaar: in beide jaren staan die op 250. Dat geldt ook voor de meeste gemiddelden van de verschillende subgroepen, op een aantal kleine verschillen na. Zo is de vaardigheid van de stratum 3-scholen in 2009 iets gestegen ten opzicht van 2008. Een nog iets grotere stijging is te vinden voor de leerlingen met formatiegewicht 1.90. Dat leverde ook een significante interactie op in het model. De richting van de interactie was vergelijkbaar met die bij Woordenschat (F1.90 omhoog, F1.00 iets omlaag in vaardigheid), maar de verschillen waren bij Spelling wel kleiner. Bij de vaardigheid Spelling zijn de leerlingen met formatiegewicht 1.90 nog dichterbij de leerlingen met gewicht 1.00 gekomen, en zijn ook verder van de leerlingen met gewicht 1.25 (de minst vaardige formatiegewichtsgroep) afgeraakt. Deze laatste groep blijft ongeveer op hetzelfde vaardigheidsniveau als in 2008.

Bij de aanvullende achtergrondkenmerken blijven de verschillende groepen op hetzelfde niveau als in 2008. De enige variabele waarbij er nog wel verschuivingen plaats heeft gevonden is thuistaal. De leerlingen die alleen Nederlands spreken blijven op hetzelfde niveau, terwijl de leerlingen die een andere taal spreken dichterbij dat niveau komen. De leerlingen die een andere taal spreken - al dan niet met Nederlands erbij - zitten echter nog steeds onder het gemiddelde. Bij de analyse van de verschillende talen is te zien dat de leerlingen die thuis, eventueel naast Nederlands, ook Turks of Arabisch spreken in 2009 de meest vooruitgang geboekt hebben in spellingsvaardigheid: ze komen dichterbij het niveau van de leerlingen die thuis alleen Nederlands spreken. De leerlingen die thuis Surinaams, Antilliaans of een West-Europese taal spreken, blijven op hetzelfde niveau als in 2008. Al deze groepen zitten in vaardigheidsniveau nog wel onder het gemiddelde.

Tabel 5.2.3 Verdelingen van de leerlingen op de vaardigheid Nederlandse taal – Spelling

Jaar Variabele	Niveau	2008		2009	
		Gem.	SD	Gem.	SD
Jaar		250	50	250	50
Geslacht	Jongens	242	49	241	49
	Meisjes	258	49	258	49
Leertijd	Regulier	257	48	256	48
	Vertraagd	221	48	221	48
Stratum	Stratum 1	253	49	252	49
	Stratum 2	247	50	247	50
	Stratum 3	239	51	243	50
Formatiegewicht	F100	254	49	253	49
	F125	230	50	231	50
	F190	236	50	243	50
Toets	EB	251	49	251	49
	NT	187	51	190	50
Advies VO	vmbo-BB	195	40	195	40
	vmbo-KB	215	40	215	40
	vmbo-GT	231	39	230	40
	Havo	262	39	261	39
	VWO	318	39	316	39
Regio	Noord	246	50	245	50
	Oost	249	50	249	50
	West	250	50	250	50
	Zuid	253	50	253	50
Urbanisatiegraad	Zeer Sterk	248	51	250	50
	Sterk	251	50	251	50
	Matig	252	50	252	50
	Weinig	250	50	249	50
	Niet	247	50	246	49
Thuis taal	Alleen NLS	251	50	251	50
	NLS + bu	247	50	249	50
	Alleen bu	238	50	243	50
	Turks	238	50	244	50
	Arabisch	235	50	241	50
	Surinaams	230	50	231	50
	WestEuro	249	50	249	50
	Ter referentie:	Typische leerling	Schaalwaarde (geen verschil tussen de jaren)		
	Zeer laag vaardige leerling (ZLV)	186		186	
	Laag vaardige leerling (LV)	216		216	
	Standaard leerling (ST)	250		250	
	Hoog vaardige leerling (HV)	284		284	
	hoog vaardige leerling (ZHV)	314		314	

De effecten van achtergrondvariabelen

In de tabellen 5.2.4 en 5.2.5 zijn de effectschattingen weergegeven. In tabel 5.2.4 staan de effectschatting voor de variabele jaar: de jaareffecten (2009 – 2008) in het model met alleen de variabele jaar en de jaareffecten per formatiegewicht in het model met de basisvariabelen (geslacht, leertijd, stratum en formatiegewicht) en de interactie jaar*formatiegewicht. Bij Spelling is er geen jaareffect in het eerste model: het effect is 0,00 en verschilt dus niet significant van 0. Als in het tweede model naar de “gezuiverde” jaareffecten gekeken wordt per formatiegewicht, dan is te zien dat bij leerlingen met formatiegewicht 1.00 de effectschatting -0,04 is, bij 1.25 is het 0,00 en bij 1.90 is het 0,14. Al deze effecten worden geclassificeerd als “geen effect”. De eerste twee zijn daarnaast ook niet significant verschillend van 0. De tweede is echter wel significant ($\alpha = ,0001$) en positief. Bij de allochtone leerlingen, bij wie thuis vaak geen Nederlands gesproken wordt en waarvan de ouders vaak een laag opleidingsniveau hebben, is er dus een zeer kleine vooruitgang in vaardigheid waar te nemen. Bij de overige leerlingen is er geen significante verandering waargenomen.

Tabel 5.2.4 Jaareffect Taal – Spelling

Variabele: Contrast	Model	Effectgrootte	Effectkwalificatie
Jaar: 2009 – 2008	Alleen jaareffect	0,00	geen effect
Jaar: 2009 – 2008	Basismodel taal (G L S F J JF)		
	Per formatiegewicht: F1.00	-0,04	geen effect
	F1.25	0,00	geen effect
	F1.90	0,14	geen effect**

** Is geclassificeerd als “geen effect” (niet groter dan 0,2 en niet kleiner dan -,2), maar wel significant ($\alpha = ,0001$)

Net als bij Woordenschat kan hier opgemerkt worden dat de leerlingen met formatiegewicht 1,90 nog steeds slechter presteren dan de autochtone leerlingen, maar dat het verschil tussen de groepen kleiner is geworden. Dat is ook te zien als kijken naar de effecten bij formatiegewicht in tabel 5.2.5. Daar zijn vanwege de interactievariabele ‘jaar*formatiegewicht’ de effecten voor formatiegewicht per jaar weergegeven. Ook daarin is te zien dat het verschil tussen de F1.90 en de F1.00 leerlingen, en tussen de F1.90 en de F1.25 leerlingen, in 2009 kleiner is dan in 2008.

Verder is te zien dat het verschil tussen jongens en meisjes bij Spelling iets kleiner geworden is. De meisjes blijven echter nog wel de beste spellers (een klein significante effect). De verschillen bij leertijd zijn nog steeds matig (vertraagde leerlingen blijven slechter spellen). Er is nog steeds geen effect bij stratum. In vergelijking met vorig jaar is het effect zelfs nog kleiner geworden.

Dat de effecten kleiner geworden zijn, geldt ook voor de aanvullende achtergrondkenmerken urbanisatiegraad en regio. Vorig jaar waren er bij regio nog een paar kleine effecten (in de noordelijke provincies werd iets slechter gespeld), maar de effecten zijn nu zo klein dat ze als “geen effect” kunnen worden geclassificeerd. Het is nog wel zo dat de noordelijke provincies iets slechter scoren bij Spelling, maar de verschillen zijn niet meer significant. Bij urbanisatiegraad waren er al geen significante effecten. Bij advies van de docent richting het Voortgezet Onderwijs zijn de verschillen in spellingsvaardigheid tussen de verschillende richtingen iets kleiner geworden, maar de effecten zijn nog steeds significant en klein tot matig.

De effecten van thuistaal waren in bijna alle groepen al afwezig. Dat blijven ze nog steeds, zij het dat ook bij de effectschattingen een kleine relatieve vooruitgang geboekt is bij anderstaligen. De enige groep die hier wat achter blijft is de groep die thuis Surinaams of Antilliaans spreekt. Voor deze groep blijft een klein negatief significant effect bovenop de basisvariabelen.

Tabel 5.2.5 Effectgrootte op de vaardigheid Nederlandse taal – Spelling

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	0,33	klein
Leertijd	Vertraagd – Regulier	-0,69	matig
Stratum	Stratum2 – Stratum 1	-0,06	geen effect
	Stratum 3 – Stratum 2	-0,02	geen effect
	Stratum 3 – Stratum 1	-0,07	geen effect
Formatiegewicht	F1.25 – F1.00 (2008)	-0,41	klein
	F1.25 – F1.00 (2009)	-0,38	klein
	F1.90 – F1.25 (2008)	0,22	klein
	F1.90 – F1.25 (2009)	0,36	klein
	F1.90 – F1.00 (2008)	-0,19	geen effect
	F1.90 – F1.00 (2009)	-0,02	geen effect
Toets	NT – EB	-1,04	groot
Advies VO	vmbo-KB – vmbo-BB	0,47	klein
	vmbo-GT – vmbo-KB	0,39	klein
	Havo – vmbo-GT	0,77	matig
	VWO – Havo	1,45	groot
Regio	Oost – Noord	0,05	geen effect
	West – Noord	0,06	geen effect
	Zuid – Noord	0,15	geen effect
	West – Oost	0,01	geen effect
	Zuid – Oost	0,10	geen effect
	Zuid – West	0,09	geen effect
Urbanisatiegraad	Sterk – Zeer sterk	-0,09	geen effect
	Matig – Sterk	-0,01	geen effect
	Weinig – Matig	-0,07	geen effect
	Niet – Weinig	-0,04	geen effect
Thuis taal	NLs+Ander – Alleen NLs	0,10	geen effect
	Alleen ander – Alleen NLs	0,05	geen effect
	Alleen ander – NLs+Ander	-0,05	geen effect
	Turks – Alleen NLs	0,08	geen effect
	Arabisch – Alleen NLs	0,00	geen effect
	Surinaams/Antilliaans – Alleen NLs	-0,22	klein
	West-Europees – Alleen NLs	0,07	geen effect

5.3 Nederlandse Taal: Begrijpend lezen

De vaardigheid

Leesvaardigheid betreft de vaardigheid om schriftelijke teksten te begrijpen en te gebruiken in overeenstemming met het leesdoel. Met de toetsen kan vastgesteld worden in welke mate leerlingen daar in slagen. De toetsen bestaan voor het grootste deel uit (korte) teksten met bepaalde kenmerken op woord-, zins- en tekstniveau (moeilijke woorden of niet, lange zinnen, duidelijke structuur). De opgaven proberen hierbij te achterhalen in welke mate de leerlingen de teksten begrijpen. Het begrijpen heeft betrekking op het verwerken van informatie die een schrijver expliciet vermeldt. Het gaat met andere woorden om tekstgebaseerde verwerking, dus om de verwerking van inhoud en relaties tussen tekstelementen die in de tekst terug te vinden zijn. De leerling kan het antwoord op de vraag in de tekst vinden. Ook tekstgebaseerde gevolgtrekkingen behoren tot dit aspect van leesvaardigheid. In mindere mate spelen interpreteren en reflecteren een rol in het beantwoorden van deze opgaven. De vragen bij deze opgaven zijn gesteld als meerkeuzeopgaven met vier antwoordalternatieven.

Naast het type opgaven waarbij vragen over een tekst gesteld worden, is er een klein aantal opgaven waarbij een vijftal zinnen gegeven wordt. Deze zinnen staan in willekeurige volgorde, maar vormen samen een kort verhaal. De leerling moet aangeven welke van deze vijf zinnen de eerste moet zijn van het verhaal. Dit is dus een vijfkeuze opgave waarbij begrijpen en interpreteren een belangrijke rol spelen.

Wat leerlingen kunnen

De schaal voor Begrijpend lezen bestaat uit 126 opgaven. Hiervan waren er 15 uit het Toetsboekje Extra. Deze opgaven zijn alleen gebruikt voor de koppeling tussen de eindtoetsen uit 2008 en 2009. In de schaal voor de analyses van de achtergrondvariabelen zitten zodoende hooguit 111 opgaven. Van deze opgaven zijn er 37 afkomstig uit de niveautoetsen (23 zowel 2008 als 2009, 7 alleen in 2008 en 7 alleen in 2009), en 74 uit eindtoetsen (30 in 2008, 30 in 2009 en 14 ankeropgaven). In de analyse zijn twee opgaven uit de schaal verwijderd: een uit de Eindtoets 2009 en een ankeropgave. De schaal bestaat zodoende uiteindelijk uit 109 opgaven. De 29 overgebleven opgaven uit de Eindtoets 2009 zijn in tabel 5.3.1 weergegeven.

De eerste tien opgaven voor Begrijpend lezen zitten in de eerste taak voor Taal (Taal 1), wat ook de eerste taak van de eindtoets is op de eerste toetsdag. De opgaven T1 – i11 t/m i20 zijn de opgaven die precies tussen de twee sets met spellingsopgaven zaten. De overige negentien opgaven zaten in de laatste taak voor taal (Taal 4), wat ook de laatste taak was op de laatste toetsdag.

De verdeling van de moeilijkheidsgraad bij de vijf verschillende typische leerlingen is weergegeven in tabel 5.3.2. De eerste drie kolommen vermelden achtereenvolgens de naam, het volgnummer en het label van de taak. In Bijlage 2 zijn de opgaven nogmaals weergegeven, maar nu geordend op basis van de waarde van P80 (zowel als tabel als in de vorm van een figuur). Uit tabel 5.3.2 is onder ander te achterhalen dat er, zelfs voor leerlingen die zeer hoogvaardig zijn, nog één opgave is die nog steeds moeilijk is. Van de tien opgaven die moeilijk zijn voor zeer laag vaardige leerling blijft er een over die ook voor de laag vaardige leerling moeilijk blijft. Deze opgave is overigens zelfs voor een zeer hoog vaardige leerling nog moeilijk (T4 – i12). Daarnaast is er ook een opgave die niet fraai op de vaardigheidsschaal past. Dat betekent niet dat dit een slechte opgave is, of geen Begrijpend lezen zou meten, maar het betekent dat deze opgave niet goed met het gebruikte IRT model beschreven wordt. Om misschattingen of verkeerde omschrijvingen te voorkomen, is deze opgave (T4 – i19) uit de schaal verwijderd. Dit had overigens verder geen effect op de analyses met de achtergrondvariabelen.

Tabel 5.3.1 Posities van de opgaven Woordenschat op de vaardigheidsschaal geordend naar de volgorde van afname.

Taak	Item (tk)	Label	JPON_p50	JPON_p80	ZLV	LV	ST	HV	ZHV
Taal 1 ('09)	11	T1 - i11	193	293	2	1	1	1	0
Taal 1 ('09)	12	T1 - i12	132	211	1	0	0	0	0
Taal 1 ('09)	13	T1 - i13	185	235	1	1	0	0	0
Taal 1 ('09)	14	T1 - i14	202	282	2	1	1	0	0
Taal 1 ('09)	15	T1 - i15	168	301	1	1	1	1	0
Taal 1 ('09)	16	T1 - i16	96	162	0	0	0	0	0
Taal 1 ('09)	17	T1 - i17	151	208	1	0	0	0	0
Taal 1 ('09)	18	T1 - i18	156	213	1	0	0	0	0
Taal 1 ('09)	19	T1 - i19	164	231	1	1	0	0	0
Taal 1 ('09)	20	T1 - i20	143	276	1	1	1	0	0
Taal 4 ('09)	01	T4 - i01	-131	2	0	0	0	0	0
Taal 4 ('09)	02	T4 - i02	200	280	2	1	1	0	0
Taal 4 ('09)	03	T4 - i03	112	178	0	0	0	0	0
Taal 4 ('09)	04	T4 - i04	146	226	1	1	0	0	0
Taal 4 ('09)	05	T4 - i05	101	168	0	0	0	0	0
Taal 4 ('09)	06	T4 - i06	183	263	1	1	1	0	0
Taal 4 ('09)	07	T4 - i07	132	265	1	1	1	0	0
Taal 4 ('09)	08	T4 - i08	208	308	2	1	1	1	0
Taal 4 ('09)	09	T4 - i09	157	357	1	1	1	1	1
Taal 4 ('09)	10	T4 - i10	203	403	2	1	1	1	1
Taal 4 ('09)	11	T4 - i11	155	255	1	1	1	0	0
Taal 4 ('09)	12	T4 - i12	343	476	2	2	2	2	2
Taal 4 ('09)	13	T4 - i13	187	254	2	1	1	0	0
Taal 4 ('09)	14	T4 - i14	142	186	1	0	0	0	0
Taal 4 ('09)	15	T4 - i15	179	229	1	1	0	0	0
Taal 4 ('09)	16	T4 - i16	203	260	2	1	1	0	0
Taal 4 ('09)	17	T4 - i17	205	250	2	1	0	0	0
Taal 4 ('09)	18	T4 - i18	158	224	1	1	0	0	0
Taal 4 ('09)	19	T4 - i19	past niet op de schaal						
Taal 4 ('09)	20	T4 - i20	198	255	2	1	1	0	0

Tabel 5.3.2 Aantal opgaven bij Spelling in de Eindtoets 2009 per type leerling en moeilijkheidsgraad

Moeilijkheidsgraad \Typische leerling		ZLV	LV	ST	HV	ZHV
0	Gemakkelijk	4	8	14	23	26
1	Uitdagend	15	20	14	5	2
2	Moeilijk	10	1	1	1	1

De beheersing van de vijf typische leerlingen kunnen we aan de hand van de opgaven als volgt omschrijven.

De zeer laag vaardige (ZLV) leerling beheerst vier van de 29 EB opgaven inde analyses goed: T1-i16 en T4-i1 /i3 /i5. Er zijn 10 opgaven die slecht beheerst worden op dit niveau: T1-i11 /i14 en T4-i2 /i8 /i10 /i12 /i13 /i16 /i17 /i20. De overige 15 opgaven zijn uitdagend voor een percentiel 10 leerling slecht.

De laag vaardige (LV) leerling beheerst alleen opgave T4-i12 nog slecht. Dit is een opvallende opgave aangezien ook op de hogere niveaus deze opgave nog slecht beheerst wordt. Er zijn 8 opgaven die

een LV-leerling goed beheerst. Dat zijn de 4 opgaven die een ZLV-leerling ook goed beheerst aangevuld met opgaven T1-i12 /i17 /i18 en T4-i14. De overige 20 opgaven zijn uitdagend.

De standaard (ST) leerling beheerst zoals al aangegeven alleen opgave T4-i12 nog slecht. Er zijn 14 opgaven die een dergelijke leerling goed beheerst. Dat zijn de 8 opgaven die een LV-leerling goed beheerst aangevuld met T1-i13 /i19 en T4-i4 /i15 /i17 /i18. De overige 14 opgaven zijn uitdagend.

De hoog vaardige (HV) leerling beheerst 23 opgaven goed. Er zijn 5 opgaven die uitdagend voor een dergelijke leerling: T1-i11 /i15 en T4-i8 /i9 /i10. Opgave T4-i12 is ook voor deze leerling dus moeilijk.

De zeer hoog vaardige (ZHV) leerling beheerst 26 van de EB opgaven goed. Er zijn nog maar 3 opgaven die uitdagend voor een dergelijke leerling: T4-i9 /i10. Zelfs voor een ZHV-leerling is opgave T4-i12 dus moeilijk.

Wat groepen van leerlingen kunnen.

In tabel 5.3.3 staat de verdeling van de vaardigheid over de gehele groep leerlingen in 2008 en 2009. Daarnaast wordt de verdeling gegeven per subgroep binnen de twee jaren. De posities zijn weergegeven op dezelfde schaal als de opgaven. Merk op dat dit marginale schattingen zijn: er wordt hierbij niet gecorrigeerd voor eventuele verschillen in de verdeling van de overige achtergrondvariabelen. Er zijn een paar kleine verschillen tussen de rapportage van 2008 en die van het jaar 2009 in deze rapportage. Deze verschillen zijn merendeel afrondingsverschillen vanwege de keuze voor een iets ander model. Bij thuistaal zijn de verschillen iets groter, waarover zo meer.

Bij Begrijpend lezen is een kleine stijging in vaardigheid gevonden van 2008 naar 2009. Deze stijging is bijna bij alle groepen in dezelfde mate gevonden behalve bij de variabele formatiegewicht en thuistaal, die sterk met elkaar samenhangen. Dat komt omdat het ruime merendeel van de leerlingen met gewicht 1.90 thuis, eventueel naast Nederlands, een andere taal spreekt. Bij Begrijpend lezen is te zien dat leerlingen met gewicht 1.90 in 2009 niet vooruit maar juist iets achteruit gaan, terwijl de vaardigheid van 1.00 leerlingen juist relatief sterker gestegen is. Ook bij thuistaal is bij anderstaligen een lichte achteruitgang te zien in vaardigheid. Alleen bij leerlingen die een West-Europese taal spraken, is dat niet gevonden.

De effecten van de achtergrondvariabelen

In tabel 5.3.4 staan de effectschattingen voor de variabele jaar voor het model met alleen jaar als verklarende variabele en voor het model met de basisvariabelen en het interactie-effect tussen jaar en formatiegewicht. In het eerste model met alleen jaar (2009 – 2008) is te zien dat er een zeer kleine vooruitgang in vaardigheid heeft plaatsgevonden. Ondanks dat dit effect zeer klein is, en als “geen effect” wordt geclassificeerd, is dit effect wel significant groter dan 0. Begrijpend lezen is de enige vaardigheid van de zes gemeten vaardigheden waarbij het algemene jaareffect significant van 0 verschilt (bij Woordenschat en Spelling was dat al gemeld en ook bij de drie rekenschalen wordt geen significant verschil gevonden worden).

Tabel 5.3.3 Verdelingen van de leerlingen op de vaardigheid Nederlandse taal – Begrijpend lezen

Jaar Variabele	Niveau	2008		2009		
		Gem.	SD	Gem.	SD	
Jaar		250	50	252	50	
Geslacht	Jongens	246	50	248	50	
	Meisjes	254	50	256	50	
Leertijd	Regulier	257	48	258	48	
	Vertraagd	223	49	224	50	
Stratum	Stratum 1	256	48	259	48	
	Stratum 2	246	50	248	50	
	Stratum 3	225	50	224	51	
Formatiegewicht	1.00	256	48	260	48	
	1.25	224	49	225	49	
	1.90	218	49	214	49	
Toets	EB	251	49	253	49	
	NT	163	50	167	50	
Advies VO	vmbo-BB	176	37	177	37	
	vmbo-KB	207	37	208	37	
	vmbo-GT	231	37	232	37	
	Havo	267	36	269	37	
	VWO	316	36	318	36	
Regio	Noord	246	49	248	50	
	Oost	249	50	252	50	
	West	250	50	251	51	
	Zuid	253	50	254	50	
Urbanisatiegraad	Zeer Sterk	242	52	243	53	
	Sterk	251	50	253	51	
	Matig	253	50	256	50	
	Weinig	251	49	254	49	
	Niet	248	49	251	49	
Thuis taal	Alleen NLS	254	49	256	49	
	NLS + bu	230	50	228	51	
	Alleen bu	219	50	217	51	
	Turks	209	49	205	50	
	Arabisch	215	49	212	50	
	Surinaams	212	50	210	51	
	WestEuro	251	50	251	51	
	Ter referentie:	Typische leerling	Schaalwaarde (geen verschil tussen de jaren)			
		Zeer laag vaardige leerling (ZLV)	186		186	
	Laag vaardige leerling (LV)	216		216		
	Standaard leerling (ST)	250		250		
	Hoog vaardige leerling (HV)	284		284		
	Zeer hoog vaardige leerling (ZHV)	314		314		

In het model met de basisvariabelen geslacht, leertijd, stratum en formatiegewicht en de interactievariabele jaar*formatiegewicht is een omgekeerd beeld gevonden als eerder bij Spelling gevonden was. Bij Begrijpend lezen is te zien dat de leerlingen met formatiegewicht 1.00 juist iets vooruitgegaan zijn (effect van 0,06, geclassificeerd als “geen effect”) en dat dit effect significant groter is dan 0 ($\alpha < ,0001$). Het jaareffect bij leerlingen met formatiegewicht 1.90 is echter negatief (-0,10, geclassificeerd als “geen effect”), maar dat effect is niet significant verschillend van 0. Wel is net als bij Spelling bij formatiegewicht 1.25 geen effect gevonden (effect van 0,00).

Tabel 5.3.4 Jaareffect Taal – Begrijpend lezen

Variabele: Contrast	Model	Effectgrootte	Effectkwalificatie
Jaar: 2009 – 2008	Alleen jaareffect	0,04	geen effect
Jaar: 2009 – 2008	Basismodel taal (G L S F J JF)		
	Per formatiegewicht: F1.00	0,06	geen effect**
	F1.25	0,00	geen effect
	F1.90	-0,10	geen effect

** Is geclassificeerd als “geen effect” (niet groter dan 0,2 en niet kleiner dan -,2), maar wel significant ($\alpha = ,0001$)

Bij Begrijpend lezen vinden we dus dat de afstand tussen de leerlingen met formatiegewicht 1.90 en 1.00 juist iets groter wordt (zie ook tabel 5.3.5), terwijl dit bij Woordenschat en Spelling juist kleiner werd. De afstand tussen de leerlingen met een gewicht van 1.00 en de 1.25 blijft bij de taalvaardigheden vrijwel onveranderd. Bij Woordenschat is te zien dat de afstand in vaardigheid tussen leerlingen met 1.90 en 1.25 kleiner wordt, waarbij leerlingen met 1.25 nog wel beter presteren. Het effect daalt echter van groot naar matig. Bij Spelling is te zien dat de F1.90 leerlingen nu – gezuiverd voor de andere basisvariabelen – juist nog beter scoren dan de F1.25 leerlingen. Bij Begrijpend lezen is te zien dat de leerlingen met formatiegewicht 1.25 nu juist iets beter scoren dan de leerlingen met gewicht 1.90. In beide jaren is het verschil echter niet significant en in beide jaren is de effectgrootte te classificeren als “geen effect”.

Verder is te zien dat ook bij Begrijpend lezen het verschil tussen jongens en meisjes iets kleiner geworden is. De meisjes blijven echter nog wel beter in Begrijpend lezen, maar het effect wordt nu geclassificeerd als “geen effect”, terwijl dat in 2008 nog een klein effect was. Het verschil tussen jongens en meisjes is echter nog wel significant. De verschillen bij leertijd zijn nog steeds matig (vertraagde leerlingen blijven minder vaardig). De richting en de classificatie van de effecten bij stratum zijn bij Begrijpend lezen nog steeds gelijk aan die van 2008. De verschillen van de scholen in stratum 1 en 2 met die in stratum 3 zijn alleen iets groter geworden.

Bij de variabele advies VO is voor Begrijpend lezen, net zoals bij Woordenschat, de afstand tussen vmbo-BB en vmbo-KB iets groter geworden en tussen vmbo-KB en vmbo-GT iets kleiner. De leerlingen in vmbo-KB zijn relatief iets vooruitgegaan ten opzichte van 2008, maar de verschillen zijn zeer klein (zie ook de marginalen). Net als bij de andere twee taalschalen zijn de effecten voor regio en urbanisatiegraad in absolute waarde kleiner geworden. Daar waar men in de Noordelijke provincies voor Begrijpend lezen in 2008 ook iets slechter scoorden, is dat verschil nu kleiner geworden. Ook voor urbanisatiegraad lijken de resultaten, gecorrigeerd voor de basisvariabelen, nog meer op elkaar.

Bij de variabele thuistaal worden de effecten juist groter. Het negatieve effect van het thuis een andere taal spreken dan Nederlands is sterker dan in 2008. Het bijkomend effect, boven de basisvariabelen, van alleen een andere taal thuis spreken in contrast met alleen Nederlands, is nu zelfs klein (was eerder “geen effect”). Het versterkte negatieve effect van een andere taal spreken is bij bijna alle talen groter geworden.

Tabel 5.3.5 Effectgrootte op de vaardigheid Nederlandse taal – Begrijpend lezen

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	0,16	geen effect
Leertijd	Vertraagd – Regulier	-0,60	matig
Stratum	Stratum2 – Stratum 1	-0,11	geen effect
	Stratum 3 – Stratum 2	-0,15	geen effect
	Stratum 3 – Stratum 1	-0,27	klein
Formatiegewicht	F1.25 – F1.00 (2008)	-0,55	matig
	F1.25 – F1.00 (2009)	-0,61	matig
	F1.90 – F1.25 (2008)	0,02	geen effect
	F1.90 – F1.25 (2009)	-0,08	geen effect
	F1.90 – F1.00 (2008)	-0,54	matig
	F1.90 – F1.00 (2009)	-0,69	matig
Toets	NT – EB	-1,60	groot
Advies VO	vmbo-KB – vmbo-BB	0,77	matig
	vmbo-GT – vmbo-KB	0,60	matig
	Havo – vmbo-GT	0,95	groot
	VWO – Havo	1,33	groot
Regio	Oost – Noord	0,06	geen effect
	West – Noord	0,11	geen effect
	Zuid – Noord	0,14	geen effect
	West – Oost	0,05	geen effect
	Zuid – Oost	0,07	geen effect
	Zuid – West	0,03	geen effect
Urbanisatiegraad	Sterk – Zeer sterk	-0,06	geen effect
	Matig – Sterk	-0,01	geen effect
	Weinig – Matig	-0,07	geen effect
	Niet – Weinig	-0,06	geen effect
Thuis taal	NLs+Ander – Alleen NLs	-0,17	geen effect
	Alleen ander – Alleen NLs	-0,23	klein
	Alleen ander – NLs+Ander	-0,06	geen effect
	Turks – Alleen NLs	-0,46	klein
	Arabisch – Alleen NLs	-0,33	klein
	Surinaams/Antilliaans – Alleen NLs	-0,53	matig
	West-Europees – Alleen NLs	0,10	geen effect

6 Resultaten voor Rekenen/Wiskunde in jaargroep 8

In dit hoofdstuk worden de resultaten voor Rekenen/Wiskunde per rekenvaardigheid gepresenteerd. Dat zijn achtereenvolgens Getallen en Bewerkingen (GB), Breuken, Verhoudingen en Procenten (BPV) en Meten, Meetkunde, Tijd en Geld (MMTG). In de discussie (hoofdstuk 8) worden de resultaten voor de drie rekenvaardigheden met elkaar vergeleken. Hierbij kan alvast opgemerkt worden dat de resultaten bij de drie rekenvaardigheden meer op elkaar lijken dan dat bij de taalvaardigheden het geval is.

Op iedere toetsdag van de Eindtoets Basisonderwijs is een rekentaak van 20 opgaven afgenomen. Dit is altijd de tweede taak van de dag. Bij rekentaak 1 en 3 mogen de leerlingen uitrekenpapier gebruiken, maar bij taak 2 is dat niet geoorloofd. Een verschil met taal is dat de opgaven niet in een vaste volgorde zijn afgenomen. Binnen de drie rekentoetsen komen opgaven van drie rekenschalen door elkaar voor, met uitzondering van MMTG op toetsdag 2. Er is dus geen enkele opgave bij MMTG die niet zonder uitrekenpapier gemaakt mocht worden. In tabel 6 is weergegeven hoe de rekenopgaven over de drie taken verdeeld zijn. In de nummering van opgaven wordt eerst het nummer van de rekentaak genoemd (R1, R2 of R3, voor Taak Rekenen 1, 2 en 3), gevolgd door het nummer van de opgave binnen die taak. Met behulp van deze nummering zijn de opgaven gemakkelijk terug te vinden in de opgavenboekjes van de Eindtoets 2009 (aangezien daar ook het taaknummer en itemnummer binnen die taak gegeven worden).

