

Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2008

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



Jaarlijks Peilingsonderzoek van het Onderwijsniveau

**Peiling van de rekenvaardigheid en de taalvaardigheid
in jaargroep 8 en jaargroep 4 in 2008**

Technisch rapportage

B:T. Hemker en J.J. van Weerden

© Cito B.V. Arnhem (2009)

Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van Cito B.V. worden openbaar gemaakt en/of veeelvoudigd door middel van druk, fotografie, scanning, computersoftware of andere elektronische veeelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook.

Inhoud

Voorwoord 5

Samenvatting 7

1 Inleiding 11

2 De opzet van het onderzoek 13

2.1 Relatie met Periodieke Peiling van het Onderwijsniveau in Nederland 13

2.2 Opzet van het onderzoek voor de vaardigheden van leerlingen in jaargroep 8 16

2.3 Opzet van het onderzoek voor de vaardigheden van leerlingen in jaargroep 4 20

3 Eigenschappen van de gebruikte toetsen 23

3.1 Toetseigenschappen van de toetsen in jaargroep 8 23

3.2 Toetseigenschappen van de toetsen in jaargroep 4 24

4 Analyseprocedures: opgave- en groepeeigenschappen op één schaal 27

4.1 Toepassing van een IRT-model 27

4.2 Rapportage van de opgaven en populatie 28

4.3 Rapportage van verschillen tussen groepen 32

4.4 Presentatie van de resultaten in leerjaar 4 33

5 Validiteit 37

5.1 Begripsvaliditeit 37

5.2 Criteriumvaliditeit 43

5.3 Nederlandse taal: Begrijpend lezen 49

5.4 Rekenen/Wiskunde: Getallen en bewerkingen 57

5.5 Rekenen/Wiskunde: Breuken, procenten en verhoudingen 63

5.6 Rekenen/Wiskunde: Meten, meetkunde, tijd en geld 70

6 Resultaten van leerlingen in jaargroep 4 73

6.1 Nederlandse taal: Woordenschat 73

6.2 Nederlandse taal: Spelling 81

6.3 Nederlandse taal: Begrijpend lezen 86

6.4 Rekenen 93

7 Discussie 105

7.1 Samenvatting van het onderzoek: wat is er onderzocht? 105

7.2 Samenvatting van de resultaten: jaargroep 8 vergeleken met jaargroep 4 107

7.3 Tot slot 107

Literatuur 109

Voorwoord

Eind 2007 is de kwaliteitsagenda voor het Primair Onderwijs verschenen met de titel Scholen voor morgen. Hierin wordt door de staatssecretaris een agenda voor de komende jaren geschetst. Deze agenda is na overleg met het onderwijsveld in maart 2008 aangepast en aan de Tweede Kamer aangeboden. In de jaren 2008 tot en met 2010 zal uitvoering worden gegeven aan de ambities en plannen die daarin zijn geformuleerd.

Scholen voor Morgen is een ambitieuze agenda gericht op duurzame verbetering van het primair onderwijs. Prioriteit wordt daarbij gelegd op verhoging van de taal/-lees- en rekenopbrengsten.

Ten eerste is het dan van belang meer nauwgezet te formuleren wat leerlingen moeten kennen en kunnen als zij het primair onderwijs verlaten. Dit is uitgewerkt in de aanbevelingen van de Expertgroep Doorlopende Leerlijnen (EGDLL, ofwel de commissie Meijerink). Ten tweede is het dan zaak vast te stellen wat de beginsituatie is en hoe het niveau zich ontwikkelt. Scholen beschikken doorgaans over voldoende toetsgegevens, gebaseerd op leerlingvolgsystemen en eindtoetsen. Dat kan input zijn voor verbeterplannen op groeps- en school- en bestuursniveau. Deze gegevens kunnen echter ook, onder bepaalde condities, worden ingezet voor een landelijke niveaubepaling.

Een project dat bedoeld is om gegevens op te leveren die bruikbaar zijn voor een landelijke niveaubepaling, is PPO, de periodieke peiling van het onderwijsniveau, uitgevoerd door Cito. Bij dit onderzoek is de onderzoekscyclus voor de basisvakken echter vijf jaar. Bovendien beoogt dat project een bredere doelstelling. Voor het volgen van eventuele effecten van de nieuwe beleidsimpulsen wordt een kortere periode wenselijk geacht.

Vanuit deze achtergrond heeft OCW bij Cito de opdracht neergelegd om jaarlijks een onderzoek te doen naar het niveau voor rekenen en taal/lezen. Cito kan daarvoor toetsen inzetten die al bij de meeste scholen worden gebruikt voor het in kaart brengen van onderwijsopbrengsten bij rekenen en taal. Het gaat dan om de Eindtoets in groep 8 en toetsen van het Cito Leerling- en onderwijsvolgsysteem in groep 4. Scholen die in de steekproef zitten krijgen de recente toetsen voor de onderzochte gebieden toegestuurd.

De uitkomsten van deze jaarlijkse peiling kunnen op termijn ook geïnterpreteerd worden in termen van de referentieniveaus behorende bij de doorlopende leerlijnen zoals geformuleerd in het advies van de EGDLL "Over de drempels met taal en rekenen".

In 2008 is deze peiling voor de eerste keer uitgevoerd. Daarmee is een start gemaakt met wat een reeks van jaarlijkse rapporten moet worden, waarin het niveau van taal en rekenen in het primair onderwijs in beeld wordt gebracht.

Bij de uitvoering van het project zijn diverse mensen betrokken geweest. In het bijzonder moet Bas Hemker worden genoemd, die de analyse van de gegevens en de verslaglegging voor zijn rekening heeft genomen. Veel dank gaat ten slotte uit naar de scholen die bereid bleken de toegezonden toetsen aan hun leerlingen voor te leggen en de resultaten ter beschikking te stellen.

Drs. J.J van Weerden
Hoofd Research Primair Onderwijs Voortgezet onderwijs

Samenvatting

Inleiding

In 2008 is de eerste peiling van de rekenvaardigheid en de taalvaardigheid uitgevoerd voor jaargroep 8 en jaargroep 4 van het basisonderwijs, in het kader van wat een jaarlijkse reeks van verslagen moet worden. Merk op dat hier eerst jaargroep 8 genoemd is omdat deze groep in dit onderzoek de meeste aandacht krijgt. Doel van het project is op jaarbasis de ontwikkeling van de leerprestaties van leerlingen op systeemniveau te monitoren in het kader van de evaluatie van de kwaliteitsagenda 'Scholen voor morgen'.

Het onderzoek dat voor deze jaarlijkse peiling is uitgevoerd ligt in de traditie van het periodieke peilingonderzoek van het onderwijsniveau in Nederland (PPON), maar heeft een aantal duidelijke eigen kenmerken. Een belangrijk verschil is de mate van detaillering van de opgavenverzameling. Bij PPON zijn de opgavenverzamelingen een uitwerking van de geformuleerde onderwijsinhouden uit de kerndoelen. De toetsen en opgaven in dat onderzoek zijn speciaal voor dat doel ontwikkeld. Het aantal gemeten vaardigheden is ook duidelijk groter dan in de jaarlijkse peiling. Het PPON-onderzoek heeft echter een lage frequentie: om de vijf jaar voor vaardigheden in de Nederlandse taal en rekenen/wiskunde. Daarmee is het goed mogelijk om inzicht te verkrijgen in de ontwikkelingen op de lange termijn. Het maakt het echter wel lastig om de effectschommelingen voor de taal- en rekenprestaties te monitoren en af te zetten tegen het onderwijsbeleid. Het pikt signalen pas wat later op. Daar ligt nu de aanvulling van deze jaarlijkse peiling.

Wat is er gemeten?

De gemeten taalvaardigheden betroffen zowel in jaargroep 4 als in jaargroep 8 *Begrijpend lezen*, *Spelling* en *Woordenschat*. De rekenvaardigheden verschilden voor jaargroep 8 en jaargroep 4. In jaargroep 8 zijn drie vaardigheden gemeten: *Getallen en Bewerkingen*, *Breuken*, *Procenten en Verhoudingen* en *Meten, Meetkunde, Tijd en Geld*. In jaargroep 4 zijn vier rekenvaardigheden onderzocht: *Getallen en Getalrelaties*, *Optellen en aftrekken*, *Vermenigvuldigen en Delen* en *Meten, Tijd en Geld*. Deze vaardigheden zijn gemeten met behulp van de Eindtoets Basisonderwijs voor jaargroep 8 en de LOVS-toetsen voor einde jaargroep 4. Naast de Eindtoets is in jaargroep 8 ook gebruikgemaakt van de Niveautoets.

De vaardigheden zijn beschreven door middel van de inhoud van de opgaven. Deze opgaven zijn gerelateerd aan het niveau van verschillende groepen leerlingen. De eerste wijze waarop groepen zijn gedefinieerd is door leerlingen in te delen naar vaardigheid. Door bij vijf verschillende vaardigheidsniveaus te kijken wat voor opgaven leerlingen wel en niet beheersen, krijgen we inzicht in de inhoudelijke vaardigheid van de leerlingen in de onderzoekspopulatie. De vaardigheidsniveaus die onderzocht zijn, waren percentielgroepen P10, P25, P50, P75 en P90. Die groepen komen deels overeen met de niveaus zoals die opgesteld zijn door de commissie Meijerink. Merk overigens op dat als een leerling bij de ene vaardigheid behoort tot een bepaalde vaardigheidsgroep, dit niet hoeft te betekenen dat de leerling bij een andere vaardigheid in dezelfde vaardigheidsgroep zit. Verschillende leerlingen hebben verschillende talenten.

Het tweede type indeling in groepen is gedaan op basis van achtergrondvariabelen. Leerlingen met dezelfde achtergrondkenmerken zijn samen als groep bekeken. Een vraag die op basis van een dergelijke indeling kan worden beantwoord is wat de gemiddelde vaardigheid van jongens is, of die van meisjes. De groepen worden niet alleen onderscheiden door geslacht, maar ook door leertijd, leerlinggewicht (formatiegewicht), schoolgewicht (stratum), regio en urbanisatiegraad van de plaats van de school, en de thuistaal. Voor het onderzoek in jaargroep 8 zijn ook andere variabelen meegenomen in de analyse: de toets die de leerlingen gedaan hebben en het ingeschatte niveau van het vervolgonderwijs in het VO.

Niet alleen is er gekeken naar de gemiddelden van deze verschillende groepen, maar ook naar de gecorrigeerde verschillen tussen deze groepen. Met behulp van effectschattingen van de vaardigheden is bijvoorbeeld gekeken of er verschillen zijn tussen scholen in het westen van Nederland in vergelijking met scholen in het oosten van Nederland, als gecorrigeerd wordt voor de variabelen geslacht, leertijd,

leerlinggewicht en stratum. We kijken hierbij naar de grootte van het effect en of het effect significant groter dan 0 is. Dat is onderzocht voor alle variabelen waar ook gemiddelden voor zijn berekend.

Beschrijving van wat kinderen kunnen

Een belangrijk deel van de resultaten betreft het beschrijven van wat leerlingen wel en niet kunnen. De meeste aandacht gaat uit naar het beschrijven van de vaardigheden in jaargroep 8. Dit is een belangrijk moment in de schoolcarrière aangezien dat het snijpunt is tussen het basisonderwijs en het voortgezet onderwijs. Wat kunnen en kennen leerlingen nu na de eerste jaren basisonderwijs en met welke reken- en taalvaardigheid gaan zij het voortgezet onderwijs in?

De beschrijving van de vaardigheden op dit meetmoment gebeurt met behulp van opgaven uit de Eindtoets Basisonderwijs (EB) 2008. Het is een toets waarvoor de meeste leerlingen goed gemotiveerd zijn waardoor de prestaties waarschijnlijk optimaal zijn. Voor de beschrijving van de vaardigheid van de leerlingen in jaargroep 4 wordt verwezen naar de toetsen van het Leerling- en onderwijsvolgsysteem en proeftoetsen die voor dit Leerling- en onderwijsvolgsysteem zijn uitgevoerd.

De opgaven worden zowel bij jaargroep 8 als bij jaargroep 4 afgebeeld op vaardigheidsschalen waarbij de opgaven direct vergeleken kunnen worden met de niveaus van de minder vaardige, gemiddeld vaardige en de meer vaardige leerlingen, zoals die met behulp van de percentielgroepen beschreven zijn. Zo kan een inhoudelijke vergelijking gemaakt worden: wat kan een gemiddelde leerling in jaargroep 8 en wat kan een gemiddelde leerling in jaargroep 4. Hiermee is de inhoudelijke vooruitgang van de leerlingen te zien.

De relatie tussen achtergrondvariabelen en de vaardigheden

Per achtergrondvariabele kan aangegeven worden wat de relatie is met de diverse gemeten vaardigheden. Als er verschillen bestaan voor wat er gevonden wordt bij de gemiddelde vaardigheidsniveaus en bij de gezuiverde effecten dan worden die hieronder gegeven. De relatie tussen de achtergrondvariabelen en de vaardigheden kunnen over de vaardigheden verschillen en ook tussen jaargroep 8 en jaargroep 4 kunnen verschillen worden gevonden. Merk op dat al deze opmerkingen over groepen gaan, en niet over individuele leerlingen.

Geslacht

Jongens presteren zowel in jaargroep 4 als in jaargroep 8 beter op rekenvaardigheden dan meisjes. Meisjes presteren over het algemeen beter op de taalvaardigheden. Enige uitzondering was de vaardigheid Woordenschat waarop jongens in jaargroep 8 een hogere vaardigheid toonden. Alle gevonden effecten zijn klein, op die bij breuken, procenten en verhoudingen in jaargroep 8 na, waar een matig effect gevonden werd.

Leertijd

Vertraagde leerlingen hebben op alle vaardigheidsschalen een duidelijk lagere vaardigheid dan de leerlingen die een regulier (of vervroegd) traject volgen. De verschillen tussen deze twee groepen zijn groter in jaargroep 8 dan in jaargroep 4.

Stratum (schoolgewicht)

In jaargroep 8 worden amper effecten gevonden voor de variabelen stratum (schoolgewicht). Het schoolgewicht heeft een zeer geringe toegevoegde invloed op de effecten van de leerlinggewichten in jaargroep 8. Dat geldt voor alle vaardigheden. Er werd alleen een klein effect gevonden, bij taalvaardigheden, tussen scholen met (zeer) lage en (zeer) hoge schoolgewichten. De ongecorrigeerde gemiddelden van de scholen laten nog wel meer verschillen zien. In jaargroep 4 zijn de verschillen op dit punt groter.

Formatiegewicht (leerlinggewicht)

De gezuiverde effecten bij formatiegewicht laten in jaargroep 8 verschillen over vaardigheden zien. Bij Woordenschat wordt gevonden dat hoe hoger het leerlinggewicht is, hoe lager deze vaardigheid is: de groep leerlingen met een gewicht van 1.00 presteert het best, de groep met gewicht 1.90 het slechtst en de groep met gewicht 1.25 ligt daar met hun vaardigheid Woordenschat tussenin. Bij Begrijpend lezen valt het

verschil tussen leerlingen met gewicht 1.25 en 1.90 weg. Deze twee groepen presteren wel slechter dan leerlingen met gewicht 1.00. Bij Spelling en de rekenvaardigheden wordt dat bij de meeste schalen ook gevonden, al doet de groep leerlingen met gewicht 1.25 het vaak iets slechter dan de groep met gewicht 1.90. De effecten van formatiegewicht in jaargroep 4 waren lastiger te interpreteren doordat scholen oude en nieuwe leerlinggewichten door elkaar gebruikten.

Toets

Leerlingen die de Niveautoets gemaakt hebben, waren duidelijk minder vaardig dan de leerlingen die Eindtoets gemaakt hebben. Dit effect was het sterkst bij de rekenvaardigheden.

Doorstroom naar het voortgezet onderwijs

De verschillen tussen de leerlingen die na jaargroep 8 doorstromen naar het voortgezet onderwijs waren groot.

De vaardigheid van de leerlingen die waarschijnlijk zullen doorstromen naar het vmbo – leerweg basisberoep verschilt enorm met leerlingen die naar het vwo gaan. Bij de taalvaardigheden ligt de gecorrigeerde effectgrootte iets boven de 3.0 terwijl die bij rekenen tegen de 4.0 aan ligt. Dat zijn zeer grote effecten. Wat betreft de ongecorrigeerde gemiddelden kan gesteld worden dat het gemiddelde niveau van de eerste groep rond percentiel 10 ligt, en van de tweede groep bij percentiel 90. Hierbij zijn er wel enige kleine verschillen over vaardigheden. Voor de andere doorstroomprofielen werden hiertussen liggende resultaten gevonden. Voor alle schalen lag het algemene populatiegemiddelde (250) tussen de vaardigheid van de groepen vmbo – gemengde/theoretische leerweg en havo in.

Regio

Er worden weinig verschillen tussen regio's gevonden. Als er echter kleine effecten gevonden werden, dan viel op dat zowel bij de ongecorrigeerde gemiddelden als bij de gezuiverde effecten de leerlingen op scholen in het noorden van Nederland minder goed presteerden. Die tendens werd vrijwel over alle vaardigheden en zowel in jaargroep 4 als jaargroep 8 gevonden. Een uitzondering was Woordenschat in jaargroep 8.

Urbanisatiegraad

Bij het vergelijken van de gemiddelde vaardigheid in jaargroep 8 was deze bij een aantal vaardigheden iets lager bij de scholen in zeer sterk stedelijke en in de niet-stedelijke gebieden. In jaargroep 4 was dat alleen bij de zeer sterk stedelijke gebieden het geval. Bij de gezuiverde effecten, wanneer onder meer voor leerlinggewicht gecorrigeerd is, waren die effecten geheel verdwenen. Als er in jaargroep 8 verschillen gevonden werden, dan bleek dat na correctie over het algemeen dat hoe meer stedelijk het gebied was, hoe hoger de geobserveerde vaardigheid was.

Thuis taal

In jaargroep 4 zien we bij de meeste vaardigheden een effect van thuistaal waarbij de geobserveerde vaardigheden van de leerlingen die thuis een andere thuistaal naast of in plaats van het Nederlands gebruikten lager lagen dan de leerlingen die alleen Nederlands gebruiken. In jaargroep 8 zijn die effecten echter verdwenen op alle schalen, behalve bij Woordenschat. Dit zijn gezuiverde effecten en betreft zodoende het toegevoegde effect van de thuistaal op de vaardigheid. Dat is in jaargroep 8 dus veel kleiner dan in groep 4. Als niet gecorrigeerd is voor leerlinggewicht, stratum, geslacht en leertijd zijn de gemiddelden van de leerlingen met een andere thuistaal wel bij alle schalen lager dan gemiddeld.

Tot slot

De resultaten zoals die gepresenteerd zijn geven een stand van zaken weer. Hoewel deze resultaten wellicht kunnen dienen bij het vormen van ideeën over mogelijke oorzaken en verbanden, zijn deze uiteraard met dit onderzoek niet aan te tonen. Daar zal meer onderzoek voor gedaan moeten worden. Ook is verder onderzoek noodzakelijk voor het interpreteren van de prestaties van de leerlingen. We hebben beschreven wat leerlingen van verschillende niveaus kunnen, maar daarmee is nog niet de vraag beantwoord of het al dan niet *voldoende* is wat de leerlingen beheersen. Om daar een antwoord op te geven is standaardonderzoek nodig. Daarmee kan het beoogde fundamentele niveau en het beoogde

streekniveau beschreven worden. Die resultaten kunnen dan in een later stadium vergeleken worden met de resultaten uit dit onderzoek om te zien waar in Nederland de sterke en zwakke punten liggen: bij welke vaardigheden en welke groepen.

Ander toekomstig onderzoek betreft de vergelijking met de resultaten in het komende jaar. Daarbij worden de gegevens van de komende Eindtoets vergeleken worden met die van de Eindtoets 2008. Bij de vergelijking van de resultaten over de jaren spelen de anker-items een belangrijke rol. Door die overlap in opgaven is de vergelijking mogelijk. We moeten ons wel realiseren dat te verwachten jaareffecten klein zullen zijn. We hebben gelukkig wel de beschikking over grote hoeveelheden gegevens, zodat wanneer er relevante verschillen zullen zijn, die bij de eindtoets ook opgemerkt kunnen worden.

1 Inleiding

Dit verslag is het eerste in wat een jaarlijkse reeks van onderzoeksverslagen moet worden. In dit verslag beschrijven we het niveau van het Nederlands onderwijs: Jaarlijks Peilingsonderzoek van het Onderwijsniveau (JPO). In dit onderzoek richten we ons op Nederlandse taal en rekenen/wiskunde in de jaargroepen 4 en 8 van het basisonderwijs. Hiervoor gebruiken we toetsen uit het Leerlingvolgsysteem en de Eindtoets Basisonderwijs. Doel van het project is op jaarbasis de ontwikkeling van de leerprestaties van leerlingen op systeemniveau te monitoren in het kader van de evaluatie van de kwaliteitsagenda 'Scholen voor morgen'.

Dit verslag geeft het beginpunt van de jaarlijkse peilingen aan. Het is het ankerpunt van waaruit de komende jaren de resultaten vergeleken kunnen worden.

Het verslag begint met een uiteenzetting van het onderzoeksterrein. Hierbij komen de relatie met de traditie van het periodieke peilingsonderzoek, de omschrijving van de onderzochte populatie en de steekproef, de gekozen meetinstrumenten en de gebruikte analysemethode aan bod.

Vervolgens worden de resultaten besproken. We geven hierbij aandacht aan wat de vaardigheden zijn die de leerlingen wel of juist niet goed beheersen aan de hand van de resultaten op de opgaven. Speciale aandacht wordt gegeven aan de relatie tussen achtergrondvariabelen en de resultaten op de toetsen: welke variabelen hebben wel en welke hebben geen samenhang met de resultaten.

Het laatste hoofdstuk betreft een korte samenvatting waarin onder meer de resultaten van jaargroep 8 en jaargroep 4 met elkaar vergeleken worden. Ook zal een blik op de toekomst geworpen worden voor wat betreft de komende jaarlijkse peiling.

2 De opzet van het onderzoek

Het onderzoek van het niveau van het onderwijs past in de twintigjarige traditie van de Periodieke Peiling van het Onderwijsniveau in Nederland (PPON; Van der Schoot, 2008). In zeker zin valt dit onderzoek te zien als een nieuwe stap binnen PPON. In dit hoofdstuk komen de verschillen en de overeenkomsten van de Jaarlijkse Peiling van het Onderwijsniveau (JPO) met PPON aan bod. De meetinstrumenten en de definitie van de populatie komen ook in dit hoofdstuk aan bod.

2.1 Relatie met Periodieke Peiling van het Onderwijsniveau in Nederland

JPO richt zich op het beantwoorden van vragen als:

- Wat steken de leerlingen van het onderwijs op?
- Welke verschillen zijn er in de leerresultaten tussen categorieën van leerlingen?

Een vraag die bij JPO nu nog niet beantwoord wordt, maar bij de volgende peiling wel, is:

- Welke wijzigingen voltrekken zich over de jaren in de onderwijsresultaten?

Dit zijn vragen die ook in PPON gesteld worden. Er zijn nog wat overeenkomsten met PPON, maar er zijn ook een aantal verschillen.