Tabel 6 De verdeling van de opgaven voor Rekenen/Wiskunde

		Aantal opgaven			
		GB	BPV	MMTG	Totaal
Taak Rekenen 1	R1	6	7	7	20
Taak Rekenen 2	R2	14	6	0	20
Taak Rekenen 3	R3	5	7	8	20
Totaal		25	20	15	60

GB Getallen en Bewerkingen
BPV Breuken Procenten en Verhoudingen
MMTG Meten, Meetkunde, Tijd en Geld

Taken: Percentage per vaardigheid

		GB	BPV	MMTG
Taak Rekenen 1	R1	24	35	47
Taak Rekenen 2	R2	56	30	0
Taak Rekenen 3	R3	20	35	53
Totaal		100	100	100

Vaardigheid: Percentage per taak

	GB	BPV	MMTG	Totaal
R1	30	35	35	100
R2	70	30	0	100
R3	25	35	40	100

Ook bij de rekenschalen wordt de schaal geïllustreerd met de opgaven uit de EB van 2009 aangezien dit de enige nieuwe opgaven zijn op de vaardigheidsschalen. De overige opgaven, te weten de opgaven van de Eindtoets 2008, de rekenopgaven van de NT en de ankeropgaven, zijn onveranderd ten opzichte van het vorig jaar. De eigenschappen van deze opgaven zijn al beschreven in het verslag van 2008 (Cito, 2008) en zijn ongewijzigd.

Bij rekenen moesten in de analyse iets meer opgaven verwijderd worden dan bij taal. Dit betrof vier opgaven uit 2008 die ook in de analyses van 2008 waren verwijderd. Daarnaast is er een opgave uit het Ankerboekje verwijderd. Er zijn dus geen opgaven uit de Eindtoets 2009 verwijderd. Opgemerkt wordt dat het verwijderen van de opgaven alleen gebeurt in de context van dit onderzoek. De door ons verwijderde opgaven waren wel gewoon onderdeel van de Eindtoets. Wij hebben deze opgaven dan ook niet verwijderd omdat het slechte opgaven zouden zijn, maar vanwege de statistische eigenschappen van deze opgaven.

Het verwijderen van de opgaven had overigens weinig effect op de betrouwbaarheid van de afzonderlijke toetsen of op de conclusies.

Behalve de moeilijkheidsgraad van de opgaven rapporteren we ook over de mate waarin de leerlingen deze opgaven beheersen. Hoewel deze twee direct aan elkaar gerelateerd zijn, geeft de mate van beheersing beter aan wat leerlingen nu werkelijk kennen en kunnen dan de moeilijkheidsgraad van de opgaven.

Net als bij de taalvaardigheden zijn de populatieverdelingen op de vaardigheidsschaal geplaatst. Na de omschrijving van de vaardigheidsschaal worden de populatieverdelingen op deze vaardigheidsschaal weergegeven. Merk op dat deze populatieverdelingen niet gecorrigeerd zijn voor de eventuele invloed van de achtergrondkenmerken. Voor de interpretatie van deze verschillen kan beter gekeken worden naar de gecorrigeerde effectschattingen (zie hoofdstuk 4). Merk overigens op dat het gebruik van de term effect niet veronderstelt dat de achtergrondkenmerken de *oorzaak* van het verschil zijn. Er worden geen veronderstellingen gedaan over de causaliteit.

Ook bij rekenen gaat extra aandacht uit naar het jaareffect. Dit jaareffect is het verschil tussen 2009 en 2008. Als deze waarde positief is, dan is de prestatie in 2009 beter dan in 2008, en is de waarde negatief dan is het omgekeerde het geval. Het jaareffect is onder verschillende modellen uitgerekend. De eerste analyses betreffen het model met jaar als enige variabele. In dit model is het ongecorrigeerde jaareffect berekend. Daarna wordt het gecorrigeerde jaareffect gegeven, binnen een model waarin ook de basisvariabelen geslacht, leertijd, stratum en formatiegewicht zijn opgenomen. Hierbij moet opgemerkt worden dat het gecorrigeerde jaareffect het echte jaareffect niet per definitie beter weergeeft (zie Hoofdstuk 4).

Bij rekenvaardigheid hoefde de correctie alleen volgens het hoofdeffecten model uitgevoerd te worden. Het was immers niet nodig om een interactiefactor op te nemen. Het hoofdeffecten model is vervolgens uitgebreid met de aanvullende achtergrondkenmerken type Toets, advies VO, regio, urbanisatiegraad en thuistaal. Ook voor deze achtergrondkenmerken zijn populatieschattingen berekend.

6.1 Rekenen/Wiskunde: Getallen en Bewerkingen

De vaardigheid

De vaardigheid bij het onderdeel Getallen en Bewerkingen betreft een aantal verschillende onderwerpen. De onderwerpen en de opgaven bij deze onderwerpen worden als volgt omschreven:

Getallen en getalrelaties. Hierbij staat centraal het doorzien van de structuur van de telrij, de structuur van getallen en de relaties tussen getallen.

Hoofdrekenen: optellen en aftrekken. De opgaven bij dit onderwerp betreffen de vaardigheid om de bewerkingen optellen en aftrekken vlot, handig en inzichtelijk te kunnen uitvoeren. Daarbij kan de leerling kennis van getallen, basisoperaties en eigenschappen van bewerkingen inzetten. De leerling moet de opgaven 'uit het hoofd' (dat is zonder uitrekenpapier) oplossen.

Hoofdrekenen: vermenigvuldigen en delen. De opgaven bij dit onderwerp betreffen de vaardigheid om de bewerkingen vermenigvuldigen en delen vlot, handig en inzichtelijk uit te voeren. Daarbij kan de leerling kennis van getallen, basisoperaties en eigenschappen van bewerkingen inzetten. De opgaven moeten 'uit het hoofd' (dat is zonder uitrekenpapier) worden opgelost.

Schattend rekenen. Ook bij schattend rekenen spelen eigenschappen van bewerkingen, het kunnen uitvoeren van basisoperaties en het inzicht in getallen (onder andere in de orde van grootte, de ligging in de getallenrij en de structuur) een belangrijke rol. Vooral afronden en weten wat de orde van grootte van een getal is, zijn bij dit onderdeel erg belangrijk. Bij schattend rekenen wordt van leerlingen verwacht dat zij bewerkingen met afgeronde getallen uitvoeren om de orde van grootte van de uitkomst aan te geven. Ook deze opgaven moeten 'uit het hoofd' (dat is zonder uitrekenpapier) worden opgelost.

Bewerkingen: optellen en aftrekken. Dit onderwerp betreft de bewerkingen optellen en aftrekken, waarbij de leerlingen wel uitrekenpapier mogen gebruiken. De getallenkeuze bij de opgaven van dit onderdeel is meestal ook zodanig dat het nodig of wenselijk is tussenuitkomsten te noteren of een standaardcijferprocedure toe te passen.

Bewerkingen: vermenigvuldigen en delen. Dit onderwerp betreft de bewerkingen vermenigvuldigen en delen waarbij de leerlingen wel uitrekenpapier mogen gebruiken. De getallenkeuze bij de opgaven is ook weer zodanig dat het nodig of wenselijk is tussenuitkomsten te noteren of een standaardcijferprocedure uit te voeren.

Samengestelde bewerkingen. Bij de opgaven van dit onderdeel moeten meerdere operaties (bijvoorbeeld zowel een optelling als een deling) uitgevoerd worden. Daarbij mogen de leerlingen uitrekenpapier gebruiken om tussenuitkomsten te noteren of kunnen de leerlingen een of meer standaardcijferprocedures uitvoeren.

Wat leerlingen kunnen

De schaal Getallen en Bewerkingen bestaat uit 85 opgaven. Hiervan waren er vijf afkomstig uit het Toetsboekje Extra. Deze opgaven zijn alleen gebruikt voor de koppeling tussen de Eindtoetsen uit 2008 en 2009. In de schaal voor de analyses van de achtergrondvariabelen zitten zodoende hooguit 80 opgaven. Van deze opgaven zijn er 25 afkomstig uit de Niveautoetsen (alle zijn zowel in 2008 als in 2009 afgenomen), en 55 uit Eindtoetsen (25 in 2008, 25 in 2009 en 5 ankeropgaven). Uit de analyses zijn vier opgaven verwijderd: drie opgaven die ook voor alleen onze analyses in 2008 verwijderd waren uit de eindtoets van 2008, en een ankeropgave. Doordat de itemkalibratie is uitgevoerd op de gehele rekenschaal blijft de koppeling echter groot genoeg. De gebruikte schaal voor de meting van de vaardigheid Getallen en Bewerkingen vond dus plaats met behulp van 75 opgaven. De 25 opgaven uit de Eindtoets 2009 worden hieronder beschreven.

Tabel 6.1.1 Posities van de opgaven Getallen en Bewerkingen op de vaardigheidsschaal geordend op volgorde van afname, en de typering van de opgaven naar moeilijkheidsgraad voor de vijf typische leerlingen.

Taak	Item (tk)	Label	JPON_p50	JPON_p80	ZLV	LV	ST	HV	ZHV
Rekenen 1 ('09)	01	R1 - i01	150	240	1	1	0	0	0
Rekenen 1 ('09)	02	R1 - i02	162	216	1	0	0	0	0
Rekenen 1 ('09)	04	R1 - i04	118	208	1	0	0	0	0
Rekenen 1 ('09)	12	R1 - i12	252	320	2	2	2	1	1
Rekenen 1 ('09)	15	R1 - i15	144	234	1	1	0	0	0
Rekenen 1 ('09)	17	R1 - i17	117	253	1	1	1	0	0
Rekenen 2 ('09)	01	R2 - i01	180	247	1	1	0	0	0
Rekenen 2 ('09)	03	R2 - i03	185	239	1	1	0	0	0
Rekenen 2 ('09)	04	R2 - i04	181	272	1	1	1	0	0
Rekenen 2 ('09)	06	R2 - i06	187	255	2	1	1	0	0
Rekenen 2 ('09)	08	R2 - i08	212	266	2	1	1	0	0
Rekenen 2 ('09)	10	R2 - i10	158	248	1	1	0	0	0
Rekenen 2 ('09)	11	R2 - i11	159	226	1	1	0	0	0
Rekenen 2 ('09)	12	R2 - i12	207	298	2	1	1	1	0
Rekenen 2 ('09)	13	R2 - i13	229	319	2	2	1	1	1
Rekenen 2 ('09)	15	R2 - i15	222	276	2	2	1	0	0
Rekenen 2 ('09)	16	R2 - i16	239	307	2	2	1	1	0
Rekenen 2 ('09)	18	R2 - i18	173	240	1	1	0	0	0
Rekenen 2 ('09)	19	R2 - i19	198	243	2	1	0	0	0
Rekenen 2 ('09)	20	R2 - i20	222	312	2	2	1	1	0
Rekenen 3 ('09)	01	R3 - i01	118	253	1	1	1	0	0
Rekenen 3 ('09)	05	R3 - i05	232	286	2	2	1	1	0
Rekenen 3 ('09)	09	R3 - i09	179	234	1	1	0	0	0
Rekenen 3 ('09)	13	R3 - i13	212	266	2	1	1	0	0
Rekenen 3 ('09)	18	R3 - i18	145	199	1	0	0	0	0

Tabel 6.1.2 Aantal opgaven Getallen en Bewerkingen per moeilijkheidsgraad in de Eindtoets 2009 voor de vijf typische leerlingen.

Moeilijkheidsgraad \Typische leerling	ZLV	LV	ST	HV	ZHV
0 Gemakkelijk ($P > .80$)	0	3	12	19	23
1 Uitdagend ($.50 \leq P \leq .80$)	14	16	12	6	2
2 Moeilijk ($P < .50$)	11	6	1	0	0

Tabel 6.1.1 beschrijft de vaardigheid in termen van de mate waarin de typische leerlingen verschillende opgaven beheersen.

De zeer laag vaardige (ZLV) leerling beheerst geen van de EB opgaven goed. Van de 25 opgaven beheerst een percentiel 10 leerling er 11 opgaven slecht (De kans om de opgave goed te maken is kleiner dan een half). Dit betreft de opgaven R1-i12, R2-i6 /i8 /i12 /i13 /i15 /i16 /i19 /i20 en R3-i5 /i13. De overige opgaven zijn uitdagend voor deze leerling.

De laag vaardige (LV) leerling beheerst 3 van de EB uit 2009 opgaven goed (de kans op het correct maken van de opgave is groter dan ,80): R1-i2 /i4 en R3-i18, en 6 opgaven relatief slecht: R1-i12, R2- i13 /i15 /i16 /i20 en R3-i5. Merk op dat deze 6 uiteraard ook tussen de 11 opgaven zitten die door een ZLV- leerling slecht gedaan worden. De overige opgaven zijn uitdagend voor een LV- leerling.

De standaard (ST) leerling beheerst 12 van de 25 EB opgaven goed: dat zijn de 3 opgaven die een P-25 leerling ook goed maakt, aangevuld met R1-i1 /i15, R2-i1 /i3 /i10 /i11 /i18 /i19 en R3-i9. De enige opgave die voor een ST-leerling moeilijk is, is opgave R1-i12. De overige opgaven zijn uitdagend voor deze leerling.

De hoog vaardige (HV) leerling heeft geen moeite met een van de 25 opgaven; 6 van de opgaven zijn nog wel uitdagend voor deze leerling: R1-i12, R2-i12 /i13 /i16 /i20 en R3-i5. De overige 19 opgaven zijn gemakkelijk voor deze leerling.

De zeer hoog vaardige (ZHV) leerling beheerst 23 van de EB opgaven goed (NRS). De overige twee opgaven R1-i12 en R2-i13 zijn uitdagend voor deze leerling.

Wat groepen van leerlingen kunnen

Naast een beschrijving in termen van opgaven en wat leerlingen op de verschillende vaardigheidsniveaus wel en niet kunnen, is de vaardigheid ook te beschrijven in termen van deelpopulaties, gecategoriseerd op achtergrondvariabelen. De resultaten van die analyses staan voor Getallen en Bewerkingen in tabel 6.1.3. De gemiddelden van de deelpopulaties zijn op dezelfde schaal gegeven als de opgaven. Merk op dat dit marginale schattingen zijn. Er wordt hierbij namelijk niet gecorrigeerd voor eventuele verschillen in de verdeling van de overige achtergrondvariabelen. Ook zijn de standaarddeviaties van de verschillende deelpopulaties weergegeven. Als de standaarddeviatie kleiner is dan 50, dan is de groep homogener dan de gehele populatie. De leerlingen lijken dan wat betreft hun vaardigheid dan meer op elkaar.

Bij Getallen en Bewerkingen zijn de resultaten voor 2008 en 2009 zeer vergelijkbaar. De gemiddelde vaardigheid is voor 2009 zelfs gelijk aan die in 2008. Omdat er geen interactie gevonden was tussen jaar en een van de basisvariabelen, zijn ook de gemiddelden van de verschillende subgroepen in 2009 vrijwel gelijk aan die van 2008. De opmerkingen die daarover gemaakt zijn in 2008 gelden zodoende nog steeds voor 2009 (Cito, 2008).

Net als in 2008 geldt in 2009 dat het gemiddelde van leerlingen op de stratum 1-scholen, de jongens, de niet-vertraagde ('reguliere') leerlingen, en de leerlingen met formatiegewicht 1.00 alle iets boven het algemene gemiddelde liggen. Al deze waarden zijn groter dan 250. Voor de leerlingen van de stratum 2-scholen ligt het gemiddelde daar iets onder en daar nog iets onder liggen de gemiddelden van de stratum 3-scholen. Ook het gemiddelde van de meisjes ligt onder de 250. De verschillen zijn hier niet zeer groot. Groter zijn de verschillen van deze groepen met de groepen van vertraagde leerlingen, de leerlingen met formatiegewicht 1.25 en de leerlingen met formatiegewicht 1.90 ('allochtone kinderen'). Al deze gemiddelden liggen nog wel boven het niveau van een LV-leerling. In tegenstelling tot wat bij de taalschalen gevonden werd, ligt de gemiddelde vaardigheid van de 1.90 leerling boven die van de 1.25 leerling. De mogelijke taalachterstand lijkt bij deze opgaven een minder grote rol te spelen.

Het is geen verrassende bevinding dat de standaarddeviatie bij de groepen met een verschillend VO-advies duidelijk kleiner is dan 50. De leerlingen krijgen immers een verschillend advies juist omdat zij wat betreft hun vaardigheid op elkaar lijken. De ZHV-leerlingen komen eerder bij elkaar in de groep advies VWO en de ZLV-leerlingen eerder in de groep advies vmbo-BB. Dit zijn duidelijk homogener groepen dan de gehele groep: leerlingen met eenzelfde VO-advies lijken meer op elkaar wat betreft hun vaardigheidsniveau dan twee willekeurige leerlingen. Dit resultaat wordt overigens ook bij de andere rekenvaardigheden gevonden (evenals bij de taalvaardigheden).

Tabel 6.1.3 Verdelingen van de leerlingen op de vaardigheid Rekenen/Wiskunde – Getallen en Bewerkingen

Jaar Variabele	Niveau	2008		2009		
		Gem.	SD	Gem.	SD	
Jaar		250	50	250	50	
Geslacht	Jongens	259	49	259	49	
	Meisjes	241	49	241	49	
Leertijd	Regulier	257	48	256	48	
	Vertraagd	224	49	224	48	
Stratum	Stratum 1	254	49	253	49	
	Stratum 2	246	50	246	50	
	Stratum 3	238	50	238	51	
Formatiegewicht	1.00	255	49	254	49	
	1.25	227	50	226	49	
	1.90	234	50	234	50	
Toets	EB	252	49	251	48	
	NT	157	50	159	49	
Advies VO	vmbo-BB	177	37	176	37	
	vmbo-KB	209	37	208	37	
	vmbo-GT	232	37	231	37	
	Havo	268	37	266	37	
	VWO	314	37	313	37	
Regio	Noord	244	50	243	50	
	Oost	249	50	249	50	
	West	250	50	250	50	
	Zuid	254	50	254	50	
Urbanisatiegraad	Zeer Sterk	247	51	247	50	
	Sterk	250	50	250	50	
	Matig	252	50	252	50	
	Weinig	251	50	251	50	
	Niet	249	50	249	50	
Thuis taal	Alleen NLs	252	50	251	50	
	NLs + bu	244	51	244	51	
	Alleen bu	239	50	239	51	
	Turks	240	50	240	51	
	Arabisch	235	50	236	50	
	Surinaams	219	51	218	51	
	WestEuro	247	51	245	51	
	Ter referentie:	Typische leerling	Schaalwaarde (geen verschil tussen de jaren)			
		Zeer laag vaardige leerling (ZLV)	186		186	
	Laag vaardige leerling (LV)	216		216		
	Standaard leerling (ST)	250		250		
	Hoog vaardige leerling (HV)	284		284		
	Zeer hoog vaardige leerling (ZHV)	314		314		

De effecten van de achtergrondvariabelen

Behalve de gemiddelden van de onderscheiden groepen willen we ook weten of de gevonden verschillen relevant zijn: hebben zij enig effect? Ook willen we de verschillen tussen de groepen corrigeren voor de andere variabelen. De effectschattingen geven deze informatie en de resultaten zijn weergegeven in tabel 6.1.4 voor de jaareffecten en 6.1.5 voor de overige effecten. In hoofdstuk 4 (tabel 4.2) is ook te vinden hoe deze getallen zijn te interpreteren.

Tabel 6.1.4 Jaareffect Rekenen/Wiskunde – GB: Getallen en Bewerkingen

Variabele: Contrast	Model	Effectgrootte	Effectkwalificatie
Jaar: 2009 – 2008	Alleen jaareffect	0,00	geen effect
Jaar: 2009 – 2008	Basismodel rekenen (G L S F J)	-0,01	geen effect

Net als bij de marginalen geven de effectschattingen in het model met alleen de variabele jaar aan dat er geen verschil is tussen de twee jaren: het effect is 0,00. Dit is (uiteraard) niet significant verschillend van 0. De correctie voor andere achtergrondkenmerken heeft amper effect. Corrigeren we ook voor de variabele advies VO dan wordt het effect -0,03, wat nog steeds als “geen effect” geclassificeerd wordt. Wat het probleem is als er echter voor die extra variabele gecorrigeerd wordt, is duidelijk gemaakt in hoofdstuk 4.

De resultaten van het onderzoek naar de relevantie van de gecorrigeerde effecten is gegeven in tabel 6.1.5. Merk op dat alle gevonden effecten (klein, matig en groot) significant zijn. De effecten van de overige variabelen lijken ook erg op de effecten zoals die in 2008 gevonden zijn.

Bij de vaardigheid Rekenen/Wiskunde - Getallen en Bewerkingen scoren jongens nog steeds hoger dan meisjes (klein effect) en ‘reguliere’ leerlingen over het algemeen een hogere vaardigheid hebben dan leerlingen die vertraagd zijn (matig effect). Deze effecten zijn vrijwel even groot als in 2008. Als gecorrigeerd wordt voor het formatiegewicht van leerlingen op die scholen, valt op dat er geen effect te vinden is voor stratum. Voor formatiegewicht zijn er wel effecten, die vergelijkbaar zijn met die van 2008. Merk op dat het effect van leerlingen met formatiegewicht 1.25 en 1.90 aangeeft dat leerlingen met een gewicht 1.90 beter presteren dan die met gewicht 1.25. Doordat de leerlingen met gewicht 1.90 iets meer vooruit zijn gegaan wordt het effect nu echt geclassificeerd als een “klein effect” (was in 2008 “geen effect”).

Zoals bij de taalvaardigheden ook gevonden is, zien we dat ook rekenvaardigheid dat de leerlingen die de EB doen uiteraard duidelijk vaardiger zijn dan de leerlingen die de NT doen. Het effect is hier groter dan bij taalvaardigheden en onveranderd over de jaren. De verschillen tussen de diverse niveaus van het VO waar de leerlingen op ingeschat zijn, zijn behoorlijk groot, zij het dat de verschillen iets kleiner zijn dan in 2008. Het verschil tussen vmbo-BB en vwo (de som van alle effecten van de variabele advies VO) blijft echter met een waarde van 3,67 zeer groot: het verschil tussen een gemiddelde vmbo-BB leerling en een gemiddelde vwo-leerling volgens het advies van de leraar is ligt tussen de 3 en 4 standaarddeviaties.

Regio en urbanisatiegraad hebben bijna geen effect als er gecorrigeerd is op formatiegewicht en stratum. Er is nog wel steeds een klein effect gevonden tussen leerlingen in het zuiden en het noorden van Nederland, waarbij leerlingen in het zuiden, gecorrigeerd voor andere variabelen beter presteren. Bij de variabele Thuis taal valt op dat er geen effecten zijn, zodra er gecorrigeerd wordt voor geslacht, leertijd, stratum en formatiegewicht. Bij de specifieke talen is alleen een klein effect te zien bij leerlingen die Surinaams of Antilliaans spreken. Ook dit is vergelijkbaar met 2008.

Tabel 6.1.5 Effectgrootte op de vaardigheid Rekenen/Wiskunde – Getallen en Bewerkingen

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	-0,43	klein
Leertijd	Vertraagd – Regulier	-0,67	matig
Stratum	Stratum2 – Stratum 1	-0,06	geen effect
	Stratum 3 – Stratum 2	-0,02	geen effect
	Stratum 3 – Stratum 1	-0,08	geen effect
Formatiegewicht	F1.25 – F1.00	-0,48	klein
	F1.90 – F1.25	0,24	klein
	F1.90 – F1.00	-0,24	klein
Toets	NT – EB	-1,75	groot
Advies VO	vmbo-KB – vmbo-BB	0,86	groot
	vmbo-GT – vmbo-KB	0,61	matig
	Havo – vmbo-GT	0,93	groot
	VWO – Havo	1,27	groot
Regio	Oost – Noord	0,11	geen effect
	West – Noord	0,14	geen effect
	Zuid – Noord	0,24	klein
	West – Oost	0,03	geen effect
	Zuid – Oost	0,12	geen effect
	Zuid – West	0,09	geen effect
Urbanisatiegraad	Sterk – Zeer sterk	-0,07	geen effect
	Matig – Sterk	0,01	geen effect
	Weinig – Matig	-0,04	geen effect
	Niet – Weinig	-0,01	geen effect
Thuis taal	NLs+Ander – Alleen NLs	0,06	geen effect
	Alleen ander – Alleen NLs	0,10	geen effect
	Alleen ander – NLs+Ander	0,03	geen effect
	Turks – Alleen NLs	0,15	geen effect
	Arabisch – Alleen NLs	0,02	geen effect
	Surinaams/Antilliaans – Alleen NLs	-0,43	klein
	West-Europees – Alleen NLs	0,03	geen effect

6.2 Rekenen/Wiskunde: Breuken, Procenten en Verhoudingen

De vaardigheid

Verhoudingen kunnen beschreven worden in verhoudingentaal (één op de tien kinderen), in breukentaal (een kwart van de bevolking) of met procenten (20% van de aanwezigen). Begrip van verhoudingen houdt in dat de relatie kan worden gelegd tussen die verschillende beschrijvingen van verhoudingen. Bij het onderdeel Verhoudingen, breuken en procenten komen opgaven uit de volgende onderwerpen voor:

- *Breuken.* Bij dit onderwerp gaat het om basiskennis en elementaire begrippen die nodig zijn om met breuken en gemengde getallen te kunnen werken en rekenen. Concreet betekent dat onder andere: breuken op een getallenlijn plaatsen, breuken omzetten in kommagetallen, breuken vereenvoudigen en breuken als gemengd getal schrijven. Daarnaast moeten leerlingen elementaire operaties (optellen, aftrekken, vermenigvuldigen en delen) met breuken kunnen uitvoeren en die vaardigheid in contexten kunnen toepassen. De breuken en gemengde getallen die daarbij voorkomen hebben een hoge gebruikswaarde.
- *Verhoudingen.* Bij dit onderwerp moeten kinderen elementaire verhoudingsproblemen oplossen, waarbij ook berekeningen uitgevoerd moeten worden.
- *Procenten.* Bij dit onderwerp staat allereerst het begrijpen van wat procenten zijn centraal. Dat betekent, onder andere, inzien dat het geheel 100% is en de grootte van een deel van een geheel met procenten kunnen aanduiden. Ook moet de relatie tussen procenten enerzijds en breuken en verhoudingen anderzijds doorzien worden. Percentages worden gebruikt in allerlei contexten. Daarbij staat niet alleen centraal het begrip van en de vaardigheid in het rekenen met percentages, maar ook kennis van begrippen en afspraken in bepaalde sectoren. Naast het toepassen van de procedure waarbij eerst één procent uitgerekend wordt, is het soms efficiënter het percentage om te zetten naar een breuk of gebruik te maken van verhoudingen.

Wat leerlingen kunnen

De schaal Breuken, Procenten en Verhoudingen is gemaakt met behulp van 72 opgaven. Hiervan waren er 5 uit het Toetsboekje Extra. Deze opgaven zijn alleen gebruikt voor de koppeling tussen eindtoetsen uit 2008 en 2009. In de schaal voor de analyses van de achtergrondvariabelen zitten zodoende op ten hoogste 67 opgaven. Van deze opgaven zijn er 20 afkomstig uit de niveautoetsen (allen zijn zowel in 2008 als in 2009 afgenomen), en 47 uit eindtoetsen (20 in 2008, 20 in 2009 en 7 ankeropgaven). Uit de analyses is een opgave verwijderd, namelijk een opgave die ook (alleen voor deze analyses) in 2008 verwijderd was uit de eindtoets van 2008. De gebruikte schaal voor de meting van de vaardigheid Breuken, Procenten en Verhoudingen vond dus plaats met behulp van 66 opgaven. De 20 opgaven uit de eindtoets 2009 worden hieronder beschreven.

In tabel 6.2.1 staan de locaties van de opgaven weergegeven evenals de classificatie naar moeilijkheidsgraad voor de vijf typische leerlingen. De opgaven en de figuren staan in Bijlage 2. Op basis van deze tabel kunnen de aantallen opgaven bepaald worden die als moeilijk, uitdagend of gemakkelijk gezien worden voor de typische leerlingen die worden gegeven in tabel 6.2.2. Ook vormt tabel 6.2.1 de basis van het bepalen van het beheersingsniveau bij de verschillende opgaven voor de verschillende typische leerlingen. Hoe erg het is dat op bepaalde niveaus de opgaven niet beheerst worden, is een inhoudelijke discussie.

Tabel 6.2.1 Posities van de opgaven Breuken, Procenten en Verhoudingen op de vaardigheidsschaal geordend op volgorde van afname, en de typering van de opgaven naar moeilijkheidsgraad voor de vijf typische leerlingen.

Taak	Item (tk)	Label	JPON_p50	JPON_p80	ZLV	LV	ST	HV	ZHV
Rekenen 1 ('09)	03	R1 - i03	195	252	2	1	1	0	0
Rekenen 1 ('09)	05	R1 - i05	209	280	2	1	1	0	0
Rekenen 1 ('09)	06	R1 - i06	213	283	2	1	1	0	0
Rekenen 1 ('09)	08	R1 - i08	184	230	1	1	0	0	0
Rekenen 1 ('09)	09	R1 - i09	241	298	2	2	1	1	0
Rekenen 1 ('09)	13	R1 - i13	222	292	2	2	1	1	0
Rekenen 1 ('09)	16	R1 - i16	211	251	2	1	1	0	0
Rekenen 2 ('09)	02	R2 - i02	191	238	2	1	0	0	0
Rekenen 2 ('09)	05	R2 - i05	195	289	2	1	1	1	0
Rekenen 2 ('09)	07	R2 - i07	202	249	2	1	0	0	0
Rekenen 2 ('09)	09	R2 - i09	224	318	2	2	1	1	1
Rekenen 2 ('09)	14	R2 - i14	161	218	1	1	0	0	0
Rekenen 2 ('09)	17	R2 - i17	155	249	1	1	0	0	0
Rekenen 3 ('09)	02	R3 - i02	210	267	2	1	1	0	0
Rekenen 3 ('09)	06	R3 - i06	238	285	2	2	1	1	0
Rekenen 3 ('09)	10	R3 - i10	170	240	1	1	0	0	0
Rekenen 3 ('09)	14	R3 - i14	213	269	2	1	1	0	0
Rekenen 3 ('09)	17	R3 - i17	169	240	1	1	0	0	0
Rekenen 3 ('09)	19	R3 - i19	219	289	2	2	1	1	0
Rekenen 3 ('09)	20	R3 - i20	213	253	2	1	1	0	0

Tabel 6.2.2 Aantal opgaven Breuken, Procenten en Verhoudingen per moeilijkheidsgraad in de Eindtoets 2009 voor de vijf typische leerlingen.

Aantal opgaven per type en niveau	ZLV	LV	ST	HV	ZHV
0 Gemakkelijk ($P > .80$)	0	0	7	14	19
1 Uitdagend ($.50 \leq P \leq .80$)	5	15	13	6	1
2 Moeilijk ($P < .50$)	15	5	0	0	0

De beheersingsniveaus van de typische leerlingen zijn de volgende.

De zeer laag vaardige (ZLV) leerling beheerst geen van de EB-opgaven goed. Van de 20 opgaven zijn er 5 uitdagend (kans goed te maken tussen ,5 en ,8): R1-i8, R2-i14 /i17 en R3-i10 /i17. De overige 15 beheerst een percentiel 10 leerling slecht.

De laag vaardige (LV) leerling beheerst 5 opgaven slecht: R1-i9 /i13, R2-i9 en R3-i6 /i19. De overige 15 opgaven zijn uitdagend voor een LV-leerling. Geen van de opgaven wordt door een dergelijke leerling dus goed beheerst.

De standaard (ST) leerling beheerst 7 opgaven goed: R1-i8, R2-i2 /i7 /i14 /i17 en R3-i10 /i17. De overige 13 zijn uitdagend voor de een dergelijke leerling. Geen van de opgaven wordt door een dergelijke leerling dus slecht beheerst. Dit betekent dat ook voor alle leerlingen van een hoger percentiel, er dus ook geen van de opgaven slecht beheerst wordt.

De hoog vaardige (HV) leerling beheerst 14 opgaven goed. De overige 6 opgaven zijn uitdagend voor een dergelijke leerling: R1-i9 /i13, R2-i5 /i9 en R3-i6 /i19.

De zeer hoog vaardige (ZHV) leerling beheerst 19 van de EB opgaven goed. De enige uitdagende opgaven voor een dergelijk leerling bij deze breuken, procenten verhoudingen is R2-i9.

Tabel 6.2.3 Verdelingen van de leerlingen op de vaardigheid Rekenen/Wiskunde – Breuken, Procenten en Verhoudingen

Jaar Variabele	Niveau	2008		2009	
		Gem.	SD	Gem.	SD
Jaar		250	50	250	50
Geslacht	Jongens	261	49	261	49
	Meisjes	239	49	239	48
Leertijd	Regulier	257	48	256	48
	Vertraagd	222	48	222	48
Stratum	Stratum 1	255	49	255	49
	Stratum 2	246	50	246	50
	Stratum 3	233	50	233	51
Formatiegewicht	1.00	256	49	255	49
	1.25	226	49	226	49
	1.90	228	50	228	50
Toets	EB	252	49	251	48
	NT	157	50	159	49
Advies VO	vmbo-BB	177	36	176	37
	vmbo-KB	207	36	207	36
	vmbo-GT	231	36	230	36
	Havo	269	36	267	36
	VWO	317	36	316	36
Regio	Noord	245	50	245	50
	Oost	249	50	250	50
	West	250	50	250	50
	Zuid	253	50	253	50
Urbanisatiegraad	Zeer Sterk	244	51	245	51
	Sterk	249	50	249	50
	Matig	252	50	252	50
	Weinig	252	50	251	49
	Niet	250	49	249	49
Thuis taal	Alleen NLs	252	49	252	49
	NLs + bu	239	51	239	51
	Alleen bu	233	50	233	51
	Turks	230	50	230	51
	Arabisch	231	50	231	50
	Surinaams	213	51	213	51
	WestEuro	246	51	244	51
	Ter referentie:	Typische leerling	Schaalwaarde (geen verschil tussen de jaren)		
	Zeer laag vaardige leerling (ZLV)	186		186	
	Laag vaardige leerling (LV)	216		216	
	Standaard leerling (ST)	250		250	
	Hoog vaardige leerling (HV)	284		284	
	Zeer hoog vaardige leerling (ZHV)	314		314	

Wat groepen van leerlingen kunnen

Naast een beschrijving in termen van opgaven en wat leerlingen op de verschillende vaardigheidsniveaus wel en niet kunnen, is de vaardigheid ook te beschrijven in termen van de geobserveerde populatieverdelingen. Dat wordt gedaan in tabel 6.2.3.

In tabel 6.2.3 staat de verdeling van de vaardigheid over de gehele groep leerlingen in 2008 en 2009. Daarnaast wordt de verdeling gegeven per subgroep binnen de twee jaren. De posities zijn gegeven op dezelfde schaal als de opgaven. Merk op dat dit marginale schattingen zijn: er wordt hierbij niet gecorrigeerd voor eventuele verschillen in de verdeling van de overige achtergrondvariabelen. De resultaten zijn vergelijkbaar met die bij de rekenschaal Getallen en Bewerkingen.

Bij BPV geldt net zoals bij GB dat de resultaten voor 2008 en 2009 zeer vergelijkbaar zijn. Ook hier zijn is de algemene gemiddelde vaardigheid bij voor 2009 gelijk aan die in 2008. Doordat ook hier er geen interactie gevonden was tussen jaar en een van de basisvariabelen, zijn ook de gemiddelden van de verschillende subgroepen vrijwel gelijk in 2008 en 2009. De opmerkingen die daarover gemaakt zijn in 2008 gelden zodoende nog steeds voor 2009 (Cito, 2008). Doordat de resultaten van de rekenschalen zo op elkaar lijken geldt in grote lijnen de beschrijving van de marginalen bij “getallen en bewerkingen” ook voor “breuken procenten en verhoudingen”.

De effecten van de achtergrondvariabelen

Ook bij het schatten van de effecten zijn de verschillen tussen BPV en GB gering. De gevonden jaareffecten (ongecorrigeerd en gecorrigeerd) bij BPV zijn bijvoorbeeld gelijk aan die bij GB.

Tabel 6.2.4 Jaareffect Rekenen/Wiskunde – BPV: Breuken, Procenten en Verhoudingen

Variabele: Contrast	Model	Effectgrootte	Effectkwalificatie
Jaar: 2009 – 2008	Alleen jaareffect	0,00	geen effect
Jaar: 2009 – 2008	Basismodel rekenen (G L S F J)	-0,01	geen effect

Als er ook gecorrigeerd wordt voor andere factoren, dan heeft dit amper effect. Als in het model ook gecorrigeerd wordt voor advies VO dan wordt het effect -0,03, wat nog steeds als “geen effect” geclassificeerd wordt. Bij extra correcte bovenop het basismodel voor leertijd dan wordt het jaareffect -0,02.

Dat deze resultaten zeer op elkaar lijken is niet verbazingwekkend aangezien deze vaardigheden hoog met elkaar correleren. Als we kijken naar de latente correlaties dan zijn de correlaties tussen de drie rekenvaardigheden allen groter dan 0,90. Dat is ook niet vreemd omdat voor het rekenen vaak ook combinaties van rekenvaardigheden gebruikt worden om een vraag op te lossen. De beschrijving van de effecten bij de schaal Getallen en Bewerkingen is zodoende ook hier in grote lijnen van toepassing. Dit geldt overigens ook voor de schaal “Meten, Meetkunde, Tijd en Geld”.

Tabel 6.2.5 Effectgrootte op de vaardigheid Rekenen/Wiskunde – Breuken, Procenten en Verhoudingen

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	-0,51	matig
Leertijd	Vertraagd – Regulier	-0,71	matig
Stratum	Stratum2 – Stratum 1	-0,09	geen effect
	Stratum 3 – Stratum 2	-0,07	geen effect
	Stratum 3 – Stratum 1	-0,16	geen effect
Formatiegewicht	F1.25 – F1.00	-0,50	matig
	F1.90 – F1.25	0,15	geen effect
	F1.90 – F1.00	-0,35	klein
Toets	NT – EB	-1,76	groot
Advies VO	vmbo-KB – vmbo-BB	0,83	groot
	vmbo-GT – vmbo-KB	0,63	matig
	Havo – vmbo-GT	1,02	groot
	VWO – Havo	1,37	groot
Regio	Oost – Noord	0,10	geen effect
	West – Noord	0,13	geen effect
	Zuid – Noord	0,19	geen effect
	West – Oost	0,02	geen effect
	Zuid – Oost	0,08	geen effect
	Zuid – West	0,06	geen effect
Urbanisatiegraad	Sterk – Zeer sterk	-0,07	geen effect
	Matig – Sterk	0,02	geen effect
	Weinig – Matig	-0,04	geen effect
	Niet – Weinig	-0,03	geen effect
Thuis taal	NLs+Ander – Alleen NLs	0,00	geen effect
	Alleen ander – Alleen NLs	0,03	geen effect
	Alleen ander – NLs+Ander	0,02	geen effect
	Turks – Alleen NLs	0,01	geen effect
	Arabisch – Alleen NLs	-0,01	geen effect
	Surinaams/Antilliaans – Alleen NLs	-0,50	klein
	West-Europees – Alleen NLs	0,02	geen effect

6.3 Rekenen/Wiskunde: Meten, Meetkunde, Tijd en Geld

De vaardigheid

In zekere zin is deze schaal een verzamelschaal. Bij de vaardigheid *Meten* zijn verschillende onderwerpen te onderscheiden, waar dan de onderdelen meetkunde, tijd en geld aan toegevoegd kunnen worden. Bij de meeste van deze onderdelen kan gesteld worden dat het hier om specifieke vormen van toepassingen van rekenen gaat. De opgaven kunnen als volgt onderverdeeld worden:

- *Lengte*. Bij dit onderwerp gaat het om basiskennis en begrip, zoals het aflezen van meetinstrumenten en het onderling herleiden van lengtematen, notie van lengtematen en toepassen.
- *Oppervlakte*. Bij dit onderwerp gaat het om basiskennis en begrip met betrekking tot het meetaspect oppervlakte, zoals het afpassen met natuurlijke oppervlaktematen, onderling herleiden van enkele veel voorkomende oppervlaktematen, notie van maten en het kunnen uitvoeren van oppervlakteberekeningen.
- *Inhoud*. Bij dit onderwerp gaat het om basiskennis en begrip, afpassen met natuurlijke maten, onderling herleiden van enkele veel voorkomen inhoudsmaten, notie van maten en toepassen.
- *Gewicht*. Bij dit onderwerp gaat het om basiskennis en begrip, aflezen van meetinstrumenten, onderling herleiden van maten, notie van maten en toepassen.
- *Meetkunde*. Hierbij gaat het om eenvoudige noties en begrippen waarmee de ruimte meetkundig geordend, beschreven en verklaard kan worden. Centraal bij dit onderwerp staat de vaardigheid 'ruimtelijk redeneren'.
- *Tijd*. Bij dit onderwerp gaat het om het rekenen met tijd in toepassingssituaties.
- *Geld*. Hierbij gaat het om toepassingsgericht rekenen met geld waarbij specifieke handelingen met munten en bankbiljetten uitgevoerd moeten worden.
- *Toepassingen*. Bij dit onderwerp overschrijden we de afzonderlijke meetgebieden. Dat is bijvoorbeeld het geval als een relatie gelegd moet worden tussen tijd en afstand of tussen omtrek en oppervlakte of oppervlakte en prijs.

Wat leerlingen kunnen

De schaal Meten, Meetkunde, Tijd en Geld is gemaakt met behulp van 63 opgaven. Hiervan waren er 10 uit het Toetsboekje Extra. Deze opgaven zijn alleen gebruikt voor de koppeling tussen eindtoetsen uit 2008 en 2009. In de schaal voor de analyses van de achtergrondvariabelen zitten zodoende op ten hoogste 53 opgaven. Van deze opgaven zijn er 15 afkomstig uit de niveautoetsen (allen zijn zowel in 2008 als in 2009 afgenomen), en 38 uit eindtoetsen (15 in 2008, 15 in 2009 en 8 ankeropgaven). Uit de analyses is een opgaven uit de eindtoets van 2008 verwijderd. De opgave is alleen verwijderd uit de analyses voor JPON en was ook in de analyses van 2008 verwijderd. De gebruikte schaal voor de meting van de vaardigheid Meten, Meetkunde, Tijd en Geld vond dus plaats met behulp van 52 opgaven. De 15 opgaven uit de eindtoets 2009 worden hieronder beschreven.

Net als bij de andere twee rekenvaardigheden worden de posities van de opgaven gegeven, de classificaties van de moeilijkheidsgraad voor de vijf typische leerlingen en de beheersing van de vijf Typische leerlingen in termen van de opgaven. De beheersingsniveaus staan beschreven onder tabel 6.3.1 en 6.3.2. De opgaven en figuren zijn te vinden in Bijlage 2. Merk op dat geen van de opgaven voor MMTG in Rekentaak 2 zaten. Dat betekent dat alle opgaven bij deze vaardigheid gemaakt mochten worden met hulp van kladpapier.

Tabel 6.3.1 Posities van de opgaven Meten, Meetkunde, Tijd en Geld op de vaardigheidsschaal geordend op volgorde van afname, en de typering van de opgaven naar moeilijkheidsgraad voor de vijf typische leerlingen.

Taak	Item (tk)	Label	JPON_p50	JPON_p80	ZLV	LV	ST	HV	ZHV
Rekenen 1 ('09)	07	R1 - i07	226	273	2	2	1	0	0
Rekenen 1 ('09)	10	R1 - i10	151	244	1	1	0	0	0
Rekenen 1 ('09)	11	R1 - i11	155	249	1	1	0	0	0
Rekenen 1 ('09)	14	R1 - i14	197	243	2	1	0	0	0
Rekenen 1 ('09)	18	R1 - i18	149	206	1	0	0	0	0
Rekenen 1 ('09)	19	R1 - i19	235	306	2	2	1	1	0
Rekenen 1 ('09)	20	R1 - i20	190	284	2	1	1	0	0
Rekenen 3 ('09)	03	R3 - i03	258	314	2	2	2	1	0
Rekenen 3 ('09)	04	R3 - i04	194	241	2	1	0	0	0
Rekenen 3 ('09)	07	R3 - i07	218	265	2	2	1	0	0
Rekenen 3 ('09)	08	R3 - i08	162	232	1	1	0	0	0
Rekenen 3 ('09)	11	R3 - i11	218	259	2	2	1	0	0
Rekenen 3 ('09)	12	R3 - i12	186	280	1	1	1	0	0
Rekenen 3 ('09)	15	R3 - i15	201	257	2	1	1	0	0
Rekenen 3 ('09)	16	R3 - i16	215	285	2	1	1	1	0

Tabel 6.3.2 Aantal opgaven Meten, Meetkunde, Tijd en Geld per moeilijkheidsgraad in de Eindtoets 2009 voor de vijf typische leerlingen.

Aantal opgaven per type en niveau		ZLV	LV	ST	HV	ZHV
0	Gemakkelijk ($P > .80$)	0	1	6	12	15
1	Uitdagend ($.50 \leq P \leq .80$)	5	9	8	3	0
2	Moeilijk ($P < .50$)	10	5	1	0	0

De beheersingsniveaus van de typische leerlingen zijn de volgende.

De zeer laag vaardige (ZLV) leerling beheerst geen van de EB opgaven goed. Van de 15 opgaven zijn er 5 uitdagend: R1-i10 /i11 /i18 en R3-i8 /i12. De overige 10 beheerst een dergelijke leerling slecht.

De laag vaardige (LV) leerling beheerst 5 opgaven slecht: R1-i7 /i19 en R3-i3 /i7 /i11. Opgaven R1-i18 wordt goed beheerst door een LV-leerling, terwijl de overige 9 opgaven zijn uitdagend zijn.

De standaard (ST) leerling beheerst 6 opgaven goed: R1-i10 /i11 /i14 /i18 en R3-i4 /i8. Alleen opgave R3-i3 wordt nog slecht beheerst. De overige 8 opgaven zijn uitdagend voor een dergelijke leerling.

De hoog vaardige (HV) leerling beheerst 12 opgaven goed. De overige 3 opgaven zijn uitdagend: R1-i9 en R3-i3 /i16. Er is geen enkele opgave die slecht beheerst wordt.

De zeer hoog vaardige (ZHV) leerling beheerst alle 15 opgaven goed.

Tabel 6.3.3 Verdelingen van de leerlingen op de vaardigheid Rekenen/Wiskunde – Meten, Meetkunde, Tijd en Geld

Jaar Variabele	Niveau	2008		2009		
		Gem.	SD	Gem.	SD	
Jaar		250	50	249	50	
Geslacht	Jongens	261	49	260	49	
	Meisjes	239	49	239	49	
Leertijd	Regulier	257	48	256	48	
	Vertraagd	223	49	222	49	
Stratum	Stratum 1	255	49	254	49	
	Stratum 2	246	50	245	50	
	Stratum 3	231	50	231	50	
Formatiegewicht	1.00	256	48	255	48	
	1.25	225	49	224	49	
	1.90	226	50	226	50	
Toets	EB	252	48	251	48	
	NT	153	50	155	49	
Advies VO	vmbo-BB	178	36	177	36	
	vmbo-KB	208	36	207	36	
	vmbo-GT	230	36	228	36	
	Havo	268	36	266	36	
	VWO	320	35	319	35	
Regio	Noord	245	50	244	49	
	Oost	249	50	248	50	
	West	249	50	249	50	
	Zuid	254	50	253	50	
Urbanisatiegraad	Zeer Sterk	242	51	242	51	
	Sterk	250	50	249	50	
	Matig	252	50	252	50	
	Weinig	252	50	251	49	
	Niet	251	49	250	49	
Thuis taal	Alleen NLs	252	49	252	49	
	NLs + bu	237	51	236	51	
	Alleen bu	232	50	232	51	
	Turks	229	50	228	50	
	Arabisch	227	50	227	50	
	Surinaams	211	51	211	51	
	WestEuro	246	51	243	51	
	Ter referentie:	Typische leerling	Schaalwaarde (geen verschil tussen de jaren)			
		Zeer laag vaardige leerling (ZLV)	186		186	
	Laag vaardige leerling (LV)	216		216		
	Standaard leerling (ST)	250		250		
	Hoog vaardige leerling (HV)	284		284		
	Zeer hoog vaardige leerling (ZHV)	314		314		

Wat groepen van leerlingen kunnen

Net als bij de andere schalen wordt ook een beschrijving gegeven in termen van de geobserveerde populatieverdelingen. Dat wordt gedaan in tabel 6.3.3. In die tabel staan ook weer de verdeling van de vaardigheid over de gehele groep leerlingen in 2008 en 2009. Daarnaast wordt de verdeling gegeven per subgroep binnen de twee jaren. De posities zijn gegeven op dezelfde schaal als de opgaven. Merk op dat dit marginale schattingen zijn: er wordt hierbij niet gecorrigeerd voor eventuele verschillen in de verdeling van de overige achtergrondvariabelen.

De effecten van de achtergrondvariabelen

Bij MMTG geldt net als bij de andere twee rekenschalen dat de resultaten voor 2008 en 2009 zeer vergelijkbaar zijn. Hier is echter wel gevonden dat de gemiddelde score van 2009 iets lager ligt dan die van 2008. Doordat ook hier er geen interactie gevonden is tussen jaar en een van de basisvariabelen, is dit kleine verschil over een aantal verschillende subgroepen te zien. Soms is het verschil iets groter, en soms is het afwezig, maar enorme variatie is hier niet te zien. Als er een interactievariabele voor jaarformatiegewicht opgenomen zou zijn, dan zou te zien zijn dat de daling voor de leerlingen met formatiegewicht 1.25 iets groter geweest is, dan nu in de marginalen gevonden is, maar dit verschil is zo klein dat deze interactie niet significant is. Die factor is dan ook niet meegenomen.

Bij het berekenen van een jaareffect zonder andere variabele is, net als bij de marginalen, een zeer kleine daling te zien; zie tabel 6.3.4. Deze daling is echter zo klein dat deze niet significant van 0 verschilt. Met meer dan 30.000 waarnemingen in de twee jaargroepen was er in ieder geval genoeg *power* om een significant verschil waar te nemen als het er zou zijn.

Tabel 6.3.4 Jaareffect Rekenen/Wiskunde – MMTG: Meten, Meetkunde, Tijd en Geld

Variabele: Contrast	Model	Effectgrootte	Effectkwalificatie
Jaar: 2009 – 2008	Alleen jaareffect	-0,01	geen effect
Jaar: 2009 – 2008	Basismodel rekenen (G L S F J)	-0,02	geen effect

Als er ook gecorrigeerd wordt voor andere factoren, dan heeft dit amper effect. Als in het model ook gecorrigeerd wordt voor advies VO dan wordt het effect -0,04. Bij correctie voor type toets of de 3 deling naar thuistaal wordt het jaareffect -0,02, en bij correctie naar leertijd bovenop het basismodel wordt het effect -0,03. Al deze verschillen worden nog steeds geclassificeerd als “geen effect”. Zoals gezegd lijken de andere effecten ook zo op de resultaten bij de andere rekenschalen dat deze naast de tabel niet meer apart beschreven worden.

Tabel 6.3.5 Effectgrootte op de vaardigheid Rekenen/Wiskunde – Meten, Meetkunde, Tijd en Geld

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	-0,51	matig
Leertijd	Vertraagd – Regulier	-0,69	matig
Stratum	Stratum2 – Stratum 1	-0,10	geen effect
	Stratum 3 – Stratum 2	-0,10	geen effect
	Stratum 3 – Stratum 1	-0,20	klein
Formatiegewicht	F1.25 – F1.00	-0,52	matig
	F1.90 – F1.25	0,16	geen effect
	F1.90 – F1.00	-0,36	klein
Toets	NT – EB	-1,86	groot
Advies VO	vmbo-KB – vmbo-BB	0,83	groot
	vmbo-GT – vmbo-KB	0,58	matig
	Havo – vmbo-GT	1,03	groot
	VWO – Havo	1,51	groot
Regio	Oost – Noord	0,09	geen effect
	West – Noord	0,14	geen effect
	Zuid – Noord	0,22	klein
	West – Oost	0,05	geen effect
	Zuid – Oost	0,13	geen effect
	Zuid – West	0,08	geen effect
Urbanisatiegraad	Sterk – Zeer sterk	-0,02	geen effect
	Matig – Sterk	0,00	geen effect
	Weinig – Matig	-0,04	geen effect
	Niet – Weinig	-0,01	geen effect
Thuis taal	NLs+Ander – Alleen NLs	-0,02	geen effect
	Alleen ander – Alleen NLs	0,03	geen effect
	Alleen ander – NLs+Ander	0,06	geen effect
	Turks – Alleen NLs	0,00	geen effect
	Arabisch – Alleen NLs	-0,06	geen effect
	Surinaams/Antilliaans – Alleen NLs	-0,53	matig
	West-Europees – Alleen NLs	0,01	geen effect

7 Resultaten voor taalvaardigheid en rekenvaardigheid in jaargroep 4

De resultaten op de taal- en rekenvaardigheden worden voor jaargroep 4 in een hoofdstuk beschreven. De beschrijvingen zijn korter dan voor jaargroep 8 het geval is. De taalvaardigheden zijn gelijk aan die gemeten zijn in jaargroep 8. Bij rekenvaardigheden zijn er wel kleine verschillen tussen jaargroep 4 en jaargroep 8. De vaardigheden die gemeten worden bij rekenen zijn echter gelijk aan wat er gemeten is bij jaargroep 4 in 2008. Voor de taalvaardigheden worden, net zoals bij jaargroep 8, de resultaten per vaardigheid gegeven. De samenvatting van belangrijkste resultaten wordt in de discussie gegeven. Bij rekenen hangen de gemeten vaardigheden onderling zodanig samen dat deze samen gepresenteerd worden: correlaties tussen de vier vaardigheidsschalen zijn met een gemiddelde van 0,92 hoog te noemen. De gevonden verschillen tussen de vier rekenschalen in jaargroep 4 zijn wat betreft de verschillen tussen de groepen en de effecten daardoor ook gering. Merk op dat dit ook bij de analyses in 2008 het geval was.

De vaardigheden hoeven ook niet meer beschreven te worden in termen van de opgaven aangezien de gebruikte toetsen voor de vaardigheden hetzelfde zijn gebleven in vergelijking met de afnamen in 2008. Dat betekent ook dat de schalen en de posities van de opgaven vast liggen omdat de populatie van 2008 de referentiepopulatie blijft. Er was ook gevonden dat de opgaven in 2009 hetzelfde functioneren als in 2008. Voor de beschrijving van de vaardigheden in termen van opgaven wordt dan ook verwezen naar het verslag van 2008. De enige uitzondering is de vaardigheid Woordenschat waarvoor in de 2009 een nieuwe toets gebruikt is. Hoewel er wel overlap was met de gebruikte opgaven in 2008 zijn er dusdanig verschillen gevonden dat deze niet op dezelfde schaal geplaatst konden worden.

De verdeling van de referentiepopulatie (2008) is zo gesteld dat het gemiddelde van die populatie is gelijk gesteld op 250 en de standaarddeviatie op 50. Op deze schaal zijn ook weer de "typische leerlingen" af te beelden. De definities van deze leerlingen op de vaardigheidsschaal zijn gelijk aan die voor jaargroep 8 (zie ook Hoofdstuk 4 en Hoofdstuk 5).

Code	Label typische leerling	Definitie (positie op vaardigheidsschaal)
ZLV	Zeër laag vaardige leerling	186
LV	Laag vaardige leerling	216
ST	Standaard leerling	250
HV	Hoog vaardige leerling	284
ZHV	Zeër hoog vaardige leerling	314

Hierbij moet wel opgemerkt worden dat deze waarden niet dezelfde mate van *vaardigheid* aangeeft als die in jaargroep 8. Een leerling met een vaardigheid van 250 in jaargroep 8 kan, in absolute zin, vaardiger dan een leerling met een vaardigheid van 250 in jaargroep 4. Deze vaardigheidsschaal is dan ook te vergelijken met een IQ-schaal: er wordt ook niet verwacht dat een 7 jarige met een IQ van 100 precies hetzelfde kan als een 11 jarige met een IQ van 100. Net zoals het IQ bruikbaar is om te spreken over verschillen binnen een leeftijdsgroep, zo is de vaardigheidsindex die nu hier gebruikt wordt zinvol om binnen leerjaren groepen te vergelijken.

Het gevolg van deze relatieve schalen is overigens dat jaargroep 4 en jaargroep 8 echter niet direct met elkaar vergelijkbaar zijn. Zo kan een ZLV-leerling in jaargroep 8 in absolute zin mogelijk meer dan ZHV-leerling in jaargroep 4. De definities van de typische leerlingen zijn dus relatief en gebonden aan de groep waarin de leerling zit, net zoals de IQ relatief is per leeftijdsgroep bij kinderen.

Anders dan bij de vergelijking voor jaargroep 8 is het hier niet zinvol de resultaten van de opeenvolgende jaren zonder meer naast elkaar te zetten. De steekproef is te klein om aan te nemen dat veranderingen in

verdeling van de achtergrondvariabelen daadwerkelijke veranderingen representeren. Het zal hier eerder steekproeffluctuaties betreffen. De te observeren jaareffecten zijn zo klein dat kleine wijzigingen in de samenstelling van steekproef op achtergrondvariabelen door deze steekproeffluctuaties al snel tot misinterpretaties van de gevonden jaarverschillen zou kunnen leiden (zie hoofdstuk 4). Om die reden worden voor jaargroep 4 alleen gecorrigeerde verschillen gegeven. We berekenen in dat geval de jaarverschillen na een correctie op de belangrijkste achtergrondvariabelen, te weten geslacht, formatiegewicht, leertijd en stratum.

Voor de schattingen van de effecten bij de rekvaardigheden was naast dit hoofdeffectenmodel ook een interactiemodel nodig. Zowel bij de taal- als bij de rekvaardigheden was in het model een interactievariabele tussen stratum en afnamejaar nodig: de jaareffecten verschilden voor groepen uit de verschillende strata. Het jaareffect wordt daarom hier (ook) per stratum berekend.

Net als bij jaargroep 8 analyseren we de verschillen per achtergrondvariabele bij de diverse vaardigheden. Naast de effecten van de variabelen van het basismodel (geslacht, leertijd, formatiegewicht, en stratum) zijn ook modellen met andere variabelen gebruikt. Dat zijn modellen aangevuld met de variabele thuistaal en de variabelen regio en urbanisatiegraad. De variabelen type toets (NT of EB) en advies VO die wel bij de analyses voor jaargroep 8 voorkomen, zijn uiteraard niet opgenomen voor jaargroep 4.

7.1 Taalvaardigheid: Woordenschat

Bij de gemeten vaardigheden in 2009 is “woordenschat” een bijzonder geval. Daar waar bij alle overige vaardigheden met exact dezelfde toets gemeten werd, was dat niet het geval bij woordenschat. In 2008 was de woordenschattoets voor LOVS Einde 4 in de fase waarin deze vervangen werd. Daardoor was de vaardigheid in dat jaar gemeten met vier proeftoetsen, aangevuld met de resultaten van een halve proeftoets. In 2009 is deze toets afgenomen met de nieuwe toets Woordenschat van het LOVS Einde jaargroep 4. Deze toets bevat opgaven die ook in de proeftoetsen zaten, maar verschilt op een aantal punten van de proeftoets. Om te beginnen is de vormgeving van een aantal opgaven wat aangepast. Er was ook een aantal opgaven dat in de proeftoets anders functioneerde dan in de uiteindelijke toets. Dat was het geval bij de opgaven waarbij gevraagd werd naar wat het tegengestelde was van een bepaald woord.

Een ander verschil met de overige vaardigheden die in 2008 gemeten waren was dat de meting in dat jaar was gedaan bij een kleiner aantal leerlingen. Ook waren er minder achtergrondgegevens bekend over deze leerlingen. Zo was van de leerlingen in 2008 niet bekend wat hun formatiegewicht was en wat hun thuistaal was. Hierdoor was deze meting in 2008 al een uitbijter in vergelijking met de overige vaardigheden. Het is door het verschil in functioneren en het ontbreken van deze achtergrondgegevens lastig om een jaarvergelijking met 2009 te maken. Hierdoor is besloten om van 2009 de referentie populatie te maken. Ook de beschrijving van de toets in 2009 zal nu op deze schaal plaats vinden.

De gemiddelde score op de LOVS toets Woordenschat einde jaargroep 4, zoals gemeten bij de 2576 leerlingen in dit onderzoek is 37,5. Met 50 opgaven betekent dit dat 75% van de opgaven goed was beantwoord. De gemiddelde p-waarde van de opgaven is dan 75 voor deze groep leerlingen. De toets heeft een betrouwbaarheid van 0,87. Dat is voor het gebruik van deze toets voor het bepalen van de vaardigheid van individuele leerlingen goed te noemen, en dus ook zeker goed genoeg voor het meten van groepen, wat in dit verslag gebeurt.

In tabel 7.1.1 is de verdeling gegeven van de hoeveelheid moeilijke, uitdagende en gemakkelijke opgaven van deze toets per type leerling. De posities van de individuele opgaven zijn gegeven in tabel, 7.1.2, waarbij ook een indicatie gegeven is van de moeilijkheidsgraad van de opgaven voor de verschillende typische leerlingen.

Tabel 7.1.1 Aantal opgaven Woordenschat per moeilijkheidsgraad in de LOVS-toets Einde jaargroep 4 voor de vijf typische leerlingen.