Overeenkomsten tussen JPO en PPON

Een belangrijke overeenkomst met PPON is dat JPO ook wordt uitgevoerd in opdracht van het Ministerie van Onderwijs, Cultuur en Wetenschap, vanuit haar verantwoordelijkheid voor de kwaliteit van het onderwijs. Ook de doelgroepen voor de rapportage komen overeen, namelijk de diverse actoren in het onderwijsveld zoals het Ministerie van OCW, de Vaste Commissie voor Onderwijs van de Tweede Kamer en de Onderwijsraad, de Inspectie van het Onderwijs, en vakinhoudelijke experts en onderzoekers. Ook een aantal doelen van PPON is gelijk aan dat van JPO. Het verschaffen van een empirische basis voor de algemeen maatschappelijke discussie over inhoud, kwaliteit en niveau van het onderwijs is voor beide een belangrijk doel.

Om dit doel te bereiken beschrijven we de vaardigheden van de leerlingen, zowel bij PPON als bij JPO, in termen van een vaardigheidsschaal. Een belangrijke reden om in termen van een vaardigheidsschaal te rapporteren is dat niet de toetsscore zelf van belang is, maar de onderliggende vaardigheid. Deze vertaling van toetsscore naar schaal gebeurt met behulp van *Item Response Theorie* (IRT; Lord, 1980; Hambleton, Swaminathan & Rogers, 1991; Van der Linden & Hambleton, 1997) en aan IRT gerelateerde methoden. Toetsscores kunnen in principe alleen geïnterpreteerd kan worden als de leerlingen dezelfde toets gemaakt hebben. De interpretatie is dan op een ordinaal meetniveau. Met het gebruik van IRT-technieken is het mogelijk ook afstanden op de vaardigheidsschaal te interpreteren. Dit heeft tot gevolg dat ook leerlingen die verschillende toetsen gemaakt hebben, mits aan een aantal voorwaarden voldaan is, met elkaar vergeleken kunnen worden.

Een ander voordeel is dat zowel opgaven als leerlingen op dezelfde vaardigheidsschaal geplaatst kunnen worden. Aan de ene kant kan hierdoor de vaardigheidsschaal inhoudelijk beschreven worden in termen van opgaven. Deze voorbeeldopgaven illustreren wat leerlingen op verschillende niveaus van vaardigheid goed, matig of in onvoldoende mate weten of kunnen. Hoewel vrijwel alle leerinhouden van de leergebieden – rekenen/wiskunde in het bijzonder – een hiërarchische structuur hebben, is dat vaak onvoldoende om alleen op die basis te beschrijven in hoeverre de leerlingen bepaalde vaardigheden wel of niet beheersen. Om die reden is de beschrijving van de vaardigheden van leerlingen vooral gebaseerd op deze voorbeelden van opgaven die aan de leerlingen zijn voorgelegd.

De voorbeeldopgaven maken het vaardigheidsniveau op de schaal concreet. Bij de beschrijving van deze opgaven krijgen twee punten op de vaardigheidsschaal speciale aandacht. Ten eerste is dat het vaardigheidsniveau waaronder de opgave als 'moeilijk' geclassificeerd kan worden. Dat is gedefinieerd als

het punt van de vaardigheidsschaal waaronder een leerling een kans heeft van minder dan een half om de opgave goed te beantwoorden. In plaats van deze kans van een half te zien als een kans die een enkele persoon heeft om een opgave goed te beantwoorden, is het ook mogelijk deze kans te beschrijven als de kans binnen een groep personen. Het p50-punt van een opgave kan zo gedefinieerd worden als het punt op de vaardigheidsschaal, waarbij de kans exact een half is om binnen de groep personen met precies die vaardigheid, iemand te vinden die de opgave goed kan maken. Onder dat vaardigheidsniveau is de kans kleiner, waarmee de opgave voor die groep personen een moeilijke opgave is.

Het andere niveau dat speciale aandacht krijgt in de beschrijving van de opgaven is het beheersingsniveau. Dat is het niveau op de vaardigheidsschaal waar precies 80% van de leerlingen op dat niveau die opgave goed kan beantwoorden. In termen van individuele personen: het niveau op de vaardigheidsschaal waarop een persoon een kans van 0,80 heeft om de opgave goed te maken. Wanneer we de opgaven voor een onderwerp op moeilijkheidsgraad rangschikken, dan verkrijgen we een hiërarchische opbouw van rekeninhoudelijke opgavenkenmerken waardoor opgaven moeilijker worden. Dergelijke ontwikkelingslijnen vertonen uiteraard een sterke overeenkomst met de opbouw van de rekenvaardigheid in de gehanteerde onderwijsmethoden. Door deze ordening is er voor ieder punt van de schaal een beschrijving van het niveau van de vaardigheid.

Naast deze inhoudelijke beschrijving van de schaal en daarmee de vaardigheid, kan de schaal beschreven worden in termen van de verdeling van de populatie van leerlingen. Wat is het niveau van de 10% slechtst presterende leerlingen en wat dat van de 10% best presterende? Welke opgaven zijn moeilijk voor een gemiddelde leerling en welke beheerst deze leerling? Behalve het beschrijven van de gehele onderzoekspopulatie, is het mogelijk de populatie op te delen in subpopulaties. Zo kan gekeken worden hoe deze subgroepen presteren op de schaal. Wat is de gemiddelde vaardigheid voor jongens, en wat die voor meisjes? Dit kan in een directe vergelijking, maar ook gecorrigeerd voor diverse variabelen. Zo kunnen bijvoorbeeld regio-effecten gecorrigeerd worden voor de verdeling op grond van het formatiegewicht van de leerlingen. Dat laatste gebeurt met behulp van gecorrigeerde effectschattingen in een structureel model waarbij de vaardigheid beschreven wordt in termen van achtergrondvariabelen. Deze methoden en technieken die nu gebruikt worden binnen JPO, worden ook gebruikt binnen PPON.

Verschillen tussen JPO en PPON

Het is duidelijk dat de methoden en technieken en de wijze van presenteren van JPO en PPON vergelijkbaar zijn. Toch is er ook een groot aantal verschillen, die het naast elkaar bestaan van de twee rechtvaardigen. Een heel belangrijk verschil is de mate van detaillering van de opgavenverzameling. In PPON zijn de opgavenverzamelingen een uitwerking van de geformuleerde onderwijsinhouden uit de kerndoelen. Deze verzamelingen of overzichten zijn dus van essentieel belang voor het beschrijven van de leeropbrengsten van het onderwijs. Om nu een precies beeld te krijgen van wat leerlingen wel en niet kunnen betreft dit niet een algemene 'Rekenschaal', maar wordt rekenen opgedeeld in meer dan 20 subvaardigheden (Kraemer, et al., 2005; Janssen, Van der Schoot, & Hemker, 2005). De opgaven voor PPON worden speciaal voor dit doel geconstrueerd. Dit gaat dan uiteraard ook gepaard met een zeer grote opgavenverzameling die al deze schalen exact kan beschrijven, en waarmee de leerlingen nauwkeurig op deze subschalen gemeten kunnen worden. Uiteraard geldt dit ook voor de taalpeilingen, waar binnen PPON over verschillende vaardigheden verschillende balansen worden uitgebracht (Sijtstra, Van der Schoot & Hemker, 2002; Van Berkel, et al 2002; Heesters, et al, 2007; Moelands, et al, 2007). Doordat de ruime dekking van de leer- en vormingsgebieden in het PPON-ontwerp voorop staat, is het goed mogelijk om op basis van de resultaten een steekhoudende inhoudelijke discussie over het onderwijs te voeren, zoals op basis van PPON-gegevens binnen het rekenen gevoerd wordt door Van Putten (in Janssen, Van der Schoot, & Hemker, 2005; pp 125-132). Binnen JPO is de opdeling in subvaardigheden binnen rekenen veel minder gedetailleerd.

Door de hoge mate van detaillering en de relatie met de kerndoelen is ook goed mogelijk om een sterk inhoudelijke *standaard setting*-procedure toe te passen met behulp van de voorbeeldopgaven. Deze standaarden kunnen gegeven worden op basis van compacte, didactisch betekenisvolle onderwerpen of modules gebaseerd op de domeinbeschrijving. Dat is een structurele beschrijving van een leergebied aan de hand van een geordende lijst van leer- en vormingsdoelen en is gebaseerd op de eindtermen

respectievelijk de kerndoelen basisonderwijs, methodeanalyses en vakinhoudelijke en vakdidactische publicaties. De domeinbeschrijving vormt de basis voor de opgaven waarmee een leergebied wordt gemeten. Met behulp van deze opgaven wordt het dan ook mogelijk aan te geven of de leerlingen in kwalitatief opzicht het gewenste kennis- of vaardigheidsniveau bereiken: 'Wat zouden de leerlingen aan het einde van het basisonderwijs over een bepaald onderwerp moeten kennen, weten of kunnen wil er sprake zijn van voldoende beheersing van de kerndoelen' (Van der Schoot, 2008). Binnen JPO zal de standaard niet gebaseerd worden op een didactisch inhoudelijk criterium, zoals bij PPON het geval is, maar op basis van de verdeling van de populatie.

Doordat het PPON-onderzoek een lage frequentie heeft (om de vijf jaar voor vaardigheden in de Nederlandse taal en rekenen/wiskunde) is het met dat onderzoek goed mogelijk om inzicht te verkrijgen in de ontwikkelingen op de lange termijn. Het heeft echter als nadeel dat deze lange termijnen het lastig maken om de effectschommelingen voor de taal- en rekenprestaties te monitoren en af te zetten tegen het onderwijsbeleid. Dat is ook waar JPO een aanvulling geeft. Er kan sneller dan bij PPON gereageerd worden op de actualiteit en er kan sneller door de beleidsmaker gereageerd worden op gesignaleerde trends in het onderwijs. Het is aannemelijk dat de veranderingen over de tijd kleiner zullen zijn, naarmate de tijdsspanne waarover gemeten wordt kleiner is. Dat is een reden dat we bij JPO bij leerjaar 8 graag gebruikmaken van een (zeer) grote hoeveelheid leerlingen, om deze mogelijke veranderingen toch te kunnen observeren. Dat wordt gedaan door gebruik te maken van de gegevens van de Eindtoets Basisonderwijs (EB).

De meetmomenten van PPON en JPO verschillen ook enigszins. Bij PPON worden taal en rekenen gemeten aan het einde van leerjaar 8 (omstreeks mei) en medio leerjaar 5. Bij JPO ligt het meetmoment iets eerder: medio leerjaar 8 en einde leerjaar 4. De meeste aandacht zal hierbij uitgaan naar leerjaar 8. Dat is een meetmoment dat ook in ander onderzoek, zoals PRIMA en COOL., veel aandacht krijgt. Dat is ook begrijpelijk aangezien dat het snijpunt is tussen het basisonderwijs en het voortgezet onderwijs. Het geeft antwoord op de vragen: wat kunnen en kennen kinderen na de eerste jaren basisonderwijs en met welke reken- en taalvaardigheid gaan zij het voortgezet onderwijs in?

JPO geeft een aantal aanvullingen op andere onderzoeken die jaargroep 8 betreffen. Ten eerste is geen van deze andere onderzoeken zo frequent. Ten tweede is het aantal leerlingen dat gemeten wordt zeer hoog. Ten derde is de prestatie-motivatie van de leerlingen waarschijnlijk hoger dan bij ander onderzoek: we krijgen echt te zien wat de leerling 'op de toppen van zijn of haar kunnen' aan vaardigheid heeft. Ten vierde zal het door de toepassing van IRT goed mogelijk zijn opgaven en leerlingen aan elkaar te relateren zoals dat ook in PPON gebeurt. Dat gebeurt weliswaar op een minder gedetailleerd niveau, maar zowel de reken- als de taalvaardigheden worden in drie subschalen verdeeld, waarbij op opgaveniveau naar de vaardigheden gekeken wordt. Het is daarmee gedetailleerder dan veel ander onderzoek.

Als nadeel zou gezien kunnen worden dat de leerlingen die de EB maken niet representatief zijn voor de Nederlandse onderwijspopulatie. Hiervoor is echter in het onderzoek gecontroleerd door de eigenschappen van de 'steekproef' te relateren aan de landelijke verdeling. Het bezwaar dat de zwakkere leerlingen door scholen bewust weggehouden worden uit de EB, wordt ondervangen doordat de resultaten van de Niveautoets (NT) ook gebruikt worden voor dit onderzoek. De NT is een toets die bedoeld is voor 'zwakkere leerlingen', voor wie de EB als te moeilijk gezien wordt. Hierbij liggen de opgaven op dezelfde schalen als bij de EB. De leerlingen en de opgaven van de EB en NT zijn hierdoor direct met elkaar vergelijkbaar. In paragraaf 2.2. wordt dieper in gegaan op de opzet van het onderzoek voor de leerlingen in jaargroep 8.

De resultaten voor het niveau aan het einde van leerjaar 4 zijn gebaseerd op de gegevens die verkregen zijn met behulp van toetsen die deel uitmaken van het leerling- en onderwijsvolgsysteem van Cito (LOVS). Hierbij is gebruikgemaakt van resultaten van bestaande reken- en taaltoetsen. De boekjes waarin de toetsen gemaakt zijn, zijn naar Cito teruggestuurd waardoor het mogelijk werd de resultaten van jaargroep 4 op deze vaardigheden op het niveau van de opgave te analyseren. Ook is er gebruikgemaakt van gegevens uit extra onderzoek binnen het LOVS naar de vaardigheid Woordenschat. Zowel in het geval van de bestaande toetsen als bij de nieuwe toetsen is er gewerkt met gestratificeerde steekproeven waarbij rekening is gehouden met schoolgewicht en regio van de scholen. Net zoals bij het onderzoek naar leerjaar 8 worden hier de leerlingen uit het Speciaal (Basis) Onderwijs buiten beschouwing gelaten. Het aantal

waarnemingen per vaardigheid ligt bij deze toetsen wel lager dan bij de EB, en er zijn ook minder achtergrondgegevens beschikbaar. De opzet van het onderzoek voor de leerlingen in jaargroep 4 wordt verder uitgewerkt in paragraaf 2.3.

2.2 Opzet van het onderzoek voor de vaardigheden van leerlingen in jaargroep 8

De vaardigheden en de gebruikte meetinstrumenten

In het onderzoek van JPO wordt gerapporteerd over subschalen binnen de hoofdvaardigheden Nederlandse taal en rekenen/wiskunde. Voor jaargroep 8 zijn daarbij de volgende deelvaardigheden onderscheiden:

- Nederlandse taal:
- Begrijpend lezen
 - Spelling
 - Woordenschat
- Rekenen/wiskunde:
- Getallen en bewerkingen
 - Breuken, procenten en verhoudingen
 - Meten, tijd en geld.

In de Eindtoets Basisonderwijs (EB) zijn opgaven opgenomen die deze vaardigheden meten, waardoor dit een geschikt meetinstrument is om voor dit onderzoek te gebruiken. De gegevens zijn verzameld tijdens de afname van 12 tot en met 14 februari 2008. 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. Op het ogenblik wordt er ook onderzoek gedaan naar exacte inhoudelijke interpretatie van deze schaal *Schrijven van teksten*. Ook technisch blijkt dat deze vaardigheid niet goed op dezelfde schaal geplaatst kan worden als de vaardigheid *Begrijpend lezen*. Om die reden zijn de opgaven behorend bij 'Schrijven van teksten' buiten beschouwing gelaten. Een ander onderdeel dat wel deel uit maakt van de standaardscore op de EB, maar dat nu ook buiten beschouwing gelaten is, betreft de vaardigheid 'studievaardigheden'.

Er is ook interesse in de vaardigheid Technisch lezen. Het onderzoek naar die vaardigheid is op zoveel verschillende punten anders dan het onderzoek dat hier nu beschreven is, dat het niet in dit rapport past. Het type toetsen, de wijze van afname, het type opgaven, het type gegevens, het soort analyses en in totaal daarmee ook de presentatie van de gegevens verschillen. Binnenkort zal een kort verslag over de resultaten bij technisch lezen verschijnen.

Het onderzoek naar die vaardigheid is op zoveel verschillende punten anders dan het onderzoek dat hier nu beschreven is, dat het niet in dit rapport past. Het type toetsen, de wijze van afname, het type opgaven, het type gegevens, het soort analyses en in totaal daarmee ook de presentatie van de gegevens verschillen. Wat betreft het type analyse kan zelfs opgemerkt worden dat afgelopen maanden nieuwe modellen uitgewerkt zijn en nieuwe programmatuur is ontwikkeld om deze opgaven beter te kunnen analyseren.

Een mogelijk nadeel van het gebruik van de EB is dat mogelijk minder vaardige leerlingen deze toets niet maken. Om dat probleem te ondervangen zijn ook de resultaten op de Niveautoets (NT) gebruikt om de vaardigheid van de leerlingen in jaargroep 8 te meten.

- **Niveautoets** De Niveautoets is bestemd voor leerlingen met een leerachterstand van tenminste 1½ jaar over vrijwel de gehele linie. Het betreft leerlingen van wie veruit de meesten in aanmerking komen voor het praktijk- of het leerwegondersteunend onderwijs. De Niveautoets wordt op de computer gemaakt. De toets bevat dezelfde onderdelen en hetzelfde aantal opgaven als de reguliere, papieren Eindtoets. De opgaven van de Niveautoets zijn echter aangepast aan het niveau van leerlingen met een grote leerachterstand

Ondanks dat deze toets bestaat uit andere, (veel) eenvoudiger opgaven voor deze zes vaardigheden, is het met behulp van Item Response Theorie mogelijk de resultaten op de NT direct te vergelijken met de EB.

Naast de gegevens van de reguliere EB is er ook gebruikgemaakt van de gegevens van een controlevariant van de EB, die tegelijkertijd en onder dezelfde omstandigheden met de EB is afgenomen. Het is ondermeer met behulp van deze 'ankertoets' (EB-Anker) mogelijk de normering van de EB over jaren heen vergelijkbaar te houden. Deze normering is dan niet afhankelijk van een mogelijke daling of stijging in de vaardigheid van de leerlingen. Om die reden is deze ankertoets ook belangrijk in het kader van JPO: met behulp van deze toets kan onderzocht worden of de vaardigheid van leerlingen verandert. De EB en EB-Anker verschillen alleen op de derde dag van de afname en verschillen op deze dag slechts bij een deel van de opgaven.

Rond de campagne van de Eindtoets Basisonderwijs 2008 zijn ook twee andere aan de Eindtoets gerelateerde toetsen afgenomen, te weten de Inhaaltoets (IHT) en de digitale eindtoets (DEB).

- **Digitale Eindtoets** Net als voorgaande jaren konden scholen in 2008 de Eindtoets Basisonderwijs digitaal afnemen. De Digitale Eindtoets bestaat uit dezelfde onderdelen en hetzelfde aantal opgaven als de reguliere, papieren Eindtoets. De opgaven van de Digitale Eindtoets zijn anders; de moeilijkheid van de toets is echter gelijk aan de moeilijkheid van de reguliere Eindtoets. De afname van de Digitale Eindtoets vindt plaats in de periode tussen 28 januari en 7 maart 2008. Scholen kunnen zelf bepalen wanneer zij die toets afnemen.
- **Inhaaltoets** De Inhaaltoets is specifiek bestemd voor leerlingen die bijvoorbeeld wegens ziekte op één of meer vaste afnamedagen van de reguliere, papieren Eindtoets niet hebben kunnen deelnemen. Ook leerlingen van wie de leerkracht zeker weet dat er tijdens de afname van de reguliere Eindtoets iets misgegaan moet zijn, kunnen eventueel de Inhaaltoets maken. De Inhaaltoets wordt dan gebruikt als herkansingstoets. Net als de Digitale Eindtoets en de Niveautoets wordt de Inhaaltoets op de computer gemaakt. De Inhaaltoets bestaat uit dezelfde onderdelen en hetzelfde aantal opgaven als de reguliere, papieren Eindtoets. Ook is de moeilijkheid van de Inhaaltoets gelijk aan de moeilijkheid van de reguliere Eindtoets. Er kon voor de Inhaaltoets ingeschreven worden tot en met 14 maart 2008. Afnames waren mogelijk tot en met 4 april 2008.

Deze toetsen worden bij een zeer beperkt aantal leerlingen afgenomen, in totaal minder dan 1% van de leerlingen die de Eindtoets maakte (714 leerlingen IHT; 213 leerlingen DEB). In tegenstelling tot bij 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. Ook de periode van afname verschilt duidelijk van die van de EB. Om deze redenen zijn de resultaten van deze toetsen niet opgenomen in de analyses. Deze toetsen zijn ook niet opgenomen in de analyses rond de Terugblik en resultaten Eindtoets Basisonderwijs 2008 (Van der Lubbe, 2008).

Opgemerkt kan worden dat een deel van de data die nu onderzocht wordt, ook geanalyseerd is bij de Terugblik en resultaten Eindtoets Basisonderwijs 2008. Er zijn echter ook verschillen. Ten eerste zijn de vaardigheden waarover gerapporteerd wordt in JPO gedetailleerder dan bij de terugblik het geval is. De belangrijkste analyses in de Terugblik betreffen de totaalscore en de standaardscore. Voor Taal en Rekenen-Wiskunde worden alleen samengevatte resultaten gegeven, en niet per subschaal. Ten tweede is de methode van analyse duidelijk anders dan in het geval van JPO. Door het gebruik van IRT kunnen de schalen beter geïllustreerd worden door de opgaven af te beelden op de vaardigheidsschalen. Ook de effecten van achtergrondvariabelen worden gecorrigeerd voor de aanwezigheid van storende variabelen. De achtergrondvariabelen verschillen ook enigszins over de studies, waarbij het aantal variabelen hoger ligt in de analyses van JPO. Het doel van JPO is ook anders dan dat van de terugblik.

De onderzoekspopulatie

De onderzoekspopulatie van JPO betreft de reguliere leerlingen die in 2008 in het basisonderwijs in Nederland zaten. Voor de analyses van de gegevens van de EB en de NT betekent dit dat een aantal leerlingen dat een van deze toetsen gemaakt heeft niet in de onderzoekspopulatie thuishoort. De EB wordt namelijk ook afgenomen bij leerlingen in het voortgezet onderwijs (474 leerlingen in 2008) en bij leerlingen in het buitenland (218 leerlingen). Er is ook een groep instellingen waarbij het niet duidelijk is of het een reguliere basisschool betreft. In deze groep zitten bijvoorbeeld ook leerlingen uit het speciaal (basis)onderwijs. Omdat de leerlingen van deze scholen niet tot de onderzoekspopulatie behoren, zijn ze niet meegenomen in de analyses, net zomin als de leerlingen die de DEB of de IHT gemaakt hebben.

Er waren in totaal 146608 leerlingen, afkomstig van 6252 scholen en instellingen, die deel hebben genomen aan de Eindtoets 2008. Daarvan waren er 140337 (afkomstig van 5850 basisscholen) die zeker deel uitmaken van de te meten populatie. In tabel 2.1 wordt per toets het aantal leerlingen gegeven die al dan niet in de analyses zijn opgenomen.

Tabel 2.1 Aantal leerlingen in de Eindtoetscampagne in 2008, verdeeld over de onderzoekspopulatie en over de toetsen

Toets	BO NL	Uit analyses	Totaal
EB	136811	5236	142047
NT	2315	78	2393
IHT	0	714	714
DEB	0	243	243
EB Anker	1211	0	1211
Totaal	140337	6271	146608

Er zijn geen CBS-gegevens beschikbaar per leerjaar, maar op basis van andere gegevens zaten er in 2008 naar schatting tussen de 180.000 en de 185.000 leerlingen in jaargroep 8 van het reguliere basisonderwijs. Dat betekent dat met het aantal leerlingen dat nu gemeten is, tussen de 75% en 80% van de leerlingen van de onderzoekspopulatie gemeten is. Als daarbij ook nog de leerlingen geteld worden waar we gegevens over hebben, maar die mogelijk wel opgenomen hadden kunnen worden – inclusief de leerlingen die de DEB of de IHT gemaakt hebben –, dan ligt dit percentage eerder richting de 80%.