Aantal opgaven per type en niveau		ZLV	LV	ST	HV	ZHV
0	Gemakkelijk ($P > .80$)	0	10	29	40	46
1	Uitdagend ($.50 \leq P \leq .80$)	29	33	21	10	4
2	Moeilijk ($P < .50$)	21	7	0	0	0

Tabel 7.1.2 Posities van de opgaven Woordenschat op de vaardigheidsschaal met de leerlingen
Einde jaargroep 4 in 2009 als referentiepopulatie, geordend op volgorde van afname, en de
typering van de opgaven naar moeilijkheidsgraad voor de vijf typische leerlingen

Taak	Item	Label	p50	p80	ZLV	LV	St	HV	ZHV
Woordenschat LOVS E4	1	E4 WS - i01	224	314	2	2	1	1	1
Woordenschat LOVS E4	2	E4 WS - i02	168	233	1	1	0	0	0
Woordenschat LOVS E4	3	E4 WS - i03	196	287	2	1	1	1	0
Woordenschat LOVS E4	4	E4 WS - i04	205	295	2	1	1	1	0
Woordenschat LOVS E4	5	E4 WS - i05	128	203	1	0	0	0	0
Woordenschat LOVS E4	6	E4 WS - i06	131	221	1	1	0	0	0
Woordenschat LOVS E4	7	E4 WS - i07	144	220	1	1	0	0	0
Woordenschat LOVS E4	8	E4 WS - i08	143	233	1	1	0	0	0
Woordenschat LOVS E4	9	E4 WS - i09	100	191	1	0	0	0	0
Woordenschat LOVS E4	10	E4 WS - i10	181	272	1	1	1	0	0
Woordenschat LOVS E4	11	E4 WS - i11	171	246	1	1	0	0	0
Woordenschat LOVS E4	12	E4 WS - i12	179	269	1	1	1	0	0
Woordenschat LOVS E4	13	E4 WS - i13	145	258	1	1	1	0	0
Woordenschat LOVS E4	14	E4 WS - i14	193	284	2	1	1	0	0
Woordenschat LOVS E4	15	E4 WS - i15	151	264	1	1	1	0	0
Woordenschat LOVS E4	16	E4 WS - i16	163	228	1	1	0	0	0
Woordenschat LOVS E4	17	E4 WS - i17	145	209	1	0	0	0	0
Woordenschat LOVS E4	18	E4 WS - i18	194	285	2	1	1	1	0
Woordenschat LOVS E4	19	E4 WS - i19	196	272	2	1	1	0	0
Woordenschat LOVS E4	20	E4 WS - i20	214	278	2	1	1	0	0
Woordenschat LOVS E4	21	E4 WS - i21	131	244	1	1	0	0	0
Woordenschat LOVS E4	22	E4 WS - i22	158	248	1	1	0	0	0
Woordenschat LOVS E4	23	E4 WS - i23	155	212	1	0	0	0	0
Woordenschat LOVS E4	24	E4 WS - i24	169	245	1	1	0	0	0
Woordenschat LOVS E4	25	E4 WS - i25	141	216	1	0	0	0	0
Woordenschat LOVS E4	26	E4 WS - i26	222	335	2	2	1	1	1
Woordenschat LOVS E4	27	E4 WS - i27	199	274	2	1	1	0	0
Woordenschat LOVS E4	28	E4 WS - i28	130	195	1	0	0	0	0
Woordenschat LOVS E4	29	E4 WS - i29	136	212	1	0	0	0	0
Woordenschat LOVS E4	30	E4 WS - i30	138	251	1	1	1	0	0
Woordenschat LOVS E4	31	E4 WS - i31	73	224	1	1	0	0	0
Woordenschat LOVS E4	32	E4 WS - i32	191	342	2	1	1	1	1
Woordenschat LOVS E4	33	E4 WS - i33	152	228	1	1	0	0	0
Woordenschat LOVS E4	34	E4 WS - i34	106	219	1	1	0	0	0
Woordenschat LOVS E4	35	E4 WS - i35	221	334	2	2	1	1	1
Woordenschat LOVS E4	36	E4 WS - i36	114	205	1	0	0	0	0
Woordenschat LOVS E4	37	E4 WS - i37	237	294	2	2	1	1	0
Woordenschat LOVS E4	38	E4 WS - i38	202	266	2	1	1	0	0
Woordenschat LOVS E4	39	E4 WS - i39	201	246	2	1	0	0	0
Woordenschat LOVS E4	40	E4 WS - i40	204	239	2	1	0	0	0
Woordenschat LOVS E4	41	E4 WS - i41	214	246	2	1	0	0	0
Woordenschat LOVS E4	42	E4 WS - i42	227	277	2	2	1	0	0
Woordenschat LOVS E4	43	E4 WS - i43	199	240	2	1	0	0	0
Woordenschat LOVS E4	44	E4 WS - i44	190	235	2	1	0	0	0
Woordenschat LOVS E4	45	E4 WS - i45	218	294	2	2	1	1	0
Woordenschat LOVS E4	46	E4 WS - i46	221	311	2	2	1	1	0
Woordenschat LOVS E4	47	E4 WS - i47	167	224	1	1	0	0	0
Woordenschat LOVS E4	48	E4 WS - i48	151	196	1	0	0	0	0
Woordenschat LOVS E4	49	E4 WS - i49	161	237	1	1	0	0	0
Woordenschat LOVS E4	50	E4 WS - i50	154	211	1	0	0	0	0

De beheersingsniveaus van de typische leerlingen zijn de volgende:

De zeer laag vaardige (ZLV) leerling beheerst geen van de EB opgaven goed. Van de 50 opgaven beheerste een ZLV-leerling er zijn er 21 slecht (kans goed te maken tussen ,5 en ,8): 1, 10, 12-15 (12 tot en met 15), 22 - 29, 32 - 33, 35 – 37 en 39. De overige 29 zijn uitdagend voor een dergelijke leerling.

De laag vaardige (LV) leerling beheerst 10 opgaven: 3, 7, 34, 40 – 43, 45, 49-50. Van de 50 opgaven beheerste een LV leerling er 7 slecht: 25, 27, 29, 32, 36-37, 39. De overige 33 opgaven zijn uitdagend voor deze leerling.

De standaard (ST) leerling beheerst geen enkele opgave slecht. Dit betekent ook dat een HV en een ZHV geen enkele opgave slacht beheersen. Van de 50 opgaven zijn er 21 uitdagend: 1, 6, 8-10, 12, 22-25, 27-33, 35-37 en 39. De overige 29 opgaven beheerst een dergelijke leerling goed.

De hoog vaardige (HV) leerling beheerst 40 opgaven goed. De overige 10 opgaven zijn uitdagend voor een dergelijke leerling: 1, 22, 25, 29, 32-33, 35-37 en 39.

De zeer hoog vaardige (ZHV) leerling beheerst 46 opgaven goed van de LOVS toets woordenschat E4. De enige 4 uitdagende opgaven voor een dergelijk leerling zijn: 29, 32, 33 en 36.

Het eerste verschil met de overige schalen in dit verslag is dat Woordenschat in jaargroep 4 de enige schaal is waarbij 2009 de referentiepopulatie definieert: het gemiddelde van 250 is het gemiddelde in 2009. Bij de overige vaardigheden is dat het gemiddelde van 2008. Het tweede verschil is, dat woordenschat in jaargroep 4 de enige vaardigheid is die met een nieuwe toets is gemeten. Dat betekent ook dat dit de enige schaal is in jaargroep 4 waarvoor de posities van de opgaven nog niet vooraf gegeven waren. Het is dan ook de enige schaal voor jaargroep 4 met tabellen zoals 7.1.1 en 7.1.2. Het is ook de enige schaal voor jaargroep 4 die weergegeven wordt in Bijlage 2 met daarin de informatie van tabel 7.1.1 in de volgorde van de items geordend naar positie van de p80 van de opgaven en de bijbehorend figuur. De posities van de opgaven zijn voor de overige schalen onveranderd. Voor die schalen wordt verwezen naar het verslag van 2008 (Hemker en Van Weerden, 2008).

Nu de schaal gegeven is in termen van de opgaven kunnen ook de deelpopulaties weergegeven worden. In tegenstelling tot de andere schalen in dit verslag ontbreekt hier de vergelijking met 2008. In tabel 7.1.3 staan het gemiddelde en de standaarddeviatie van verschillende subgroepen op de vaardigheidsschaal van jaargroep 4. Een belangrijk verschil met de vergelijkbare tabel in het verslag van 2008 is dat nu de subgroepen voor de verschillende formatiegewichten en de verschillende groepen op basis van hun thuistaal te geven zijn. De populatie gegevens van deze verdelingen zijn echter ongecorrigeerde verdelingen. Dat is ook de belangrijkste verklaring waarom het gemiddelde van de verdeling van leerlingen in zeer sterk geurbaniseerde gebieden zeer laag ligt: in die gebieden wonen relatief veel anderstaligen en achterstandsléerlingen. Dit zien we ook bij het analyseren van de effecten van de achtergrondvariabelen.

Bij de effectschattingen worden de gecorrigeerde effecten berekend. Het basismodel bevat de variabelen geslacht, leertijd, stratum en formatiegewicht. Dit is in een extra analyse aangevuld met de variabele regio en urbanisatiegraad, en in een volgende analyse met de variabele thuistaal. De effectschattingen zijn gegeven in tabel 7.1.4. Ook hier ontbreekt de variabele afnamejaar. Het effect van urbanisatiegraad valt weg zodra er gecorrigeerd wordt voor formatiegewicht. Dit ondersteunt de verklaring dat het lagere vaardigheidsniveau bij de zeer sterk stedelijke gebieden deels te wijten is aan de grotere proporties leerlingen met een hoger leerlinggewicht. Het is ook duidelijk dat formatiegewicht de variabele is met de sterkste effecten, gevolgd door thuistaal. Thuistaal heeft bij de vaardigheid Woordenschat nog een additief effect bovenop het effect van formatiegewicht: ook als er gecorrigeerd wordt voor de opleiding van de ouders heeft thuistaal effect op de vaardigheid bij Woordenschat in jaargroep 4.

Het ontbreken van de twee meest verklarende achtergrondvariabelen (leerlinggewicht en thuistaal) in 2008 maakt het lastig een goede vergelijking te maken tussen 2008 en 2009. Ieder gevonden verschil tussen de twee jaren valt te verklaren uit een ongelijkheid in de verdeling op deze achtergrondvariabelen waarop niet gecontroleerd kan worden in de jaarvergelijking. Dat dit grote effecten kan hebben is al duidelijk bij de achtergrondvariabele urbanisatiegraad. Een andere reden waarom de vergelijking lastig is, is het deels anders functioneren van een aantal opgaven.

Tabel 7.1.3 Gemiddelde (gem.) en standaarddeviatie (SD) op de vaardigheidsschaal van jaargroep 4 per achtergrondvariabelegroep in 2009 op de vaardigheid Woordenschat

		<u>2009</u>	
		Gem.	SD
Jaar verdeling	Gehele groep	250	50
	Subgroep		
Geslacht	Jongens	249	50
	Meisjes	252	50
Leertijd	Regulier	254	49
	Vertraagd	233	52
Stratum	Stratum 1	258	47
	Stratum 2	248	48
	Stratum 3	213	50
Formatiegewicht	F 1	256	47
	F 2	235	48
	F 3	197	48
Regio	Noord	258	48
	Oost	252	48
	West	243	51
	Zuid	254	49
Urbanisatiegraad	Zeer sterk	215	51
	Sterk	245	50
	Matig	252	48
	Weinig	255	48
	Niet	262	47
Thuistaal*	Alleen Nederlands (NLs)	256	47
	NLs en een andere taal	226	50
	Alleen een andere taal	196	49
Ter referentie:	Typische leerling	Schaalwaarde (2009 is referentie)	
	Zeer laag vaardige leerling (ZLV)	186	
	Laag vaardige leerling (LV)	216	
	Standaard leerling (ST)	250	
	Hoog vaardige leerling (HV)	284	
	Zeer hoog vaardige leerling (ZHV)	314	

Tabel 7.1.4 Effectgrootte op de vaardigheid Woordenschat

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Stratum	Stratum 2 – Stratum 1	-0,17	geen effect
	Stratum 3 – Stratum 2	-0,36	klein
	Stratum 3 – Stratum 1	-0,53	matig
Formatiegewicht	F 2 (.30) – F 1 (.00)	-0,29	klein
	F 3 (1.20) – F 2 (.30)	-0,52	matig
	F 3 (1.20) – F 1 (.00)	-0,80	groot
Geslacht	Meisjes – Jongens	0,04	niet significant
Leertijd	Vertraagd – Regulier	-0,27	klein
Regio	Oost – Noord	-0,05	geen effect
	West – Noord	-0,02	geen effect
	Zuid – Noord	0,04	geen effect
	West – Oost	0,03	geen effect
	Zuid – Oost	0,08	geen effect
	Zuid – West	0,05	geen effect
Urbanisatiegraad	Sterk – Zeer sterk	-0,02	geen effect
	Matig – Sterk	0,02	geen effect
	Weinig – Matig	0,06	geen effect
	Niet – Weinig	0,15	geen effect
Thuis taal	NLs+Ander – Alleen NLs	-0,34	klein
	Alleen ander – Alleen NLs	-0,62	matig
	NLs+Ander – Alleen ander	0,28	klein

Om toch een indruk te krijgen van een mogelijk jaareffect zijn de opgaven die het meest afwijkend zijn uit de analyses verwijderd. Dat zijn de opgaven 17 en 37 tot en met 44. Merk op dat opgave 37 tot en met 44 in het tweede deel van de toets zitten. Binnen de delen zijn de opgaven voor de leerlingen opnieuw genummerd. In de toets zelf zijn dit opgave 12 tot en met 19. De basisvraag bij deze zeven opgaven is “Wat is het tegengestelde van ...”.

Als gekeken wordt naar het effect van afnamejaar bij de set van 42 opgaven dat hetzelfde functioneert over beide afname jaren dan is het een negatief effect: -0.23. Dat betekent dat de leerlingen in 2009 een vijfde standaarddeviatie slechter presteerden dan in 2008 bij deze steekproeven, wat een klein effect zou zijn. Het effect is echter niet significant. Behalve dat dit niet significant is, ontbreekt hier ook informatie over de verdeling van het formatiegewicht en thuis taal in 2008. Al met al is het gevonden effect dusdanig dat er niet al te veel staat opgemaakt kan worden. We zullen tot de peiling van volgend jaar moeten wachten om hier meer over te weten te komen.

7.2 Taalvaardigheid: Spelling

Bij de vaardigheid Spelling is wel een jaarvergelijking mogelijk. De toets Spelling die in 2008 gebruikt is, is exact dezelfde als in 2009 gebruikt is en ook de achtergrondvariabelen zijn gelijk. Eveneens zijn de aantallen leerlingen bij beide metingen vergelijkbaar (ongeveer 2500 leerlingen). De aantallen zijn echter te klein om ongecorrigeerde jaareffecten goed te kunnen schatten. Als er kleine verschuivingen in vaardigheid verwacht worden dan kunnen kleine verschuivingen in de verdeling van de achtergrondvariabelen bij de steekproeven die verschillen al versterken dan wel teniet doen. Om die reden moet voor de verschillen in de verdeling van de achtergrondvariabelen gecorrigeerd worden.

Het gebruikte model voor de jaareffecten is het model waarin gecorrigeerd wordt voor de verdeling van geslacht, leertijd, formatiegewicht en stratum. Er is ook een variabele voor de interactie opgenomen tussen stratum en afname jaar. De analyses zijn aangevuld met modellen waarbij ook thuistaal, regio en urbanisatiegraad zijn opgenomen.

Bij de IRT-analyses is gebruikgemaakt van het model met precies dezelfde parameters in 2008 en 2009. Twee opgaven functioneren iets anders. De passing is niet echt heel goed, maar gegeven de hoeveelheid observaties per opgave (>5000) mag verwacht worden dat de passing niet sterk is. De schattingen zijn echter robuust en als naar de figuren gekeken wordt dan valt de misfit nog wel mee. De parameters verschillen weinig van de schattingen in het geval met aangepast schattingen gewerkt zou worden. Uitzondering zijn twee opgaven uit de startmodule, waarbij het meervoud van een woord gevraagd werd. In beide gevallen moest het meervoud leiden tot het verdubbelen van de laatste letter van de stam: mug – muggen, blok – blokken.

De analyses zijn op twee wijzen uitgevoerd, te weten met de opgaven als gelijk over de twee jaren en gecorrigeerd waarbij de twee opgaven als nieuwe opgaven gezien werden. De schattingen van de populaties zijn gegeven met deze opgaven als gelijk, terwijl die bij de effectschattingen gegeven worden met de opgaven als twee nieuwe opgaven. Overigens zijn ook de populaties geschat met de variant waarbij deze opgaven als nieuw gezien werden, maar de verschillen zijn bij de belangrijkste variabelen niet groot.

Opgemerkt moet worden dat de schattingen van de gemiddelden van de verschillende groepen met modellen zijn bepaald met extra interacties. De gemiddelden voor de groepen die zich onderscheiden in formatiegewicht en in thuistaal zijn berekend met een interactiemodel, waarbij de interactie tussen jaar en formatiegewicht, dan wel thuistaal is opgenomen. Het effect van deze interactie zit ook deels in het interactie effect bij stratum, die ook in dat model is opgenomen. Bij het weergegeven van de jaareffecten zal dan ook het interactie effect binnen stratum gegeven worden.

De schattingen bij regio Noord en urbanisatiegraad “Niet stedelijk” onder 2008 verschillen iets van de schattingen in de rapportage van 2008. Dit heeft te maken met het gebruik van het model en de relatief kleine aantallen leerlingen in deze twee groepen. Door nu gebruik te maken van meer leerlingen (ook 2009 erbij) is het mogelijk een gecorrigeerd beeld te krijgen voor de gehele populatie. De aantallen in deze twee groepen waren ook zo klein dat de gevonden interactie niet significant was. Merk ook op dat de analyses instabiel waren bij schattingen van de gemiddelden in het geval gewerkt werd met het model waarin de variabelen regio en verstedelijking waren opgenomen..

Tabel 7.2.1 Gemiddelde (gem.) en standaarddeviatie (SD) op de vaardigheidsschaal van jaargroep 4 per achtergrondvariabelegroep per afname jaar op de vaardigheid Spelling

		<u>2008</u>		<u>2009</u>	
		Gem.	SD	Gem.	SD
Jaar verdeling	Gehele groep	250	50	251	50
	Subgroep				
Stratum	Stratum 1	254	50	252	50
	Stratum 2	248	50	252	50
	Stratum 3	235	50	247	51
Formatiegewicht	F 1	252	50	253	50
	F 2	238	50	239	50
	F 3	235	50	241	51
Geslacht	Jongens	243	50	245	49
	Meisjes	257	49	259	49
Leertijd	Regulier	254	49	256	49
	Vertraagd	225	49	229	49
Thuis taal*	Alleen Nederlands (NLs)	254	49	250	49
	NLs en een andere taal	235	50	247	50
	Alleen een andere taal	233	50	244	50
Regio	Noord	244	50	246	50
	Oost	258	50	259	50
	West	247	50	246	50
	Zuid	254	50	256	49
Urbanisatiegraad	Zeer sterk	240	50	245	50
	Sterk	251	50	252	50
	Matig	247	50	246	50
	Weinig	256	49	255	50
	Niet	254	50	255	50
Ter referentie:	Typische leerling	Schaalwaarde (geen verschil tussen de jaren)			
	Zeer laag vaardige leerling (ZLV)	186		186	
	Laag vaardige leerling (LV)	216		216	
	Standaard leerling (ST)	250		250	
	Hoog vaardige leerling (HV)	284		284	
	Zeer hoog vaardige leerling (ZHV)	314		314	

Tabel 7.2.2 Effectgrootte op de vaardigheid Spelling in jaargroep 4

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Jaar	2009 – 2008	0,07	geen effect
Jaar (Stratum 1)	2009 – 2008	-0,02	niet significant
Jaar (Stratum 2)	2009 – 2008	0,13	geen effect
Jaar (Stratum 3)	2009 – 2008	0,36	klein
Stratum (2008)	Stratum2 – Stratum1	-0,09	niet significant
	Stratum3 – Stratum2	-0,15	niet significant
	Stratum3 – Stratum1	-0,24	klein
Stratum (2009)	Stratum2 – Stratum1	0,06	niet significant
	Stratum3 – Stratum2	0,08	niet significant
	Stratum3 – Stratum1	0,14	klein
Formatiegewicht	F1 – F2	-0,24	klein
	F3 – F2	0,08	niet significant
	F3 – F1	-0,16	geen effect
Geslacht	Meisjes – Jongens	0,27	klein
Leertijd	Vertraagd – Regulier	-0,53	matig
Thuis taal	NLs+Ander – Alleen NLs	-0,08	niet significant
	Alleen ander – Alleen NLs	-0,01	niet significant
	NLs+Ander – Alleen ander	-0,07	niet significant
Regio	Oost – Noord	0,29	klein
	West – Noord	0,10	niet significant
	Zuid – Noord	0,27	klein
	West – Oost	-0,20	geen effect
	Zuid – Oost	-0,02	niet significant
	Zuid – West	0,17	geen effect
Urbanisatiegraad	sterk – zeer sterk	0,02	niet significant
	matig – sterk	-0,07	niet significant
	weinig – matig	0,16	geen effect
	niet – weinig	0,05	niet significant

Het gevonden effect tussen 2009 en 2008 is significant, maar zodanig klein dat het als “geen effect” gekwalificeerd wordt. Als het model gekozen wordt waarbij de twee opmerkelijke opgaven als gelijk gezien worden dan wordt ook gevonden dat de prestatie in 2009 beter was dan in 2008, maar het effect is wel kleiner. In plaats van een effect van 0,07 wordt een effect gevonden van 0,05. Dit effect is ook niet significant verschillend van 0.

7.3 Taalvaardigheid: Begrijpend lezen

De toetsen voor Begrijpend lezen bestaan uit een startmodule en een vervolgmodule. Er zijn twee versies van de vervolgmodule, die van elkaar verschillen in moeilijkheid. Op basis van de resultaten op de startmodule wordt een van de twee vervolgmodes voorgelegd. Voor de beschrijving van de opgaven en de relaties tussen de opgaven en de typische leerlingen wordt net als bij Spelling verwezen naar het verslag van 2008.

Tabel 7.3.1 Gemiddelde (gem.), standaarddeviatie (SD) per afnamejaar in jaargroep 4 op de vaardigheid Begrijpend lezen

		<u>2008</u>		<u>2009</u>	
		Gem.	SD	Gem.	SD
Jaar verdeling	Gehele groep	250	50	246	49
	Subgroep				
Stratum	Stratum 1	260	47	253	47
	Stratum 2	242	48	244	48
	Stratum 3	215	49	217	49
Formatiegewicht	F 1	256	48	251	47
	F 2	227	48	227	48
	F 3	205	48	205	48
Geslacht	Jongens	245	50	242	49
	Meisjes	255	50	251	49
Leertijd	Regulier	253	49	251	48
	Vertraagd	229	51	226	50
Thuis taal	Alleen Nederlands (NLs)	256	48	252	47
	NLs en een andere taal	201	49	206	48
	Alleen een andere taal	229	50	228	49
Regio *	Noord	240	48	243	48
	Oost	261	48	253	48
	West	248	50	240	50
	Zuid	250	50	247	49
Urbanisatiegraad *	Zeer sterk	228	52	224	51
	Sterk	252	49	241	49
	Matig	256	49	247	48
	Weinig	257	48	249	48
	Niet	250	48	256	48
Ter referentie:	Typische leerling	Schaalwaarde (geen verschil tussen de jaren)			
	Zeer laag vaardige leerling (ZLV)	186		186	
	Laag vaardige leerling (LV)	216		216	
	Standaard leerling (ST)	250		250	
	Hoog vaardige leerling (HV)	284		284	
	Zeer hoog vaardige leerling (ZHV)	314		314	

Bij het toevoegen van de gegevens van 2009 wordt het aantal waarnemingen per opgave verdubbeld. De modelpassing voor OPLM (het gebruikte IRT-model) is bij Begrijpend lezen zelfs met 5000 waarnemingen goed te noemen. Alle opgaven gedragen zich vrijwel hetzelfde in 2008 als in 2009. De marginalen voor regio en urbanisatiegraad zijn hier geschat met een interactiemodel waarbij de Jaar-Urbanisatiegraad interactie ook geschat is. In de discussie wordt verder ingegaan op de interpretatie van de resultaten in de tabellen 7.3.1 en 7.3.2.

Tabel 7.3.2 Effectgrootte op de vaardigheid Begrijpend lezen in jaargroep 4

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Jaar	2009 – 2008	-0,05	niet significant
Jaar (Stratum 1)	2009 – 2008	-0,13	geen effect
Jaar (Stratum 2)	2009 – 2008	0,04	niet significant
Jaar (Stratum 3)	2009 – 2008	0,12	niet significant
Stratum (2008)	Stratum2 - Stratum1	-0,31	klein effect
	Stratum3 - Stratum2	-0,31	klein effect
	Stratum3 - Stratum1	-0,62	matig effect
Stratum (2009)	Stratum2 - Stratum1	-0,14	geen effect
	Stratum3 - Stratum2	-0,23	klein effect
	Stratum3 - Stratum1	-0,37	klein effect
Formatiegewicht	F1 - F2	-0,40	klein effect
	F3 - F2	-0,23	klein effect
	F3 - F1	-0,64	matig effect
Geslacht	Meisjes – Jongens	0,19	geen effect
Leertijd	Vertraagd - Regulier	-0,37	klein effect
Thuis taal	NLs+Ander - Alleen NLs	-0,29	klein effect
	Alleen ander - Alleen NLs	-0,48	klein effect
	NLs+Ander - Alleen ander	0,19	geen effect
Regio	Oost - Noord	0,30	klein effect
	West - Noord	0,19	geen effect
	Zuid - Noord	0,21	klein effect
	West - Oost	-0,12	geen effect
	Zuid - Oost	-0,10	geen effect
	Zuid - West	0,02	niet significant
Urbanisatiegraad	sterk - zeer sterk	0,07	niet significant
	matig - sterk	0,05	niet significant
	weinig - matig	-0,02	niet significant
	niet - weinig	0,09	niet significant

7.4 Rekenvaardigheid

Bij Rekenen/Wiskunde zijn er in jaargroep 4 vier verschillen de vaardigheden onderscheiden, te weten Getallen en Getalsrelaties (G/G), Optellen en Aftrekken (O/A), Vermenigvuldigen en Delen (V/D) en Meten, Tijd en Geld (MTG). Deze vier vaardigheden zijn gemeten met behulp van een rekentoets met 54 opgaven. De posities van de opgaven zijn niet veranderd ten opzichte van 2008. Voor de omschrijving van de opgaven wordt dan ook verwezen naar het verslag van 2008 (Hemker en Van Weerden, 2008). Daarin wordt ook de kwaliteit van het meetinstrument verder beschreven.

Tabel 7.4.1a Gemiddelde (gem.) en standaarddeviatie (SD) van de groepen in jaargroep 4 op de rekenschaal Getallen en getalsrelaties (G/G) in 2008 en in 2009

		<u>2008</u>		<u>2009</u>	
		Gem.	SD	Gem.	SD
Jaar verdeling	Gehele groep	250	50	249	48
	Subgroep				
Stratum	Stratum 1	260	47	255	47
	Stratum 2	241	48	242	48
	Stratum 3	214	48	229	49
Formatiegewicht	F1	256	48	253	47
	F2	226	49	233	48
	F3	211	49	219	48
Geslacht	Jongens	259	49	258	48
	Meisjes	241	49	239	48
Leertijd	Regulier	253	49	252	48
	Vertraagd	233	51	231	49
Thuis taal	Alleen Nederlands (NLs)	257	48	254	47
	NLs en een andere taal	212	49	216	48
	Alleen een andere taal	223	49	234	48
Regio	Noord	245	48	254	48
	Oost	263	48	248	48
	West	247	50	245	49
	Zuid	247	50	253	48
Urbanisatiegraad	Zeer sterk	232	52	233	49
	Sterk	251	49	246	49
	Matig	253	50	251	48
	Weinig	258	47	253	48
	Niet	250	48	253	47
Ter referentie:	Typische leerling	Schaalwaarde (geen verschil tussen de jaren)			
	Zeer laag vaardige leerling (ZLV)	186		186	
	Laag vaardige leerling (LV)	216		216	
	Standaard leerling (ST)	250		250	
	Hoog vaardige leerling (HV)	284		284	
	Zeer hoog vaardige leerling (ZHV)	314		314	

Net als in het verslag van 2008 worden de resultaten voor deze vier schalen niet in afzonderlijke alinea's gepresenteerd. De gemeten vaardigheden hangen hoog met elkaar samen. De correlaties tussen de vaardigheden variëren tussen de 0,82 (tussen O/A en MTG) en 0,97 (tussen V/D en MTG). De gemiddelde correlatie tussen de rekenvaardigheden is 0,92. Deze correlaties zijn redelijk hoog te noemen, maar niet zo hoog dat er in het geheel geen verschillen tussen de vaardigheden gevonden worden. Merk overigens op dat dit de correlaties betreft tussen de latente schalen en niet de correlaties tussen de toetsscores. Die correlaties zijn lager aangezien hier met korte toetsen gewerkt is. De gevonden verschillen tussen de vier rekenschalen in jaargroep 4 zijn wat betreft de verschillen tussen de groepen en de effecten daardoor ook gering. Bij de analyses in 2008 was dit ook het geval.

Tabel 7.4.1b Gemiddelde (gem.) en standaarddeviatie (SD) van de groepen in jaargroep 4 op de rekenschaal Optellen/Aftrekken (O/A) in 2008 en in 2009

		<u>2008</u>		<u>2009</u>	
		Gem.	SD	Gem.	SD
Jaar verdeling	Gehele groep	250	50	250	49
	Subgroep				
Stratum	Stratum 1	258	48	251	49
	Stratum 2	243	49	251	49
	Stratum 3	223	49	240	49
Formatiegewicht	Fg 1	254	49	252	48
	Fg 2	230	49	237	49
	Fg 3	222	49	234	49
Geslacht	Jongens	257	50	257	48
	Meisjes	243	49	242	48
Leertijd	Regulier	253	49	253	48
	Vertraagd	234	50	234	49
Thuis taal	Alleen Nederlands (NLs)	255	49	252	48
	NLs en een andere taal	227	50	235	49
	Alleen een andere taal	226	50	240	49
Regio	Noord	241	49	250	49
	Oost	258	49	251	49
	West	249	50	249	49
	Zuid	249	50	250	49
Urbanisatiegraad	Zeer sterk	237	51	242	49
	Sterk	250	50	248	49
	Matig	250	50	249	49
	Weinig	258	49	254	49
	Niet	250	49	250	48
Ter referentie:	Typische leerling	Schaalwaarde (geen verschil tussen de jaren)			
	Zeer laag vaardige leerling (ZLV)	186		186	
	Laag vaardige leerling (LV)	216		216	
	Standaard leerling (ST)	250		250	
	Hoog vaardige leerling (HV)	284		284	
	Zeer hoog vaardige leerling (ZHV)	314		314	

De waarden van de ongecorrigeerde gemiddelden zijn gegeven in tabel 7.4.1a tot en met tabel 7.4.1d. De gecorrigeerde effecten zijn gegeven in tabel 7.4.2a tot en met tabel 7.4.2d. In de discussie (hoofdstuk 8) wordt verder ingegaan op een interpretatie van de resultaten.

Tabel 7.4.1c Gemiddelde (gem.) en standaarddeviatie (SD) van de groepen in jaargroep 4 op de rekenschaal Vermenigvuldigen/Delen (V/D) in 2008 en in 2009

		<u>2008</u>		<u>2009</u>	
		Gem.	SD	Gem.	SD
Jaar verdeling	Gehele groep	250	50	246	48
	Subgroep				
Stratum	Stratum 1	261	47	251	47
	Stratum 2	240	48	245	47
	Stratum 3	213	49	225	49
Formatiegewicht	Fg 1	256	48	250	47
	Fg 2	228	49	233	47
	Fg 3	207	48	213	47
Geslacht	Jongens	258	49	254	47
	Meisjes	242	49	238	47
Leertijd	Regulier	253	49	249	47
	Vertraagd	233	51	229	48
Thuis taal	Alleen Nederlands (NLs)	256	48	250	47
	NLs en een andere taal	212	49	215	48
	Alleen een andere taal	223	50	234	48
Regio	Noord	247	48	252	47
	Oost	262	48	250	47
	West	248	50	243	49
	Zuid	245	51	247	48
Urbanisatiegraad	Zeer sterk	233	52	233	49
	Sterk	249	50	242	49
	Matig	255	50	248	47
	Weinig	258	47	252	47
	Niet	251	47	251	47
Ter referentie:	Typische leerling	Schaalwaarde (geen verschil tussen de jaren)			
	Zeer laag vaardige leerling (ZLV)	186		186	
	Laag vaardige leerling (LV)	216		216	
	Standaard leerling (ST)	250		250	
	Hoog vaardige leerling (HV)	284		284	
	Zeer hoog vaardige leerling (ZHV)	314		314	

Tabel 7.4.1d Gemiddelde (gem.) en standaarddeviatie (SD) van de groepen in leerjaar op de rekenschaal Meten, tijd en geld (MTG) in 2008 en in 2009.

		<u>2008</u>		<u>2009</u>	
		Gem.	SD	Gem.	SD
Jaar verdeling	Gehele groep	250	50	250	49
	Subgroep				
Stratum	Stratum 1	259	48	257	48
	Stratum 2	244	48	246	48
	Stratum 3	217	49	226	50
Formatiegewicht	Fg 1	255	48	255	48
	Fg 2	228	49	234	48
	Fg 3	211	49	215	48
Geslacht	Jongens	259	49	259	48
	Meisjes	242	49	242	48
Leertijd	Regulier	253	49	253	48
	Vertraagd	234	51	233	50
Thuis taal	Alleen Nederlands (NLs)	257	48	255	48
	NLs en een andere taal	213	49	215	49
	Alleen een andere taal	224	50	237	49
Regio	Noord	245	48	258	48
	Oost	263	49	250	48
	West	246	50	247	50
	Zuid	248	50	254	49
Urbanisatiegraad	Zeer sterk	236	52	234	50
	Sterk	251	50	246	50
	Matig	255	50	255	48
	Weinig	253	48	253	48
	Niet	250	48	257	48
Ter referentie:	Typische leerling	Schaalwaarde (geen verschil tussen de jaren)			
	Zeer laag vaardige leerling (ZLV)	186		186	
	Laag vaardige leerling (LV)	216		216	
	Standaard leerling (ST)	250		250	
	Hoog vaardige leerling (HV)	284		284	
	Zeer hoog vaardige leerling (ZHV)	314		314	

Tabel 7.4.2a Effectgrootte op de vaardigheid Rekenen: Getallen en getalsrelaties in jaargroep 4

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Jaar	2009 – 2008	-0,01	niet significant
Jaar (Stratum 1)	2009 – 2008	-0,11	geen effect
Jaar (Stratum 2)	2009 – 2008	0,00	niet significant
Jaar (Stratum 3)	2009 – 2008	0,35	klein effect
Stratum (2008)	Stratum2 – Stratum1	-0,35	klein effect
	Stratum3 – Stratum2	-0,39	klein effect
	Stratum3 – Stratum1	-0,75	matig effect
Stratum (2009)	Stratum2 – Stratum1	-0,24	klein effect
	Stratum3 – Stratum2	-0,05	niet significant
	Stratum3 – Stratum1	-0,29	klein effect
Formatiegewicht	F1 – F2	-0,34	klein effect
	F3 – F2	-0,11	niet significant
	F3 – F1	-0,45	klein effect
Geslacht	Meisjes – Jongens	-0,42	klein effect
Leertijd	Vertraagd – Regulier	-0,36	klein effect
Regio	Oost – Noord	0,09	niet significant
	West – Noord	0,00	niet significant
	Zuid – Noord	0,05	niet significant
	West – Oost	-0,09	niet significant
	Zuid – Oost	-0,03	niet significant
	Zuid – West	0,05	niet significant
Urbanisatiegraad	sterk – zeer sterk	0,04	niet significant
	matig – sterk	0,01	niet significant
	weinig – matig	0,00	niet significant
	niet – weinig	-0,05	niet significant
Thuistaal	NLs+Ander – Alleen NLs	-0,33	klein effect
	Alleen ander – Alleen NLs	-0,35	klein effect
	NLs+Ander – Alleen ander	0,02	niet significant

Niet significant is niet significant bij $\alpha = 0,05$; NB: * dan niet significant bij $\alpha = 0,01$

Tabel 7.4.2b Effectgrootte op de vaardigheid Rekenen: Optellen en Aftrekken in jaargroep 4

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Jaar	2009 – 2008	0,02	niet significant
Jaar (Stratum 1)	2009 – 2008	-0,13	geen effect
Jaar (Stratum 2)	2009 – 2008	0,16	geen effect *
Jaar (Stratum 3)	2009 – 2008	0,37	klein effect
Stratum (2008)	Stratum2 – Stratum1	-0,27	klein effect
	Stratum3 – Stratum2	-0,27	klein effect
	Stratum3 – Stratum1	-0,54	matig effect
Stratum (2009)	Stratum2 – Stratum1	0,02	niet significant
	Stratum3 – Stratum2	-0,05	niet significant
	Stratum3 – Stratum1	-0,03	niet significant
Formatiegewicht	F1 – F2	-0,30	klein effect
	F3 – F2	0,02	niet significant
	F3 – F1	-0,27	klein effect
Geslacht	Meisjes – Jongens	-0,32	klein effect
Leertijd	Vertraagd – Regulier	-0,36	klein effect
Regio	Oost – Noord	0,14	niet significant
	West – Noord	0,11	niet significant
	Zuid – Noord	0,13	niet significant
	West – Oost	-0,03	niet significant
	Zuid – Oost	-0,01	niet significant
	Zuid – West	0,02	niet significant
Urbanisatiegraad	sterk – zeer sterk	0,02	niet significant
	matig – sterk	-0,01	niet significant
	weinig – matig	0,08	niet significant
	niet – weinig	-0,08	niet significant
Thuistaal	NLs+Ander – Alleen NLs	-0,27	niet significant
	Alleen ander – Alleen NLs	-0,13	klein effect
	NLs+Ander – Alleen ander	-0,15	niet significant

Niet significant is niet significant bij $\alpha = 0,05$; NB: * dan niet significant bij $\alpha = 0,01$

Tabel 7.4.2c Effectgrootte op de vaardigheid Rekenen: Vermenigvuldigen en Delen in jaargroep 4

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Jaar	2009 – 2008	-0,06	niet significant
Jaar (Stratum 1)	2009 – 2008	-0,22	klein effect
Jaar (Stratum 2)	2009 – 2008	0,10	niet significant
Jaar (Stratum 3)	2009 – 2008	0,29	klein effect
Stratum (2008)	Stratum2 – Stratum1	-0,38	klein effect
	Stratum3 – Stratum2	-0,37	klein effect
	Stratum3 – Stratum1	-0,75	matig effect
Stratum (2009)	Stratum2 – Stratum1	-0,06	niet significant
	Stratum3 – Stratum2	-0,19	geen effect *
	Stratum3 – Stratum1	-0,25	klein effect
Formatiegewicht	F1 – F2	-0,31	klein effect
	F3 – F2	-0,22	klein effect
	F3 – F1	-0,54	matig effect
Geslacht	Meisjes – Jongens	-0,36	klein effect
Leertijd	Vertraagd – Regulier	-0,34	klein effect
Regio	Oost – Noord	0,09	niet significant
	West – Noord	-0,01	niet significant
	Zuid – Noord	-0,02	niet significant
	West – Oost	-0,10	geen effect
	Zuid – Oost	-0,11	geen effect
	Zuid – West	0,00	niet significant
Urbanisatiegraad	sterk – zeer sterk	-0,04	niet significant
	matig – sterk	0,07	niet significant
	weinig – matig	-0,03	niet significant
	niet – weinig	-0,04	niet significant
Thuistaal	NLs+Ander – Alleen NLs	-0,23	klein effect
	Alleen ander – Alleen NLs	-0,23	klein effect
	NLs+Ander – Alleen ander	0,00	niet significant

Niet significant is niet significant bij $\alpha = 0,05$; NB: * dan niet significant bij $\alpha = 0,01$

Tabel 7.4.2d Effectgrootte op de vaardigheid Meten, Tijd en Geld in jaargroep 4

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Jaar	2009 – 2008	0,02	niet significant
Jaar (Stratum 1)	2009 – 2008	-0,04	niet significant
Jaar (Stratum 2)	2009 – 2008	0,04	niet significant
Jaar (Stratum 3)	2009 – 2008	0,22	klein effect *
Stratum (2008)	Stratum2 – Stratum1	-0,25	klein effect
	Stratum3 – Stratum2	-0,35	klein effect
	Stratum3 – Stratum1	-0,60	matig effect
Stratum (2009)	Stratum2 – Stratum1	-0,18	geen effect
	Stratum3 – Stratum2	-0,17	geen effect *
	Stratum3 – Stratum1	-0,35	klein effect
Formatiegewicht	F1 – F2	-0,35	klein effect
	F3 – F2	-0,16	niet significant
	F3 – F1	-0,52	matig effect
Geslacht	Meisjes – Jongens	-0,38	klein effect
Leertijd	Vertraagd – Regulier	-0,33	klein effect
Regio	Oost – Noord	0,08	niet significant
	West – Noord	-0,04	niet significant
	Zuid – Noord	0,04	niet significant
	West – Oost	-0,12	geen effect
	Zuid – Oost	-0,04	niet significant
	Zuid – West	0,08	niet significant
Urbanisatiegraad	sterk – zeer sterk	-0,06	niet significant
	matig – sterk	0,07	niet significant
	weinig – matig	-0,11	geen effect
	niet – weinig	0,04	niet significant
Thuisstaal	NLs+Ander – Alleen NLs	-0,26	klein effect
	Alleen ander – Alleen NLs	-0,33	klein effect
	NLs+Ander – Alleen ander	0,07	niet significant

Niet significant is niet significant bij $\alpha = 0,05$; NB: * dan niet significant bij $\alpha = 0,01$

8 Discussie

In deze discussie volgt een samenvatting van het in de vorige hoofdstukken beschreven onderzoek naar de resultaten op de taal- en rekenvaardigheid in jaargroep 8 en jaargroep 4 in het reguliere basisonderwijs. Na een korte herhaling van het doel en de context van het onderzoek volgt een beknopte beschrijving van de onderzoeksopzet en de onderzoeksmethode. Voor de omschrijving van de vaardigheden wordt verwezen naar hoofdstuk 5 tot en met 7. Meer aandacht zal echter uitgaan naar de relatie tussen de achtergrondvariabelen en de gevonden resultaten. Daar waar de resultaten in de voorgaande hoofdstukken geordend waren per vaardigheid, worden in deze discussie voor de twee leerjaren afzonderlijk de resultaten geordend per achtergrondvariabele. Speciale aandacht zal daarbij uitgaan naar de achtergrondvariabele afnamejaar, om te zien waar de resultaten van 2008 en 2009 van elkaar verschilden.

8.1 Doel en context

In het kader van de kwaliteitsagenda 'Scholen voor morgen' is in 2008 het Jaarlijks Peilingsonderzoek van het Onderwijsniveau (JPON) van start gegaan. JPON is gericht op het monitoren van het onderwijsniveau op het gebied van taal- en rekenvaardigheid in het basisonderwijs. JPON staat niet op zichzelf. Het heeft raakvlakken met onder meer het Periodieke Peilingsonderzoek van het Onderwijsniveau in Nederland (PPON) en het Cohort Onderzoek Onderwijsloopbanen onder leerlingen van 5 tot 18 jaar (COOL 5-18). Voor meer informatie verwijzen wij u naar het verslag van het eerste onderzoek voor JPON uit 2008 (Hemker & Van Weerden, 2009).

In het verslag van het onderzoek in 2008 wordt gerapporteerd over de eerste jaarlijkse niveaupeiling van taal en rekenen in jaargroep 8 en 4 (Hemker & Van Weerden, 2009). In het voorliggende verslag worden de resultaten op deze vaardigheden bij deze groepen in 2009 beschreven, waarbij die resultaten vergeleken worden met die van 2008.

Het belangrijkste doel van JPON is het nauwkeurig vaststellen van veranderingen over de afnamejaren in de taal- en rekenvaardigheden van leerlingen in jaargroep 8 en 4. Daarnaast is het doel na te gaan in hoeverre de prestatiekloof tussen bepaalde groepen leerlingen en scholen gelijk gebleven is dan wel groter of kleiner is geworden. Naast de variabele **Afnamejaar** (2008 of 2009), zijn de volgende achtergrondkenmerken in het onderzoek betrokken:

- **Geslacht:**
- **Leertijd:** is een leerling die al dan niet is vertraagd;
- **Formatiegewicht:** leerlinggewicht dat gebruikt wordt door CFI (Centrale Financiën Instellingen) voor de financiering van scholen. Hieruit kan worden afgeleid of een leerling een achterstandsleerling is;
- **Stratum.** Scholen met bijna uitsluitend allochtone leerlingen behoren tot stratum 3, scholen met overwegend leerlingen met de gewichten 1.25 en 1.90 tot stratum 2 en de overige scholen met weinig 'gewichtleerlingen' tot stratum 1.

Bovenstaande kenmerken zijn, naast het afnamejaar, de basisvariabelen. Deze variabelen zijn in de meeste analyses gebruikt in dit onderzoek. Overige kenmerken die zijn meegenomen:

- **Thuis taal:** Nederlands dan wel een andere taal;
- **Regio:** Op basis van de plaats waar de school staat, onderverdeeld in 4 regio's (Noord, Oost, West, en Zuid; CBS-definitie op basis van provincie);
- **Urbanisatiegraad:** Op basis van de plaats waar de school staat (postcode), onderverdeeld in 5 niveaus, variërend van Zeer sterk stedelijk tot Niet stedelijk.

Bij de analyses bij jaargroep 8 zijn extra twee variabelen meegenomen die niet relevant waren in jaargroep 4:

- **Advies VO:** Welk voortgezet onderwijs wordt voor de leerling adviseert de leraar voor de leerling.
- **Toets:** Reguliere Eindtoets basisonderwijs, dan wel de gemakkelijker Niveautoets.

8.2 Onderzoekopzet en de onderzoeksmethode

8.2.1 Wat is er gemeten?

De gemeten taalvaardigheden in jaargroep 4 en jaargroep 8 waren in 2008 en 2009 Begrijpend lezen, Spelling en Woordenschat. Voor Rekenen/Wiskunde verschilde de indeling in onderscheiden vaardigheden voor jaargroep 4 en jaargroep 8. In jaargroep 8 zijn dat Getallen en Bewerkingen, Breuken, Procenten en Verhoudingen en Meten, tijd en geld. In jaargroep 4 zijn dat er niet drie maar vier: Getallen en getalsrelaties, Optellen en aftrekken, Vermenigvuldigen en delen, en Meten, tijd en geld.

Om te zorgen dat de toetsdruk op scholen niet te groot wordt is er zo veel mogelijk gebruikgemaakt van toetsen die toch al op de scholen worden afgenomen. De scholen hoefden daardoor niet veel extra tijd vrij te maken. Voor de meting van de taal- en rekenvaardigheid in jaargroep 8 zijn eerder verzamelde gegevens uit de Eindtoets Basisonderwijs 2008 en 2009 opnieuw gebruikt. Daarnaast is gebruikgemaakt van opgaven uit de bijbehorende Niveautoets en het zogeheten boekje Extra. In jaargroep 4 is gebruikgemaakt van toetsen uit LOVS. Dat is het leerlingvolgsysteem van Cito. De resultaten op de toetsen worden bij reguliere afnamen niet terug gestuurd naar Cito, maar dat is voor dit onderzoek wel gedaan, te samen met de relevante achtergrondkenmerken van de leerlingen die de toetsen gemaakt hebben. Ook is een aantal scholen benaderd dat normaal niet het LOVS gebruikt.

In jaargroep 8 was de steekproef van scholen en leerlingen in beide jaren aanzienlijk groter dan in jaargroep 4. De responsgroepen voor jaargroep 8 en 4 zijn beide representatief ten opzichte van de desbetreffende populaties in Nederland.

8.2.2 Statistische analyse

Om de vaardigheden van 2008 en 2009 direct met elkaar te kunnen vergelijken, moet er aan ten minste twee voorwaarden zijn voldaan:

- A De vaardigheden moeten in 2009 op dezelfde schaal gemeten zijn als in 2008;
- B De samenstelling van de steekproef mag niet gewijzigd zijn (bijvoorbeeld meer of minder zittenblijvers bevatten).

Aan voorwaarde A is voldaan als de steekproef van 2009 precies dezelfde toetsen heeft gemaakt als de steekproef van 2009. Voor jaargroep 4 is aan deze voorwaarde voldaan, aangezien de leerlingen in beide gevallen ongewijzigde toetsen uit het reguliere LOVS hebben gemaakt. De voor jaargroep 8 gebruikte Eindtoets Basisonderwijs wordt echter omwille van de geheimhouding ieder jaar volledig ververs. Wij hebben de vergelijkbaarheid echter op een aantal manieren gewaarborgd. Een eerste mogelijkheid was gebruikmaking van een Ankerboekje bij de Eindtoets dat bij een deel van de achtstegroepers in 2008 en 2009 was afgenomen en dat in beide jaren deels dezelfde opgaven bevat. De tweede mogelijkheid was het analyseren van een aantal extra opgaven die een deel van de leerlingen maken uit het zogeheten toetsboekje Extra. Met behulp van een speciale analysetechniek – een zogeheten itemresponsemodel (IRT-model) – konden de prestaties van 2009 toch met die van 2008 op dezelfde meetschaal met elkaar worden vergeleken (vgl. Hemker & Van Weerden, 2008, 2009).

Aan voorwaarde B is voldaan als de samenstelling van steekproef in 2009 vergelijkbaar is met die in 2008. Veranderingen in de samenstelling van een steekproef kunnen een gevolg zijn van zogenaamde steekproeffluctuaties en 'echte' verandering in de samenstelling van de populatie. Het probleem van steekproeffluctuaties zal zich voor jaargroep 8 niet zo gauw voordoen. Zowel in 2008 als in 2009 is namelijk een zeer grote steekproef 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 scholen bijvoorbeeld overgaan op een strenger doubleerbeleid. De responsgroep in 2009 zou dan minder zittenblijvers bevatten dan die in 2008. We zouden dan ten onrechte kunnen concluderen dat de vaardigheid van de leerlingen vooruit is gegaan. Gelukkig zijn er tegenwoordig 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 jaargroep 4 is de steekproef veel minder groot dan in jaargroep 8 (namelijk 'slechts' ongeveer 2500 leerlingen van ongeveer 100 scholen). Daar waar in jaargroep 8 geobserveerde veranderingen in samenstelling van de steekproef opgevat kunnen worden als veranderingen in de populatie, is de omvang van de beide steekproeven in jaargroep 4 te klein om een dergelijke aanname te kunnen doen. Zodoende zal bij de analyses in jaargroep 4 voor de jaarvergelijking alleen gebruikgemaakt moeten worden van de gecorrigeerde analyses om jaarverschillen te kunnen interpreteren. Overigens moet opgemerkt worden dat bij de vaardigheid Woordenschat in jaargroep 4 voorwaarden A en B gevaar lopen: er is in beide jaren een andere toets gebruikt, en ook de achtergrondkenmerken die bekend waren verschilden in de twee jaren (zie hoofdstuk 7). Hierdoor was een jaarvergelijking tussen 2008 en 2009 voor die vaardigheid in jaargroep 4 niet goed mogelijk.

8.2.3 De gekozen rapportageschaal

Doordat de vaardigheden in 2009 op dezelfde schaal gemeten zijn als in 2008 is het mogelijk met behulp van het eerder genoemde IRT-model de opgaven en de leerlingen op een en dezelfde vaardigheidsschaal te plaatsen. De vaardigheidsschaal wordt nu gespecificeerd door de inhoud van de opgaven. Door te beschrijven wat "typische leerlingen" kunnen, variërend van zeer laag vaardig tot zeer hoog vaardig, ontstaat een duidelijk beeld van de gemeten vaardigheid en van het beheersingsniveau binnen het leerjaar. Voor de kwaliteit van de geconstrueerde meetschalen en de uitkomsten op itemniveau verwijzen we naar hoofdstuk 5, 6 en 7.

De getalswaarden die aan deze vaardigheidsschalen gegeven worden, zijn verkregen door middel van een transformatie naar een schaal met een gemiddelde van 250 en een standaarddeviatie van 50. Dit is conform de werkwijze bij PPON (zie bijv. Janssen, Van der Schoot & Hemker, 2005). De startwaarde is voor elke schaal het gemiddelde dat we in 2008 hebben aangetroffen. Dat gemiddelde is arbitrair 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 met elkaar kunnen vergelijken.

De getalswaarde van de schaal functioneert net zoals de waarden van de IQ-schaal. Daar waar IQ per leeftijd is gedefinieerd, zo is de vaardigheidsschaal per leerjaar bepaald. Zo representeert de waarde 250 in jaargroep 8 een hoger absoluut niveau van vaardigheid dan in jaargroep 4, net zoals een kind van 7 met een IQ van 100 andere dingen kan dan een kind van 11. Ook moet opgemerkt worden dat een vaardigheidsniveau van een enkele leerling op bijvoorbeeld Begrijpend lezen niet betekent dat deze leerling ook dat niveau heeft bij getallen en getalsrelaties. Verschillende leerlingen hebben verschillende talenten.

Naast het plaatsen van de opgaven en de individuele leerlingen op de vaardigheidsschaal, is het ook mogelijk verdelingen van de groepen leerlingen op die schaal te plaatsen. Dit zijn dan de ongecorrigeerde gemiddelden van de groepen. Echter, om verschillen tussen groepen zinvol te kunnen interpreteren is het zuiverder om de effectschattingen te interpreteren, in plaats van deze ongecorrigeerde gemiddelden op de vaardigheidsschaal. Vandaar dat in deze discussie de nadruk ligt op de effectschattingen.

8.2.4 Gebruikte analysemodellen voor effectschattingen

Bij de vergelijking van de resultaten over de jaren is gebruikgemaakt van verschillende analysemodellen die de relatie tussen de achtergrondvariabelen en de resultaten op de vaardigheidsschalen beschrijven. In het geval van jaargroep 8 was het mogelijk een "ongecorrigeerde" jaarvergelijking te doen. De resultaten in

2008 en 2009 zijn hier direct met elkaar vergeleken, waarbij verschillen in verdeling van de achtergrondkenmerken als veranderingen in de populatie worden opgevat. Zoals eerder aangegeven, was dat niet mogelijk in jaargroep 4.

Het model dat zowel in jaargroep 4 als in jaargroep 8 is gebruikt, is het hoofdeffectenmodel met als variabele afnamejaar. De achtergrondkenmerken waarvoor gecorrigeerd wordt zijn geslacht, leertijd, stratum en formatiegewicht. In het hoofdeffectenmodel wordt er van uitgegaan dat het jaareffect over alle variabelen gelijk is. Zo spreken we van een hoofdeffect van geslacht als het vaardigheidsverschil tussen jongens en meisjes in beide jaren gelijk is. Het is echter mogelijk dat er sprake van een interactie is tussen een correctievariabele en afnamejaar. Een interactie tussen geslacht en jaareffect wil dan zeggen dat prestatieverschil tussen jongens en meisjes het ene jaar groter of kleiner is dan het andere jaar. Bij alle schalen is gekeken of er sprake was van een significante interactie tussen afnamejaar en een van de corrigerende variabelen. In die gevallen is er gebruikgemaakt van een interactiemodel. Bij de taalvaardigheden in jaargroep 8 was er sprake van een interactie tussen afnamejaar en formatiegewicht. Bij de vaardigheden in jaargroep 4 was er sprake van een interactie tussen afnamejaar en stratum.

Of een gemiddeld vaardigheidsverschil tussen 2009 en 2008 statistisch significant is, 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 in jaargroep 4 en 8 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 2008 en 2009 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 8.1.

Tabel 8.1 *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

8.3 De resultaten voor jaargroep 8

In hoofdstuk 5 en 6 zijn de resultaten gegeven per vaardigheidschaal voor de taal- en rekenvaardigheden afzonderlijk. In de discussie worden deze samengevat per achtergrondkenmerk, beginnend met het afnamejaar. In paragraaf 8.4 gebeurt dat ook voor de resultaten van hoofdstuk 7 met betrekking tot jaargroep 4.

8.3.1 De vergelijking van 2009 met 2008

Ervan uitgaande dat steekproeffluctuaties geen rol spelen (zie paragraaf 8.2.4) vergelijken we in deze paragraaf de resultaten van jaargroep 8 in 2008 met die in 2009. Eerst presenteren we de ongecorrigeerde en vervolgens de gecorrigeerde verschillen (d.w.z. gezuiverd voor eventuele verschillen in de verdeling van de achtergrondkenmerken). In het geval van de taalvaardigheden was er zoals gezegd sprake van interactie met de variabele formatiegewicht. Zodoende bekijken we bij taal ook de jaareffecten per formatiegewicht.

Ongecorrigeerd

In het model met ongecorrigeerde jaareffecten is afnamejaar de enige verklarende variabele. Binnen dit model worden verschuivingen in de verdeling van de achtergrondvariabelen geïnterpreteerd als

populatieverschuivingen. Als dit effect heeft op de vaardigheidsverdeling in de populatie wordt dit direct meegenomen in de analyses als een daadwerkelijke verschuiving in vaardigheid. Daar zou je dan niet voor willen corrigeren. Deze ongecorrigeerde effecten zijn gegeven in tabel 8.2.

Tabel 8.2 Effecten voor jaarvergelijking (Jaar: 2008-2009) jaargroep 8 – ongecorrigeerd

Taalvaardigheden	Effectgrootte	Effectkwalificatie
Begrijpend Lezen	0,04	geen effect
Spelling	0,00	niet significant
Woordenschat	-0,02	niet significant
Rekenvaardigheden		
Getallen en Bewerkingen	0,00	niet significant
Breuken, Procenten en Verhoudingen	0,00	niet significant
Metten, Meetkunde, Tijd en Geld	-0,01	niet significant

“Niet significant” : bij $\alpha = ,01$; “Geen effect” als effect niet groter dan 0,2 en niet kleiner dan -,2.

Er zijn drie effectgroottes die afwijken van 0, één positief en twee negatief. Alleen bij Begrijpend lezen is er sprake van een in statistisch opzicht significant verschil. De gevonden 0.04 ligt echter nog ver onder de grens van 0.20 om betekenisvol te mogen worden genoemd (zie tabel 8.1).

In het geval van de ongecorrigeerde effecten zijn de populatieverdelingen vergelijkbaar met de gevonden effecten. In tabel 8.3 worden deze gegeven. Het gemiddelde voor 2008 is per definitie 250 en de standaarddeviatie is 50 (zie verder Hemker & Van Weerden, 2009). De schaalwaarde van 250 uit 2008 vormt het criterium waarmee we de uitkomsten van 2009 vergelijken. Ook is weergegeven welke percentielscore daarbij hoort, dat wil zeggen het percentage leerlingen met een gelijke of lagere score dan 250. Als het verschil tussen 2008 en 2009 statistisch significant is, zoals gevonden bij de effectschattingen (tabel 8.2), is het gemiddelde van 2009 vetgedrukt.

Tabel 8.3 Jaarvergelijking jaargroep 8 – ongecorrigeerd (vet: significant)

Jaar	2008			2009		
	Gem*	SD	perc	Gem	SD	perc
Taal						
Begrijpend lezen	250	50	(50)	252	50	(52)
Spelling	250	50	(50)	250	50	(50)
Woordenschat	250	50	(50)	249	49	(49)
Rekenen						
Getallen en Bewerkingen	250	50	(50)	250	50	(50)
Breuken, Procenten en Verhoudingen	250	50	(50)	250	50	(50)
Metten, Meetkunde, Tijd en Geld	250	50	(50)	249	50	(49)

* Gem = gemiddelde; sd = standaarddeviatie; perc = gemiddelde percentielscore basis 2008

Bij Taal zien we dat jaargroep 8 in Begrijpend lezen een iets hoger gemiddelde bereikt dan in 2008, en bij de effectschatting statistisch significant. Spelling levert dezelfde waarde op als in 2008 en Woordenschat komt een punt lager uit. We zien deze verschuivingen ook terug in de percentielscores. Bij Rekenen zien we alleen een verandering bij Metten, Meetkunde, Tijd en Geld. Het verschil tussen 2008 met 2009 is één punt op de meetschaal.

Gecorrigeerd

De gecorrigeerde effecten zijn weergegeven in tabel 8.4. Dat wil zeggen dat de resultaten gezuiverd zijn voor eventuele verschillen in de samenstelling van de steekproeven naar geslacht, leertijd, stratum en formatiegewicht. Het model dat gebruikt is om de waarden van tabel 8.4 te verkrijgen gaat er vanuit dat geen van de interacties van de achtergrondkenmerken met het jaar van afname significant was. Het betreft dus de effecten in het zogeheten hoofdeffectenmodel (zie paragraaf 8.2.4).

Tabel 8.4 Jaareffect (Jaar: 2009 – 2008) in hoofdeffectenmodel

	Effectgrootte	Effectkwalificatie
Taalvaardigheden		
Begrijpend lezen	0,03	geen effect
Spelling	-0,01	niet significant
Woordenschat	-0,03	geen effect
Rekenvaardigheden		
Getallen en Bewerkingen	-0,01	niet significant
Breuken, Procenten en Verhoudingen	-0,01	niet significant
Meten, Meetkunde, Tijd en Geld	-0,02	niet significant

“Niet significant” : bij $\alpha=,01$; “Geen effect” als effect niet groter dan 0,2 en niet kleiner dan -,2.

In de vergelijking met tabel 8.3 valt op dat de effecten allen iets kleiner zijn dan in het ongecorrigeerde model. Er is verder op de effecten van de basisvariabelen ingegaan door de interacties te onderzoeken. Het bleek dat de interacties bij de rekenvaardigheden niet significant waren, en dat de gevonden effectgroottes ook binnen de diversie subgroepen dicht bij elkaar lagen, en vergelijkbaar zijn met de waarden zoals gegeven in tabel 8.4.

Bij de taalvaardigheden was er wel sprake van een interactie-effect tussen afnamejaar en formatiegewicht. Per formatiegewicht zijn er verschillen gevonden voor wat betreft het jaareffect. Deze resultaten van het jaareffect in het interactiemodel (zie paragraaf 2.1) zijn weergegeven in tabel 8.5.

Tabel 8.5 Jaareffect (Jaar: 2009 – 2008) bij taal, per formatiegewicht (in interactiemodel)

Formatiegewicht	Taalvaardigheid	Effectgrootte	Effectkwalificatie
F1.00	Begrijpend Lezen	0,06	geen effect
F1.00	Spelling	-0,04	geen effect
F1.00	Woordenschat	-0,07	geen effect
F1.25	Begrijpend Lezen	0,00	niet significant
F1.25	Spelling	0,00	niet significant
F1.25	Woordenschat	-0,09	niet significant
F1.90	Begrijpend Lezen	-0,10	geen effect
F1.90	Spelling	0,14	geen effect
F1.90	Woordenschat	0,22	klein effect

“Niet significant” : bij $\alpha = ,01$; “Geen effect” als effect niet groter dan 0,2 en niet kleiner dan -,2.

De resultaten van de vergelijking tussen 2008 en 2009 kunnen als volgt worden beschreven:

- Bij de 1.00-leerlingen is er sprake van een significante vooruitgang in Begrijpend lezen en een significante achteruitgang in Spelling en Woordenschat. Bij geen van de taalvaardigheden is het effect groter dan 0,2 of kleiner dan -0,2, en is dus geclassificeerd als afwezig.
- Bij de 1.25-leerlingen zijn de prestaties voor alle drie onderdelen niet aantoonbaar veranderd, als is er bij Woordenschat een negatieve tendens (zij het niet significant).
- Bij de 1.90-leerlingen is er sprake van een significante vooruitgang in Spelling en Woordenschat en een achteruitgang bij Begrijpend lezen. Alleen bij Woordenschat is het effect groot genoeg voor een interpretatie, zij het dat het effect klein is.

Gezien de hoogte van de effectgrootte kan slechts een van de verschillen als betekenisvol worden gekarakteriseerd. Het betreft het onderdeel Woordenschat waar 1.90-leerlingen in 2009 een deel van hun achterstand ten opzichte van 1.00-leerlingen hebben goedgehaakt.

8.3.2 De vergelijking van de overige achtergrondvariabelen

Bij de samenvattingen van de effecten bij de andere achtergrondvariabelen richten we ons hier op de effectschattingen aangezien die beter te interpreteren zijn dan de ongecorrigeerde schattingen van de verdelingen van de groepen op de vaardigheidsschaal. De effectschattingen kunnen geïnterpreteerd worden door middel van hun effectgrootte en worden direct getoetst op significantie. Voor de schattingen van de verdelingen verwijzen we naar de hoofdstukken 5, 6 en 7. De effecten zijn berekend in modellen waarin altijd de basisvariabelen (stratum, formatiegewicht, geslacht en leertijd) voorkomen. Later zijn daar de overige variabelen aan toegevoegd.

Formatiegewicht

Dat er bij de taalvaardigheden een interactie gevonden is tussen afnamejaar en formatiegewicht heeft ook gevolgen voor de interpretatie van het formatiegewicht bij deze vaardigheden. De effecten van de formatiegewichten verschillen per afnamejaar. Dit betekent dat als we de contrasten bekijken voor de formatiegewichten er bij de taalvaardigheden onderscheid gemaakt wordt tussen 2008 en 2009. Bij de rekenvaardigheden was dat niet nodig: daar is de situatie in 2009 niet significant veranderd ten opzichte van 2008.