Naast de toetsresultaten zijn van deze leerlingen ook achtergrondkenmerken verzameld. Deze achtergrondvariabelen kunnen gebruikt worden om verschillen tussen leerlingen te kunnen vergelijken op basis van hun verschillen in eigenschappen. De gegevens betreffen:

- geslacht;
- leeftijd;
- formatiegewicht;
- land van herkomst;
- thuistaal;
- doorstroomkenmerk.

De leeftijd van de leerlingen wordt vervangen door de variabele leertijd. In deze variabele onderscheiden we twee categorieën:

regulier: de leerlingen in jaargroep 8 die op 1 oktober 2007 niet ouder zijn dan 11 jaar.

vertraagd: de leerlingen die op dat moment 12 jaar of ouder zijn

Als toelichting op de variabele formatiegewicht merken we het volgende op. Op het ogenblik worden de leerlinggewichten binnen het basisonderwijs gefaseerd aangepast. In het schooljaar 2007-2008 vallen de leerlingen in jaargroep 8 nog onder de oude gewichtenregeling. Dat houdt in dat de volgende gewichten gehanteerd zijn voor jaargroep 8:

- (nu factor 0.25) voor Nederlandse arbeiderskinderen (in termen van opleidings- en/of beroepsniveau van de ouders);
- (nu factor 0.40) voor schipperskinderen in een internaat of pleeggezin;
- (nu factor 0.70) voor kinderen uit de reizende en trekkende bevolking;
- (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);
- (nu factor 0.00) voor alle andere kinderen.

De gewichten tussen haakjes zijn de gewichten die tegenwoordig gebruikt worden *naast* de oude gewichten. Ondanks de grote aantallen leerlingen in de analyses komen de gewichten 1.40 en 1.70 zeer weinig voor. Slechts 191 leerlingen hadden een leerlinggewicht van 1.40 en 202 hadden een leerlinggewicht van 1.70. Om die reden zijn in de analyses de leerlingen met gewichten 1.25, 1.40 en 1.70 samengenomen als een groep.

Met doorstroomkenmerk wordt bedoeld het (ingeschatte) niveau van de opleiding die de leerling na jaargroep 8 in het voortgezet onderwijs gaat volgen. 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 tussenliggende categorieën: BB/KB, KB/GT, GT/havo en havo/vwo. In de rapportage van de resultaten worden alleen de resultaten van de hoofdcategorieën gegeven. De resultaten van de tussenliggende categorieën zijn conform wat verwacht mag worden: de resultaten van een tussencategorie liggen ook tussen 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 voor de analyses van deze groepen de leerlingen die alleen deze andere taal spreken samen genomen met de leerlingen die deze andere taal samen met Nederlands thuis spreken. Bij 30% van de leerlingen die thuis Turks spreekt, wordt thuis ook Nederlands gesproken, en bij leerlingen die thuis Arabisch spreken geldt voor 36% dat zij thuis ook Nederlands spreken. Voor de leerlingen die thuis Surinaams/Antilliaans of een West-Europese taal spreken geldt dat meer dan de helft thuis ook Nederlands thuis spreekt.

Naast deze leerlingkenmerken zijn ook andere kenmerken opgenomen. Deze kenmerken zijn gerelateerd aan de school:

- stratumindeling;
- regio;
- urbanisatiegraad,

De stratumindeling is een indeling van scholen op basis van schoolgewicht. De schoolscore is hierbij 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. 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 die gehanteerd wordt bij 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 in CBS-publicaties gebruikt wordt:

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

2.3 Opzet van het onderzoek voor de vaardigheden van leerlingen in jaargroep 4

Voor het onderzoek in jaargroep 4 worden de deelnemende scholen gefaciliteerd in het gebruik van de LOVS-toetsen. Deze toetsen worden momenteel vernieuwd. In de toekomst zal onderzocht moeten worden in hoeverre eventuele bekendheid met deze nieuwe toetsen een zelfstandig positief effect heeft op de toetsresultaten.

De vaardigheden, de gebruikte meetinstrumenten en de onderzoekspopulatie

De afzonderlijke vaardigheden die gemeten worden voor Nederlandse taal en voor rekenen in leerjaar 4 verschilt enigszins van die van die in leerjaar 8. Dat verschil wordt veroorzaakt door het verschil in curriculum in de twee leerjaren. Een aantal basisvaardigheden wordt niet meer direct getoetst in leerjaar 8 omdat vrijwel iedereen die dan beheerst. Aan de andere kant worden er in leerjaar 8 een aantal vaardigheden gemeten die niet onderwezen werden in leerjaar 4, waardoor het meten daarvan in die groep ook niet zinvol is.

Voor jaargroep 4 zijn de volgende deelvaardigheden onderscheiden:

- | | |
|-------------------|-----------------------------|
| Nederlandse taal: | – Begrijpend lezen |
| | – Spelling |
| | – Woordenschat |
| Rekenen/Wiskunde: | – Getallen en getalrelaties |
| | – Optellen en aftrekken |
| | – Vermenigvuldigen en delen |
| | – Meten, tijd en geld |

Voor de vaardigheden Begrijpend lezen en Spelling zijn de gegevens afkomstig van de bestaande startmodules en vervolgmodes binnen het LOVS. Voor de vaardigheid Woordenschat zijn de gegevens afkomstig van de proeftoetsen voor nieuw te ontwikkelen toetsen binnen het LOVS. Voor de gegevens voor

rekenen is er voor einde leerjaar 4 één toets beschikbaar. Deze toets bestaat uit twee delen die alle leerlingen maken.

De onderzoekspopulatie betreft jaargroep 4 in het reguliere basisonderwijs. Net zoals bij het onderzoek naar jaargroep 8 worden hier de leerlingen uit het speciaal (basis)onderwijs buiten beschouwing gelaten. Voor het onderzoek in jaargroep 4 zijn verschillende steekproeven getrokken. De eerste steekproef betrof de leerlingen die bestaande LOVS-toetsen gemaakt hebben. Hierbij is een steekproef getrokken van 60 basisscholen. Deze steekproef is getrokken op basis van cfi-gegevens aangaande regio en schoolgewicht. Een aparte steekproef is gebruikt voor de toetsen voor Woordenschat.

3 Eigenschappen van de gebruikte toetsen

In deze studie zijn verschillende meetinstrumenten gebruikt. Voor jaargroep 8 zijn de Eindtoets Basisonderwijs (EB) en een variant van de Eindtoets (EB-anker) gebruikt, evenals de Niveautoets. De resultaten betreffen echter niet de gehele toets, maar subschalen uit deze toets. In dit hoofdstuk worden de eigenschappen van de gebruikte subtoetsen beschreven voor jaargroep 8.

Als meetinstrument voor de studie in leerjaar 4 zijn LOVS-toetsen gebruikt. Voor zes van de zeven schalen zijn bestaande toetsen gebruikt. Voor de vaardigheid Woordenschat zijn proeftoetsen gebruikt. Bij de bestaande taaltoetsen kon een onderscheid gemaakt worden tussen startmodules en vervolgoetsen. Een verschil met de rekentoets is dat in het reguliere gebruik deze toets één score oplevert. In dit onderzoek kunnen we de toets gebruiken om vier verschillende rekenvaardigheden te beschrijven. In dit hoofdstuk worden de eigenschappen van deze (sub)toetsen beschreven.

3.1 Toetseigenschappen van de toetsen in jaargroep 8

De drie toetsen die gebruikt zijn voor de bepaling van de vaardigheden zijn de EB, de NT en de EB-Anker. Aangezien de vaardigheid niet gemeten wordt met behulp van de gehele toets, maar slechts met een deel van de opgaven uit die toets, zijn niet de toetseigenschappen van de gehele toets van belang, maar die van de subtoetsen waarmee de vaardigheden gemeten worden. Hieronder worden de eigenschappen gegeven van deze zes subtoetsen voor de Eindtoets.

Tabel 3.1.1 Toetseigenschappen van de 3 subtoetsen Nederlandse Taal bij de Eindtoets (N = 136.858)

	Aantal items	Gem.Score	SD	GemP	Alpha	GLB
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

Tabel 3.1.2 Toetseigenschappen van de 3 subtoetsen Rekenen/Wiskunde bij de Eindtoets (N = 136.858)

	Aantal items	Gem.Score	SD	GemP	Alpha	GLB
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

Naast de gemiddelde score en de standaarddeviatie van de scores is ook de gemiddelde P-waarde gegeven. Dat 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 populatie. Het 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, betreft Cronbach's alpha. Dat is een veel gebruikte maat waarmee men een indicatie heeft van de betrouwbaarheid van een toets. Deze alpha is echter een ondergrens van de feitelijke betrouwbaarheid van een toets. 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 dichter bij de feitelijke betrouwbaarheid ligt. Volgens het toetsbeoordelingsstelsel 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,

Van Vliet-Mulder & Groot, 2000; pp 1409-1415). De enige uitzondering is de subtoets Woordenschat, die echter nog altijd de kwalificatie voldoende krijgt. Opgemerkt moet worden dat bij de Eindtoets geen scores voor deze subschalen geleverd worden, omdat dat beoordelingen scores op individueel niveau zijn, en niet op groepsniveau.

Voor de schalen bij de Eindtoets-Anker is geen GLB geleverd. Het aantal leerlingen (1164) bij deze toetsen is aanzienlijk lager dan bij de EB. De subtoetsen in de EB-Anker zijn voldoende betrouwbaar voor analyses op groepsniveau voor de subschalen.

Tabel 3.1.3 Toetseigenschappen van de drie subtoetsen Nederlandse Taal bij de Eindtoets-Anker (N = 1164)

	Aantal items	GemSc	StDev	GemP	Alpha
Spelling	20	15,0	3,1	74,9	0,66
Woordenschat	19	13,4	2,9	70,7	0,61
Begrijpend lezen	30	22,4	4,6	74,5	0,77

Tabel 3.1.4 Toetseigenschappen van de drie subtoetsen Rekenen/Wiskunde bij de Eindtoets-Anker (N = 1164)

	Aantal items	GemSc	StDev	GemP	Alpha
Getallen en Bewerkingen	25	18,5	4,3	73,9	0,79
Breuken, Procenten en Verhoudingen	20	14,1	4,1	70,7	0,81
Metten, Meetkunde, Tijd en Geld	15	11,0	2,6	73,1	0,64

De gemiddelde P-waarden verschillen zowel bij taal als bij rekenen iets van die bij de EB en dat geldt ook voor de betrouwbaarheid. De verschillen zijn echter niet enorm groot. De ankertoetsen zijn nog steeds voldoende betrouwbaar om uitspraken op te doen voor de afzonderlijke vaardigheden. Dat de verschillen niet erg groot zijn, valt ook te verwachten gezien de grote mate van overlap tussen de toetsen.

De resultaten van de Niveautoets zijn gebaseerd op de gegevens van alle 2315 leerlingen van het regulier basisonderwijs die deze toets gemaakt hebben op een reguliere basisschool. In totaal zaten zij op 921 verschillende basisscholen, waarbij op 919 van die scholen ook leerlingen de Eindtoets hebben gemaakt. Ook voor de schalen van de Niveautoets is geen GLB gegeven, gezien het aantal leerlingen.

Tabel 3.1.5 Toetseigenschappen van de drie subtoetsen Nederlandse Taal bij de Niveautoets (N = 2315)

	Aantal items	GemSc	StDev	GemP	Alpha
Spelling	20	12,8	3,7	64,0	0,72
Woordenschat	20	15,2	3,1	76,1	0,67
Begrijpend lezen	30	20,6	5,0	68,6	0,79

Tabel 3.1.6 Toetseigenschappen van de drie subtoetsen Rekenen/Wiskunde bij de Niveautoets (N = 2315)

	Aantal items	GemSc	StDev	GemP	Alpha
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
Metten, Meetkunde, Tijd en Geld	15	9,9	2,8	66,0	0,63

De betrouwbaarheid van de subtoetsen is vergelijkbaar met die van de EB en EB-Anker. De alpha is relatief iets hoger bij taal bij de NT dan bij de EB(-Anker) en de alpha's zijn bij de rekentoetsen net iets lager. Heel veel scheelt dit echter niet. De toetsen zijn ook hier goed bruikbaar om groepen leerlingen met elkaar te vergelijken.

De gemiddelde P-waarden liggen, op die bij Woordenschat na, lager bij de NT. Aan de ene kant is dit opmerkelijk en aan andere kant ligt dit in de lijn der verwachtingen. De opgaven in de toetsen zijn aanzienlijk gemakkelijker dan die van de EB en EB-Anker. Daarmee zou verwacht kunnen worden dat de P-waarden hoger zouden zijn. Echter, de groep leerlingen is dusdanig minder vaardig dan de leerlingen die de EB of de EB-Anker) maken; dat zij ondanks dat verschil in moeilijkheid gemiddeld een lager percentage opgaven correct maken. Met klassieke testtheorie zijn de opgaven niet direct met elkaar te vergelijken, terwijl dat met behulp van IRT wel mogelijk is.

3.2 Toetseigenschappen van de toetsen in jaargroep 4

Nederlandse taal

Aan de leerlingen bij wie de toetsen voor de vaardigheden Spelling en Begrijpend lezen zijn afgenomen, is eerste de startmodule voorgelegd. Naar aanleiding van de score op die toets kregen de leerlingen een vervolgmodule voorgelegd. Die was bij een lagere score op de startmodule relatief gemakkelijk (Vervolg 1) en bij een hogere score juist relatief moeilijk (Vervolg 2). De metingen op zowel de startmodule als de vervolgmodule zijn gebruikt om de vaardigheid te bepalen. De toetsonderdelen samen vormen ze het meetinstrument om de vaardigheid te bepalen.

Tabel 3.2.1 Toetseigenschappen van de toetsen Spelling

	N	aantal items	GemSc	StDev	GemP	Alpha
Startmodule + Vervolg 1	696	50	27,5	6,3	55,0	0,75
Startmodule + Vervolg 2	1636	50	36,1	7,0	72,3	0,83

Tabel 3.2.2 Toetseigenschappen van de toetsen Begrijpend lezen

	N	aantal items	GemSc	StDev	GemP	Alpha
Startmodule + Vervolg 1	205	50	24,6	6,6	49,2	0,76
Startmodule + Vervolg 2	2147	50	35,2	7,4	70,5	0,85

Volgens het toetsbeoordelingssysteem (Evers, Van Vliet-Mulder & Groot, 2000) zijn al deze toetsen voldoende betrouwbaar voor metingen op individueel niveau en op groepsniveau. De startmodule samen met een vervolgmodule lijkt in beide gevallen moeilijker omdat de P-waarden lager zijn, maar de groepen die vervolgmodule 1 maken zijn ook duidelijk minder vaardig. Als alleen naar de startmodule gekeken wordt dan is het verschil in P-waarden ook duidelijk groter. Bij Spelling heeft de minder vaardige jaargroep 43,8% van de startmodule goed (gemiddelde score 10,9 op 25 opgaven) terwijl de vaardiger groep 77,4% goed heeft (gemiddelde score 19,4) op die toets. Bij de startmodule Begrijpend lezen is het verschil nog groter: 41,7% van de startmodule goed (gemiddelde score van 10,4 op 25 opgaven) bij leerlingen die vervolgmodule 1 maken, tegen 78,1% (gemiddelde score 19,5) bij de leerlingen die vervolgmodule 2 maken. De groepen die de verschillende toetsen voor dezelfde vaardigheid gemaakt hebben worden voor verdere analyses met behulp van IRT vergelijkbaar gemaakt. Voor de uiteindelijke analyses om de vaardigheid van de leerlingen te meten zijn ook de leerlingen meegenomen die alleen de startmodule gemaakt hebben, alleen een vervolgmodule gemaakt hebben of beide vervolgmodes gemaakt hebben, al dan niet met een startmodule vooraf. In totaal zijn de analyses voor Spelling gedaan op 2394 leerlingen en voor Begrijpend lezen op 2401 leerlingen.

IRT is ook noodzakelijk om de leerlingen te vergelijken die de toetsen Woordenschat gemaakt hebben. Deze leerlingen hebben meegedaan aan een proefonderzoek waarbij vijf verschillende toetsen zijn afgenomen. Deze zijn in verschillende combinaties afgenomen.

Tabel 3.2.3 Toetseigenschappen van de proeftoetsen Woordenschat

	N	aantal items	GemSc	StDev	GemP	Alpha
Combinatie A [toets 1 en 2]	194	60	54,6	4,8	90,9	0,82
Combinatie B [toets 2 en 3]	215	60	53,7	6,6	89,4	0,89
Combinatie C [toets 4 en 5]	171	60	47,7	7,6	79,5	0,86
Combinatie D [toets 1 en 4]	147	60	48,6	8,0	81,1	0,88
Combinatie E [toets 5]	18	30	26,3	3,9	87,6	0,82

De eerste vier combinaties bestonden uit twee toetsen van ieder 30 opgaven. De laatste set was geen combinatie, maar betrof één proeftoets.

Het aantal waarnemingen per opgave varieert van 189 bij toets 5 (in combinatie C en E) tot 409 bij toets 2 (combinatie A en B). Alle toetsen zien er relatief gemakkelijk uit: bijna 80% tot ruim 90% van de opgaven in de toetsen wordt goed gemaakt. Deze percentages zouden goed vergelijkbaar kunnen zijn aangezien niet hoeft te worden aangenomen dat de groepen die de verschillende toetsen gemaakt hebben duidelijk in vaardigheid verschillen. De betrouwbaarheden zijn hoog genoeg voor individuele vergelijkingen. In totaal zijn de gegevens van 745 leerlingen gebruikt voor de analyses.

Rekenen

De vier verschillende vaardigheden rekenen zijn allen met één toets bepaald, te weten de LOVS-rekentoets toets voor einde leerjaar 4 (E4). Deze toets bestaat uit 54 opgaven. Dit levert vier relatief korte subtoetsen op.

Tabel 3.2.4 Toetseigenschappen van de toetsen Rekenen (startmodule subschalen; N = 2407)

	Aantal items	GemSc	StDev	GemP	Alpha
Getallen en Getalrelaties	13	10,2	2,3	78,8	0,66
Optellen en Aftrekken	12	8,3	2,8	69,5	0,76
Vermenigvuldigen en Delen	14	9,5	3,5	68,2	0,83
Metten, Tijd en Geld	15	11,0	3,2	73,0	0,77

Ondanks het feit dat deze toetsen aan de korte kant zijn, zijn betrouwbaarheden van de subtoetsen goed te noemen voor de vergelijking van groepen, op die voor Getallen en Getalrelaties na. De betrouwbaarheid van die subtoets krijgt de kwalificatie voldoende.

De toetsen zijn niet moeilijk te noemen voor de leerling aan het einde van jaargroep 4, met gemiddeld 72% van de opgaven correct beantwoord. Vooral de opgaven voor Getallen en Getalrelaties zijn aan de makkelijke kant. Voor de vergelijking van de leerlingen is IRT hier niet noodzakelijk, maar met behulp van IRT zijn wel de opgaven en de populaties direct met elkaar te vergelijken. Daarover meer in het volgende hoofdstuk.

4 Analyseprocedures: opgave- en groepseigenschappen op één schaal

In dit hoofdstuk wordt een beknopte technische beschrijving van de analyses gegeven. Na een introductie van het in dit onderzoek gebruikte *item response theorie* (IRT) model, volgt een uitleg over de presentatie van de resultaten. Hierin wordt uiteengezet hoe groepen en opgaven direct met elkaar vergeleken kunnen worden. Tot slot wordt beschreven hoe de verschillen tussen de groepen geanalyseerd kunnen worden, en hoe verschillende resultaten geïnterpreteerd kunnen worden. In de presentatie van de resultaten wordt hier gebruik van gemaakt. In hoofdstuk 5 wordt inhoudelijk ingegaan op waar een leerling uit jaargroep 8 meer of minder moeite mee heeft, en hoe de verschillende groepen binnen dat leerjaar van elkaar verschillen. De resultaten bij leerjaar 4 worden op eenzelfde wijze gepresenteerd in hoofdstuk 6, zij het dat deze presentatie een meer summier karakter heeft.

4.1 Toepassing van een IRT-model

Om verschillende toetsen en de groepen leerlingen die verschillende toetsen maken met elkaar te kunnen vergelijken is IRT een nuttige techniek. In deze studie hebben we gebruikgemaakt van het *One-Parameter Logistic Model* (OPLM) en het bijbehorende computerprogramma (Verhelst, Glas & Verstralen, 1995). Er is binnen dat programma gebruik gemaakt van CML-schattingsprocedures (Verhelst & Eggen, 1989). Daarbij is het noodzakelijk dat toetsen overlap hebben. De EB en de EB-Anker hebben minstens de helft van de opgaven overeenkomstig, maar er is geen enkele van die opgaven die daarvan in de NT voorkomt. Echter, door een *design* van proeftoetsen is het toch mogelijk deze toetsen op een en dezelfde schaal te beschouwen. In eerdere analyses waren de eigenschappen van de opgaven van de EB en de NT gegeven met behulp van itemparameters. Deze itemparameters waren bepaald op een steekproef van leerlingen. In deze studie zijn die parameters getoetst op de gehele dataset.

De gehele dataset bestaat uit een zo groot aantal leerlingen dat een directe analyse met alle leerlingen niet mogelijk was: het programma OPLM kan geen dataset aan waarbij er meer dan 32.000 observaties zijn per opgave. Om die reden is de gehele dataset met behulp van een *random* procedure verdeeld in vijf niet-overlappende groepen van ongeveer gelijke grootte. Al deze groepen waren vergelijkbaar voor wat betreft de verdeling van de achtergrondvariabelen en voor wat betreft de verdeling van de leerlingen over de drie toetsen. Deze vijf datasets zijn allen 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 statische *power* zeer groot. Een statische passing is in een dergelijk geval moeilijk te verkrijgen. We kunnen ook kijken naar de grootte van de verschillen van de geschatte en de geobserveerde beschrijving van de opgaven. Dat wordt 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 dan mogelijk wel significant kunnen zijn, maar dusdanig klein zijn 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. Bij rekenen zijn een aantal opgaven verwijderd (zie hoofdstuk 5), maar bij de taalvaardigheden geen enkele. Bij de unieke opgaven van de EB-Anker- en de NT-opgaven waren de verschillen tussen de geschatte en de geobserveerde psychometrische beschrijvingen van de opgaven vrijwel geheel afwezig. Wat betreft de vergelijking van de opgaven op de vaardigheidsschaal waren er binnen een vaardigheid geen noemenswaardige verschillen tussen de vijf opdelingen van de gehele dataset.

De analyses van de resultaten voor leerjaar 4 zijn in grote lijnen vergelijkbaar met die gedaan zijn in leerjaar 8. De aantallen leerlingen waren echter aanzienlijk lager dan het geval was bij de analyses van jaargroep 8. Hier is geen gebruikgemaakt van eerdere schattingen, maar zijn de parameters voor deze set opnieuw uitgerekend om deze dataset zo goed mogelijk te beschrijven. Voor de opgaven uit de proeftoets waren er geen eerdere parameterschattingen beschikbaar.

4.2 Rapportage van de opgaven en populatie

De representatie van de opgaven op de schaal is op eenzelfde manier gedaan als gebruikelijk is bij PPON-onderzoek. Dat betekent dat ervoor gekozen is deze schaal te definiëren door de gemiddelde vaardigheid van de gehele groep en standaarddeviatie van deze groep. De verdeling van de leerlingen per vaardigheid en per onderzoekpopulatie wordt zo getransformeerd dat het gemiddelde gelijk is aan 250 en de standaarddeviatie aan 50. De aanname is dat de vaardigheid zoals die met de schaal gemeten wordt, bij benadering normaal verdeeld is in de populatie. Net zoals in PPON gebruikelijk is, gebeurt dat voor alle vaardigheden, en voor zowel jaargroep 4 als voor jaargroep 8. De opgaven en verschillende te onderscheiden groepen binnen de vaardigheid en onderzoekpopulatie zijn op één schaal geplaatst. Met behulp van de vaardigheidsverdeling kunnen ook de relatieve standaarden beschreven worden. Op deze schaal is in dit rapport speciale aandacht voor de volgende percentielpunten:

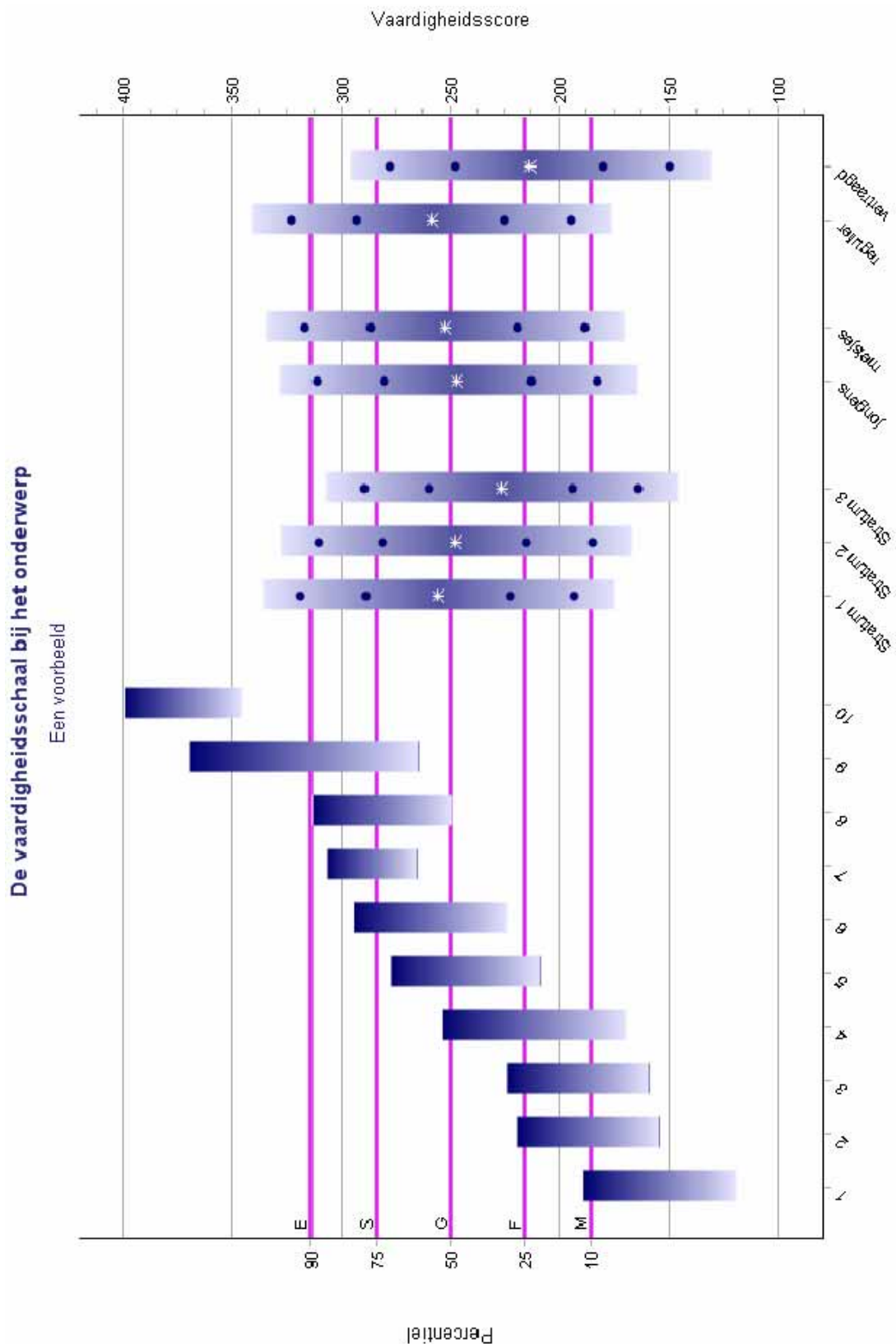
- P10 **Percentiel 10-leerling:** Het niveau van een leerling waarbij 10% van de leerlingen een gelijke of lagere vaardigheid heeft en 90% van de leerlingen een hogere vaardigheid;
- P25 **Percentiel 25-leerling:** Het niveau van een leerling waarbij 25% van de leerlingen een gelijke of lagere vaardigheid heeft en 75% van de leerlingen een hogere vaardigheid;
- P50 **Percentiel 50-leerling:** Het niveau van een leerling waarbij 50% van de leerlingen een gelijke of lagere vaardigheid heeft en 50% van de leerlingen een hogere vaardigheid;
- P75 **Percentiel 75-leerling:** Het niveau van een leerling waarbij 75% van de leerlingen een gelijke of lagere vaardigheid heeft en 25% van de leerlingen een hogere vaardigheid;
- P90 **Percentiel 90 leerling:** Het niveau van een leerling waarbij 90% van de leerlingen een gelijke of lagere vaardigheid heeft en 10% van de leerlingen een hogere vaardigheid.

Deze percentielpunten worden ook vaak in verband gebracht met referentieniveaus. Bij een vaardigheid op P25 in jaargroep 8 wordt wel gerefereerd aan de basiskwaliteit, dat ook wel het *fundamentele* niveau van die vaardigheid genoemd wordt. Dit punt wordt zowel voor taal als voor rekenen gehanteerd (Expertgroep Doorlopende Leerlijnen, 2008). Bij taal wordt in die zelfde context de vaardigheid op het niveau van een P75 leerling het *streefniveau* genoemd. Dat is het niveau waarbij de streefkwaliteit gerealiseerd wordt. Bij rekenen wordt dit niveau gerealiseerd op het P50-niveau. In de normale verdeling is dit punt gelijk aan het gemiddelde punt van de verdeling. Het P50-punt op de vaardigheidsverdeling geeft zodoende het *gemiddelde* niveau in jaargroep 8 aan. Ook de extremere percentielpunten zouden als niveaupunten benoemd kunnen worden. De P90-vaardigheid zou bijvoorbeeld het *expertniveau* genoemd kunnen worden en het niveau van een percentiel 10-leerling een *minimumniveau*. Merk op dat deze kwalificatie van de percentielpunten hier geheel gerelateerd is aan de geobserveerde eigenschappen van de vaardigheidsverdeling en niet aan de didactische interpretatie van de vaardigheidsschaal. Die interpretatie zal moeten plaatsvinden met behulp van de opgaven die op die schaal geplaatst worden. Als die relatie onderzocht wordt, kunnen uitspraken gedaan worden over of bijvoorbeeld het fundamentele niveau ook didactisch als fundamenteel beschouwd kan worden.

De percentielpunten liggen in deze studie vast doordat gesteld is dat de verdeling van de gehele groep zodanig getransformeerd wordt dat het gemiddelde op 250 ligt op de vaardigheidsschaal en de standaarddeviatie gelijk aan 50 is. Hierdoor weten we dat de positie op de vaardigheidsschaal van P10 gelijk is aan het punt 186 op de vaardigheidsschaal. De waarde van P25 ligt op 216, van P50 uiteraard op 250, van P75 op 284 en van P90 op 314. In de resultaten worden deze punten geïllustreerd door middel van de opgaven om inhoudelijk inzicht te krijgen in het vaardigheidsniveau.

In hoofdstuk 5 en 6 beschrijven we per onderwerp de resultaten van de leerlingen. De voorbeeldopgaven worden afgebeeld in een diagram en worden ook beschreven voor de vijf genoemde percentielpunten. In het diagram zijn behalve de percentielpunten ook de verdelingen gegeven van verschillende groepen. Enerzijds wordt het diagram daardoor complex, anderzijds illustreert het de samenhang tussen de verschillende resultaten. In figuur 4.1 staat een voorbeeld van een dergelijk diagram. Voor iedere vaardigheid worden in hoofdstuk 5 en 6 vergelijkbare diagrammen gegeven. Op de volgende pagina's geven we een toelichting op figuur 4.1.

Figuur 4.1 Een voorbeeld van een vaardigheidsschaal



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 figuur heeft zodoende een bereik van drie standaardafwijkingen boven en drie onder het

gemiddelde van 250. De bredere horizontale lijnen geven de vijf onderscheiden percentielniveaus aan: P10, P25, P50, P75 en P90. In dit figuur is aan de percentiepunten ook een interpretatie meegeven, respectievelijk een minimum niveau (M), een fundamenteel niveau (F), een gemiddeld niveau (G) een streefniveau (S) en een expertniveau (E). In de figuren in hoofdstuk 5 en 6 worden die interpretaties van de vijf percentiepunten niet meer vermeld.

Op deze horizontale lijnen liggen twee typen kolommen. De kolommen aan de linkerzijde representeren de opgaven. Deze kolommen zijn onder aan het diagram genummerd en verkleuren van boven naar beneden van licht naar donker. De moeilijkheid van de opgaven is uit deze kolommen af te leiden. De kolommen aan de rechterzijde representeren de verschillen de groepen. Deze kolommen hebben onder aan het diagram een label, hebben in het middel een asterisk en hebben zowel boven als onder die asterisk twee puntjes. In het voorbeeld diagram zijn de opgaven genummerd van 1 tot en met 10, en zijn de verdelingen van de drie strata gegeven, evenals die van de jongens en meisjes en die van de vertraagde en reguliere leerlingen.

De moeilijkheidsgraad van de opgaven

Een bekende manier om de moeilijkheidsgraad van een opgave aan te geven, is de zogenoemde p-waarde. Een p-waarde van 0,80 betekent dat 80% van de leerlingen die opgave correct heeft beantwoord. Een opgave met een p-waarde van 0,50 is moeilijker, omdat nu slechts de helft van de leerlingen de opgave juist heeft gemaakt.

Een opgave is echter niet voor alle leerlingen even moeilijk te maken. Over het algemeen zal gelden dat naarmate een leerling een onderwerp beter beheerst, hij of zij een grotere kans heeft om een opgave over dat onderwerp goed te beantwoorden. Die relatie wordt voor een aantal opgaven afgebeeld in de linkerkolom van de figuur met verticale balkjes. Het verticale balkje begint op het punt op de vaardigheidsverdeling waarop de kans om die opgave goed te maken een half is. Dat is het onderste (lichtste) punt van bij iedere opgave-kolom in figuur 4.1. Leerlingen op dit vaardigheidsniveau zullen gemiddeld vijf van de tien opgaven van precies dit type 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 0,80 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 0,80 te hebben om de opgave correct te maken altijd boven dat om een kans van 0,50 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.

Een opgave kan op een bepaald percentiepunt als *moeilijk* geclassificeerd worden als de kans op het juist maken van deze opgave bij die vaardigheid onder de 0,50 ligt. Een opgave wordt als *beheerst* beschouwd als een leerling van een bepaald vaardigheidsniveau een kans heeft van 0,80 of groter. De opgaven zijn geordend op de grenspunten tussen een matige en goede beheersing, dus op basis van het benodigde vaardigheidsniveau om een kans van 0,80 te hebben om de opgave goed te beantwoorden.

Laten we ter verdere illustratie opgave 8 uit figuur 4.1 nemen. 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 percentielijnen 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 percentiepunt 90 (expertniveau) wordt deze opgave goed beheerst: vanaf het vaardigheidsniveau 314 wordt de kans om de opgave goed te beantwoorden groter dan 0,80. Opgemerkt moet worden dat hiermee niet automatisch bepaald is dat het als problematisch gezien moet worden dat deze opgave niet beheerst wordt door een grote groep leerlingen. Dat is een inhoudelijke discussie die alleen beantwoord kan worden door naar de inhoud van de opgaven te kijken.

Een voorbeeld van een makkelijker opgave is opgave 2. Die wordt goed beheerst door leerlingen op een percentiepunt 25 (fundamenteel niveau). Leerlingen op P10-niveau hebben een matige beheersing van deze opgave. Alleen leerlingen met een extreem lage vaardigheid van onder de 155, wat ongeveer twee standaarddeviaties onder het gemiddelde is, beheersen deze opgave onvoldoende

De afgebeelde opgaven in hoofdstuk 5 vormen een selectie van alle opgaven op de schaal en zijn met zorg gekozen. Zij vormen enerzijds een goede afspiegeling van de inhoudelijke aspecten die met de opgaven worden gemeten. Anderzijds bestrijken zij een groot bereik op de vaardigheidsschaal, dat wil zeggen dat zij een goed beeld geven van de spreiding van de moeilijkheidsgraad van de opgaven over de gehele schaal. De afgebeelde opgaven zijn allen afkomstig uit de EB. De opgaven uit de NT zijn niet afgebeeld omdat dit computeropgaven zijn, die niet als zodanig afgebeeld kunnen worden. Deze opgaven zijn allen aan de makkelijke kant en kunnen ook gerepresenteerd worden door makkelijke opgaven uit de EB. De opgaven uit de EB-Ankertoets zijn niet gerepresenteerd omdat deze opgaven niet openbaar gemaakt mogen worden.

De vaardigheidsverdelingen van groepen leerlingen

In het rechtergedeelte van de figuur zijn de vaardigheidsverdelingen van verschillende groepen leerlingen afgebeeld. In deze figuur betreft het de vergelijking tussen leerlingen afkomstig van scholen uit verschillende strata, geslacht en leertijd. Voor iedere onderscheiden groep leerlingen wordt de geschatte vaardigheidsverdeling afgebeeld. Bij deze vaardigheidsverdelingen is niet gecorrigeerd voor andere factoren die mogelijk van invloed zijn op de resultaten. De wijze van afbeelding laat een vergelijking toe tussen de prestaties van de leerlingen wat betreft de variabelen:

- stratum, met de niveaus stratum1, stratum 2 en stratum 3;
- geslacht, met de niveaus jongen en meisje; en;
- leertijd, met de niveaus regulier en vertraagd.

We onderscheiden voor iedere groep leerlingen vijf percentiepunten op de vaardigheidsschaal. De gemiddelde vaardigheidsscore van een groep (percentiel 50) is met een wit sterretje aangeduid. De punten geven, van boven naar beneden percentiepunten 10, 25, 75 en 90 aan binnen die subgroepverdeling. De onderkant van de kolom wordt gegeven bij percentiel 5, en de bovenkant bij percentiel 95 van die groep.

In dit geval leert figuur 4.1 ons bijvoorbeeld dat de gemiddelde vaardigheidsscore van leerlingen die nog nooit zijn blijven zitten (regulier) een gemiddeld vaardigheidsniveau hebben van de 259, terwijl leerlingen die (minstens) één klas opnieuw hebben gedaan, of op oudere leeftijd groep 3 zijn ingestroomd, (vertraagd) een gemiddelde vaardigheid hebben van 214. Merk op dat het gemiddelde van deze twee getallen geen 250 is. Als gewogen wordt naar groeps grootte is het gemiddelde wel gelijk aan 250.

De verschillen in vaardigheidsniveaus tussen de onderscheiden groepen leerlingen kunnen vervolgens inhoudelijke betekenis krijgen aan de hand van de gestelde niveaus en de voorbeeldopgaven. Een gemiddelde reguliere leerling zit op een niveau dat wat boven het algemene gemiddelde ligt (P50). Een gemiddelde vertraagde leerling zit rond het P25-niveau. In dit voorbeeld zitten iets minder dan 25% van de vertraagde leerlingen op of boven het algemeen gemiddelde niveau van 250: het eerste puntje boven de asterisk zit net iets onder het P50-niveau. Een gemiddelde reguliere leerling beheerst de eerste vier opgaven goed, en beheerst drie opgaven onvoldoende (opgaven 7, 9 en 10). Voor een gemiddelde vertraagde leerling geldt dat deze in dit voorbeeld alleen de eerste opgave echt goed beheerst (opgaven 2 en 3 bijna) en vijf opgaven onvoldoende beheerst. Op een vergelijkbare manier illustreert de afbeelding ook de verschillen tussen de leerlingen afkomstig van verschillende strata en tussen jongens en meisjes.

In het geval van de resultaten voor jaargroep 8 worden ook de verdelingen van de leerlingen met verschillende formatiegewichten afgebeeld (op drie niveaus: 1.00, 1.25 en 1.90). Voor de leerlingen in

jaargroep 8 zijn naast deze vier typen verdelingen ook de verdelingen van andere achtergrondgegevens bepaald. Om de figuren overzichtelijk te houden zijn die niet afgebeeld, maar worden de waarden hoofdstuk 5 en 6 in een tabel gegeven. Op basis daarvan kunnen vergelijkbare conclusies getrokken worden als op basis van de figuren. Naast de gegevens over stratum, formatiegewicht, leertijd en geslacht staan daar ook de gevonden gemiddelden en de standaarddeviaties in van de leerlingen die de verschillende toetsen hebben gemaakt (op drie niveaus: EB, EB-Anker en NT). Ook de verdelingen op de andere variabelen die genoemd zijn komen hierbij aan bod: Doorstroom naar VO (de vijf hoofdniveaus: BB, KB, GT, havo en vwo), regio (vier niveaus: noord, oost, west en zuid), urbanisatiegraad (vijf niveaus: zeer sterk, sterk, matig, weinig en niet stedelijk) en thuistaal (drie hoofdniveaus: alleen Nederlands, Nederlands plus een andere taal en alleen een andere taal, en de vier extra niveaus)

Merk op dat bij Thuistaal de categorieën met een uitsplitsing naar specifieke andere talen dan het Nederlands het om dezelfde leerlingen gaat als in de categorie *Nederlands plus een andere taal*, dan wel *Alleen een andere taal* vallen. Merk ook op dat de er ook gegevens bekend zijn voor leerlingen die tussen twee categorieën in het VO in liggen. Voor de overzichtelijkheid worden hier alleen de hoofdcategorieën gegeven.

Naast de gemiddelden op de vaardigheidsverdelingen van de groepen worden ook de gemiddelden van die groepen gegeven op de percentielschaal. De percentielwaarde die gegeven wordt, is het percentielpunt in de gehele verdeling van de leerlingen in de onderzoekspopulaties. Daarmee is het gemiddelde van de groep vergelijkbaar met wat de leerlingen kunnen zoals dat gepresenteerd is bij de beschrijving van wat een P10-, een P25-, een P50-, een P75- en een P90-leerling kan.

4.3 Rapportage van verschillen tussen groepen

De verschillen in prestaties tussen groepen leerlingen kunnen zoals hierboven aangegeven gerapporteerd worden op basis van de geschatte vaardigheidsverdelingen van deze groepen. Deze verdelingen laten de verschillen tussen groepen zien zonder dat we ons afvragen of die groepen naar samenstelling op andere relevante variabelen wel vergelijkbaar zijn. Het is echter ook mogelijk te onderzoeken wat de specifieke bijdrage is van een aantal variabelen op verschillen in prestaties tussen leerlingen. Het betreft dan het differentiële effect van categorieën binnen de verschillende variabelen. We spreken van gezuiverde, of gecorrigeerde effecten omdat de andere kenmerken van de leerlingen, voor zover die ons bekend zijn, constant worden gehouden.

Het verschil in vaardigheid tussen de groepen wordt statistisch getoetst op significantie. Deze toetsing geeft echter geen informatie over de grootte van het verschil. En zeker bij grote aantallen subjecten, zoals in dit peilingsonderzoek, kunnen relatief kleine verschillen al gauw een significant effect genereren. We rapporteren de verschillen daarom in termen van effectgrootten. De effectgrootte is het quotiënt van het verschil tussen de gemiddelden en de standaardafwijking van de twee groepen die onderling worden vergeleken. Voor de interpretatie van de effectgrootten volgen we in de literatuur gebruikelijke kwalificaties.

Tabel 4.1 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

Er is in dit onderzoek een aantal verschillende modellen onderzocht voor jaargroep 8. Het eerste onderzochte model is dat waarin de hoofdvariabelen geslacht, leertijd, stratum en formatiegewicht zijn

opgenomen. Het effect van ieder van deze variabelen is onderzocht, gecorrigeerd voor de overige variabelen. Daarna is een aantal uitbreidingen dit model onderzocht met telkens vijf in plaats van vier variabelen, doordat er telkens een andere variabele werd toegevoegd. De toegevoegde variabelen waren dezelfde als waarvoor de gemiddelden waren uitgerekend: gemaakte toets, doorstroomgegevens, regio, urbanisatiegraad en thuistaal.

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.

Merk op dat van de overige variabelen niet alle effecten zijn afgebeeld. Zo is nergens het effect van EB-Anker gegeven. Die ontbreekt omdat nergens de resultaten tussen EB en EB-Anker significant verschillen en het effect van EB-Anker en de NT vergelijkbaar was met dat van EB en NT. Bij Doorstroomgegevens zijn alleen de effecten gegeven van de opeenvolgende hoofdniveaus. 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 opeenvolgende niveaus van deze variabele. Wanneer bij de variabele thuistaal de effecten gegeven worden per specifieke thuistaal, dan worden die alleen gegeven in vergelijking met de leerlingen die alleen Nederlands spreken. De effecten tussen de andere talen onderling zijn daaruit af te leiden.

In de rapportage van de resultaten wordt naast het gevonden effect ook direct een kwalificatie van het effect gegeven. Hierbij worden kwalificaties zoals gegeven in tabel 4.1 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 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 $p < .01$ allen significant zijn. Voor de effecten die als matig of groot gekwalificeerd kunnen worden, geldt dat zij bij een overschrijdingskans $p < .000001$ allen significant zijn.

4.4 Presentatie van de resultaten in leerjaar 4

De presentatie van de resultaten van de leerlingen in jaargroep 4 komt in grote lijnen overeen met hoe de resultaten voor jaargroep 8 weergegeven worden. Zo worden vergelijkbare figuren gegeven zoals figuur 4.1, worden de verdelingen van de verschillende groepen gegeven, en worden de gezuiverde effecten gepresenteerd. Er zijn echter ook wat verschillen met jaargroep 8.

De verslaglegging van de zeven verschillende vaardigheden die gemeten zijn in jaargroep 4 is minder uitgebreid dan die van jaargroep 8. Voor het overzicht van de opgaven, verdelingen van de te onderscheiden groepen binnen jaargroep 4 en de geschatte effecten worden de figuren en tabellen gegeven. In plaats van voorbeeldopgaven worden in de figuren alle opgaven weergegeven. Voor de inhoud van de opgaven wordt gerefereerd aan de LOVS-toetsen voor E4. Alleen voor de woordenschattoets zijn proeftoetsen gebruikt.

De vaardigheden zijn voor leerjaar 4 net zoals bij leerjaar 8 op een schaal gezet waarbij het gemiddelde van de schaal gelijkgesteld is aan 250, en de standaarddeviatie gelijk aan 50. Dat maakt de gevonden schaalwaarden gemakkelijk interpreteerbaar in verhouding tot de populatie. De overeenkomst tussen beide waarden is dat ze het gemiddelde representeren binnen de groepen, maar de inhoudelijke vaardigheid van

de leerlingen in jaargroep 8 ligt bij 250 uiteraard hoger dan in jaargroep 4 bij 250. Om het verschil tussen beide waarden aan te geven worden de waarden voor jaargroep 4 cursief gegeven. Het alternatief is om een andere schaaltransformatie te nemen, bijvoorbeeld een waarbij het gemiddelde van leerjaar 4 bij 100 zou liggen. Dat zou echter de indruk wekken dat hiermee ook de inhoudelijke afstand tussen de vaardigheden bij leerjaar 4 en leerjaar 8 bekend zou zijn. Ook zou dan de indruk gewekt kunnen worden dat deze afstand over alle vaardigheden identiek zou zijn. Door het gemiddelde van beide populaties gelijk aan 250 te stellen wordt die suggestie niet gewekt, omdat evident is dat de inhoudelijke vaardigheden in de verschillende groepen niet gelijk zijn. Wat de afstand op de vaardigheidsschaal is tussen deze twee leerjaren is niet bekend in deze analyses.