Tabel 8.6.1 Effectgrootten bij het achtergrondkenmerk formatiegewicht (F)

	F1.25–F1.00	F1.90–F1.25	F1.90–F1.00
Taalvaardigheden (per afnamejaar)			
Begrijpend lezen (2008)	-0,55	0,02	-0,54
Begrijpend lezen (2009)	-0,61	-0,08	-0,69
Spelling (2008)	-0,41	0,22	-0,19
Spelling (2009)	-0,38	0,36	-0,02
Woordenschat (2008)	-0,54	-0,51	-1,05
Woordenschat (2009)	-0,56	-0,20	-0,76
Rekenvaardigheden (gelijk over afnamejaar)			
Getallen en Bewerkingen	-0,48	0,24	-0,24
Breuken, Procenten en Verhoudingen	-0,50	0,15	-0,35
Meten, Meetkunde, Tijd en Geld	-0,52	0,16	-0,36

Bij dit achtergrondkenmerk zien we een gevarieerd beeld. Bij elke taalvaardigheid is de situatie anders. De belangrijkste resultaten vatten we als volgt samen:

- Het verschil tussen een 1.00-leerling ('reguliere leerling') en een 1.25 leerling ('Nederlandse achterstandsleerling') is significant en het effect ligt rond de 0,50. Daarbij past de kwalificatie 'matig'. Het grootste verschil wordt gevonden bij Begrijpend lezen en het verschil is in 2009 groter geworden in vergelijking met 2008. Bij Spelling is het verschil het kleinst en is het juist kleiner geworden.
- Het verschil tussen een 1.25- en een 1.90-leerling ('achterstandsleerling met niet-Nederlandse achtergrond') varieert zeer over de vaardigheden heen. Bij Woordenschat is de vaardigheid groter bij de 1.25-leerlingen, al is het verschil tussen de formatiegewichten duidelijk afgenomen van 2008 naar 2009. Bij Spelling en de rekenvaardigheden presteren de 1.90-leerlingen relatief beter. Bij Spelling is het verschil groter geworden in vergelijking met 2008.
- Het verschil tussen de 1.00- en de 1.90-leerlingen varieert ook over de vaardigheden. Daarbij scoren echter de 1.00-leerlingen altijd hoger dan de 1.90 leerlingen. Het grootst is het verschil bij Woordenschat. Het verschil tussen de 1.00 en de 1.90-leerlingen is in 2009 echter wel kleiner geworden in vergelijking met 2008. Dat geldt ook voor Spelling waar het verschil tussen de formatiegewichten in 2009 vrijwel nihil is. Bij Begrijpend lezen is het verschil echter groter geworden. Bij de rekenvaardigheden wordt een effect gevonden dat tussen matig en klein geassocieerd kan worden.

Stratum

Ten aanzien van de achtergrondvariabele stratum is er bij de leerlingen in jaargroep 8 amper verschil tussen de resultaten in 2008 en 2009. Een uitsplitsing naar afnamejaar is zodoende ook niet nodig. Dat geldt overigens ook voor alle andere variabelen in jaargroep 4 die hieronder gepresenteerd worden.

Scholen met bijna uitsluitend allochtone leerlingen behoren tot stratum 3, scholen met overwegend leerlingen met de gewichten 1.25 en 1.90 tot stratum 2 en de overige scholen met weinig 'gewicht-leerlingen' tot stratum 1. In alle gevallen zijn de gevonden effecten zodanig dat stratum 1-scholen beter presteren dan stratum 2-scholen, die op hun beurt weer beter presteren dan stratum 3-scholen. Hierbij moet opgemerkt worden dat het hier om effecten gaat die gecorrigeerd zijn voor formatiegewicht. Het gaat dus om een bijkomend effect van de school, naast de leerlingeffecten die bij *formatiegewicht* beschreven zijn. Als naar de ongecorrigeerde effecten (de populatieverdelingen) gekeken wordt dan zullen de verschillen groter zijn.

Tabel 8.6.2 Effectgrootten bij het achtergrondkenmerk stratum (S)

Taalvaardigheden	S2 – S1	S3 – S2	S3 – S1
Begrijpend lezen	-0,11	-0,15	-0,27
Spelling -0,06	-0,02	-0,07	
Woordenschat	-0,08	-0,15	-0,23
Rekenvaardigheden			
Getallen en Bewerkingen	-0,06	-0,02	-0,08
Breuken, Procenten en Verhoudingen	-0,09	-0,07	-0,16
Meten, Meetkunde, Tijd en Geld	-0,10	-0,10	-0,20

De belangrijkste resultaten vatten we als volgt samen:

- Het verschil tussen stratum 2- en stratum 1-scholen is zo klein dat het als "geen effect" geassocieerd wordt. Dat geldt ook voor het verschil tussen de stratum 3- en stratum 2-scholen.
- Het effect tussen de stratum 3- en stratum 1-scholen kan bij Begrijpend lezen, Woordenschat en de rekenvaardigheid met betrekking tot meten, meetkunde, tijd en geld kan als klein geassocieerd worden. Bij de andere vaardigheden is er sprake van "geen effect".

Merk op dat de verschillen over de scholen veel groter zijn als in het model niet gecorrigeerd wordt voor de variabele formatiegewicht. De verschillen tussen de stratum 2- en stratum 1-scholen worden dan ongeveer twee keer zo groot, en de verschillen tussen de stratum 3- en de stratum 2-scholen worden 2 tot 3 keer zo groot. Dat geldt ook voor de verschillen tussen de stratum 3- en de stratum 1-scholen.

Geslacht en leertijd

Met betrekking tot het vaardigheidsverschil tussen jongens en meisjes en de verschillen tussen regulier en vertraagde leerlingen is er niet veel veranderd. De resultaten van 2008 en 2009 lijken erg op elkaar.

Tabel 8.6.3 Effectgrootten bij het achtergrondkenmerk geslacht (meisjes – jongens)

Taalvaardigheden	Effectgrootte	Effectkwalificatie
Begrijpend lezen	0,16	geen effect
Spelling	0,33	klein
Woordenschat	-0,23	klein
Rekenvaardigheden		
Getallen en Bewerkingen	-0,43	klein
Breuken, Procenten en Verhoudingen	-0,51	matig
Meten, Meetkunde, Tijd en Geld	-0,51	matig

Tabel 8.6.4 Effectgrootten bij het achtergrondkenmerk leertijd (vertraagd – regulier)

Taalvaardigheden	Effectgrootte	Effectkwalificatie
Begrijpend lezen	-0,60	matig
Spelling	-0,69	matig
Woordenschat	-0,54	matig
Rekenvaardigheden		
Getallen en Bewerkingen	-0,67	matig
Breuken, Procenten en Verhoudingen	-0,71	matig
Meten, Meetkunde, Tijd en Geld	-0,71	matig

Samengevat zijn de belangrijkste resultaten:

- Meisjes presteren beter bij Begrijpend lezen en Spelling, terwijl jongens beter presteren bij Woordenschat en de rekenvaardigheden. De grootste verschillen worden gevonden bij de rekenvaardigheden.
- Vertraagde leerlingen, d.w.z. leerlingen die ten minste één keer zijn blijven zitten, behalen aanzienlijke lagere scores dan reguliere leerlingen. Het verschil is het kleinst bij Woordenschat. Bij de rekenvaardigheden en Spelling is het effect het grootst.

Extra variabele thuistaal

Bij de extra variabelen gaat het telkens om het additionele effect van deze variabele. Dat is het effect bovenop het effect van de basisvariabelen. Bij de extra variabele thuistaal gaat het bijvoorbeeld om het toegevoegde effect van thuistaal. Het gaat om de vraag of er een taaleffect is binnen de groep leerlingen met een gelijk formatiegewicht, gelijk geslacht en gelijke leertijd binnen een school van hetzelfde stratum.

Tabel 8.6.5 Effectgrootten bij het achtergrondkenmerk thuistaal

Thuistaal	Taalvaardigheden			Rekenvaardigheden		
	BegrLez	Spell	WoSch	GB	BPV	MMTG
Thuistaal algemeen						
NLs+ander – Alleen NLs	-0,17	0,10	-0,28	0,06	0,00	-0,02
Alleen ander – Alleen NLs	-0,23	0,05	-0,47	0,10	0,03	0,03
Alleen ander – NLs+ander	-0,06	-0,05	-0,19	0,03	0,02	0,06
Thuistaal specifiek						
Turks – Alleen NLs	-0,46	0,08	-0,84	0,15	0,01	0,00
Arabisch – Alleen NLs	-0,33	0,00	-0,48	0,02	-0,01	-0,06
Surinaams/Antilliaans – Alleen NLs	-0,53	-0,22	-0,44	-0,43	-0,50	-0,53
West-Europees – Alleen NLs	0,10	0,07	0,01	0,03	0,02	0,01

De belangrijkste resultaten voor de variabele thuistaal algemeen vatten we als volgt samen:

- De thuistaal heeft in algemene zin geen additioneel effect op de hoofdvariabelen bij Spelling en de rekenvaardigheden.
- Het spreken van een andere taal naast het Nederlands heeft bij Woordenschat een klein additioneel effect als dit vergeleken wordt met leerlingen die alleen Nederlands spreken. Bij Begrijpend lezen ligt het effect tegen klein aan.
- Thuis alleen een ander taal spreken vergeleken met alleen Nederlands spreken heeft bij Woordenschat een matig additioneel effect. Bij Begrijpend lezen is dit een klein effect.
- Het al dan niet Nederlands spreken naast een andere taal heeft alleen bij Woordenschat een additioneel effect. Dit is een (bijna) klein effect.

Voor specifieke thuistalen geldt het volgende:

- Als er naar de specifieke andere (eventueel naast Nederlands) gebruikte taal gekeken wordt, valt op dat kinderen die Surinaams / Antilliaans spreken bij alle vaardigheden duidelijk minder goed scoren.
- Voor andere specifieke thuistalen zijn er geen additionele effecten bij Spelling en rekenvaardigheid.
- Leerlingen die thuis Turks spreken (al dan niet met Nederlands erbij) presteren bij Woordenschat slechter. Het additionele effect is groot te noemen. Bij Begrijpend lezen is het effect matig .
- Leerlingen die thuis een Arabische taal spreken (al dan niet met Nederlands erbij) hebben daar bij Woordenschat ook last van, zij het dat het additionele effect matig is. Het effect is nog wat kleiner bij Begrijpend lezen.
- Als leerlingen thuis een andere West-Europese taal spreken heeft dit geen negatief additioneel effect.

Extra variabelen regio en urbanisatiegraad

De samenvatting voor het effect van regio is zeer bondig:

- Regio heeft geen additioneel effect op de basisvariabelen, op een enkele uitzondering na.
- De uitzonderingen betreffen een verschil bij de rekenvaardigheden. Leerlingen in de zuidelijke provincies (Noord-Brabant en Limburg) in vergelijking met leerlingen uit in de noordelijke provincies (Friesland Groningen, Drenthe) een positief additioneel effect laten zien bovenop de basisvariabelen.

De verschillen met de resultaten in 2008 zijn miniem. Een klein verschil is dat in 2009 in de noordelijke provincies op Begrijpend lezen en Spelling iets beter gescoord werd dan in 2008, maar er was geen sprake van een significant interactie-effect.

Tabel 8.6.6 *Effectgrootten bij het achtergrondkenmerk regio*

Regio	Taalvaardigheden			Rekenvaardigheden		
	BegrLez	Spell	WoSch	GB	BPV	MMTG
Oost – Noord	0,06	0,05	-0,02	0,11	0,10	0,09
West – Noord	0,11	0,06	0,07	0,14	0,13	0,14
Zuid – Noord	0,14	0,15	0,03	0,24	0,19	0,22
West – Oost	0,05	0,01	0,09	0,03	0,02	0,05
Zuid – Oost	0,07	0,10	0,05	0,12	0,08	0,13
Zuid – West	0,03	0,09	-0,04	0,09	0,06	0,08

De samenvatting bij urbanisatiegraad is vergelijkbaar met die bij regio:

- Urbanisatiegraad heeft per urbanisatiestap geen additioneel effect op de basisvariabelen.

Merk op dat het hier additionele effecten betreft. Doordat in zeer sterk stedelijke gebieden meer achterstandsleerlingen wonen, is de gemiddelde vaardigheid in stedelijke gebieden wel lager op Woordenschat, Begrijpend lezen en de rekenvaardigheden. Echter, als er gecorrigeerd wordt voor het formatiegewicht valt op dat dit juist omgekeerd is. Als we leerlingen in zeer stedelijke gebieden en niet-stedelijke gebieden vergelijken dan is er zelfs een klein positief effect van verstedelijking bij Begrijpend lezen en Spelling. Dat is te zien door bij deze vaardigheden de som van alle effecten te nemen: die levert een waarde van Niet stedelijk – Zeer sterk stedelijk van rond de -0,20 op. Dat betekent dat leerlingen met vergelijkbare andere achtergrondkenmerken het in zeer sterk stedelijke gebieden iets beter doen op deze twee vaardigheden.

Tabel 8.6.7 *Effectgrootten bij het achtergrondkenmerk urbanisatiegraad*

Urbanisatiegraad	Taalvaardigheden			Rekenvaardigheden		
	BegrLez	Spell	WoSch	GB	BPV	MMTG
Sterk – zeer sterk	-0,06	-0,09	-0,04	-0,07	-0,07	-0,02
Matig – sterk	-0,01	-0,01	-0,01	0,01	0,02	0,00
Weinig – matig	-0,07	-0,07	-0,06	-0,04	-0,04	-0,04
Niet – weinig	-0,06	-0,04	-0,04	-0,01	-0,03	-0,01

Extra variabelen advies Voortgezet Onderwijs en gemaakte toets

De variabelen advies VO en de gemaakte toets zijn twee variabelen die alleen relevant zijn bij de analyses voor jaargroep 8.

Bij het advies VO wordt onderscheid gemaakt tussen drie niveaus binnen het vmbo, havo en vwo. De drie onderscheiden niveaus in het vmbo zijn de basisberoepsgerichte leerweg (BB), de kaderberoepsgerichte leerweg (KB) en de gemene leerweg en de theoretische leerweg die samengenomen worden (GT). Dit laatste gebeurt omdat deze twee wat betreft het niveau zeer weinig van elkaar verschillen.

Tabel 8.6.8 *Effectgrootten bij het achtergrondkenmerk advies VO*

Advies VO	Taalvaardigheden			Rekenvaardigheden		
	BegrLez	Spell	WoSch	GB	BPV	MMTG
Vmbo-KB – vmbo-BB	0,77	0,47	0,64	0,86	0,83	0,83
Vmbo-GT – vmbo-KB	0,60	0,39	0,46	0,61	0,63	0,58
Havo – vmbo-GT	0,95	0,77	0,77	0,93	1,02	1,03
Vwo – havo	1,33	1,45	1,24	1,27	1,37	1,51

Bij de variabele gemaakte toets wordt onderscheid gemaakt tussen de leerlingen die de reguliere Eindtoets basisonderwijs gemaakt hebben en de leerlingen die de Niveautoets gemaakt hebben. De leerlingen die de Niveautoets maken doen dat omdat van hen verwacht wordt dat zij niet vaardig genoeg zijn om de reguliere Eindtoets basisonderwijs te maken.

Tabel 8.6.9 Effectgrootten bij het achtergrondkenmerk afgenomen toets

Taalvaardigheden	Effectgrootte	Effectkwalificatie
Begrijpend lezen	-1,60	groot
Spelling	-1,04	groot
Woordenschat	-1,42	groot
Rekenvaardigheden		
Getallen en Bewerkingen	-1,75	groot
Breuken, Procenten en Verhoudingen	-1,76	groot
Meten, Meetkunde, Tijd en Geld	-1,86	groot

De resultaten van 2009 verschillen niet substantieel van die van 2008. De belangrijkste punten bij deze additionele variabelen zijn de volgende:

- Een variabele die duidelijk een additioneel effect heeft, is die van het advies VO. Het (niet verrassende) resultaat is dat hoe hoger het advies is, hoe hoger de vaardigheid.
- Opvallend is dat het verschil in vaardigheid tussen havo- en vwo-leerlingen zo groot is.
- Als verder gekeken wordt dan de twee elkaar opvolgende groepen, dan valt op dat de grootste verschillen gevonden worden bij de rekenvaardigheden Breuken, procenten, en verhoudingen en Meten, meetkunde, tijd, en geld. Het verschil tussen twee leerlingen die alleen verschillen op het kenmerk advies VO en op de overige kenmerken exact gelijk zijn vinden we dat een leerling met een advies vmbo-basisberoeps en een leerling met een vwo-advies bijna 4 standaarddeviaties van elkaar verschillen.
- Bij de vaardigheden met het kleinste verschil tussen deze twee groepen (Spelling en Woordenschat) is het verschil nog steeds ruim 3 standaarddeviaties.
- De leerlingen die de Niveautoets maken, presteren (uiteraard) slechter dan zij die de reguliere Eindtoets maken. De additionele effecten zijn het grootst bij de rekenvaardigheden.

8.4 De resultaten voor jaargroep 4

De resultaten van de analyses voor jaargroep 4 zijn beschreven in hoofdstuk 7. Hieronder volgt de samenvatting van de resultaten. In tegenstelling tot hoofdstuk 7 zijn de resultaten nu niet geordend per vaardigheid maar per achtergrondvariabele.

Een verschil met de analyses van jaargroep 8 is dat bij jaargroep 4 er met een kleinere steekproef gewerkt is. Bij jaargroep 4 waren dat ongeveer 2500 leerlingen per vaardigheid per jaar, terwijl dit aantal voor jaargroep 8 rond de 140.000 lag. Dit heeft tot gevolg dat voor het berekenen van jaareffecten in jaargroep 4 niet met ongecorrigeerde effecten gerekend kan worden in een hoofdeffectenmodel. Het jaareffect is altijd een effect dat gecorrigeerd is voor de basisvariabelen stratum, leertijd, geslacht en formatiegewicht.

Een ander verschil is dat naast het hoofdeffectenmodel een interactiemodel met een interactievariabele voor "stratum per jaar" nodig was. Dit was nodig bij alle vaardigheden omdat het jaareffect niet voor ieder stratum gelijk is. Dat betekent ook dat de effecten van de strata in 2008 en 2009 van elkaar verschilden.

Een derde verschil was dat het bij Woordenschat niet mogelijk was een jaarvergelijking te maken, omdat deze vaardigheid in 2008 in andere omstandigheden is afgenomen. Het laatste verschil is dat de extra variabelen advies VO en type toets niet in de modellen met extra variabelen zijn opgenomen. Ten slotte kan nog opgemerkt worden dat het significantieniveau in de tabellen van jaargroep 4 $\alpha=,05$ is in plaats van $\alpha=,01$ aangezien de steekproeven kleiner zijn.

8.4.1 De vergelijking van 2009 met 2008 in jaargroep 4

De jaarvergelijking is zowel met een hoofdeffectenmodel uit te rekenen als met een interactiemodel.

De belangrijkste resultaten bij het hoofdeffectenmodel zijn:

- Het valt op dat alleen bij Spelling een significant resultaat gevonden is. De leerlingen in 2009 presteren beter dan in 2008. Het verschil is echter zodanig dat het geassocieerd wordt als “geen effect”.
- De sterkste achteruitgang is gevonden bij de rekenvaardigheid Vermenigvuldigen. Dit is echter geen significante achteruitgang en wordt ook geassocieerd als “geen effect”.

Tabel 8.7.1 Jaareffect (Jaar: 2009 – 2008) in hoofdeffectenmodel

Taalvaardigheden	Effectgrootte	Effectkwalificatie
Spelling	0,07	geen effect
Begrijpend lezen	-0,05	niet significant
Rekenvaardigheden		
Getallen en Getalsrelaties	-0,01	niet significant
Optellen en aftrekken	0,02	niet significant
Vermenigvuldigen en delen	-0,06	niet significant
Metten, tijd en geld	0,02	niet significant

Bij de analyses valt op dat de verschillen tussen de jaren niet gelijk waren over de strata. Per stratum werden bij de verschillende vaardigheden afwijkende resultaten gevonden. De belangrijkste zijn:

- Bij stratum 1 (scholen zonder veel achterstandsleerlingen) wordt bij alle vaardigheden een achteruitgang geconstateerd. Deze worden echter geassocieerd als “geen effect”, niet significant of beiden. De enige uitzondering is de rekenvaardigheid Vermenigvuldigen en delen. Daar wordt een klein significant negatief effect gevonden.
- Bij stratum 2 wordt zowel achteruitgang als vooruitgang gevonden. Deze zijn in de meeste gevallen echter niet significant. Bij de twee gevallen waar significante vooruitgang geconstateerd wordt (Spelling en Optellen en aftrekken), is deze zo klein dat deze geassocieerd is als “geen effect”.
- In stratum 3 wordt bij alle vaardigheden een significante vooruitgang gevonden en kan het effect als een klein effect geassocieerd worden. De enige uitzondering is de vaardigheid Begrijpend lezen.

Tabel 8.7.2 Jaareffect (Jaar: 2009 – 2008) per stratum (S) (interactiemodel)

	Taalvaardigheden			Rekenvaardigheden			
	BegrLez	Spell	WoSch	G/G	O/A	V/D	MTG
S1	-0,13	-0,02	.	-0,11*	-0,13**	-0,22	-0,04
S2	0,04	0,13*	.	0,00	0,16*	0,10	0,04
S3	0,12	0,36	.	0,35	0,37	0,29**	0,22*

Een conclusie op basis van deze resultaten is dat de prestaties van de achterstandsscholen (S3) en de niet-achterstandsscholen dicht bij elkaar zijn komen te liggen. Dat wordt ook duidelijk als naar de effecten bij de achtergrondvariabele stratum gekeken wordt.

8.4.2 De vergelijking van de overige achtergrondvariabelen

Basisvariabele stratum

Aangezien de effecten bij de variabele stratum varieerden voor 2008 en 2009, was het nodig het interactiemodel te gebruiken. Zodoende worden de resultaten voor de stratumeffecten voor de twee jaren afzonderlijk gegeven. Merk op dat er voor Woordenschat geen geschikte gegevens zijn voor 2008 om de vergelijking te maken.

Tabel 8.7.3 Effecten tussen strata per jaar in interactiemodel

Stratum (per jaar)	Taalvaardigheden			Rekenvaardigheden			
	BegrLez	Spell	WoSch	G/G	O/A	V/D	MTG
S2 – S1 (2008)	-0,31	-0,09	.	-0,35	-0,27	-0,38	-0,25
S2 – S1 (2009)	-0,14	0,06	-0,17	-0,24	0,02	-0,06	-0,18
S3 – S2 (2008)	-0,31	-0,15	.	-0,39	-0,27	-0,37	-0,35
S3 – S2 (2009)	-0,23	0,08	-0,36	-0,05	-0,05	-0,19	-0,17
S3 – S1 (2008)	-0,62	-0,24	.	-0,75	-0,54	-0,75	-0,60
S3 – S1 (2009)	-0,37	0,14	-0,53	-0,29	-0,03	-0,25	-0,35

De belangrijkste resultaten zijn

- De verschillen tussen de strata zijn afgenomen in 2009 in vergelijking met 2008. Waar in 2008 nog kleine effecten gevonden werden, worden de effecten in 2009 over het algemeen geclassificeerd als “geen effect”.
- Het grootste gevonden effect in 2009 was bij Woordenschat tussen stratum 1 en stratum 3, waarbij nog een matig effect gevonden was (leerlingen op S1-scholen presteerden beter).
- Bij Begrijpend lezen en de rekenschalen, op die voor Optellen en aftrekken na, wordt in 2009 een klein effect gevonden tussen stratum 1 en stratum 3.

Basisvariabele formatiegewicht

Stratum is een schoolkenmerk. De overige drie basisvariabelen zijn persoonskenmerken. Een basisvariabele die samenhangt met stratum is formatiegewicht. De gevonden effecten bij de formatiegewichten zijn groter dan die bij de strata.

Tabel 8.7.4 Effecten voor de verschillende formatiegewichten.

Formatiegewicht	Taalvaardigheden			Rekenvaardigheden			
	BegrLez	Spell	WoSch	G/G	O/A	V/D	MTG
F1 – F2	-0,40	-0,24	-0,29	-0,34	-0,30	-0,31	-0,35
F3 – F2	-0,23	0,08	-0,52	-0,11	0,02	-0,22	-0,16
F3 – F1	-0,64	-0,16	-0,80	-0,45	-0,27	-0,54	-0,52

De belangrijkste resultaten zijn

- Het verschil tussen een reguliere leerling (F1; nieuw formatiegewicht 0.0) en een F2-leerling (formatiegewicht 0.3; leerling met milde achterstand) levert een klein effect op bij alle vaardigheden.
- Het verschil tussen een F2-leerling en een F3-leerling (formatiegewicht 1.2; leerling met grote achterstand) geeft een gevarieerder beeld: bij Woordenschat is het effect matig. Bij Begrijpend lezen en bij de rekenvaardigheidschaal vermenigvuldigen en delen is het effect klein. Bij de andere vaardigheden is er geen serieus verschil.
- Het verschil tussen een F1-leerling en een F3-leerling levert bij Woordenschat een groot effect, bij de helft van de rekenvaardigheden en bij Begrijpend lezen een matig effect en bij de andere helft een klein effect. Bij Spelling is het effect zo klein dat het geclassificeerd wordt als "geen effect".

Basisvariabelen geslacht en leertijd

De belangrijkste resultaten bij geslacht zijn

- Meisjes presteren beter bij de taalvaardigheden. Bij Woordenschat is dit verschil echter niet significant. Bij de andere twee taalvaardigheden is het effect (bijna) klein.
- De jongens presteren beter bij de rekenvaardigheden. Het effect wordt bij alle rekenvaardigheden geclassificeerd als klein. Het effect is (in absolute zin) wel groter dan bij de taalvaardigheden.

Het belangrijkste resultaat bij leertijd is:

- Reguliere leerlingen presteren op alle vaardigheden beter dan de vertraagde leerlingen. Het effect is bij alle vaardigheden klein (bij de meeste vaardigheden rond $-,35$) behalve bij Spelling. Daar is het effect matig.

Tabel 8.7.5 Effecten voor geslacht

Geslacht: Meisjes – Jongens

Taalvaardigheden	Effectgrootte	Effectkwalificatie
Woordenschat	0,04	niet significant
Spelling	0,27	klein
Begrijpend lezen	0,19	geen effect
Rekenvaardigheden		
Getallen en Getalsrelaties	-0,42	klein effect
Optellen en aftrekken	-0,32	klein effect
Vermenigvuldigen en delen	-0,36	klein effect
Metten, tijd en geld	-0,38	klein effect

Tabel 8.7.6 Effecten voor leertijd

Leertijd: Vertraagd – Regulier

Taalvaardigheden	Effectgrootte	Effectkwalificatie
Woordenschat	-0,27	klein
Spelling	-0,53	matig
Begrijpend lezen	-0,37	klein effect
Rekenvaardigheden		
Getallen en Getalsrelaties	-0,36	klein effect
Optellen en aftrekken	-0,36	klein effect
Vermenigvuldigen en delen	-0,34	klein effect
Metten, tijd en geld	-0,33	klein effect

Extra variabelen regio en verstedelijking

Naast het model waarin de basisvariabelen zijn opgenomen zijn ook modellen met extra variabelen gebruikt. Daarbij zijn de extra variabelen toegevoegd aan het basismodel. De extra variabelen zijn daarmee additionele effecten, waarbij de effecten gecorrigeerd zijn voor de effecten van stratum, formatiegewicht, geslacht en leertijd.

De belangrijkste resultaten voor de variabele regio zijn:

- Bij de rekenvaardigheden en bij Woordenschat worden geen serieuze regio-effecten gevonden.
- Regio Noord presteert slechter bij Begrijpend lezen en Spelling (het verschil tussen de regio's West en Noord is echter verwaarloosbaar bij Spelling).
- Tussen de overige regio's zijn geen substantiële verschillen, op die tussen West en Oost na bij Spelling; Oost presteert beter dan West.

Het belangrijkste resultaat bij de variabele verstedelijking is dat het, gecorrigeerd voor de basisvariabelen, geen serieus effect heeft.

Tabel 8.7.7 Effecten voor regio en verstedelijking

Regio	Taalvaardigheden			Rekenvaardigheden			
	BegrLez	Spell	WoSch	G/G	O/A	V/D	MTG
Oost – Noord	0,30	0,29	-0,05	0,09	0,14	0,09	0,08
West – Noord	0,19	0,10	-0,02	0,00	0,11	-0,01	-0,04
Zuid – Noord	0,21	0,27	0,04	0,05	0,13	-0,02	0,04
West – Oost	-0,12	-0,20	0,03	-0,09	-0,03	-0,10	-0,12
Zuid – Oost	-0,10	-0,02	0,08	-0,03	-0,01	-0,11	-0,04
Zuid – West	0,02	0,17	0,05	0,05	0,02	0,00	0,08

Verstedelijking	Taalvaardigheden			Rekenvaardigheden			
	BegrLez	Spell	WoSch	G/G	O/A	V/D	MTG
Sterk – zeer sterk	0,07	0,02	-0,02	0,04	0,02	-0,04	-0,06
Matig – sterk	0,05	-0,07	0,02	0,01	-0,01	0,07	0,07
Weinig – matig	-0,02	0,16	0,06	0,00	0,08	-0,03	-0,11
Niet – weinig	0,09	0,05	0,15	-0,05	-0,08	-0,04	0,04

Opgemerkt moet worden dat het ontbreken van effecten gecorrigeerd is voor de basisvariabelen. Bij de schattingen van de verdelingen wordt bijvoorbeeld nog wel gevonden dat in zeer verstedelijkte gebieden de gemiddelde vaardigheid lager ligt. Dat wordt dan bijvoorbeeld veroorzaakt door een grotere concentratie achterstandsleerlingen in zeer stedelijke gebieden. Het feit dat de leerlingen op zich in een verstedelijkt gebied wonen heeft echter geen extra additioneel negatief effect.

Extra variabele thuistaal

Opgemerkt moet worden dat het ook bij de variabele thuistaal om additionele effecten gaat. De effecten zijn gecorrigeerd voor de basisvariabelen. De belangrijkste resultaten bij thuistaal zijn

- Bij Spelling heeft de thuistaal geen effect.
- Leerlingen die thuis alleen Nederlands spreken, presteren bij de meeste overige vaardigheden beter dan de leerlingen die thuis een andere taal spreken, eventueel ook naast het Nederlands. Alleen bij de rekenvaardigheid Optellen en aftrekken en Spelling werd dit effect niet gevonden. Al deze effecten waren klein, behalve bij Woordenschat: daar werd een matig effect gevonden bij het verschil tussen alleen een andere taal en alleen Nederlands.
- Of er thuis alleen een andere taal gesproken wordt of naast een andere taal ook Nederlands, heeft bij rekenen een geen effect. Bij Woordenschat en Begrijpend lezen presteren leerlingen beter als ze ook Nederlands spreken. Dit effect is klein bij Woordenschat en bijna klein bij Begrijpend lezen.

Tabel 8.7.8 Effecten voor thuistaal

Thuistaal	Taalvaardigheden			Rekenvaardigheden			
	BegrLez	Spell	WoSch	G/G	O/A	V/D	MTG
NLs+ander – Alleen NLs	-0,29	-0,08	-0,34	-0,33	-0,27	-0,23	-0,26
Alleen ander – Alleen NLs	-0,48	-0,01	-0,62	-0,35	-0,13	-0,23	-0,33
NLs+ander – Alleen ander	0,19	-0,07	0,28	0,02	-0,15	0,00	-0,07

Literatuur

- Berkel, S. van, & N. Alberts (2009). *Primair onderwijs Leerling- en onderwijsvolgsysteem: Woordenschat Groep 4*. Arnhem, Cito.
- Berkel, S. van, F. van der Schoot, R. Engelen & G. Maris (2002) *Balans van het taalonderwijs halverwege de basisschool 3. Uitkomsten van de derde peiling in 1999*. PPON-reeks 20. Arnhem, Citogroep.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (second ed.). Lawrence Erlbaum Associates.
- Expertgroep Doorlopende Leerlijnen (2008). *Over de drempels met taal en rekenen*. Eindrapport Expertgroep doorlopende leerlijnen Taal en Rekenen.
- Evers, A., Van Vliet-Mulder, J. C., & Groot, C. J. (2000). *Documentatie van tests en testresearch in Nederland. Deel II: Testresearch*. Van Gorcum, Assen.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Press
- Heesters, K., S. van Berkel, F. van der Schoot & B. Hemker (2007) *Balans van het leesonderwijs aan het einde van de basisschool 5. Uitkomsten van de vierde peiling in 2005*. PPON-reeks nr. 33. Arnhem, Cito.
- 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 (<http://www.minocw.nl/documenten/133682d.pdf>)
- Janssen, J., F. van der Schoot, B. Hemker (2005) *Balans van het reken- wiskundeonderwijs aan het einde van de basisschool 4. Uitkomsten van de vierde peiling in 2004*. PPON-reeks nr. 32. Arnhem, Citogroep.
- Kraemer, J.-M., J. Janssen, F. van der Schoot & B. Hemker (2005). *Balans van het reken- wiskundeonderwijs halverwege de basisschool 4. Uitkomsten van de vierde peiling in 2003*. PPON-reeks nr. 31. Arnhem, Citogroep.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Moelands, F., I. Jongen, F. van der Schoot & B. Hemker (2007) *Balans over leesstrategieën in het primair onderwijs. Uitkomsten van de eerste peiling in 2005*. PPON-reeks nr. 35. Arnhem, Cito
- Van der Linden, W.J. & Hambleton, R.K. (Eds.) (1997). *Handbook of modern item response theory*. New York: Springer.
- Van der Lubbe, M. (2008). *Terugblik en resultaten Eindtoets Basisonderwijs 2008*. Arnhem, The Netherlands: Cito.
- Van der Schoot, F. (2008). *Onderwijs op peil? Een samenvattend overzicht van 20 jaar PPON*. Arnhem, The Netherlands: CITO
- Verhelst, N.D., & Eggen, T.J.H.M. (1989). *Psychometrische en Statistische Aspecten van Peilingsonderzoek*. Arnhem, The Netherlands: CITO.

Verhelst, N.D., Glas, C.W., & Verstralen, H.H.F.M. (1995). *OPLM: Computer program and manual*. Arnhem, The Netherlands: CITO.

Verhelst, N. D. (1998). *Estimating the Reliability of a Test from A Single Test Administration*. Measurement & Research Department Reports. Cito, Arnhem.

Sijtstra J., F. van der Schoot & B. Hemker, B (2002). *Balans van het taalonderwijs aan het einde van de basisschool 3. Uitkomsten van de derde peiling in 1998*. PPON-reeks 19. Arnhem: Citogroep.

Bijlage 1: Verdeling van de leerlingen in de onderzoekspopulatie in 2009

Leerlingen in 2009	Aantal		
in de onderzoekspopulatie	141931		
	Leerlingen		
Geslacht	Aantal	Percentage	% bekend
Jongen	70358	49,6	49,8
Meisje	71020	50	50,2
Onbekend	553	0,4	

	Leerlingen		
Formatie Gewicht	Aantal	Percentage	% bekend
1.00	97400	68,6	78,9
1.25-1.70	11692	8,2	9,4
1.90	14364	10,1	11,6
Onbekend	18475	13	

	Leerlingen		
Leeftijd* / Leertijd	Aantal	Percentage	% bekend
10	5458	3,8	3,8
11	108748	76,6	77,1
12	26217	18,5	18,6
13	684	0,5	0,5
Regulier (10 en 11 jaar)	114206	80,4	80,9
Vertraagd (12 en 13 jaar)	26901	19,0	19,1
onbekend*	824	0,6	

* leeftijd op 1 oktober 2008

	Leerlingen		
TT 6 deling	Aantal	Percentage	% bekend
Alleen NLs	123322	86,9	89,3
NLs + Ander	5282	3,7	3,8
Alleen Ander	9465	6,7	6,9
Turks(+NLs)	4208	3	3,1
Arab(+NLs)	3695	2,6	2,7
Suri(+NLs)	1075	0,8	0,8
WEur(+NLs)	1773	1,2	1,2
Ander(+NLs)	3800	2,7	2,8
Meer dan 1 andere (+NLs)	196	0,1	0,1
Onbekend	3862	2,7	

Advies leerkracht	Leerlingen		
	Aantal	Percentage	% bekend
BB	6430	4,5	6,3
BB/KB	5128	3,6	5,0
KB	9561	6,7	9,3
KB/GT	1107	0,8	1,1
GT	19984	14,1	19,6
GT/Havo(Vwo)	16375	11,5	16,0
Havo	14387	10,1	14,1
Havo/VWO	14997	10,6	14,8
VWO	13997	9,9	13,8
Onbekend	39965	28,2	

Stratum	Leerlingen		
	Aantal	Percentage	% bekend
Stratum 1	87354	61,5	61,5
Stratum 2	35321	24,9	24,9
Stratum 3	19245	13,6	13,6
Onbekend	11	0	

CBS-regio 4 niveaus	Leerlingen		
	Aantal	Percentage	% bekend
Noord	13914	9,8	9,8
Oost	33371	23,5	23,5
West	60190	42,4	42,4
Zuid	34456	24,3	24,3

Urbanisatiegraad	Leerlingen		
	Aantal	Percentage	% bekend
Zeer sterk stedelijk	21242	15,0	15,0
Sterk stedelijk	34659	24,4	24,4
Matig stedelijk	29870	21,0	21,0
Weinig stedelijk	36954	26,0	26,0
Niet stedelijk	19206	13,5	13,5

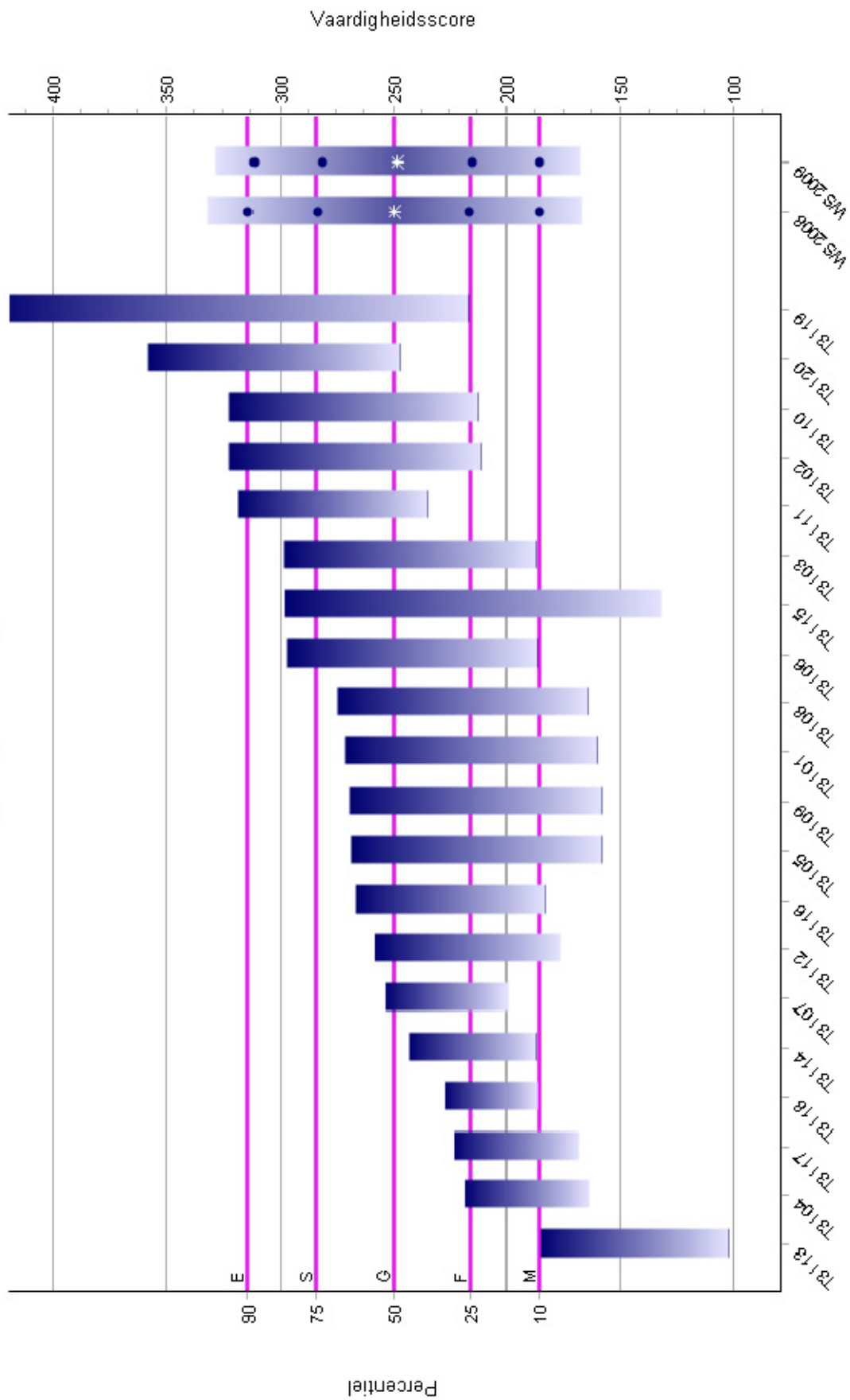
Bijlage 2: De opgaven in jaargroep 9 (inclusief tabellen en figuren)

Taal: Woordenschat

Opgaven per type en niveau	
0	Gemakkelijk ($P > .80$)
1	Uitdagend ($.50 \leq P \leq .80$)
2	Moeilijk ($P < .50$)

Label	itembanknr	JPON_p50	JPON_p80	ZLV	LV	ST	HV	ZHV
T3 - i13	563t313E	103	186	0	0	0	0	0
T3 - i04	554t3_4E	164	219	1	1	0	0	0
T3 - i17	567t317E	169	225	1	1	0	0	0
T3 - i18	568t318E	187	228	2	1	0	0	0
T3 - i14	564t314E	188	244	2	1	0	0	0
T3 - i07	557t3_7E	200	255	2	1	1	0	0
T3 - i12	562t312E	177	259	1	1	1	0	0
T3 - i16	566t316E	184	267	1	1	1	0	0
T3 - i05	555t3_5E	159	269	1	1	1	0	0
T3 - i09	559t3_9E	159	270	1	1	1	0	0
T3 - i01	551t3_1E	161	272	1	1	1	0	0
T3 - i08	558t3_8E	165	275	1	1	1	0	0
T3 - i06	556t3_6E	187	298	2	1	1	1	0
T3 - i15	565t315E	133	299	1	1	1	1	0
T3 - i03	553t3_3E	188	299	2	1	1	1	0
T3 - i11	561t311E	236	319	2	2	1	1	1
T3 - i02	552t3_2E	212	323	2	1	1	1	1
T3 - i10	560t310E	213	323	2	1	1	1	1
T3 - i20	570t320E	248	359	2	2	1	1	1
T3 - i19	569t319E	217	549	2	2	1	1	1

Taal: Woordenschat
opgaven gesorteerd op p80

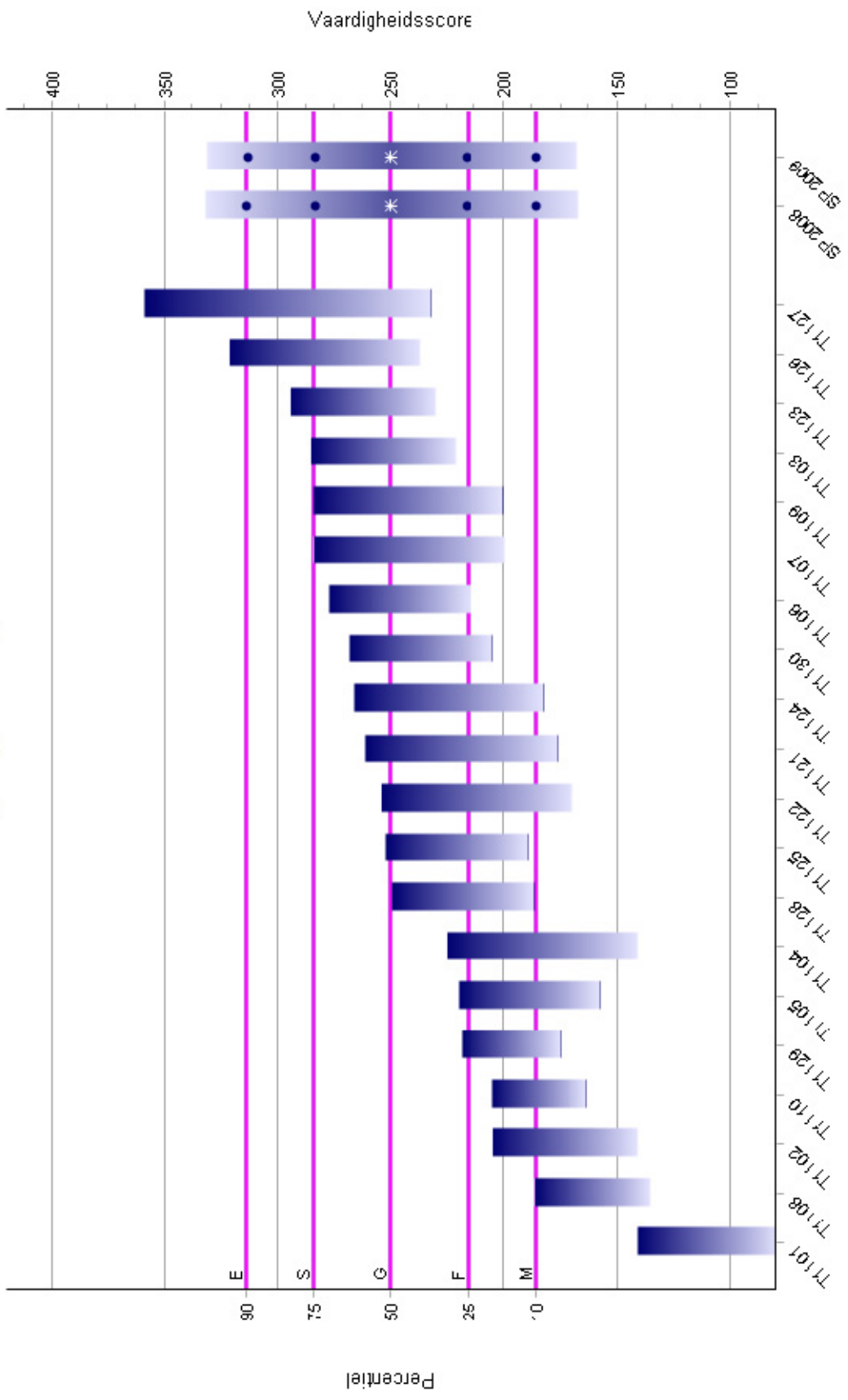


Taal: Spelling

Opgaven per type en niveau	
0	Gemakkelijk (P > .80)
1	Uitdagend (.50 =< P =< .80)
2	Moeilijk (P < .50)

Label	itembanknr	JPON_p50	JPON_p80	ZLV	LV	ST	HV	ZHV
T1 - i01	501t1_1C	79	142	0	0	0	0	0
T1 - i08	508t1_8C	136	187	1	0	0	0	0
T1 - i02	502t1_2C	142	206	1	0	0	0	0
T1 - i10	510t110C	164	206	1	0	0	0	0
T1 - i29	529t129B	176	219	1	1	0	0	0
T1 - i05	505t1_5C	158	221	1	1	0	0	0
T1 - i04	504t1_4C	142	226	1	1	0	0	0
T1 - i28	528t128B	187	250	2	1	1	0	0
T1 - i25	525t125B	190	253	2	1	1	0	0
T1 - i22	522t122B	170	255	1	1	1	0	0
T1 - i21	521t121B	177	262	1	1	1	0	0
T1 - i24	524t124B	183	267	1	1	1	0	0
T1 - i30	530t130B	206	269	2	1	1	0	0
T1 - i06	506t1_6C	215	278	2	1	1	0	0
T1 - i07	507t1_7C	200	285	2	1	1	1	0
T1 - i09	509t1_9C	201	285	2	1	1	1	0
T1 - i03	503t1_3C	222	286	2	2	1	1	0
T1 - i23	523t123B	231	295	2	2	1	1	0
T1 - i26	526t126B	238	322	2	2	1	1	1
T1 - i27	527t127B	233	360	2	2	1	1	1

Taal: Spelling
opgaven gesorteerd op p80



Taal: Begrijpend lezen

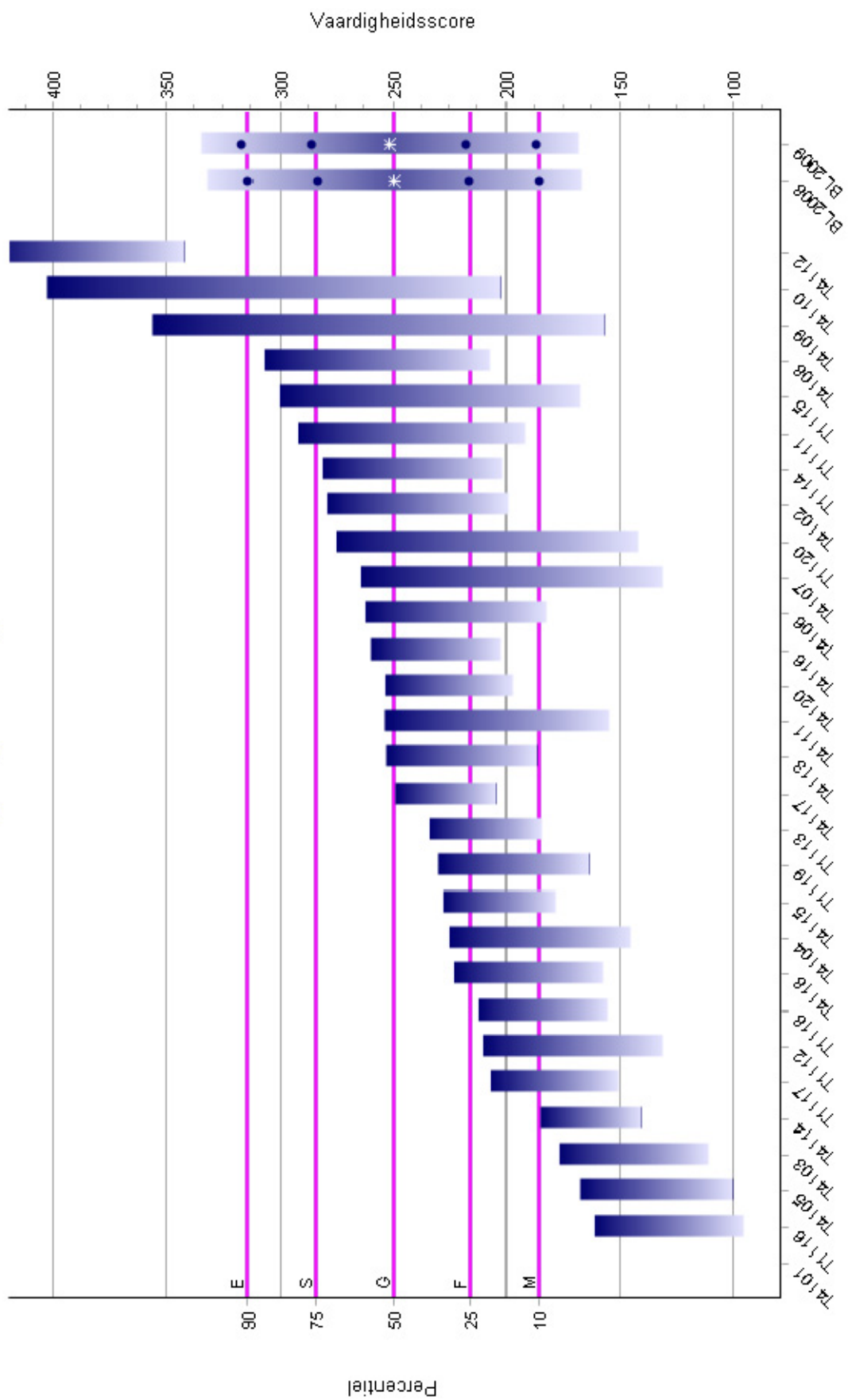
Opgaven per type en niveau	
0	Gemakkelijk (P > .80)
1	Uitdagend (.50 =< P =< .80)
2	Moeilijk (P < .50)

Label	itembanknr	JPON_p50	JPON_p80	ZLV	LV	ST	HV	ZHV
T4 - i01	581t4_1D	-131	2	0	0	0	0	0
T1 - i16	516t116D	96	162	0	0	0	0	0
T4 - i05	585t4_5D	101	168	0	0	0	0	0
T4 - i03	583t4_3D	112	178	0	0	0	0	0
T4 - i14	594t414D	142	186	1	0	0	0	0
T1 - i17	517t117D	151	208	1	0	0	0	0
T1 - i12	512t112D	132	211	1	0	0	0	0
T1 - i18	518t118D	156	213	1	0	0	0	0
T4 - i18	598t418D	158	224	1	1	0	0	0
T4 - i04	584t4_4D	146	226	1	1	0	0	0
T4 - i15	595t415D	179	229	1	1	0	0	0
T1 - i19	519t119D	164	231	1	1	0	0	0
T1 - i13	513t113D	185	235	1	1	0	0	0
T4 - i17	597t417D	205	250	2	1	0	0	0
T4 - i13	593t413D	187	254	2	1	1	0	0
T4 - i11	591t411D	155	255	1	1	1	0	0
T4 - i20	600t420D	198	255	2	1	1	0	0
T4 - i16	596t416D	203	260	2	1	1	0	0
T4 - i06	586t4_6D	183	263	1	1	1	0	0
T4 - i07	587t4_7D	132	265	1	1	1	0	0
T1 - i20	520t120D	143	276	1	1	1	0	0
T4 - i02	582t4_2D	200	280	2	1	1	0	0
T1 - i14	514t114D	202	282	2	1	1	0	0
T1 - i11	511t111D	193	293	2	1	1	1	0
T1 - i15	515t115D	168	301	1	1	1	1	0
T4 - i08	588t4_8D	208	308	2	1	1	1	0
T4 - i09	589t4_9D	157	357	1	1	1	1	1
T4 - i10	590t410D	203	403	2	1	1	1	1
T4 - i12	592t412D	343	476	2	2	2	2	2

T3 - i19	598t419D	pas niet op de schaal
----------	----------	-----------------------

Taal: Begrijpend lezen

opgaven gesorteerd op p80



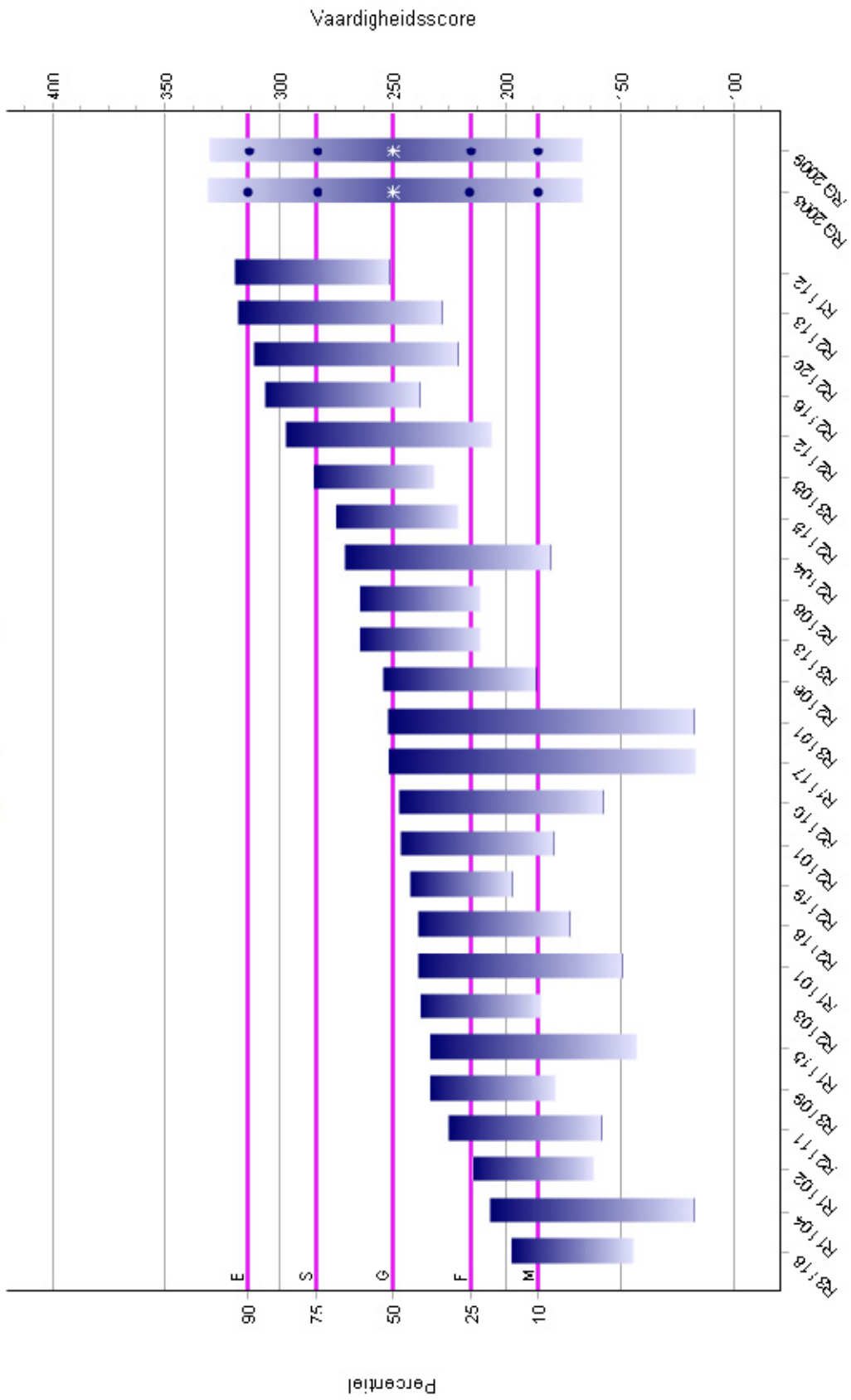
Rekenen: Getallen en Bewerkingen

Opgaven per type en niveau	
0	Gemakkelijk ($P > .80$)
1	Uitdagend ($.50 \leq P \leq .80$)
2	Moeilijk ($P < .50$)

Label	itembanknr	JPON_p50	JPON_p80	ZLV	LV	ST	HV	ZHV
R3 - i18	658R318F	145	199	1	0	0	0	0
R1 - i04	604R1_4F	118	208	1	0	0	0	0
R1 - i02	602R1_2F	162	216	1	0	0	0	0
R2 - i11	631R211F	159	226	1	1	0	0	0
R3 - i09	649R3_9F	179	234	1	1	0	0	0
R1 - i15	615R115F	144	234	1	1	0	0	0
R2 - i03	623R2_3F	185	239	1	1	0	0	0
R1 - i01	601R1_1F	150	240	1	1	0	0	0
R2 - i18	638R218F	173	240	1	1	0	0	0
R2 - i19	639R219F	198	243	2	1	0	0	0
R2 - i01	621R2_1F	180	247	1	1	0	0	0
R2 - i10	630R210F	158	248	1	1	0	0	0
R1 - i17	617R117F	117	253	1	1	1	0	0
R3 - i01	641R3_1F	118	253	1	1	1	0	0
R2 - i06	626R2_6F	187	255	2	1	1	0	0
R3 - i13	653R313F	212	266	2	1	1	0	0
R2 - i08	628R2_8F	212	266	2	1	1	0	0
R2 - i04	624R2_4F	181	272	1	1	1	0	0
R2 - i15	635R215F	222	276	2	2	1	0	0
R3 - i05	645R3_5F	232	286	2	2	1	1	0
R2 - i12	632R212F	207	298	2	1	1	1	0
R2 - i16	636R216F	239	307	2	2	1	1	0
R2 - i20	640R220F	222	312	2	2	1	1	0
R2 - i13	633R213F	229	319	2	2	1	1	1
R1 - i12	612R112F	252	320	2	2	2	1	1

Rekenen: Getallen en bewerkingen

opgaven gesorteerd op p80



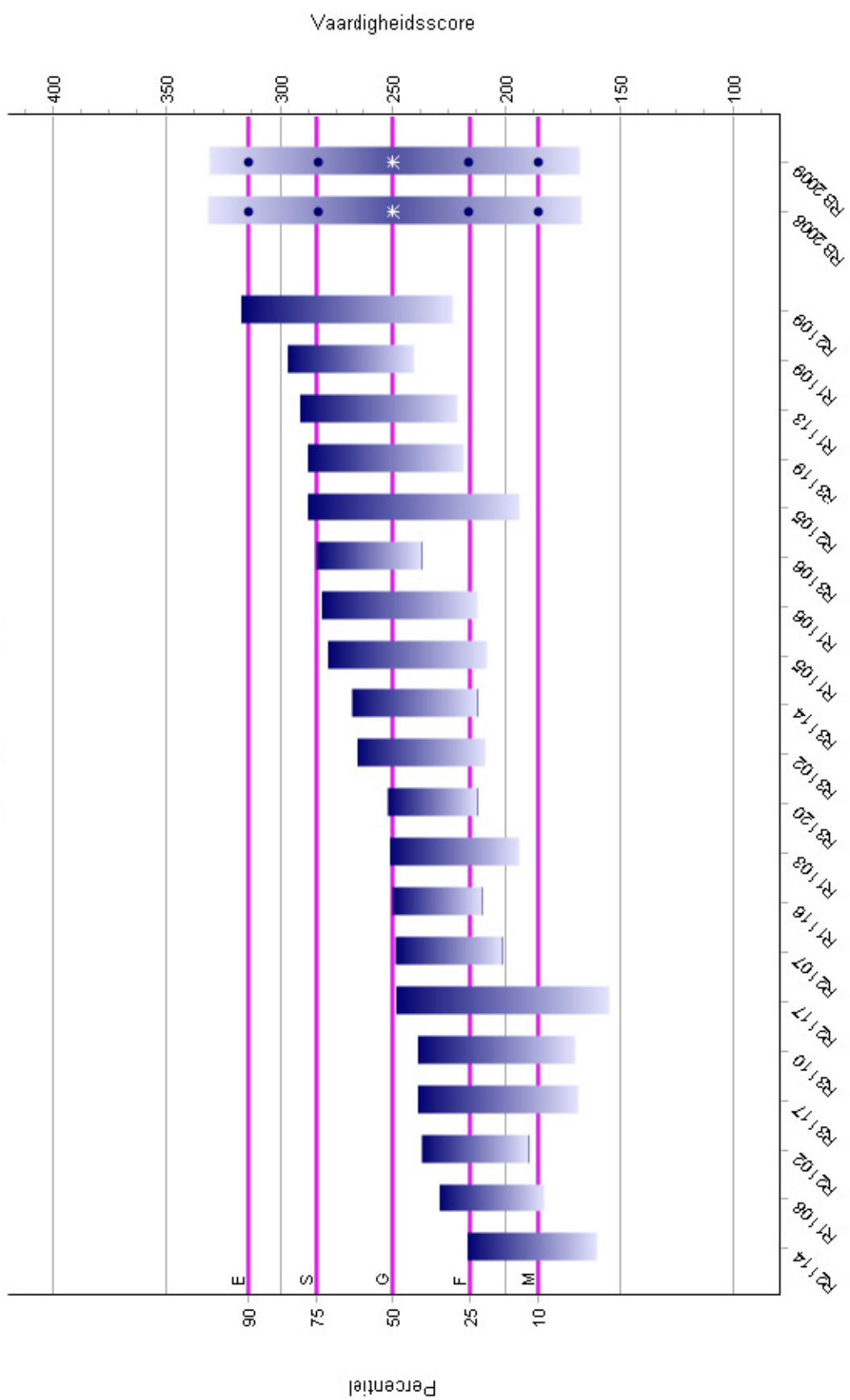
Rekenen: Breuken, Procenten en Verhoudingen