Dat deze twee onderzoekspopulaties beiden op 250 gezet zijn, is in zeker mate vergelijkbaar met de situatie voor Woordenschat, Spelling en rekenen, meten, tijd en geld, waarbij de gemiddelden allen ook gelijk zijn. Ook daarbij is evident dat deze 250 niet exact dezelfde inhoudelijke vaardigheid 250 betreft. Een vergelijking tussen jaargroep 4 en jaargroep 8 is wel mogelijk door de gezuiverde effecten van de achtergrondvariabelen met elkaar te vergelijken op de verschillende schalen in beide onderzoekspopulaties. Die vergelijking wordt gegeven in hoofdstuk 7.

Een ander verschil tussen jaargroep 4 en jaargroep 8 is dat het aantal variabelen dat in het onderzoek is opgenomen, verschilt. De variabelen leertijd, stratum, geslacht en regio zijn hetzelfde gebleven, maar de variabelen 'toets: EB versus NT' en 'doorstroom naar voortgezet onderwijs (VO)' zitten uiteraard niet in dit onderzoek. De variabele thuistaal is in dit onderzoek ook iets anders gedefinieerd dan in jaargroep 8. Doordat de aantallen leerlingen aanzienlijk lager liggen dan bij het onderzoek naar jaargroep 8, wordt hier niet het verschil gemaakt tussen de diverse specifieke talen. Bij thuistaal is ook een niveau voor streektaal opgenomen. De aantallen leerlingen voor de drie niveaus anders dan het Nederlands, te weten: streektaal, een buitenlandse taal en zowel Nederlands als een buitenlandse taal, zijn aan de lage kant.

Een ander belangrijk verschil met jaargroep 8 is dat de variabele formatiegewicht voor de meeste leerlingen in jaargroep 4 een nieuwe waarde behoort te krijgen door de veranderde leerlinggewichten. De nieuwe gewichten 0.0, 0.3 en 1.2 hebben iets andere definities dan de oude gewichten 1.00, 1.25 en 1.90. 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; hieronder de versimpelde definities daarvan):

- Categorie 1: Ouder heeft maximaal (speciaal) basisonderwijs gehad.
- Categorie 2: 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: Ouder heeft meer dan twee jaar mavo, havo, vwo dan wel vmbo gemengde of theoretische leerweg gevolgd.

Een leerling krijgt nieuw gewicht 0.0 als deze (minstens) één ouder heeft uit categorie 3. Als de leerling *niet* gewicht 0.0 heeft krijgt deze een gewicht van 1.2 als er (minstens) één ouder in categorie 1 zit. De overige leerlingen krijgen gewicht 0.3

In het verleden zijn door scholen nog wel oude gewichten gebruikt, wanneer de nieuwe gewichten gebruikt zouden moeten worden. In de gegevens over de leerlingen uit de steekproef zijn door sommigen ook nog oude gewichten gebruikt. In een aantal gevallen was dat juist omdat deze leerlingen vertraagd waren en daardoor nog tot de leeftijdscategorie behoorden voor de oude gewichten. Ook is gevonden dat op sommige scholen ook voor die oudere kinderen al de nieuwe gewichten stonden. Het gevolg is dat in de beschikbare data de oude en de nieuwe gewichten door elkaar gebruikt worden. Het aantal leerlingen per leerlinggewicht 0.3, 1.2, 1.25 en 1.90 varieerde van 75 tot 90 binnen leerjaar 4. Het aantal leerlingen met gewicht 0.0 dan wel 1.00 was aanzienlijk groter.

Er zijn nu analyses uitgevoerd waarbij de leerlingen met gewicht 0.0 en 1.00 als een groep genomen zijn, 0.30 en 1.25 als een andere groep en 1.2 en 1.90 als een derde groep. Deze samenvoegingen zijn gebaseerd op de gegevens uit het PRIMA-onderzoek waarmee een was-wordt-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 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, niet significant of beide. De gevonden tendensen zullen wel vermeld worden bij de resultaten, maar vanwege de lage aantallen hebben deze alleen maar een indicatief karakter. Wat betreft de effecten op de overige variabelen maakte het geen verschil of met de samengevoegde formatiegewichten gewerkt werd, dan wel met de oude en nieuwe gewichten naast elkaar. Bij de resultaten worden de verdelingen en effecten van de samengestelde formatiegewichten gepresenteerd in de figuren en tabellen.

Voor de vaardigheid Woordenschat waren minder variabelen beschikbaar dan voor de andere vaardigheden die gemeten zijn in leerjaar 4. Er waren hier geen gegevens voor thuistaal of formatiegewicht. Doordat deze gegevens niet beschikbaar waren, kon er niet voor geconditioneerd worden. Hierdoor zijn de effecten voor stratum op die schaal niet direct te vergelijken met de effecten voor stratum bij de andere schalen.

Een ander verschil met de resultaten voor jaargroep 8 is dat een tabel met classificaties van opgaven wordt gegeven. Hierbij wordt per toets, per vaardigheid en per percentielniveau P10, P25, P50, P75 en P90 aangegeven wat het aantal opgaven is dat (te) moeilijk is, dan wel beheerst wordt. Een opgave is (te) moeilijk voor een percentielniveau als een leerling van dat niveau een kans kleiner dan een half heeft om de opgave goed te maken ($P < 0,50$). Een opgave wordt beheerst als een leerling op een specifiek percentielniveau een kans groter dan 0,80 heeft op een goed antwoord. Deze classificatie is voor jaargroep 8 in hoofdstuk 5 telkens in woorden gegeven.

In de tabellen met de effecten worden ook significantieniveaus gegeven. Bij de gebruikte significantieniveaus zijn de overschrijdingskansen $p < .05$, $p < .01$ en $p < .000001$ gebruikt. Deze laatste twee overschrijdingskansen waren ook gebruikt bij jaargroep 8. Het aantal waarnemingen is niet, zoals bij leerjaar 8, zo groot dat als er een effect van enig formaat gevonden wordt, dat ook meteen significant is. Als een effect niet *significant* groter dan 0 is, dan is het niet echt relevant of het een groot of een klein effect betreft.

Het laatste verschil betreft de ordening in de figuren van de opgaven op de vaardigheidsschaal. Voor jaargroep 8 waren deze geordend naar de benodigde vaardigheid om een kans te hebben van 0,80 om de opgave correct te maken. Voor jaargroep 4 zijn de opgaven per vaardigheid geordend naar volgorde van afname binnen een toets. De reden daarvoor is dat daarmee gemakkelijk de inhoudelijk bijbehorende opgaven gevonden kunnen worden in de toets.

5 Resultaten van leerlingen in jaargroep 8

In dit hoofdstuk worden voor de zes verschillende vaardigheden de resultaten in jaargroep 8 gegeven. Eerst wordt iedere vaardigheid op een schaal geplaatst. De relatie tussen de populatieverdelingen en de opgaven wordt gegeven, evenals de relatie tussen de opgaven en de standaarden. Ten slotte worden per vaardigheid de gecorrigeerde effecten gegeven om inzicht te krijgen in de relatie van de leerling- en schoolkenmerken en de vaardigheden.

5.1 Nederlandse taal: Woordenschat

De vaardigheid

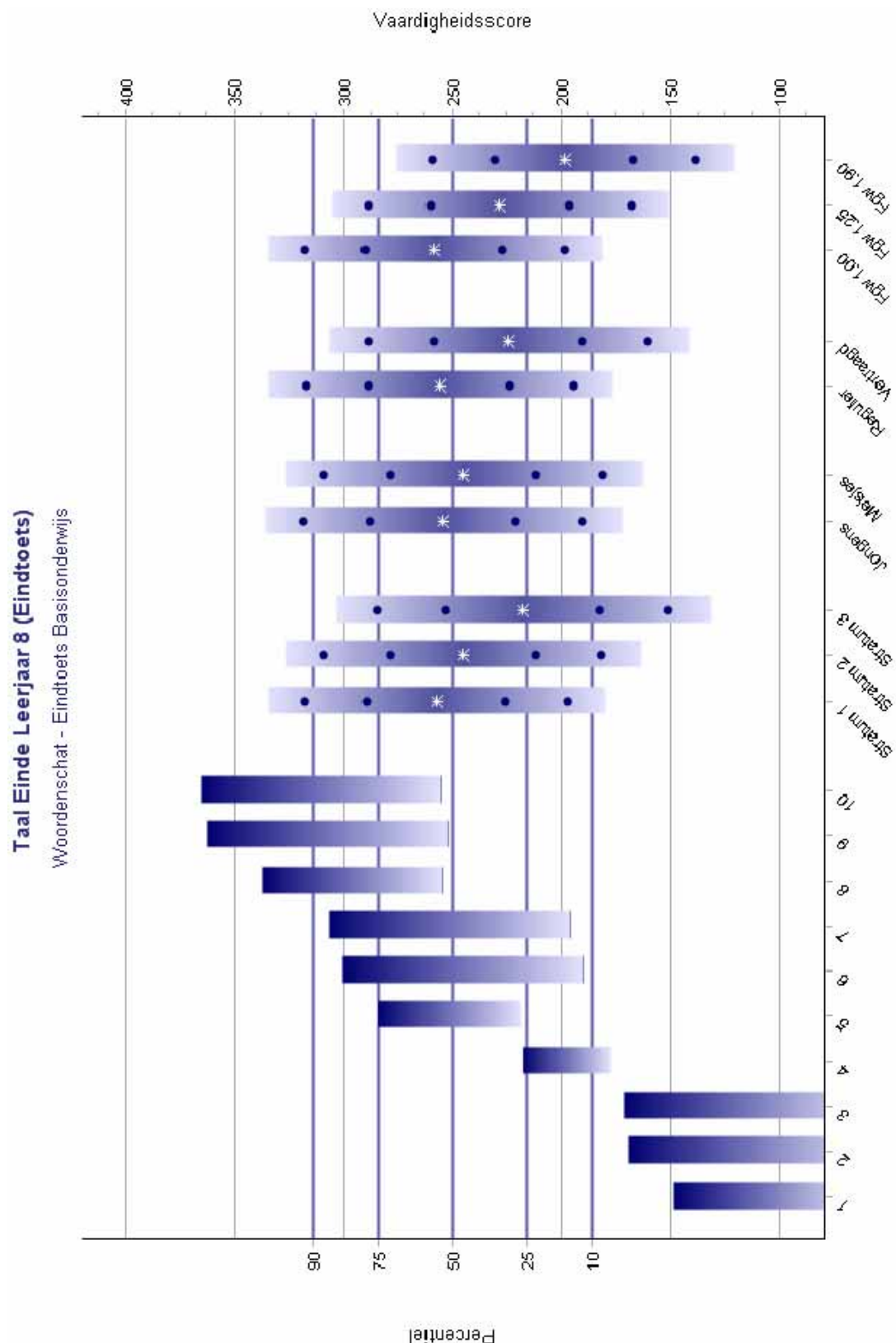
Woorden vormen de bouwstenen van onze taal en maken allerlei soorten van kennis toegankelijk. Zo vervult het hebben van een uitgebreide woordenschat een sleutelfunctie bij het verwerven van kennis via teksten. Goed begrijpend kunnen lezen hangt voor een groot deel af van de breedte en de diepte van de 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 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 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 vaardigheid Woordenschat is gemeten met in totaal 47 verschillende opgaven. Daarvan waren er 20 die alleen in de Niveautoets (NT) voorkwamen en 27 opgaven die of in de Eindtoets Basisonderwijs (EB) of in de EB-Anker voorkwamen. In de EB zaten 20 opgaven, waarvan er 13 ook in de EB-anker zaten. Die toets was aangevuld met 7 nieuwe opgaven die Woordenschat meten. Van de 20 EB-opgaven hadden 6 opgaven 5 antwoordcategorieën; de overige opgaven waren vierkeuze opgaven. De opgaven met 5 antwoordalternatieven zijn iets moeilijker dan de opgaven met 4 alternatieven, maar het verschil is gering.

Er is wel een duidelijk onderscheid tussen de opgaven uit de toetsen EB en de NT. De opgaven in de NT zijn duidelijk gemakkelijker dan die uit de EB. Alle opgaven uit de NT zijn zelfs voor een leerling op het P10-niveau niet moeilijk: op alle opgaven heeft een P10-leerling een kans groter dan 50% om deze goed te beantwoorden. Wat betreft de beheersingsgraad is de helft van de NT-opgaven gemakkelijk voor een P10-leerling: op 10 van de 20 opgaven heeft een dergelijke leerling een kans van minstens 0,80 om de opgave goed te maken. Een P25-leerling beheerst vrijwel alle NT-opgaven goed. Dat betekent dus dat een leerling op P25-niveau op vrijwel alle opgaven een kans heeft van meer dan 0,80 om de opgaven goed te beantwoorden. In de EB komen ook een aantal van dergelijke opgaven voor. Die zullen ook dienen als illustratie voor de NT-opgaven. Van de 27 EB(-Anker) opgaven zijn er 9 van een vergelijkbare moeilijkheid als in de NT. Ter illustratie van deze vaardigheid worden 10 EB-opgaven gebruikt. De illustratie van de schaal wordt gedaan met EB-opgaven, omdat de NT-opgaven computeropgaven waren. De ankeropgaven worden niet gebruikt omdat die opgaven niet openbaar gemaakt kunnen worden. De opgaven zijn weergegeven in figuur 5.1. De inhoud van de tien opgaven wordt hierna beschreven.

Figuur 5.1 De vaardigheid Nederlandse Taal – Woordenschat



De betekenis van 'een notitie maken' (opgave 1 in figuur 5.1), het woord 'decor' (2) en 'koppelen' (3) in de zin van 'aan elkaar zetten' zijn ook voor leerlingen op het P10-niveau gemakkelijk. Dat geldt ook voor de woorden 'klakkeloos' en 'failliet'. Een woord als 'obstakel' (4) begint voor een P10-leerling een uitdaging te worden met een kans op goed beantwoorden van 0,50, maar is gemakkelijk voor een P25-leerling. Het

herkennen van de betekenis van het woord 'barricaderen' (5) is dan voor een leerling op P25-niveau een uitdaging, maar voor een P75-leerling gemakkelijk. Een goede beheersing wordt ook gevonden op P75-niveau voor woorden als 'klamboe' (6) en 'figurant' (7). Dat zijn echter woorden waarbij een leerling op P25-niveau een kans groter dan 0,50 heeft om het item goed te maken. Dan waren er ook nog 7 woorden die voor een gemiddelde leerling (P50) een uitdaging waren, en ook op P90-niveau niet als gemakkelijk gekwalificeerd kon worden. Zo is ook voor P90-leerlingen de kans om de juiste betekenis van 'zich iets realiseren' (8) en van de woorden 'steriel' (9) en 'garneren' (10) te herkennen kleiner dan 0,80.

In termen van typische percentiëleerlingen kan de volgende analyse gemaakt worden.

De percentiel-10-leerling beheerst vijf opgaven goed uit de EB, waarvan er drie zijn afgebeeld in figuur 5.1 (opgaven 1, 2, en 3). Van de overige opgaven zijn er 6 uitdagend (opgaven met een kans tussen de 0,50 en 0,80) voor deze leerling, en de overige zijn moeilijk of zeer moeilijk.

De percentiel-25 leerling beheerst vijf opgaven goed (waarvan 4 afgebeeld; opgave 4 erbij ten opzichte van een P10-leerling). Het aantal uitdagende opgaven waarbij de kans om het goed te maken lager is dan 0,50 is 8, waarvan er 4 zijn afgebeeld. Van die 4 opgaven is bij opgave 5 de kans op het goed beantwoorden maar net iets kleiner dan 0,50.

De percentiel-50-leerling beheerst maar één EB-opgave meer goed dan de P25-leerling. Deze opgave is niet afgebeeld. Het aantal moeilijke opgaven is voor een dergelijke leerling wel gedaald van 8 naar 5. Drie daarvan zijn hier afgebeeld (opgaven 8, 9 en 10).

De percentiel-75-leerling beheerst maar één opgave meer goed dan de P50-leerling (opgave 5), maar voor deze leerling is geen enkele opgave meer moeilijk te noemen: bij alle opgaven is de kans op het goed beantwoorden groter dan 0,50.

De percentiel-90 leerling beheerst zelfs slechts 11 van de 20 opgaven goed. Dit houdt in dat de opgaven voor Woordenschat als behoorlijk moeilijk gekwalificeerd kunnen worden. Hoe erg het is dat de woorden bij de opgaven 8, 9 en 10 niet volledig beheerst worden, zal stof voor discussie zijn. Binnen de discussie of dit een indicatie is van een laag niveau van de vaardigheid Woordenschat, zal extra standaardonderzoek plaats moeten vinden met geëigende standaardsettingprocedures.

Van alle zes onderzochte vaardigheden is de gemiddelde benodigde vaardigheid om de EB-opgaven goed te beheersen het hoogst bij Woordenschat. De positie van het P80-punt bij de vaardigheid Woordenschat is 287, terwijl voor de andere schalen die positie ligt bij het punt 254 op de vaardigheidsschaal. Wat betreft de gemiddelde grens op de schaal waaronder de echt moeilijke opgaven zitten, verschilt Woordenschat niet echt van de andere twee taalschalen. De gemiddelde positie op de vaardigheidsschaal van de EB-opgaven, waaronder de kans om een opgave goed te doen kleiner dan een half wordt, is bij de taalschalen 174.

In figuur 5.1 zijn naast de opgaven ook de verdelingen voor verschillende groepen van leerlingen weergegeven. Zo zien we dat het gemiddelden van leerlingen op de stratum 1-scholen, de jongens, de niet-vertraagde ('reguliere') leerlingen, en de leerlingen met formatiegewicht 1.00 allemaal iets boven het algemene gemiddelde liggen. Voor de leerlingen van de stratum 2-scholen en voor de meisjes liggen de gemiddelden er iets onder. De verschillen zijn hier niet zeer groot. Die zijn wel groter als we kijken naar het gemiddelde van de vertraagde leerlingen en leerlingen met formatiegewicht 1.25. Die gemiddelden liggen net iets boven het fundamenteel niveau. Voor leerlingen met formatiegewicht 1.90 ('allochtone kinderen') ligt het gemiddelde zelfs onder het niveau van de P25-leerling in de populatie. Het ziet ernaar uit dat de grootste leerwinst te behalen is bij het versterken van de vaardigheid Woordenschat bij deze leerlingen.

Tabel 5.1.1 Verdelingen van de leerlingen op de vaardigheid Nederlandse taal – Woordenschat

Niveau	Percentiel 10-leerling (90% v.d. leerlingen is vaardiger)	186
	Percentiel 25-leerling (75% v.d. leerlingen is vaardiger)	216
	Percentiel 50-leerling (50% v.d. leerlingen is vaardiger)	250
	Percentiel 75-leerling (25% v.d. leerlingen is vaardiger)	284
	Percentiel 90-leerling (10% v.d. leerlingen is vaardiger)	314

Variabele		Gemiddelde	SD	(Percentielpunt)
Geslacht	Jongens	255	50	(54)
	Meisjes	245	50	(46)
Leertijd	Regulier	256	48	(55)
	Vertraagd	225	50	(31)
Stratum	Stratum 1	258	47	(56)
	Stratum 2	246	50	(46)
	Stratum 3	218	52	(26)
Formatiegewicht	1.00	259	46	(57)
	1.25	229	47	(33)
	1.90	199	47	(15)
Toets	EB	251	49	(51)
	NT	170	51	(5)
Doorstroom naar het VO	Vmbo-BB	182	41	(9)
	Vmbo-KB	209	40	(21)
	Vmbo-GT	234	40	(37)
	Havo	264	39	(61)
	Vwo	313	38	(89)
Regio	Noord	251	48	(51)
	Oost	249	49	(49)
	West	250	51	(50)
	Zuid	251	49	(50)
Urbanisatiegraad	Zeer sterk	237	54	(40)
	Sterk	249	50	(49)
	Matig	253	49	(53)
	Weinig	253	48	(52)
	Niet	254	48	(53)
Thuis taal	Alleen Nederlands (NLs)	255	47	(54)
	NLs en een andere taal	216	51	(25)
	Alleen een andere taal	199	49	(15)
	> Turks (eventueel met NLs)	181	48	(8)
	> Arabisch (eventueel met NLs)	198	48	(15)
	> Surinaams (eventueel met NLs)	212	50	(23)
	> West-Europees (eventueel met NLs)	242	49	(44)

Naast de verdelingen van geslacht, leertijd, stratum en formatiegewicht zijn in tabel 5.1.1 ook andere gegevens weergegeven. Het gaat om de toets die de leerling gemaakt heeft, het onderwijsniveau waarnaar de leerling doorstroomt, de regio waar de leerling vandaan komt, de urbanisatiegraad van de plek waar de school staat en de thuistaal van de leerling. Behalve gemiddelde en standaarddeviatie is ook aangegeven op welk niveau het gemiddelde van de groep zit. Dit niveau wordt aangeduid als het percentiel in de verdeling van de gehele groep leerlingen, dus de verdeling met een gemiddelde 250 en een standaarddeviatie van 50. Op basis van die percentielpositie van het gemiddelde kan een groep vergeleken worden met de leerlingniveaus zoals die hierboven zijn aangegeven. De gemiddelde vaardigheid van een groep op percentiepunt 31 is vergelijkbaar met de vaardigheid van een leerling tussen P25 en P50.

De resultaten van de leerlingen die de NT gemaakt hebben liggen duidelijk lager dan die van de leerlingen die de EB gemaakt hebben; het niveau voor Woordenschat ligt anderhalve standaarddeviatie onder het gemiddelde van de gehele groep. Het niveau van die leerlingen ligt onder de vaardigheid van een leerling van P10-niveau. Dat geldt ook voor de leerlingen die een vmbo-BB advies krijgen. Deze leerlingen hebben echter wel een hoger vaardigheidsniveau voor Woordenschat dan de leerlingen die de NT maakten. Wel moet opgemerkt worden dat de vmbo-BB-groep en de NT-groep voor een (groot) deel dezelfde leerlingen betreffen. Het gemiddelde van de vmbo-KB-groep ligt onder het P25-punt, en dat van vmbo-GT groep 16 vaardigheidspunten onder het P50-punt (percentiepunt 37 in de gehele verdeling). De havo-leerlingen hebben een vergelijkbare afstand tot het P50-vaardigheidspunt, alleen dan erboven (percentiepunt 61 in de gehele verdeling). De gemiddelde vaardigheid van vwo-leerlingen ligt daar duidelijk boven, bij 313. Dat is bijna op het P90-vaardigheidsniveau. Het verschil tussen het gemiddelde van de vmbo-leerlingen bij Woordenschat en de vwo-leerlingen, wat ruim 2,5 standaarddeviatie is in de populatieverdeling. Het is overigens ook duidelijk dat de standaarddeviaties binnen de diverse doorstroomgroepen, die rond de 40 liggen, lager zijn dan die in de gehele populatie. De leerlingen binnen de doorstroomgroepen zijn homogener wat betreft Woordenschat dan de gehele populatie, wat ook te verwachten valt.