Opgaven per type en niveau	
0	Gemakkelijk ($P > .80$)
1	Uitdagend ($.50 \leq P \leq .80$)
2	Moeilijk ($P < .50$)

Label	itembanknr	JPON_p50	JPON_p80	ZLV	LV	ST	HV	ZHV
R2 - i14	634R214G	161	218	1	1	0	0	0
R1 - i08	608R1_8G	184	230	1	1	0	0	0
R2 - i02	622R2_2G	191	238	2	1	0	0	0
R3 - i17	657R317G	169	240	1	1	0	0	0
R3 - i10	650R310G	170	240	1	1	0	0	0
R2 - i17	637R217G	155	249	1	1	0	0	0
R2 - i07	627R2_7G	202	249	2	1	0	0	0
R1 - i16	616R116G	211	251	2	1	1	0	0
R1 - i03	603R1_3G	195	252	2	1	1	0	0
R3 - i20	660R320G	213	253	2	1	1	0	0
R3 - i02	642R3_2G	210	267	2	1	1	0	0
R3 - i14	654R314G	213	269	2	1	1	0	0
R1 - i05	605R1_5G	209	280	2	1	1	0	0
R1 - i06	606R1_6G	213	283	2	1	1	0	0
R3 - i06	646R3_6G	238	285	2	2	1	1	0
R2 - i05	625R2_5G	195	289	2	1	1	1	0
R3 - i19	659R319G	219	289	2	2	1	1	0
R1 - i13	613R113G	222	292	2	2	1	1	0
R1 - i09	609R1_9G	241	298	2	2	1	1	0
R2 - i09	629R2_9G	224	318	2	2	1	1	1

Rekenen: Breuken, Procenten en Verhoudingen

opgaven gesorteerd op p80



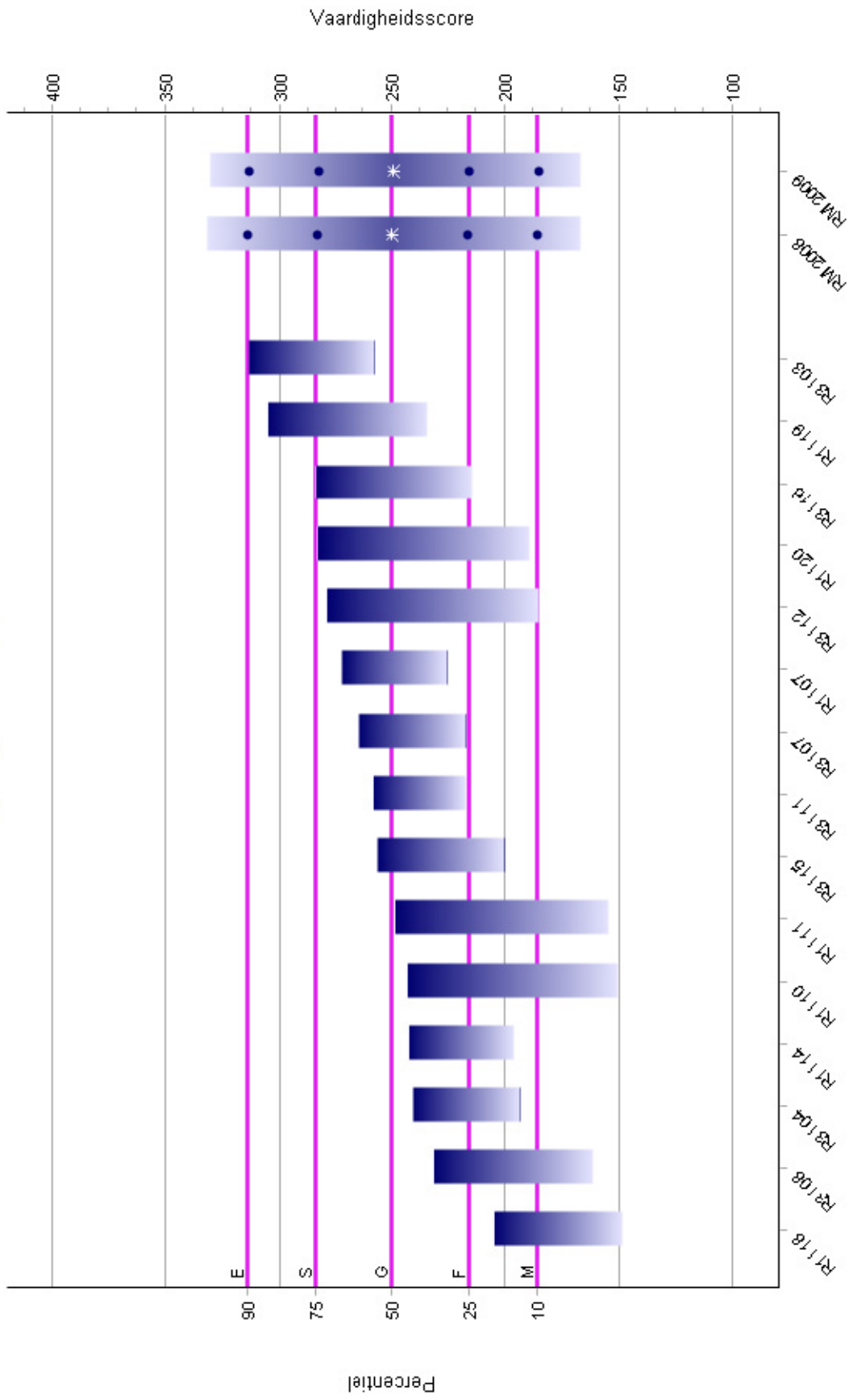
Rekenen: Meten, Meetkunde, Tijd en Geld

Opgaven per type en niveau	
0	Gemakkelijk ($P > .80$)
1	Uitdagend ($.50 \leq P \leq .80$)
2	Moeilijk ($P < .50$)

Label	itembanknr	JPON_p50	JPON_p80	ZLV	LV	ST	HV	ZHV
R1 - i18	618R118H	149	206	1	0	0	0	0
R3 - i08	648R3_8H	162	232	1	1	0	0	0
R3 - i04	644R3_4H	194	241	2	1	0	0	0
R1 - i14	614R114H	197	243	2	1	0	0	0
R1 - i10	610R110H	151	244	1	1	0	0	0
R1 - i11	611R111H	155	249	1	1	0	0	0
R3 - i15	655R315H	201	257	2	1	1	0	0
R3 - i11	651R311H	218	259	2	2	1	0	0
R3 - i07	647R3_7H	218	265	2	2	1	0	0
R1 - i07	607R1_7H	226	273	2	2	1	0	0
R3 - i12	652R312H	186	280	1	1	1	0	0
R1 - i20	620R120H	190	284	2	1	1	0	0
R3 - i16	656R316H	215	285	2	1	1	1	0
R1 - i19	619R119H	235	306	2	2	1	1	0
R3 - i03	643R3_3H	258	314	2	2	2	1	0

Rekenen: Meten, Meetkunde, Tijd en Geld

opgaven gesorteerd op p80



Bijlage 3: Een alternatieve presentatie van de vaardigheidsschalen

Het gebruik van vaardigheidsschalen is in de Item Response Theorie heel gebruikelijk. Door het gebruik van deze analysetechniek heeft een "latente" vaardigheidsschaal ook zijn weg gevonden in het presenteren van resultaten zoals in bijgaand verslag en in de PPON-rapporten. Het kan echter lastig zijn de betekenis van een vaardigheidsschaal direct te doorzien. Wat betekent welke waarde nou eigenlijk? Een oplossing hiervoor zou kunnen liggen is het transformeren van de hier gebruikte schaal waarbij het gemiddelde ligt op 250 en de standaarddeviatie gesteld is op 50 naar een van de weinige vaardigheidsschalen wel enigszins gemeengoed is, namelijk dezelfde soort schaal als de IQ-schaal.

In de IQ-schaal wordt, net als in de hier gebruikte schaal, verondersteld dat de vaardigheid normaal verdeeld is. Alleen zijn gemiddelde en standaarddeviatie niet 250 en 50, maar 100 en 15. Het voordeel is dat de meeste mensen wel weten dat een IQ van 100 staat voor een gemiddelde IQ. Ook de andere waarden op de schaal hebben in de loop der tijd een betekenis gekregen.

De indeling van IQ-waarden volgens Resing & Blok[1]

> 130 Hoogbegaafd
121 - 130 Begaafd
111 - 120 Boven gemiddeld
90 - 110 Gemiddeld
80 - 89 Beneden gemiddeld
70 - 79 Moeilijk lerend

Indien er is voldaan aan de criteria voor een verstandelijke beperking:

50 - 69 Lichte verstandelijke beperking
35 - 49 Matige verstandelijke beperking
20 - 34 Ernstige verstandelijke beperking
< 20 Diepe verstandelijke beperking

De indeling van IQ-waarden volgens Wechsler is als volgt:

145-160 Zeer hoogbegaafd
130-144 Hoogbegaafd
116-130 Begaafd
101-115 Boven gemiddeld
100 Gemiddeld begaafd
85-99 Beneden gemiddeld
70-84 Moeilijk lerend
55-69 Lichte verstandelijke beperking
40-54 Matige verstandelijke beperking
25-39 Ernstige verstandelijke beperking
< 25 Zeer ernstige verstandelijke beperking

Doordat deze betekenissen relatief bekend zijn wordt het makkelijker de resultaten met betrekking tot de vaardigheidsschalen te lezen.

Het omzetten van de schalen van de PPON-schaal naar de IQ-schaal is heel gemakkelijk: de gemiddelden worden vermenigvuldigd met 0,3 en er wordt 25 bij opgeteld. Voor de transformatie van de standaarddeviaties hoeft er alleen vermenigvuldigd te worden met 0,3.

Een voordeel van deze schaal is dat deze de vaardigheidsschaal waarschijnlijk makkelijker begripbaar maakt, door de analogie met IQ, maar er zijn ook wat nadelen. De belangrijkste is dat door het gebruik juist verwarring kan ontstaan waardoor men denkt dat de gevonden vaardigheid ook het IQ is of daar erg in de buurt moet zitten. Dat is niet direct het geval. Voor de verschillende vaardigheidsschalen kunnen juist heel verschillende resultaten gevonden worden voor dezelfde groepen. Bij een IQ is dat niet het geval.

Een ander kenmerk van een IQ-score is dat verwacht wordt dat die constant blijft bij een leerling. Iemand met een IQ van 100 wordt verondersteld zijn of haar hele leven een IQ van 100 te hebben (of in ieder geval een IQ dat daar dicht in de buurt ligt). Als dat niet het geval is dan wordt er getwijfeld aan de kwaliteit van de test. Bij de vaardigheden die we hier meten is dat echter niet het geval. Je moet juist verwachten dat beter onderwijs tot een verhoging van de vaardigheid op de schaal leidt. Dat is niet alleen een absolute stijging in vaardigheid (wat ook in een IQ-toets kan). Ook een relatieve stijging moet mogelijk zijn als gevolg van gericht onderwijs voor een specifieke leerling. Dat is iets wat bij een IQ-score niet verwacht wordt. Het feit dat een relatieve vooruitgang binnen IQ-scores niet verwacht mag worden, is vooral een probleem als we vaardigheden in het onderwijs op die schaal willen rapporteren. Een individu zou daardoor eigenlijk zijn of haar hele leven op dezelfde relatieve plek blijven, terwijl het zich met goed onderwijs ook relatief zou moeten kunnen verbeteren. Als dat niet zo zou zijn, dan zou er ook geen verschil kunnen bestaan tussen goed en minder goed onderwijs. Voor individuen is het dus zeer af te raden om vaardigheden te rapporteren op een dergelijke schaal. Als we echter kijken naar populatieverschillen, dan is dat al minder problematisch. De populatie is niet relatief.

Bij het normeren van een IQ-schaal is het gemiddelde van de populatie altijd 100. Als dat niet meer het geval is, dan wordt het tijd om de schaal te hernormeren. Als deze doctrine ook voor de vaardigheidsschalen gevolgd zou worden, dan zouden we nooit vooruitgang (of achteruitgang) kunnen waarnemen. Daarom zal de vaardigheidsschaal hier ook meer de functie van een index hebben. Het gemiddelde van 2008 is 100, en dat is de nulmeting. Net zoals bij andere indices worden de daaropvolgende scores gerelateerd aan die nulmeting.

Er is nog geen goede naam voor dit type schalen. De schaal functioneert aan de ene kant als een "Quotiënt-schaal" en aan de andere kant als een index. De transformatie van de PPON-schaal naar deze nieuwe schaal heeft gevolgen voor de waarden die gevonden worden bij de opgaven en de populatieverdelingen. In deze bijlage zijn de tabellen met deze waarden aangepast naar de nieuwe schaal. De figuren worden niet aangepast. Die zien er namelijk exact hetzelfde uit, alleen is de schaal aangepast: in plaats van de waarden 100, 150, 200, 250, 300, 350 en 400 staan in die figuren nu respectievelijk de waarden 55, 70, 85, 100, 115, 130 en 145.

Nieuwe tabellen voor de effecten zijn niet nodig aangezien de lineaire transformatie van de schaal daar geen invloed op heeft. De effectschattingen zijn "schaal-vrij" gerapporteerd. Een effect is een, al dan niet gezuiverd, verschil, tussen gemiddelden gedeeld door de standaarddeviatie, waarbij aangenomen wordt dat de standaarddeviatie van de groepen die vergeleken worden (ongeveer) gelijk zijn. Als de schaal met behulp van een lineaire transformatie verplaatst wordt, heeft dat geen invloed op de effectschatting.

$$\text{Effect} = (\text{Gem}_A - \text{Gem}_B) / \text{Stdev}$$

Als hier de genoemde transformatie op wordt toegepast, dan worden de gemiddelden met 0,3 vermenigvuldigd en wordt daar 25 bij opgeteld en wordt de standaarddeviatie alleen met 0,30 vermenigvuldigd. Dat betekent dat na transformatie:

$$\text{Effect} = ((25 + \text{Gem}_A * 0,30) - (25 + \text{Gem}_B * 0,30)) / (\text{Stdev} * 0,30) = (\text{Gem}_A - \text{Gem}_B) / \text{Stdev}$$

Het effect is dus onveranderd na transformatie.

Variabele		2008			2009		
		Gem.	SD	Pct'l	Gem.	SD	Pct'l
Jaar		100	15	(50)	100	15	(49)
Geslacht	Jongens	102	15	(54)	101	15	(53)
	Meisjes	99	15	(46)	98	15	(45)
Leertijd	Regulier	102	14	(55)	101	14	(53)
	Vertraagd	93	15	(31)	93	15	(31)
Stratum	Stratum 1	102	14	(56)	102	14	(54)
	Stratum 2	99	15	(47)	99	15	(46)
	Stratum 3	90	16	(26)	92	15	(29)
Formatiegewicht	F100	103	14	(57)	102	14	(55)
	F125	94	14	(33)	93	14	(31)
	F190	85	14	(15)	88	14	(21)
Toets	EB	100	15	(5)	100	14	(50)
	NT	77	15	(6)	77	15	(6)
Advies leraar	BB	80	12	(9)	80	12	(9)
	KB	88	12	(2)	88	12	(21)
	GT	95	12	(3)	94	12	(35)
	Havo	105	12	(62)	104	11	(60)
	VWO	119	11	(9)	118	11	(89)
Regio	Noord	100	14	(50)	99	14	(48)
	Oost	100	15	(49)	99	14	(48)
	West	100	15	(50)	100	15	(49)
	Zuid	100	15	(51)	100	14	(49)
Urbanisatiegraad	Zeer Sterk	96	16	(41)	97	16	(42)
	Sterk	100	15	(50)	100	15	(49)
	Matig	101	15	(53)	101	14	(51)
	Weinig	101	14	(52)	100	14	(50)
	Niet	101	14	(51)	100	14	(49)
Thuis taal	Alleen NLs	102	14	(54)	101	14	(52)
	NLs + bu	91	15	(28)	92	15	(30)
	Alleen bu	85	15	(16)	87	15	(20)
	Turks	80	15	(9)	82	14	(12)
	Arabisch	85	15	(16)	87	14	(20)
	Surinaams	89	15	(23)	90	15	(25)
	WestEuro	99	15	(48)	99	15	(46)

TSFJ: Taal: Spelling

Variabele		2008			2009		
		Gem.	SD	Pct'l	Gem.	SD	Pct'l
Jaar		100	15	(50)	100	15	(50)
Geslacht	Jongens	98	15	(43)	97	15	(43)
	Meisjes	102	15	(57)	102	15	(57)
Leertijd	Regulier	102	14	(56)	102	14	(55)
	Vertraagd	91	14	(28)	91	14	(28)
Stratum	Stratum 1	101	15	(53)	101	15	(52)
	Stratum 2	99	15	(47)	99	15	(48)
	Stratum 3	97	15	(41)	98	15	(44)
Formatiegewicht	1.00	101	15	(53)	101	15	(52)
	1.25	94	15	(34)	94	15	(35)
	1.90	96	15	(39)	98	15	(45)
Toets	EB	100	15	(5)	100	15	(51)
	NT	81	15	(1)	82	15	(12)
Advies leraar	BB	84	12	(1)	84	12	(14)
	KB	90	12	(2)	90	12	(24)
	GT	94	12	(3)	94	12	(35)
	Havo	104	12	(60)	103	12	(59)
	VWO	120	12	(9)	120	12	(91)
Regio	Noord	99	15	(47)	99	15	(46)
	Oost	100	15	(49)	100	15	(49)
	West	100	15	(50)	100	15	(50)
	Zuid	101	15	(53)	101	15	(52)
Urbanisatiegraad	Zeer Sterk	99	15	(49)	100	15	(50)
	Sterk	100	15	(51)	100	15	(51)
	Matig	101	15	(52)	101	15	(51)
	Weinig	100	15	(50)	100	15	(49)
	Niet	99	15	(48)	99	15	(47)
Thuis taal	Alleen NLs	100	15	(51)	100	15	(51)
	NLs + bu	99	15	(47)	100	15	(49)
	Alleen bu	96	15	(40)	98	15	(44)
	Turks	96	15	(41)	98	15	(45)
	Arabisch	96	15	(38)	97	15	(43)
	Surinaams	94	15	(34)	94	15	(35)
	WestEuro	100	15	(49)	100	15	(49)

TBFJ: Taal: Begrijpend lezen

Variabele		2008			2009		
		Gem.	SD	Pct'l	Gem.	SD	Pct'l
Jaar		100	15	(50)	101	15	(52)
Geslacht	Jongens	99	15	(47)	99	15	(48)
	Meisjes	101	15	(53)	102	15	(55)
Leertijd	Regulier	102	14	(55)	102	14	(57)
	Vertraagd	92	15	(29)	92	15	(30)
Stratum	Stratum 1	102	14	(55)	103	14	(57)
	Stratum 2	99	15	(46)	99	15	(48)
	Stratum 3	93	15	(31)	92	15	(30)
Formatiegewicht	F100	102	14	(55)	103	14	(58)
	F125	92	15	(30)	93	15	(31)
	F190	90	15	(26)	89	15	(23)
Toets	EB	100	15	(5)	101	15	(53)
	NT	74	15	(4)	75	15	(5)
Advies leraar	BB	78	11	(7)	78	11	(7)
	KB	87	11	(2)	87	11	(20)
	GT	94	11	(3)	95	11	(36)
	Havo	105	11	(64)	106	11	(65)
	VWO	120	11	(9)	120	11	(91)
Regio	Noord	99	15	(47)	99	15	(49)
	Oost	100	15	(50)	101	15	(52)
	West	100	15	(50)	100	15	(51)
	Zuid	101	15	(52)	101	15	(54)
Urbanisatiegraad	Zeer Sterk	98	16	(44)	98	16	(44)
	Sterk	100	15	(51)	101	15	(52)
	Matig	101	15	(53)	102	15	(55)
	Weinig	100	15	(51)	101	15	(53)
	Niet	99	15	(49)	100	15	(51)
Thuis taal	Alleen NLs	101	15	(53)	102	15	(55)
	NLs + bu	94	15	(35)	93	15	(33)
	Alleen bu	91	15	(26)	90	15	(25)
	Turks	88	15	(21)	87	15	(18)
	Arabisch	90	15	(24)	89	15	(22)
	Surinaams	89	15	(23)	88	15	(21)
	WestEuro	100	15	(51)	100	15	(51)

RG_0: Rekenen: getallen en bewerkingen

Variabele		2008			2009		
		Gem.	SD	Pct'l	Gem.	SD	Pct'l
Jaar		100	15	(50)	100	15	(50)
Geslacht	Jongens	103	15	(57)	103	15	(57)
	Meisjes	97	15	(43)	97	15	(42)
Leertijd	Regulier	102	14	(55)	102	14	(55)
	Vertraagd	92	15	(30)	92	14	(30)
Stratum	Stratum 1	101	15	(53)	101	15	(53)
	Stratum 2	99	15	(47)	99	15	(47)
	Stratum 3	96	15	(40)	96	15	(40)
Formatiegewicht	F100	102	15	(54)	101	15	(53)
	F125	93	15	(32)	93	15	(32)
	F190	95	15	(37)	95	15	(38)
Toets	EB	101	15	(5)	100	14	(51)
	NT	72	15	(3)	73	15	(3)
Advies leraar	BB	78	11	(7)	78	11	(7)
	KB	88	11	(2)	87	11	(20)
	GT	95	11	(3)	94	11	(35)
	Havo	105	11	(64)	105	11	(63)
	VWO	119	11	(9)	119	11	(90)
Regio	Noord	98	15	(45)	98	15	(44)
	Oost	100	15	(49)	100	15	(49)
	West	100	15	(50)	100	15	(50)
	Zuid	101	15	(53)	101	15	(53)
Urbanisatiegraad	Zeer Sterk	99	15	(47)	99	15	(47)
	Sterk	100	15	(50)	100	15	(50)
	Matig	101	15	(51)	101	15	(51)
	Weinig	100	15	(51)	100	15	(50)
	Niet	100	15	(50)	100	15	(49)
Thuis taal	Alleen NLs	101	15	(51)	100	15	(51)
	NLs + bu	98	15	(45)	98	15	(45)
	Alleen bu	97	15	(41)	97	15	(41)
	Turks	97	15	(42)	97	15	(42)
	Arabisch	96	15	(38)	96	15	(39)
	Surinaams	91	15	(26)	90	15	(26)
	WestEuro	99	15	(48)	99	15	(46)

RB_0: Rekenen: breuken , procenten en verhoudingen

Variabele		2008			2009		
		Gem.	SD	Pct'l	Gem.	SD	Pct'l
Jaar		100	15	(50)	100	15	(50)
Geslacht	Jongens	103	15	(59)	103	15	(59)
	Meisjes	97	15	(41)	97	14	(41)
Leertijd	Regulier	102	14	(56)	102	14	(55)
	Vertraagd	92	14	(29)	92	14	(29)
Stratum	Stratum 1	102	15	(54)	102	15	(54)
	Stratum 2	99	15	(47)	99	15	(47)
	Stratum 3	95	15	(37)	95	15	(37)
Formatiegewicht	F100	102	15	(54)	102	15	(54)
	F125	93	15	(31)	93	15	(31)
	F190	93	15	(33)	93	15	(33)
Toets	EB	101	15	(5)	100	14	(51)
	NT	72	15	(3)	73	15	(3)
Advies leraar	BB	78	11	(7)	78	11	(7)
	KB	87	11	(2)	87	11	(19)
	GT	94	11	(3)	94	11	(34)
	Havo	106	11	(65)	105	11	(63)
	VWO	120	11	(9)	120	11	(91)
Regio	Noord	99	15	(46)	99	15	(46)
	Oost	100	15	(50)	100	15	(50)
	West	100	15	(50)	100	15	(50)
	Zuid	101	15	(53)	101	15	(52)
Urbanisatiegraad	Zeer Sterk	98	15	(45)	99	15	(46)
	Sterk	100	15	(50)	100	15	(49)
	Matig	101	15	(52)	101	15	(52)
	Weinig	101	15	(51)	100	15	(51)
	Niet	100	15	(50)	100	15	(50)
Thuis taal	Alleen NLs	101	15	(52)	101	15	(52)
	NLs + bu	97	15	(41)	97	15	(41)
	Alleen bu	95	15	(36)	95	15	(37)
	Turks	94	15	(35)	94	15	(35)
	Arabisch	94	15	(35)	94	15	(35)
	Surinaams	89	15	(23)	89	15	(23)
	WestEuro	99	15	(47)	98	15	(46)

RM_0: Rekenen: Meten, Meetkunde, Tijd en Geld

Variabele		2008			2009		
		Gem.	SD	Pct'l	Gem.	SD	Pct'l
Jaar		100	15	(50)	100	15	(49)
Geslacht	Jongens	103	15	(59)	103	15	(58)
	Meisjes	97	15	(41)	97	15	(41)
Leertijd	Regulier	102	14	(55)	102	14	(54)
	Vertraagd	92	15	(29)	92	15	(29)
Stratum	Stratum 1	102	15	(54)	101	15	(53)
	Stratum 2	99	15	(47)	99	15	(46)
	Stratum 3	94	15	(35)	94	15	(35)
Formatiegewicht	F100	102	14	(55)	102	14	(54)
	F125	93	15	(31)	92	15	(30)
	F190	93	15	(32)	93	15	(32)
Toets	EB	101	14	(5)	100	14	(51)
	NT	71	15	(3)	72	15	(3)
Advies leraar	BB	78	11	(7)	78	11	(7)
	KB	87	11	(2)	87	11	(19)
	GT	94	11	(3)	93	11	(33)
	Havo	105	11	(64)	105	11	(62)
	VWO	121	11	(9)	121	11	(91)
Regio	Noord	99	15	(46)	98	15	(45)
	Oost	100	15	(49)	99	15	(49)
	West	100	15	(49)	100	15	(49)
	Zuid	101	15	(54)	101	15	(53)
Urbanisatiegraad	Zeer Sterk	98	15	(44)	98	15	(44)
	Sterk	100	15	(50)	100	15	(49)
	Matig	101	15	(52)	101	15	(51)
	Weinig	101	15	(51)	100	15	(51)
	Niet	100	15	(51)	100	15	(50)
Thuis taal	Alleen NLs	101	15	(52)	101	15	(51)
	NLs + bu	96	15	(40)	96	15	(39)
	Alleen bu	95	15	(36)	95	15	(36)
	Turks	94	15	(34)	93	15	(33)
	Arabisch	93	15	(32)	93	15	(33)
	Surinaams	88	15	(22)	88	15	(22)
	WestEuro	99	15	(47)	98	15	(45)

De opgaven op de "Quotiënt" schalen

Opgaven Woordenschat
op de TWQ schaal
Gesorteerd op de p80

Opgave	TSQ_St.	TSQ_p80
T3 - i13	56	81
T3 - i04	74	91
T3 - i17	76	92
T3 - i18	81	94
T3 - i14	82	98
T3 - i07	85	102
T3 - i12	78	103
T3 - i16	80	105
T3 - i05	73	106
T3 - i09	73	106
T3 - i01	73	107
T3 - i08	74	107
T3 - i06	81	114
T3 - i15	65	115
T3 - i03	82	115
T3 - i11	96	121
T3 - i02	89	122
T3 - i10	89	122
T3 - i20	99	133
T3 - i19	90	190

Opgaven Spelling
op de TSQ schaal
Gesorteerd op de p80

Opgave	TWS_St.	TWS_p80
T1 - i01	49	68
T1 - i08	66	81
T1 - i02	68	87
T1 - i10	74	87
T1 - i29	78	91
T1 - i05	72	91
T1 - i04	67	93
T1 - i28	81	100
T1 - i25	82	101
T1 - i22	76	101
T1 - i21	78	103
T1 - i24	80	105
T1 - i30	87	106
T1 - i06	89	108
T1 - i07	85	110
T1 - i09	85	111
T1 - i03	92	111
T1 - i23	94	113
T1 - i26	96	122
T1 - i27	95	133

Opgaven Begrijpend lezen
op de TBQ schaal
Gesorteerd op de p80

Opgave	TBQ_St.	TBQ_p80
T4 - i01	-14	26
T1 - i16	54	74
T4 - i05	55	75
T4 - i03	59	79
T4 - i14	67	81
T1 - i17	70	87
T1 - i12	64	88
T1 - i18	72	89
T4 - i18	72	92
T4 - i04	69	93
T4 - i15	79	94
T1 - i19	74	94
T1 - i13	80	95
T4 - i17	87	100
T4 - i13	81	101
T4 - i11	72	102
T4 - i20	84	102
T4 - i16	86	103
T4 - i06	80	104
T4 - i07	65	104
T1 - i20	68	108
T4 - i02	85	109
T1 - i14	86	110
T1 - i11	83	113
T1 - i15	76	115
T4 - i08	87	117
T4 - i09	72	132
T4 - i10	86	146
T4 - i12	128	168
T3 - i19	past niet op de schaal	

Opgaven Begrijpend Getallen
op de RGQ schaal
Gesorteerd op de p80

Opgave	RGQ_St.	RGQ_p80
R3 - i18	69	85
R1 - i04	60	87
R1 - i02	74	90
R2 - i11	73	93
R3 - i09	79	95
R1 - i15	68	95
R2 - i03	80	97
R1 - i01	70	97
R2 - i18	77	97
R2 - i19	84	98
R2 - i01	79	99
R2 - i10	72	100
R1 - i17	60	101
R3 - i01	60	101
R2 - i06	81	102
R3 - i13	89	105
R2 - i08	89	105
R2 - i04	79	106
R2 - i15	92	108
R3 - i05	94	111
R2 - i12	87	114
R2 - i16	97	117
R2 - i20	92	119
R2 - i13	94	121
R1 - i12	101	121

Opgaven Rekenen BPV
op de RBQ schaal
Gesorteerd op de p80

Opgave	RBQ_St.	RBQ_p80
R2 - i14	73	90
R1 - i08	80	94
R2 - i02	82	96
R3 - i17	76	97
R3 - i10	76	97
R2 - i17	71	100
R2 - i07	86	100
R1 - i16	88	100
R1 - i03	84	100
R3 - i20	89	101
R3 - i02	88	105
R3 - i14	89	106
R1 - i05	88	109
R1 - i06	89	110
R3 - i06	96	110
R2 - i05	84	112
R3 - i19	91	112
R1 - i13	92	113
R1 - i09	97	114
R2 - i09	92	120

Opgaven Rekenen MMTG
op de RMQ schaal
Gesorteerd op de p80

Opgave	RBQ_St.	RBQ_p80
R1 - i18	70	87
R3 - i08	74	95
R3 - i04	83	97
R1 - i14	84	98
R1 - i10	70	98
R1 - i11	72	100
R3 - i15	85	102
R3 - i11	91	103
R3 - i07	91	105
R1 - i07	93	107
R3 - i12	81	109
R1 - i20	82	110
R3 - i16	90	111
R1 - i19	96	117
R3 - i03	102	119

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

Postbus 1034

6801 MG Arnhem

T (026) 352 11 11

F (026) 352 13 56

www.cito.nl

Klantenservice

T (026) 352 11 11

F (026) 352 11 35

klantenservice@cito.nl

Fotografie: Ron Steemers