Met betrekking tot regio zijn in de vaardigheid geen grote verschillen gevonden. Wel zijn er enige verschillen bij de diverse niveaus van urbanisatiegraad. Leerlingen op scholen in zeer sterk geurbaniseerde gebieden hebben gemiddeld een lagere vaardigheid. Dit kan liggen aan de grotere concentratie kinderen met een hoger formatiegewicht. Dat kan gecontroleerd worden door naar de effecten te kijken, waarbij voor die variabele gecontroleerd wordt.

Wat betreft het effect van een andere taal, valt op dat het gemiddelde van de groep leerlingen die thuis een andere taal spreekt naast het Nederlands ligt op het niveau van de P25-leerling in de gehele populatie. Spreekt een leerling helemaal geen Nederlands thuis, dan is het gemiddelde niveau zelfs lager. Het gemiddelde van de leerlingen die thuis Turks spreken (al dan niet naast het Nederlands) ligt zelfs onder het P10-niveau. Voor de leerlingen die thuis Arabisch spreken, ook al dan niet naast het Nederlands, ligt het niveau iets hoger. Voor de leerlingen die thuis Surinaams (en eventueel ook Nederlands) spreken ligt het gemiddelde niveau voor Woordenschat net iets onder het P25-niveau. De leerlingen die thuis een andere West-Europese taal spreken ligt het niveau ook onder het gemiddelde, op het percentiepunt 44. Opgemerkt moet worden dat er geen enkele schaal is waarbij de verschillen tussen de groepen die een verschillende taal spreken zo groot zijn als bij Woordenschat. Het lijkt erop dat het minder spreken van de Nederlandse taal in de dagelijkse omvang het meeste effect heeft op de Woordenschat.

Behalve de waarden van de gemiddelden van de verschillende te 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 die in tabel 5.1.2 gegeven wordt. In hoofdstuk staat hoe deze getallen zijn te interpreteren.

Bij de vaardigheid Woordenschat zien we dat jongens hoger scoren dan meisjes. Het gevonden effect is daar klein. Bij de variabele leertijd wordt gevonden dat de vertraagde leerlingen minder vaardig zijn. Deze effectgrootte is als matig te kwalificeren. Bij stratum zien we dat het effect van stratum 1 ten opzichte van stratum 2 als afwezig gekwalificeerd kan worden en dat dit ook geldt voor het effect van stratum 2 ten opzichte van stratum 3. Die twee zeer kleine effecten zijn bij elkaar wel zo groot dat ze in de vergelijking

van stratum 1 met stratum 3 een klein effect te zien geven. Merk op dat deze effecten gecorrigeerd zijn voor de variabele formatiegewicht.

Tabel 5.1.2 Effectgrootte op de vaardigheid Nederlandse taal – Woordenschat

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	-0,23	klein
Leertijd	Vertraagd – Regulier	-0,54	matig
Stratum	Stratum 2 – Stratum 1	-0,11	geen effect
	Stratum 3 – Stratum 2	-0,14	geen effect
	Stratum 3 – Stratum 1	-0,25	klein
Formatiegewicht	F125 – F100	-0,53	matig
	F190 – F125	-0,51	matig
	F190 – F100	-1,03	groot
Toets	NT – EB	-1,49	groot
Doorstroom	KB – BB	0,59	matig
	GT – KB	0,53	matig
	havo – GT	0,71	matig
	vwo – havo	1,26	groot
Regio	Oost – Noord	-0,03	geen effect
	West – Noord	0,12	geen effect
	Zuid – Noord	0,02	geen effect
	West – Oost	0,14	geen effect
	Zuid – Oost	0,05	geen effect
	Zuid – West	-0,09	geen effect
Urbanisatiegraad	sterk – zeer sterk	-0,08	geen effect
	matig – sterk	0,02	geen effect
	weinig – matig	-0,06	geen effect
	niet – weinig	-0,02	geen effect
	geen effect		geen effect
Thuis taal	NLs+Ander – Alleen NLs	-0,41	klein
	Alleen ander – Alleen NLs	-0,56	matig
	Alleen ander – NLs+Ander	-0,15	geen effect
	Turks – Alleen NLs	-0,96	groot
	Arabisch – Alleen NLs	-0,60	matig
	Surinaams – Alleen NLs	-0,47	klein
	West-Europees – Alleen NLs	-0,12	geen effect

Bij die variabele formatiegewicht zijn de effecten groter dan bij stratum. De persoonlijke kenmerken op het gebied van formatiegewicht wegen duidelijk zwaarder dan de schooleigenschappen. Doordat er nog wel enig effect overblijft, zien we echter dat het schooleffect niet geheel wegvalt als er gecorrigeerd wordt. Bij de andere variabelen valt op dat het verschil tussen leerlingen die de NT doen die de EB maken enorm is. Ook de categorie waarnaar de leerling doorstroomt heeft duidelijk effect, zoals te verwachten. Opvallend is dat de stap in vaardigheid van havo naar vwo het grootst is. Gecorrigeerd voor geslacht, leertijd, formatiegewicht en stratum is het verschil tussen havo en vwo meer dan 1 standaarddeviatie. Als er naar grotere stappen gekeken wordt dan opvolgende categorieën, dus bijvoorbeeld van BB naar GT, dan valt op dat al die effecten groter dan 1 zijn.

Er zijn geen noemenswaardige effecten voor regio en urbanisatiegraad. De scholen in de steden of op het platteland, in het noorden, zuiden, oosten of westen van het land doen het gemiddeld gezien even goed wat betreft de vaardigheid Woordenschat. Het verschil in de verdeling bij leerlingen op scholen in de zeer sterk stedelijke gebieden en de scholen in de andere gebieden (zie tabel 5.1.1) valt weg als er gecorrigeerd wordt voor de hoofdvariabelen stratum, formatiegewicht, leertijd en geslacht.

Bij de variabele thuistaal valt op dat hoe meer leerlingen thuis Nederlands spreken, hoe beter de Woordenschat. Opvallend daarbij is dat het verschil binnen de groep die thuis een andere taal spreekt het er niet veel toe lijkt te doen of daarbij ook Nederlands gesproken wordt. Het grootste verschil in vaardigheid in woordenschat wordt gevonden tussen leerlingen die alleen Nederlands spreken en leerlingen die thuis Turks spreken, al dan niet met Nederlands erbij. De effecten zijn wel iets kleiner dan wanneer naar de ongecorrigeerde verschillen tussen de gemiddelden gekeken wordt.

5.2 Nederlandse taal: Spelling

De vaardigheid

Net als Woordenschat wordt Spelling als ondersteunende vaardigheid voor schrijfvaardigheid beschouwd. Het betreft in de vorm zoals die hier getoetst wordt een kennisaspect, waarbij de vragen in meerkeuzevorm met vier antwoordalternatieven gesteld worden. Bij de hier bevroegde vaardigheid 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 omdat de samenhang hoog is. Bij de typering van de opgaven op de schaal is het onderscheid echter wel duidelijk.

De vaardigheidsschaal Spelling is vastgesteld op basis van de antwoorden op 40 opgaven: 20 uit de EB en 20 uit de NT. Er waren voor Spelling geen nieuwe opgaven opgenomen in de EB-Anker. Van de opgaven zijn er 10 afgebeeld in figuur 5.2.

Wat leerlingen kunnen

Spelling is gemeten met 40 opgaven: 20 opgaven uit de NT en 20 opgaven uit de EB. Er waren geen nieuwe opgaven voor deze vaardigheid, wat betekende dat de leerlingen die de Ankertoets gemaakt hebben exact dezelfde opgaven hadden voor Spelling als de leerlingen die de reguliere toets gemaakt hebben. De illustratie van de vaardigheid wordt gedaan aan de hand van 10 opgaven uit de EB. De NT-opgaven waren iets makkelijker dan de EB-opgaven, maar waar het verschil bij de meeste andere vaardigheden behoorlijk groot is, verschilde de moeilijkheid van de spellingopgaven over de NT en EB minder.

Van de NT-opgaven waren er drie die voor een P10-leerling als moeilijk beschouwd kunnen worden. Twee daarvan waren ook moeilijk voor P25-leerlingen, maar niet voor de gemiddelde leerling. Het aantal spellingopgaven in de NT dat door een P10-leerling beheerst wordt (kans van 0,80 of groter om de opgave goed te maken) is met 2 van de 20 opgaven in de NT niet groot. Voor het P25-niveau is het aantal

beheerste opgaven wel 10. In de NT-toets komen nog 6 opgaven voor die de gemiddelde leerling ook nog net volledig beheerst, en zelfs één opgave waarbij een leerling op P90-niveau nog een kans onder 0,80 heeft om die goed te maken. Voorbeelden van dergelijke opgaven zullen gegeven worden aan de hand van de EB, waarbij het aantal opgaven met niet-werkwoorden en werkwoorden gelijk verdeeld is.

In de spellingopgaven heeft een leerling keuze uit vier alternatieven. De alternatieven zijn geformuleerd als zinnen, waarin een woord vetgedrukt is. In een van de vier alternatieven is het vetgedrukte woord foutief gespeld. De leerling moet aangeven welk alternatief dat is. In de voorbeelden hieronder worden de vetgedrukte woorden weergegeven, waarbij aan de werkwoorden ook enige context is meegegeven. De context in de opgaven is overigens veel ruimer dan hier is aangegeven. Het met een asterisk aangegeven woord is het foutief gespelde woord. Er is aangegeven of het de spelling van een niet-werkwoord [nww] betreft, of van een werkwoord [ww].

Opgave	1	[nww]	perfect, herfstseizoen, orgelessen*, huwelijksbootje
Opgave	2	[nww]	crimeneel*, geroezemoes, evenwichtsoefening, aanmerkelijk
Opgave	3	[nww]	velletje, stroomversneling*, verzameling, sterrenhemel
Opgave	4	[nww]	woordenlijstje*, motorraces, eigenlijk, keukenprinsessen
Opgave	5	[ww]	zij smaakten, het tochte*, hij verwisselde, we hebben getennist
Opgave	6	[ww]	hij leid*, er hipte, zij heeft opgepast, mopperend
Opgave	7	[nww]	vonkjes, grondzijk*, wafels, notitie
Opgave	8	[ww]	het verbaast, hij berekend*, hij probeert, hij heeft bedacht
Opgave	9	[ww]	hij vertelt, het was genezen, zij zijn getrouwt*, wie heeft verzonnen?
Opgave	10	[ww]	zij boden, hij lichtte, de menigte hostte* rond, zij beantwoordde

In figuur 5.2 is met [ww] achter het opgavenummer aangegeven of het een werkwoordopgave betreft. De overige opgavennummers betreffen niet-werkwoordopgaven.

Van de 20 spellingopgaven in de EB is opgave 1 de opgave waar de laagste vaardigheid nodig is voor een goede beheersing. Bij een vaardigheid van 182 op een schaal met een gemiddelde van 250 en een standaarddeviatie van 50, is de kans 0,80 om deze opgave goed te beantwoorden. Opgave 10 is de opgave waarbij de hoogste vaardigheid nodig is voor een goede beheersing (335). Van de 10 niet-werkwoorden is opgave 7 de opgave waar de hoogste vaardigheid nodig is voor een goede beheersing (298). Voor de helft van de werkwoordopgaven is voor een goede beheersing een hogere vaardigheid nodig dan 298. Van de 10 werkwoordopgaven is opgave 5 de opgave waar de laagste vaardigheid noodzakelijk is voor een goede beheersing (253). Er zijn maar twee niet-werkwoorden (waarvan opgave 7 er een is) waar een hogere vaardigheid noodzakelijk is voor een goede beheersing.

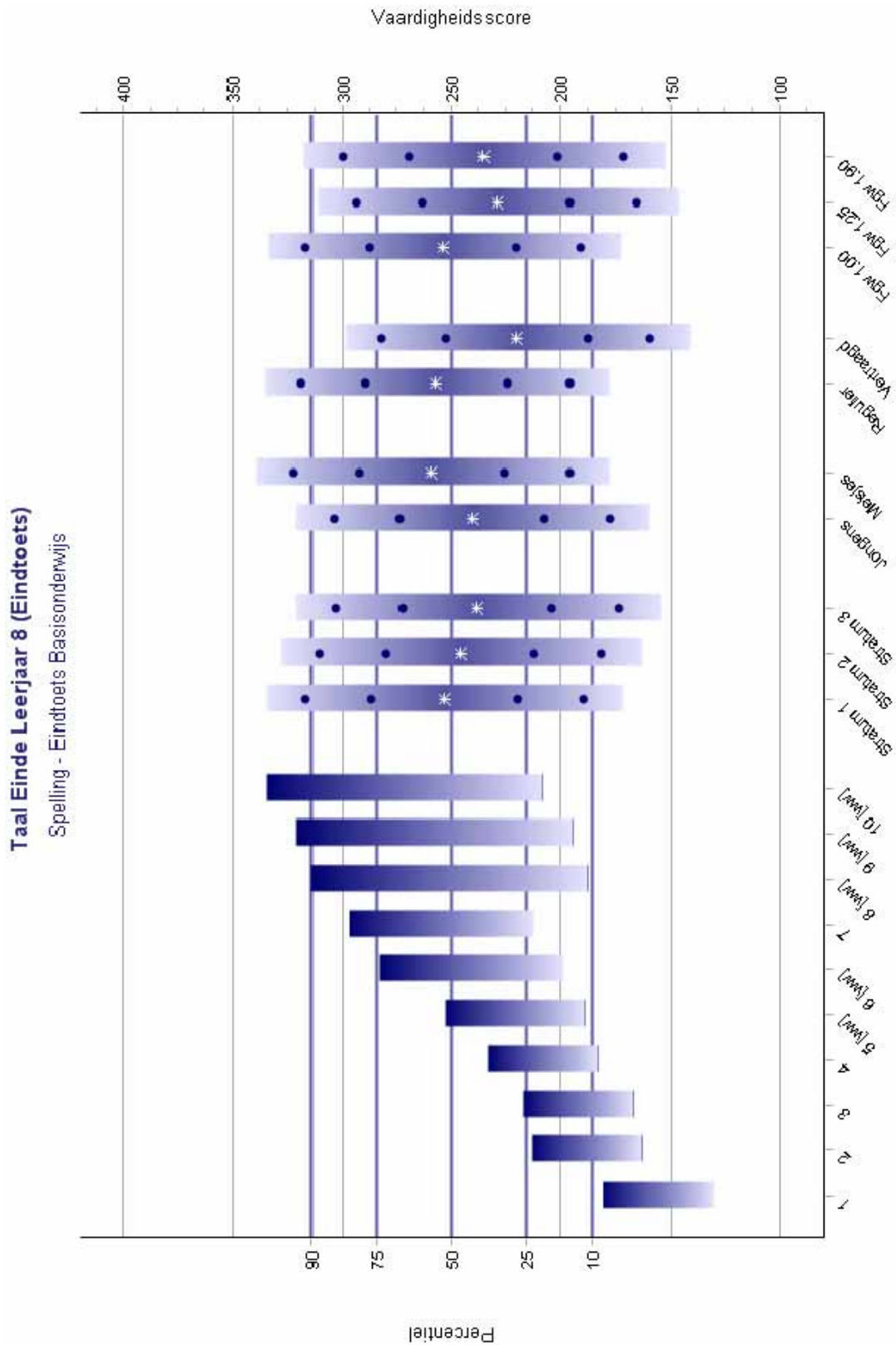
In termen van typische percentiëleerlingen kan de volgende analyse gemaakt worden.

De percentiel-10-leerling beheerst slechts één opgave in de EB goed (opgave 1), terwijl er elf opgaven zijn die (net) iets te moeilijk zijn. Voorbeelden van opgaven die net aan de moeilijke kant zijn (kans iets kleiner dan 0,50 goed te beantwoorden) zijn opgaven 5 en 8. Alle negen opgaven die voor een leerling op dit niveau niet te moeilijk waren, betroffen niet-werkwoorden.

De percentiel-25-leerling beheerst vier opgaven goed. Als opgave 3 meegerekend wordt, waarbij de kans op het goed beantwoorden voor een leerling op dit niveau net onder 0,80 ligt, beheerst een dergelijke leerling er vijf goed. Dat zijn alle niet-werkwoordopgaven. Geen van de twintig opgaven in de EB is voor een leerling op dit niveau te moeilijk: bij alle opgaven is de kans op het goed beantwoorden groter dan 0,50. Dat geldt hiermee dan ook voor alle hogere vaardigheidsniveaus, zoals P50, P75 en P90.

De percentiel-50-leerling beheerst maar zeven van de twintig EB-opgaven. Dat zijn alle niet-werkwoordopgaven. In figuur 5.2 zijn het de opgaven 1 tot en met 4.

Figuur 5.2 De vaardigheid Nederlandse taal – Spelling



De percentiel-75 leerling beheerst dertien opgaven goed. Van de niet-werkwoorden beheerst een leerling op dit niveau alleen opgave 7 niet goed. Overigens is deze opgave uiteraard ook weer niet moeilijk te noemen voor een leerling van dit niveau. Onder de werkwoordopgaven die een dergelijke leerling beheerst zijn ook opgave 5 en 6.

De percentiel-90 leerling beheerst 17 van de 20 opgaven goed. De drie EB-opgaven waarbij het kansniveau van 0,80 niet gehaald wordt, zijn alle werkwoordopgaven. Bij opgave 8 wordt dit beheersingsniveau net gehaald maar bij opgave 9 net niet. De kans om die opgave goed te beantwoorden zal echter voor een P90-leerling nog steeds behoorlijk groot zijn. Die kans is iets kleiner bij opgave 10, maar heel veel scheelt het niet.

In figuur 5.2 zijn naast de gegevens over de opgaven ook de verdelingen van verschillende groepen leerlingen weergegeven. In vergelijking met de schaal Woordenschat zien we bij Spelling een paar verschillen. Zo zijn de meisjes bij spelling vaardiger dan de jongens: de gemiddelde spellingvaardigheid van meisjes ligt iets boven het algeheel gemiddelde, en voor jongens daar net iets onder. Wat betreft de overige variabelen zijn de relatieve posities van de groepen binnen een variabele vergelijkbaar met wat er bij Woordenschat gevonden is. Wel zijn de posities enigszins verschillend. De vertraagde leerlingen zijn relatief iets slechter in Spelling dan in Woordenschat. Daar staat dan weer tegenover dat de leerlingen op stratum 3-scholen en leerlingen met formatiegewicht 1.90 juist relatief beter in Spelling zijn in vergelijking tot Woordenschat. Waar bij Woordenschat het gemiddelde niveau van de leerlingen met formatiegewicht 1.90 duidelijk onder dat van een P25-leerling lag, ligt de vaardigheid van deze groep voor Spelling daar duidelijk boven, op het niveau van een leerling uit percentiel 41.

De gemiddelden van de groepen zijn in tabel 5.2.1 ook te vinden, waar voor ieder gemiddelde de spellingvaardigheid gerelateerd is aan de percentielen in de gehele verdeling van de leerlingen. Naast de vaardigheden van de variabelen uit figuur 5.2 staan ook die van andere variabelen gegeven. Het opvallendste verschil met Woordenschat is dat het gebruik van een andere taal samen met of in plaats van het Nederlands een veel minder groot verschil maakt voor Spelling. De gemiddelden voor deze groepen liggen nog steeds onder het gemiddelde, maar bij lange na niet zo ver daaronder als bij Woordenschat. Dit resultaat zou eventueel ook kunnen verklaren waarom zeer sterk geurbaniseerde gebieden het op deze vaardigheid redelijk goed doen: het is de vaardigheid waar deze groep het best op presteert met een vaardigheid van 249. Wat betreft andere variabelen valt op dat de leerlingen uit het noorden relatief laag scoren, terwijl leerlingen uit het zuiden het vrij goed doen.

Ook voor de vaardigheid Spelling zijn de effecten van de variabelen gecorrigeerd berekend en getoetst op significantie. Deze staan voor spelling in tabel 5.2.2. De verschillen die bij Woordenschat zijn gevonden komen ook hier weer terug. Meisjes scoren ook gecorrigeerd voor andere variabelen beter dan jongens op Spelling. De effecten voor stratum vallen voor Spelling helemaal weg. Een opvallend resultaat is dat, als gecorrigeerd wordt voor de variabelen geslacht, leertijd en stratum, de leerlingen met formatiegewicht 1.90 zelfs significant beter presteren dan leerlingen met formatiegewicht 1.25 (klein effect) en het effect formatiegewicht 1.90 en 1.00 zo klein is dat het als afwezig beschouwd kan worden.

Het effect van de toets blijft wel groot maar is aanzienlijk kleiner dan bij andere vaardigheden. Het effect bij de verschillende doorstroomniveaus is onverminderd hoog. In vergelijking met de andere vaardigheden wordt het grootste effect tussen havo en vwo in dit onderzoek bij Spelling gemeten. Ook is bij Spelling een van de weinige effecten binnen regio's gevonden. Dit betreft een klein effect waarbij gecorrigeerd op geslacht, leertijd, stratus en formatiegewicht de zuidelijke provincies significant hoger scoren dan de noordelijke provincies. Bij urbanisatiegraad zijn geen effecten gevonden en ook bij thuistaal worden geen effecten gevonden, op een na: leerlingen die thuis Surinaams/Antilliaans spreken, spellen iets slechter.

Tabel 5.2.1 Verdelingen van de leerlingen op de vaardigheid Nederlandse taal – Spelling

Niveau	Percentiel 10-leerling (90% v.d. leerlingen is vaardiger)	186
	Percentiel 25-leerling (75% v.d. leerlingen is vaardiger)	216
	Percentiel 50-leerling (50% v.d. leerlingen is vaardiger)	250
	Percentiel 75-leerling (25% v.d. leerlingen is vaardiger)	284
	Percentiel 90-leerling (10% v.d. leerlingen is vaardiger)	314

Variabele		Gemiddelde	SD	(Percentiel)
Geslacht	Jongens	241	49	(43)
	Meisjes	259	49	(57)
Leertijd	Regulier	257	48	(56)
	Vertraagd	220	48	(28)
Stratum	Stratum 1	254	49	(53)
	Stratum 2	246	50	(47)
	Stratum 3	239	51	(41)
Formatiegewicht	1.00	254	49	(53)
	1.25	230	50	(34)
	1.90	236	50	(39)
Toets	EB	251	49	(51)
	EB-Anker	252	49	(51)
	NT	187	51	(11)
Doorstroom naar het VO	Vmbo-BB	193	39	(13)
	Vmbo-KB	214	39	(23)
	Vmbo-GT	231	39	(36)
	Havo	264	39	(61)
	Vwo	319	38	(92)
Regio	Noord	243	50	(44)
	Oost	249	50	(49)
	West	250	50	(50)
	Zuid	254	50	(53)
Urbanisatiegraad	Zeer sterk	249	51	(49)
	Sterk	251	50	(51)
	Matig	253	50	(53)
	Weinig	250	50	(50)
	Niet	245	50	(46)
Thuis taal	Alleen Nederlands (NLs)	252	50	(51)
	NLs en een andere taal	244	50	(45)
	Alleen een andere taal	237	50	(40)
	> Turks (eventueel met NLs)	238	50	(41)
	> Arabisch (eventueel met NLs)	233	50	(37)
	> Surinaams (eventueel met NLs)	230	50	(34)
	> West-Europees (eventueel met NLs)	249	50	(50)

Tabel 5.2.2 Effectgrootte op de vaardigheid Nederlandse taal – Spelling

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	0,37	klein
Leertijd	Vertraagd – Regulier	-0,71	matig
Stratum	Stratum2 – Stratum1	-0,08	geen effect
	Stratum3 – Stratum2	-0,02	geen effect
	Stratum3 – Stratum1	-0,10	geen effect
Formatiegewicht	F125 – F100	-0,40	klein
	F190 – F125	0,23	klein
	F190 – F100	-0,18	geen effect
Toets	NT – EB	-1,04	groot
Doorstroom	KB – BB	0,51	matig
	GT – KB	0,44	klein
	havo – GT	0,81	groot
	vwo – havo	1,48	groot
Regio	Oost – Noord	0,15	geen effect
	West – Noord	0,20	klein
	Zuid – Noord	0,26	klein
	West – Oost	0,06	geen effect
	Zuid – Oost	0,11	geen effect
	Zuid – West	0,05	geen effect
Urbanisatiegraad	sterk – zeer sterk	-0,09	geen effect
	matig – sterk	0,01	geen effect
	weinig – matig	-0,09	geen effect
	niet – weinig	-0,13	geen effect
Thuis taal	NLs+Ander – Alleen NLs	0,04	geen effect
	Alleen ander – Alleen NLs	0,03	geen effect
	Alleen ander – NLs+Ander	-0,01	geen effect
	Turks – Alleen NLs	0,05	geen effect
	Arabisch – Alleen NLs	-0,07	geen effect
	Surinaams – Alleen NLs	-0,25	klein
	West-Europees – Alleen NLs	0,06	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 moet bepaald kunnen worden in welke mate leerlingen daar in slagen. De toetsen zijn voor het grootste deel samengesteld rond (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 de opgaven. De vragen bij deze opgaven zijn gesteld als meerkeuze-opgaven met vier antwoordalternatieven.

Naast het type vragen waarbij vragen over een tekst gesteld worden, is er een klein aantal vragen 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.

De vaardigheidsschaal voor Begrijpend lezen wordt in totaal gevormd door 74 opgaven. Daarvan komen er 30 uit de EB, 30 uit de NT en zijn er 14 nieuwe opgaven opgenomen in het EB-Anker.

Wat leerlingen kunnen

Geen van de 30 NT-opgaven Begrijpend lezen is voor een leerling op P10-niveau echt moeilijk te noemen, aangezien een leerling op alle opgaven een kans groter dan een half heeft om de opgaven goed te beantwoorden. Al deze opgaven zijn vierkeuze opgaven. Een P10-leerling beheerst 14 van de 30 opgaven goed (kans op goed beantwoorden minstens 0,80). Een leerling op P25-niveau beheerst 23 van de 30 NT-opgaven, en een P50-leerling beheerst er 27. De P75- en P90-leerlingen beheersen alle opgaven uit de NT. De verdeling van de moeilijkheden van het EB-Anker is vergelijkbaar met die van de EB. De EB-opgaven worden geïllustreerd met 10 voorbeelden.

De samenhang tussen de opgaven is bij Begrijpend lezen groter dan bij de andere vaardigheden. Dit wordt veroorzaakt doordat de opgaven gegeven worden per tekst. Op de vier 'Wat is de eerste zin?'-opgaven na, refereren de opgaven altijd aan een losse tekst. Een vraag kan al dan niet moeilijk zijn door de context van die tekst. Er zijn in de EB vier verschillende teksten gebruikt, met ieder zes of zeven opgaven. In figuur 5.3.1 staan de opgaven 8 tot en met 14 van Taal Taak 4 en de bijhorende tekst. Deze taak is gemaakt op de derde toetsdag van de EB. De posities van deze opgaven worden ook gegeven in figuur 5.3.2.

Er worden ook nog drie andere opgaven ter illustratie gegeven. Ten eerste de meest gemakkelijke opgave van de vaardigheid Begrijpend lezen in de EB, Opgave 1. In deze opgave moeten de leerlingen de zin 'Hij was bezig een pees op een boog te spannen en wilde dat eerst afmaken' uit een tekst beschouwen. Hieruit kunnen kinderen concluderen dat deze persoon zijn werk serieus neemt. De alternatieven waren de mogelijke conclusies dat deze persoon blij was, niet zo handig was of opkeek tegen hooggeplaatsten.

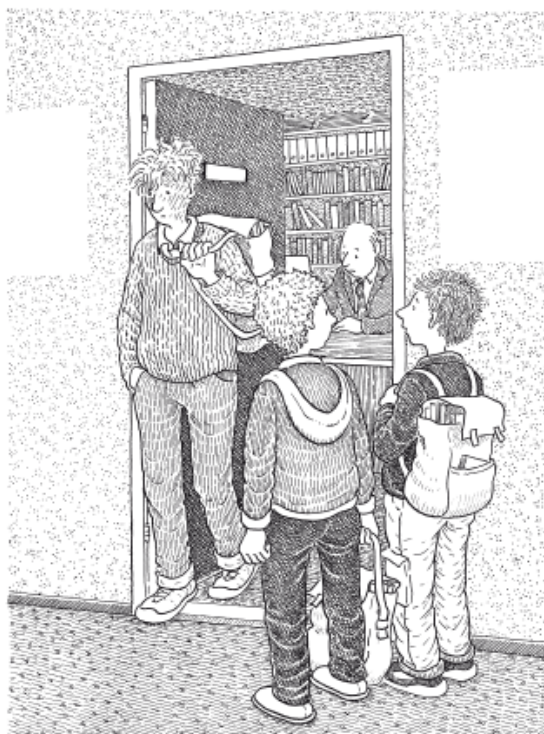
Figuur 5.6.1 De tekst en opgaven Nederlands Begrijpend lezen: Dag 3, Taak Taal 4, tekst 3, opgaven 8 tot en met 14

De opgaven 8 tot en met 14 horen bij de tekst die hieronder staat. Lees eerst de tekst en maak daarna de opgaven.

1 Brugklassers

2 "Jesse van Gent, eruit! Ga je melden bij de
3 coördinator!" Priemend wijst haar vinger naar Jesse.
4 De klas is nu niet meer te redden, dit is alleen maar
5 olie op het vuur.
6 Jesse haalt zijn schouders op en staat op om zijn
7 spullen in te pakken. Tot zijn verbazing staat Chris
8 naast hem ook op en begint ook zijn spullen in te
9 pakken.
10 "Chris Bergmans, wat doe je!", roept mevrouw
11 Prenger met overslaande stem. Van verbazing
12 wordt de klas stil. "Ik ga met Jesse mee naar de
13 coördinator. Hij heeft vandaag voor het eerst een
14 nieuwe beugel. Als u hem al niet kan verstaan, zal
15 het meneer Van Dijk ook wel niet lukken. Ik zal zijn
16 tolk zijn."
17 Jesse loopt richting deur en Chris volgt hem op de
18 voet. De klas is nog steeds muisstil en mevrouw
19 Prenger zegt niets. Als Jesse de deur achter Chris
20 en hem dichtdoet, ziet hij dat haar mond nog
21 steeds openstaat van verbazing.
22 Zwijgend lopen Jesse en Chris naar het eind van
23 de gang. Samen gaan ze door de klapdeuren en
24 daar barsten ze tegelijkertijd in lachen uit. "Sjag je
25 haar gesjigt?", lacht Jesse. En daardoor begint
26 Chris nog harder te schateren. Dan trekt hij Jesse
27 aan zijn arm. "Kom we gaan naar meneer Van Dijk,
28 dan hebben we dat maar vast gehad."
29 De deur van de kamer van meneer Van Dijk staat
30 op een kier en daarachter klinken zachte stemmen.
31 Jesse aarzelt even en dan klopt hij zachtjes. "Eén
32 momentje!", klinkt de stem van Van Dijk bars. Op
33 dezelfde toon gaat die door en foertert een leerling
34 uit: "Je blijft de hele week na en dan veeg je het
35 hele plein van voor naar achter!" Jesse en Chris
36 kijken elkaar aan, dat klinkt niet best. De deur gaat
37 open. Er komt een jongen uit met rood haar en een
38 rood hoofd. Het is een lange slungel die zeker al in
39 de derde zit, dat zie je zo.
40 "Kom verder!", roept Van Dijk. Hij voert iets in op
41 de computer en kijkt niet op als Jesse en Chris
42 binnenkomen. Zwijgend blijven ze voor zijn bureau
43 staan, wachtend op het moment dat de coördinator
44 tijd voor hen heeft. Het uur van de waarheid is
45 aangebroken. Van Dijk wendt zijn blik van het
46 scherm naar de twee jongens die voor hem staan.
47 "Brugklassers!", zegt hij verbaasd. "Dat gebeurt
48 niet vaak. Leg me eens uit wat er gebeurd is." Zijn
49 stem klinkt niet boos, eerder nieuwsgierig. "Ik heb
50 vandaag voor het eersjijt een beugel", legt Jesse
51 uit. "Van mevrouw Prenger kreeg ik de beurt om te

52 tellen vanaf sjesjstig." "Soixante", vult Chris
53 behulpzaam aan. "Dank je, jongeman. Mijn Frans
54 is prima." Van Dijk kijkt streng, maar Jesse meent
55 pretlichtjes in zijn ogen te zien. "Dat tellen klonk
56 nogal gek door mijn beugel en de hele klas moest
57 lachen. Mevrouw Prenger werd boosj en ik moesjt
58 doortellen. Bij drieënsjesjtig werd ik eruit gestuurd."
59 Jesse haalt onschuldig zijn schouders op. "En jij?",
60 vraagt Van Dijk nu aan Chris. "Ik was bang dat u
61 hem niet kon verstaan", zegt Chris en zijn stem
62 klinkt ineens klein. Van Dijk barst in schaterlachen
63 uit. "Lekker stel zijn jullie! Aan het eind van de les
64 gaan jullie naar mevrouw Prenger en bieden jullie
65 haar je excuses aan voor het verstoren van de les.
66 Laat haar zien dat je een nieuwe beugel hebt, Jesse.
67 En laat ik jullie hier de rest van het jaar niet meer
68 zien, anders zwaait er wat." Hij staat op en geeft
69 daarmee aan dat hun gesprek beëindigd is.
70 Hij loopt zowaar mee naar de deur van zijn kamer
71 om hen uit te laten. Daar legt hij nog even zijn
72 hand op Jesses schouder. "Volgens mij heb je
73 vandaag niet alleen een nieuwe beugel, maar ook
74 een nieuwe vriend." Jesse kijkt naar Chris en
75 grijnst.



In welk stukje van de tekst wordt duidelijk waarvoor Jesse uit de klas gestuurd werd?

In het stukje:

- A "Jesse ... vuur. (r. 2 t/m 5)
- B "Brugklassers ... is." (r. 47, 48)
- C "Dat ... lachen. (r. 55 t/m 57)
- D Aan ... les. (r. 63 t/m 65)

Opgave 9

Lees: *De ... vuur.* (r. 4, 5)

Wat maakt de schrijver met deze zin duidelijk?

- A De klas is stomverbaasd en wordt doodstil.
- B De klas stond al op stelten en dat wordt nog erger.
- C De klasgenoten vinden het leuk en willen ook wel weg.
- D Mevrouw Prenger is woedend en geeft nu iedereen straf.

Opgave 10

Welke zin is waar?

Meneer Van Dijk ...

- A gelooft de jongens niet, maar gaat het eerst navragen.
- B is het eens met mevrouw Prenger en straft de jongens.
- C snapt niet wat er gebeurde en doet dus maar niets.
- D vindt het niet zo erg, maar keurt het ook niet goed.

Opgave 11

Waarom wordt Jesse de klas uitgestuurd?

Mevrouw Prenger denkt dat ...

- A hij geen Frans wil leren.
- B hij haar voor de gek houdt.
- C hij zijn les niet heeft geleerd.
- D hij zijn spullen niet bij zich heeft.

Opgave 12

Lees: *Het ... aangebroken.* (r. 44, 45)

Wat wordt met deze zin bedoeld?

- A Nu kunnen ze erop rekenen dat ze naar huis gestuurd worden.
- B Nu moeten ze geen leugentjes gaan vertellen.
- C Nu weten ze spoedig wat hun te wachten staat.
- D Nu zullen ze meneer Van Dijk pas echt leren kennen.

Welke bewering is waar als je let op het stukje: "*En ... klein.*" (r. 59 t/m 62)?

- A Chris beseft opeens dat hij nog maar een brugklasser is.
- B Chris is er niet meer zo zeker van dat het goed afloopt.
- C Meneer Van Dijk gelooft niks van het verhaal van Chris.
- D Meneer Van Dijk heeft Jesse niet goed kunnen verstaan.

Opgave 14

Welke zin past het best achter: *grijnst* (r. 75)?

- A Dat heeft er niets mee te maken.
- B Hij snapt er niets van.
- C Meneer Van Dijk heeft gelijk.
- D Ze zitten alleen maar in dezelfde groep.

Opgave 10 is een opgave waarbij gevraagd wordt of de leerling aan kan geven op welk punt een alinea uit de tekst eventueel in twee verschillende alinea's opgedeeld zou kunnen worden. Een leerling moet hier herkennen wanneer een nieuw deel informatie begint. Dit voorbeeld is ter illustratie gekozen omdat voor beheersing van deze opgave de hoogste vaardigheid nodig is. Zoals bij de vorige schalen krijgt deze opgave daarom nummer 10 mee.

De laatste illustrerende opgave is een van het type 'Wat is de eerste zin?'. De zinnen werden in de volgende volgorde gegeven:

- A Daarin legde ze zo'n 20 eieren en die bedekte ze met een laag planten.
- B Dag in, dag uit bewaakte de moeder haar nest, terwijl de planten de eieren warm hielden.
- C De meest dinosauriërs legden eieren in een nest.
- D Weken later verschenen er barstjes in de eieren en kwamen de kleintjes tevoorschijn.
- E Een moederdino groef eerst een groot, rond gat, ongeveer zo groot als een tuinvijver.

Alternatief C zou de eerste zin van moeten zijn als de vijf zinnen een verhaal vormen, en is zodoende het juiste antwoord. Deze opgave wordt figuur 5.3 aangeduid met EZ ('eerste zin').

In termen van percentielniveaus is het volgende op te merken:

De percentiel-10 leerling beheerst slechts één opgave goed, en wel opgave 1. Een voorbeeld van een opgave die iets boven het niveau van goede beheersing ligt is opgave T4-11. Het aantal moeilijke opgaven is echter ook beperkt voor dit type leerlingen, met 11 van de 30 opgaven die als moeilijk gekenmerkt kunnen worden. Een voorbeeld van een opgave waar de kans op een goed antwoord voor een P10-leerling net onder 0,50 komt is T4-8.

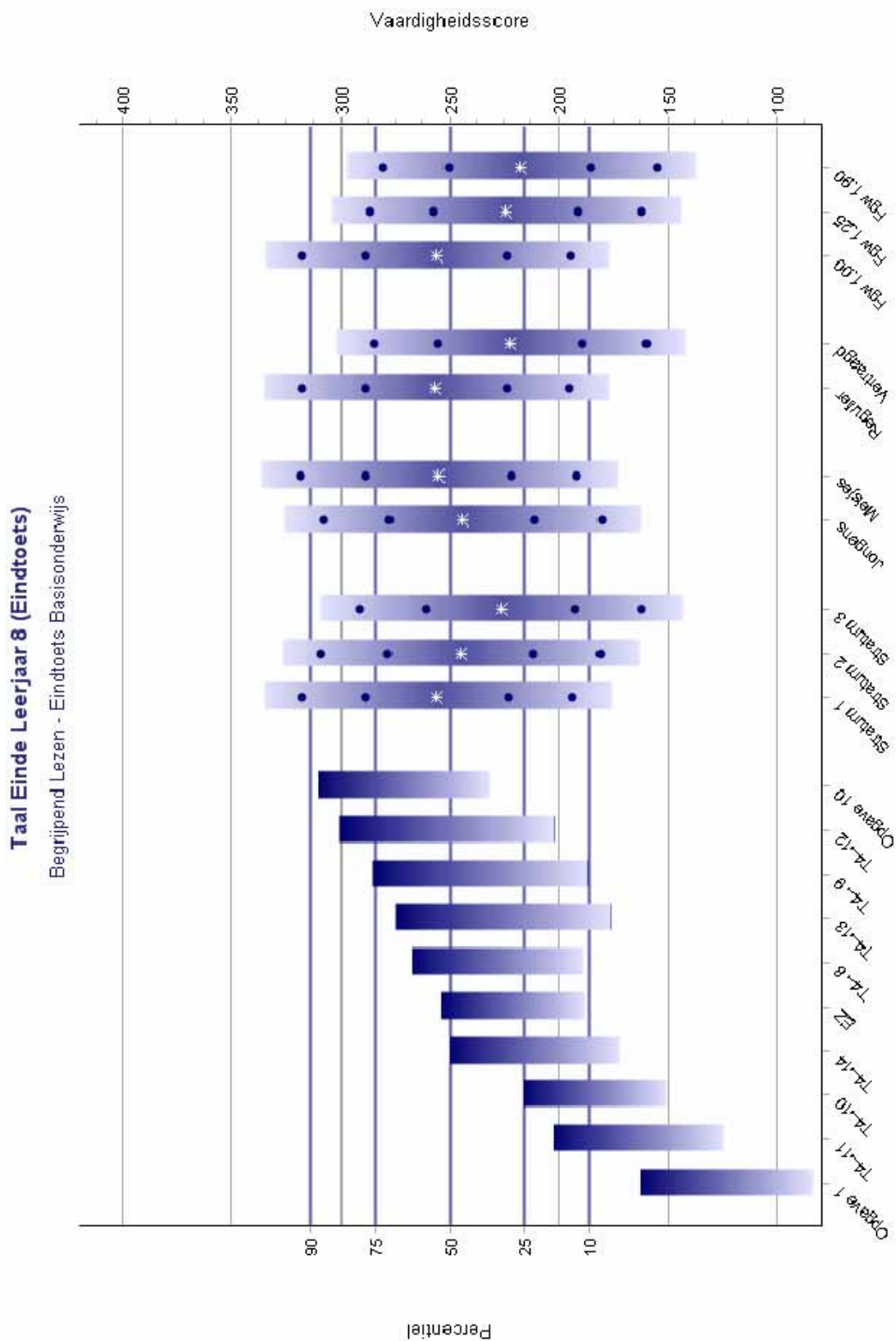
De percentiel-25 leerling beheerst 9 van de 30 opgaven goed. De kans om die opgaven goed te maken is (bijna) 0,80 of hoger voor de leerlingen van dit niveau. Een voorbeeld van een opgave waar de kans hoger is dan 0,80 is T4-11, en opgave T4-10 is een voorbeeld van een opgave waar de kans bijna 0,80 is. Er zijn slechts twee opgaven waar leerlingen van dit niveau een kans van kleiner dan een half hebben om de opgave goed te maken. Opgave 10 is een van deze twee opgaven.

De percentiel-50 leerling beheerst de helft van de opgaven goed. Hierbij wordt opgave T4-14 meegerekend, waar de kans om deze opgave goed te beantwoorden voor deze leerlingen slechts iets onder 0,80 ligt. Van de beheerste opgaven zijn er twee van het type 'Wat is de eerste zin?'. De voorbeeldopgave EZ beheerst een dergelijke leerling nog net niet goed, al is de kans op het geven van een juist antwoord al heel behoorlijk (rond de 0,75) voor een gemiddelde leerling. Geen van de opgaven is voor een gemiddelde leerling moeilijk: op alle opgaven is de kans op een goed antwoord groter dan 0,50. Dit geldt zodoende ook voor de hogere vaardigheidsniveaus.

De percentiel-75 leerling beheerst alle opgaven van het type 'Wat is de eerste zin?' goed. Een leerling van dit niveau beheerst ook opgaven T4-8 en T4-13 goed. De beheersing van T4-9 is ook bijna goed te noemen met een kans van net iets onder 0,80. Er zijn slechts een viertal opgaven dat een P75-leerling nog niet goed beheerste, waaronder T4-12 en Opgave 10.

De percentiel-90 leerling beheerst alle opgaven van deze schaal (zeer) goed.

Figuur 5.3 De vaardigheid Nederlandse Taal – Begrijpend lezen



Tabel 5.3.1 Verdelingen van de leerlingen op de vaardigheid Nederlandse taal – Begrijpend lezen

Niveau	Percentiel-10-leerling (90% v.d. leerlingen is vaardiger)	186
	Percentiel-25-leerling (75% v.d. leerlingen is vaardiger)	216
	Percentiel-50-leerling (50% v.d. leerlingen is vaardiger)	250
	Percentiel-75-leerling (25% v.d. leerlingen is vaardiger)	284
	Percentiel-90-leerling (10% v.d. leerlingen is vaardiger)	314

Variabele		Gemiddelde	SD	(Percentiel)
Geslacht	Jongens	244	50	(46)
	Meisjes	256	50	(54)
Leertijd	Regulier	257	48	(55)
	Vertraagd	222	49	(29)
Stratum	Stratum 1	256	48	(55)
	Stratum 2	245	50	(46)
	Stratum 3	227	51	(32)
Formatiegewicht	1.00	256	48	(55)
	1.25	225	49	(31)
	1.90	218	49	(26)
Toets	EB	251	49	(51)
	EB-anker	248	48	(48)
	NT	168	50	(5)
Doorstroom naar het VO	BB	177	37	(7)
	KB	206	37	(19)
	GT	233	37	(36)
	Havo	268	37	(64)
	Vwo	316	36	(91)
Regio	Noord	246	49	(47)
	Oost	249	50	(49)
	West	251	50	(50)
	Zuid	252	50	(51)
Urbanisatiegraad	Zeer sterk	243	52	(45)
	Sterk	251	50	(51)
	Matig	253	50	(52)
	Weinig	251	49	(51)
	Niet	248	49	(49)
Thuis taal	Alleen Nederlands (NLs)	253	49	(53)
	NLs en een andere taal	232	51	(36)
	Alleen een andere taal	221	50	(28)
	> Turks (eventueel met NLs)	212	49	(22)
	> Arabisch (eventueel met NLs)	216	49	(25)
	> Surinaams (eventueel met NLs)	216	50	(25)
	> West-Europees (eventueel met NLs)	252	50	(51)

De verdelingen van de verschillende groepen zijn ook in figuur 5.3 te vinden. Net als bij Spelling is de vaardigheid Begrijpend lezen bij meisjes iets hoger dan het gemiddelde, en die van jongens is lager. Wat betreft de vaardigheidsniveaus van de andere groepen in dit figuur liggen de gemiddelde niveaus op vergelijkbare hoogte als bij Woordenschat en Spelling. De gemiddelde vaardigheden voor Begrijpend lezen voor de groep leerlingen op stratum 3-scholen of de groep leerlingen met formatiegewicht 1.90 liggen minder ver onder het populatiegemiddelde dan bij Woordenschat, maar meer onder dat gemiddelde dan bij Spelling. Het gemiddelde van de groep met formatiegewicht 1.90 ligt iets boven het niveau van een leerling met een percentiel-25 niveau. De vaardigheid van de groep leerlingen met formatiegewicht 1.25 ligt net iets boven die van de groep met formatiegewicht 1.90.

Bij de verdelingen van de verschillende taalgroepen zien we iets vergelijkbaars: de vaardigheden liggen relatief hoger dan bij Woordenschat, maar relatief lager dan bij Spelling. De gemiddelde prestaties voor de leerlingen die thuis Turks, Arabisch Surinaams/Antilliaans spreken liggen allen zo rond het P25-leerlingniveau. Van alle taalvaardigheden ligt het gemiddelde van de toekomstige vmbo-BB-leerlingen het laagst bij Begrijpend lezen. Er wordt een gemiddelde vaardigheid voor deze groep gevonden van rond het niveau van een percentiel-7-leerling.

De gecorrigeerde effecten die gevonden zijn bij Begrijpend lezen zijn vergelijkbaar met de effecten die gevonden zijn bij Spelling (zie tabel 5.3.2). Een verschil was dat bij Begrijpend lezen wel een effect gevonden werd tussen formatiegewicht 1.90 en 1.00. Het effect van formatiegewicht is geheel verdwenen als leerlingen met formatiegewicht 1.25 en 1.90 met elkaar vergeleken worden. Effecten voor regio en urbanisatiegraad zijn weer vrijwel afwezig, evenals voor de algemene taalgroepen. Zo is er voor Begrijpend lezen geen effect te vinden als leerlingen thuis Nederlands spreken terwijl ze daar ook een andere taal spreken. Wel zijn er kleine effecten te vinden voor een aantal specifieke talen als deze vergeleken worden met de leerlingen die thuis alleen Nederlands gebruiken.

Tabel 5.3.2 Effectgrootte op de vaardigheid Nederlandse taal – Begrijpend lezen

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	0,22	klein
Leertijd	Vertraagd – Regulier	-0,61	matig
Stratum	Stratum2 – Stratum1	-0,11	geen effect
	Stratum3 – Stratum2	-0,12	geen effect
	Stratum3 – Stratum1	-0,23	klein
Formatiegewicht	F125 – F100	-0,55	matig
	F190 – F125	0,00	geen effect
	F190 – F100	-0,55	matig
Toets	NT – EB	-1,47	groot
Doorstroom	KB – BB	0,71	matig
	GT – KB	0,68	matig
	havo – GT	0,93	groot
	vwo – havo	1,31	groot
Regio	Oost – Noord	0,10	geen effect
	West – Noord	0,19	geen effect
	Zuid – Noord	0,15	geen effect
	West – Oost	0,10	geen effect
	Zuid – Oost	0,06	geen effect
	Zuid – West	-0,04	geen effect
Urbanisatiegraad	sterk – zeer sterk	-0,08	geen effect
	matig – sterk	-0,03	geen effect
	weinig – matig	-0,07	geen effect
	niet – weinig	-0,09	geen effect
Thuis taal	NLs+Ander – Alleen NLs	-0,11	geen effect
	Alleen ander – Alleen NLs	-0,15	geen effect
	Alleen ander – NLs+Ander	-0,04	geen effect
	Turks – Alleen NLs	-0,36	klein
	Arabisch – Alleen NLs	-0,27	klein
	Surinaams – Alleen NLs	-0,43	klein
	West-Europees – Alleen NLs	0,11	geen effect

5.4 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

In totaal is deze schaal gemeten met 52 opgaven. Hiervan waren 25 opgaven afkomstig uit de NT, 22 uit de EB en 5 opgaven waren in de EB-Anker opgenomen. Het onderdeel Getallen en bewerkingen van de Eindtoets Basisonderwijs 2008 telt overigens in totaal 25 opgaven die allen vier antwoordalternatieven hebben. Bij de analyses zijn drie opgaven, die niet goed bij de geconstrueerde schaal pasten, verwijderd. Dit ging niet ten koste van de betrouwbaarheid van de toets.

De schaal wordt geïllustreerd met EB-opgaven, omdat de NT-opgaven computeropgaven waren en ankeropgaven die niet openbaar gemaakt kunnen worden. De NT-opgaven waren allen zeer gemakkelijk; bij 24 van de 25 opgaven hadden leerlingen medio jaargroep 8 op P10-niveau meer dan 50% kans op een goed antwoord, en 16 van deze opgaven beheersten P10-leerlingen goed (kans om goed te beantwoorden is 0,80 of hoger). Voor de leerling op het P10-niveau ('minimaal niveau') was er dus in de NT één opgave die moeilijk was, en 16 die gemakkelijk waren. Van de 25 NT-opgaven is bij 23 de beheersing van de 'gemiddelde leerling' (P50) goed; er waren dus twee wat moeilijker opgaven. Deze makkelijke opgaven zijn

vergelijkbaar met de makkelijke opgaven in de EB. De moeilijkheid van de ankeropgaven is in het algemeen goed te vergelijken met de moeilijkheid van de EB opgaven.

In figuur 5.4 staan 10 van de 22 opgaven uit de EB die bij de analyses gebruikt zijn. Deze opgaven zijn geordend op het P80-kanspunt van de opgaven (de benodigde vaardigheid om een kans te hebben van 0,80 om deze opgave goed te maken). De eerste voorbeeldopgave is het gemakkelijkst en betreft de som '485 – 195' (opgave 1). Opgave 2 betreft de berekening van het verschil tussen 307 en 178 en opgave 3 de som van '0,9 + 9 + 0,09'. De daaropvolgende sommen betreffen de vraag hoeveel flesjes van 25 ml gevuld kunnen worden met 5000 ml reukwater (opgave 4) en hoe kommagetallen met een verschillend aantal cijfers achter de komma geordend moeten worden (opgave 5). Die worden gevolgd door de opgaven hoe zwaar een vracht ongeveer is van 49 dozen van 19,5 kg is (opgave 6) en de plaats van het getal 5000 aangeven op een getallenlijn van 0 tot 100 000. (opgave 7). Opgave 8 betreft het inschatten van de uitkomst van de som '60000 : 2,987 : 2,097', waarbij de alternatieven '4000', '5000', '6000' en '7500' waren. Bij opgave 9 moet aangeven worden van welk werkelijk aantal 1,2 miljoen een afronding kan zijn (opgave 9; alternatieven in deze opgaven waren '19642', '1189764', '1290361' en '12091012'). De laatste illustrerende opgave (10) betreft de som '202 x 202', dus het vermenigvuldigen van grotere getallen. Opgemerkt moet worden dat de meeste opgaven in een (verhalende) context gegeven zijn, waarbij soms een illustratie gegeven is.

De percentiel-10-leerling beheerst de eerste 4 van de in totaal 22 opgaven (nagenoeg) goed: de kans om die vragen goed te beantwoorden ligt voor ieder van deze vragen iets onder de 0,80. In figuur 5.4 afgebeelde opgaven 1, 2 en 3 zijn hiervan voorbeelden. Van de overige opgaven zijn er 13 als moeilijk voor deze leerlingen te bestempelen (kans minder dan 0,50 om deze opgaven goed te maken). De afgebeelde opgaven 5 tot en met 10 zijn hiervan voorbeelden. De opgaven die de percentiel-10-leerling goed kan maken betreffen simpele som- en verschilopgaven.

De percentiel-25-leerling beheerst 6 van de 22 opgaven goed. Opgave 4 in figuur 5.4 beheerst een dergelijke leerling bijna goed. Van de 22 opgaven zijn er 5 te moeilijk voor deze leerling. Opgave 8 en opgave 10 zijn voorbeelden van dergelijke opgaven. Het is duidelijk dat deze opgaven overigens niet veel te moeilijk zijn; de kans ligt net onder een half om deze goed te beantwoorden.

De percentiel-50-leerling beheerst 13 van de 22 opgaven. Voorbeeldopgaven 5, 6 en 7 zijn voorbeelden van opgaven die de P50-leerling wel beheerst, maar de P25-leerling nog niet. Het betreft onder andere simpele vermenigvuldig- en deelopgaven.

De percentiel-75-leerling beheerst 17 van de 22 opgaven goed. Ten opzichte van de P50-leerling is daar bijvoorbeeld voorbeeldopgave 8 bijgekomen. Geen van de opgaven kan als te moeilijk beschouwd worden voor P75-leerlingen of leerlingen met een hogere vaardigheid.

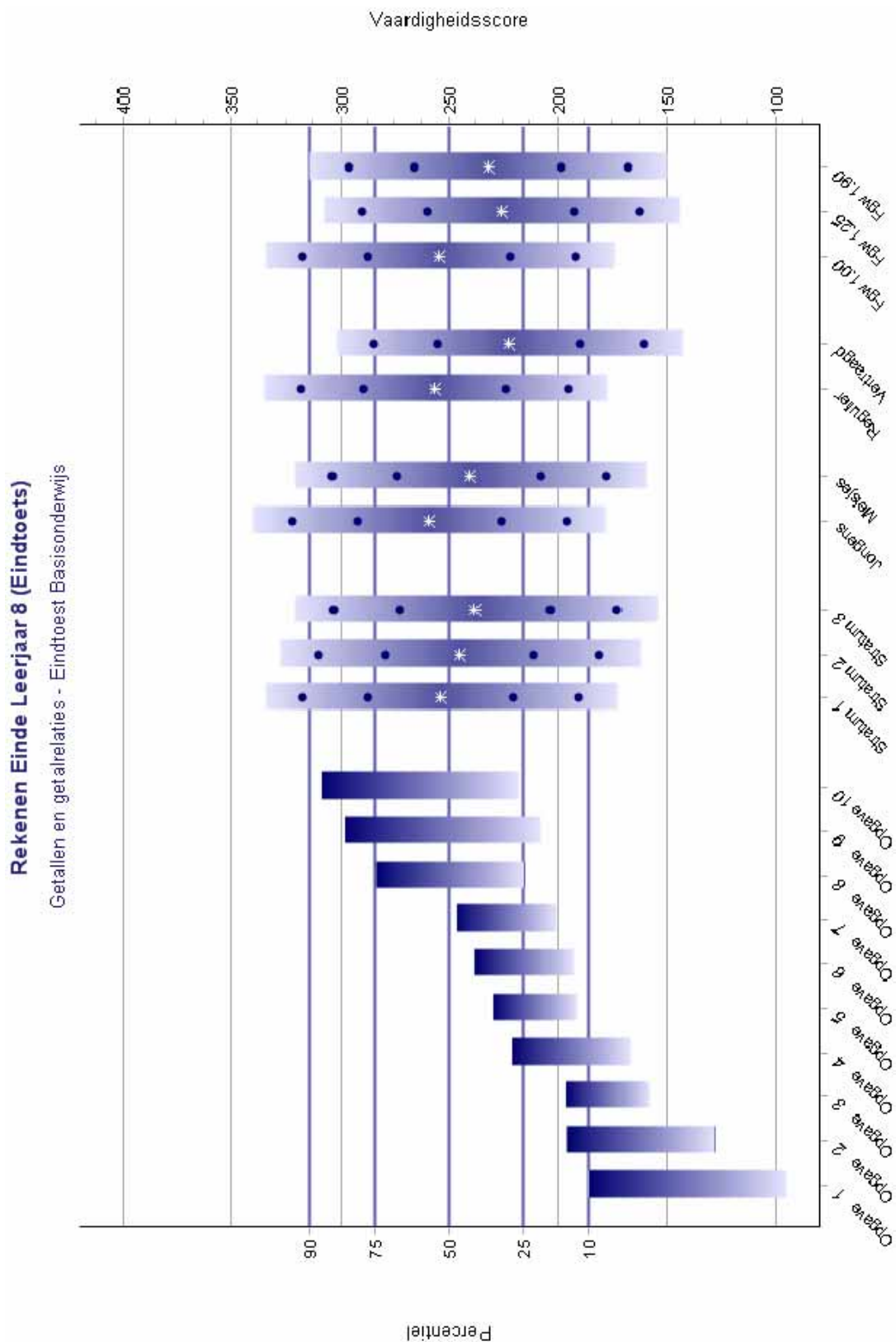
De percentiel-90-leerling beheerst alle opgaven van deze schaal (zeer) goed, dus ook opgaven 9 en 10.

Van alle drie rekenschalen is de Getallen en bewerkingenschaal de gemakkelijkste. De gemiddelde benodigde vaardigheid om de EB-opgaven goed te beheersen ligt bij deze schaal net iets onder de 250, waar het bij de andere twee schalen daar wat boven ligt. Ook de gemiddelde grens waaronder de echt moeilijke opgaven zitten (de kans kleiner dan een half om deze goed te beantwoorden), is van de rekenschalen het laagst bij Getallen en bewerkingen.

In figuur 5.4 zijn ook de gegevens voor verschillende groepen van leerlingen weergegeven. Zo zien we dat de gemiddelden van leerlingen op de stratum 1-scholen, de jongens, de niet-vertraagde ('reguliere') leerlingen, en de leerlingen met formatiegewicht 1.00 allen iets boven het algemene gemiddelde lagen. 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 en van de meisjes. 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 P25-leerling in de gehele populatie. 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 mindere rol te spelen.

Figuur 5.4 De vaardigheid Rekenen/Wiskunde – Getallen en bewerkingen



Tabel 5.4.1 Verdelingen van de leerlingen op de vaardigheid Rekenen/Wiskunde – Getallen en bewerkingen

Niveau	Percentiel 10-leerling (90% v.d. leerlingen is vaardiger)	186
	Percentiel 25-leerling (75% v.d. leerlingen is vaardiger)	216
	Percentiel 50-leerling (50% v.d. leerlingen is vaardiger)	250
	Percentiel 75-leerling (25% v.d. leerlingen is vaardiger)	284
	Percentiel 90-leerling (10% v.d. leerlingen is vaardiger)	314

Variabele		Gemiddelde	SD	(Percentiel)
Geslacht	Jongens	259	49	(57)
	Meisjes	241	49	(43)
Leertijd	Regulier	257	48	(55)
	Vertraagd	223	48	(29)
Stratum	Stratum 1	254	49	(53)
	Stratum 2	245	50	(46)
	Stratum 3	238	51	(41)
Formatiegewicht	1.00	255	49	(54)
	1.25	226	50	(32)
	1.90	232	50	(36)
Toets	EB	252	49	(51)
	NT	157	50	(3)
Doorstroom naar het VO	Vmbo-BB	175	36	(7)
	Vmbo-KB	206	36	(19)
	Vmbo-GT	232	36	(36)
	Havo	269	36	(65)
	Vwo	317	36	(91)
Regio	Noord	246	50	(47)
	Oost	249	50	(50)
	West	250	50	(50)
	Zuid	252	50	(52)
Urbanisatiegraad	Zeer sterk	245	51	(46)
	Sterk	250	50	(50)
	Matig	252	50	(52)
	Weinig	252	49	(51)
	Niet	249	49	(49)
Thuis taal	Alleen Nederlands (NLs)	252	49	(52)
	NLs en een andere taal	237	51	(40)
	Alleen een andere taal	233	51	(37)
	> Turks (eventueel met NLs)	231	50	(35)
	> Arabisch (eventueel met NLs)	230	50	(34)
	> Surinaams/Ant. (eventueel met NLs)	214	51	(24)
	> West-Europees (eventueel met NLs)	244	51	(45)

Behalve de gemiddelden die gegeven zijn in figuur 5.4 staan ook gemiddelden van de andere groepen in tabel 5.4.1. Ook daarin is te zien dat de gemiddelden voor anderstaligen zoals te verwachten minder laag liggen dan gevonden werd bij de taalschalen. Ook is te zien dat het gemiddelde van de leerlingen die direct naar vmbo-BB gaan iets onder dat van de P10-leerling in de populatie ligt en het gemiddelde van de leerling die direct naar het vwo gaat iets boven dat van de P90-leerling in de populatie.

De resultaten van het onderzoek naar de relevantie van de gecorrigeerde verschillen van die effecten is gegeven in tabel 5.4.2. Merk op dat alle effecten (klein, matig en groot) die gevonden zijn significant zijn.

Bij de vaardigheid Getallen en bewerkingen zien we dat jongens hoger scoren dan meisjes (klein effect) en 'reguliere' leerlingen over het algemeen een hogere vaardigheid hebben dan leerlingen die vertraagd zijn (matig effect). 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. 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. Het effect valt echter net onder de kwalificatie 'klein'.

Zoals bij de taalvaardigheden ook gevonden is, zien we dat ook bij rekenvaardigheid dat de leerlingen die de EB doen duidelijk vaardiger zijn dan de leerlingen die de NT doen. Het effect is hier groter dan bij taalvaardigheden. De verschillen tussen de diverse niveaus van het VO waar de leerlingen op ingeschat zijn, zijn behoorlijk groot. Merk ook op dat hier alleen verschillen tussen de opeenvolgende niveaus gegeven zijn. Het verschil tussen vmbo-BB en vwo is de som van alle effecten van de variabele 'doorstroom' (3,78; zeer groot).

Regio en urbanisatiegraad hebben bijna geen effect als er gecorrigeerd is op formatiegewicht en stratum. Er is alleen 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 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.

Tabel 5.4.2 Effectgrootte op de vaardigheid Rekenen/Wiskunde – Getallen en Bewerkingen

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	-0,43	klein
Leertijd	Vertraagd – Regulier	-0,69	matig
Stratum	Stratum2 – Stratum1	-0,09	geen effect
	Stratum3 – Stratum2	0,03	geen effect
	Stratum3 – Stratum1	-0,06	geen effect
Formatiegewicht	F125 – F100	-0,48	klein
	F190 – F125	0,19	geen effect
	F190 – F100	-0,29	klein
Toets	NT – EB	-1,74	groot
Doorstroom	KB – BB	0,90	groot
	GT – KB	0,68	matig
	havo – GT	0,94	groot
	vwo – havo	1,26	groot
Regio	Oost – Noord	0,13	geen effect
	West – Noord	0,18	geen effect
	Zuid – Noord	0,23	klein
	West – Oost	0,06	geen effect
	Zuid – Oost	0,10	geen effect
	Zuid – West	0,04	geen effect
Urbanisatiegraad	sterk – zeer sterk	-0,06	geen effect
	matig – sterk	-0,01	geen effect
	weinig – matig	-0,04	geen effect
	niet – weinig	-0,08	geen effect
Thuis taal	NLs+Ander – Alleen NLs	0,04	geen effect
	Alleen ander – Alleen NLs	0,12	geen effect
	Alleen ander – NLs+Ander	0,07	geen effect
	Turks – Alleen NLs	0,17	geen effect
	Arabisch – Alleen NLs	0,03	geen effect
	Surinaams/Antilliaans – Alleen NLs	-0,42	klein
	West-Europees – Alleen NLs	0,00	geen effect

5.5 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 bij 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.

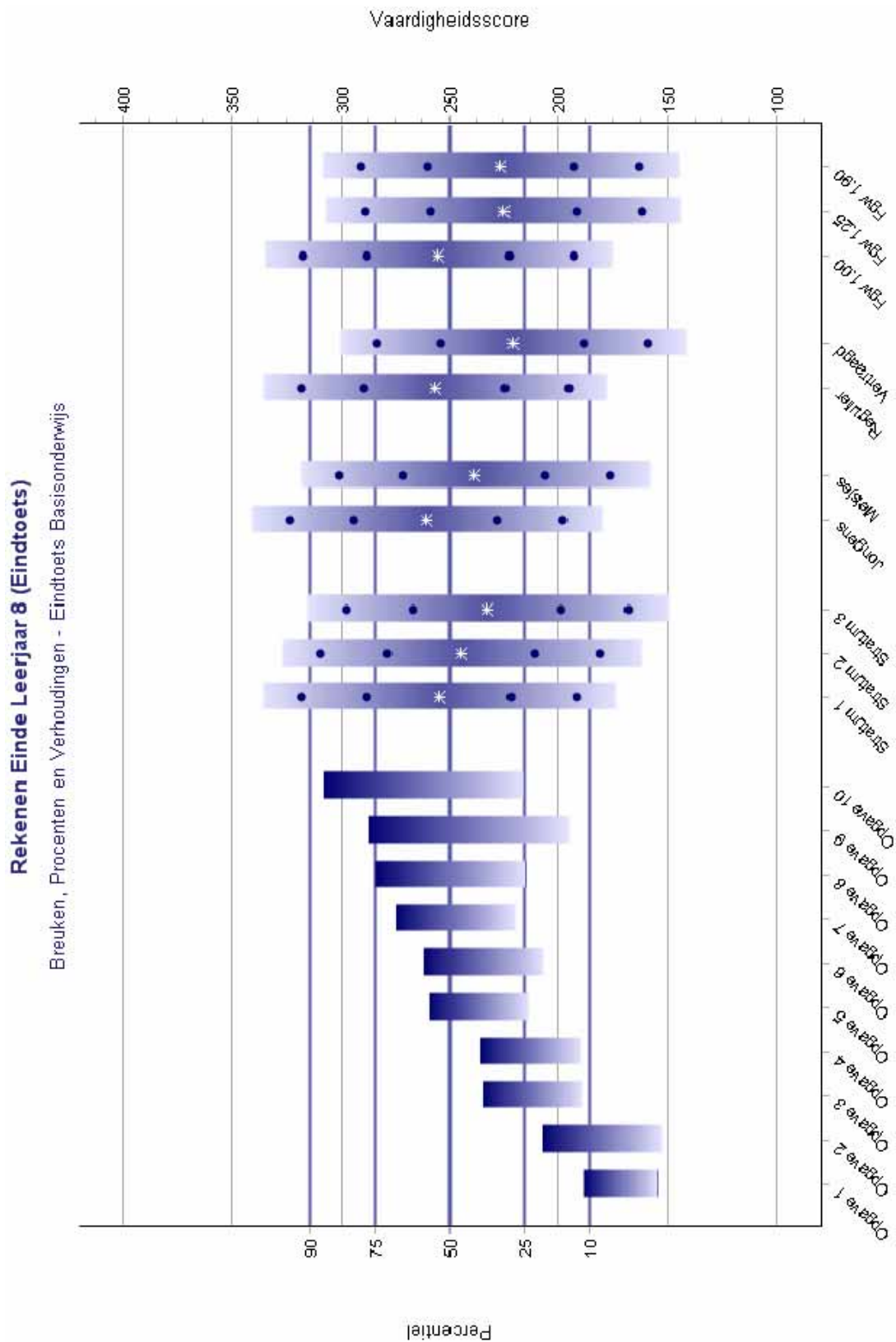
Wat leerlingen kunnen

Deze schaal is gemeten met 46 opgaven, waarvan 20 uit de NT, 19 uit de EB en 7 (nieuwe) opgaven die in de EB-anker opgenomen waren. Het onderdeel Breuken, procenten en verhoudingen telt in de EB twintig opgaven, allen met vier antwoordalternatieven. Hiervan bleek één niet goed bij de geconstrueerde schaal te passen. Deze opgave is van de schaal verwijderd, wat ook bij deze schaal niet ten koste ging van de betrouwbaarheid. Net zoals bij de schaal getallen bewerkingen zouden de conclusies overigens niet veranderen door het opnemen van deze opgave.

De schaal wordt geïllustreerd met EB-opgaven, om dezelfde reden als bij de schaal Getallen en bewerkingen. Over de NT-opgaven kan opgemerkt worden dat deze ook weer duidelijk gemakkelijker waren dan de overige opgaven. De P10-leerlingen medio jaargroep 8 hadden op alle 20 opgaven een kans groter dan een half om de opgave goed te maken en 60% van de NT-opgaven werd goed beheerst door deze leerlingen. De gemiddelde leerling (P50) beheerst 95% van deze opgaven goed. Deze makkelijke opgaven zijn vergelijkbaar met de makkelijke opgaven in de EB. De moeilijkheid van de anker opgaven is goed te vergelijken met de moeilijkheid van de EB-opgaven.

In figuur 5.5 staan ter illustratie 10 van de 19 opgaven uit de EB gegeven die voor de analyses gebruikt zijn. Deze opgaven zijn geordend op het niveau voor de goede beheersing. Dat houdt in dat de opgaven oplopen in de benodigde vaardigheid om een kans van 0,80 te hebben om de opgave goed te maken.

Figuur 5.5 De vaardigheid Rekenen/Wiskunde – Breuken, procenten en verhoudingen



Bij de eerste voorbeeldopgave moet '3 van de 4' omgezet worden in het juiste percentage. In de tweede opgave moet aangegeven worden dat 5 op de 25 gelijk is aan 1 op de 5. Bij opgave 3 moet de leerling aangeven dat 1,5 miljoen 10% is van 15 miljoen. Bij opgave 4 moet de leerling '2 van elke 3' kunnen omzetten naar het juiste percentage. Opgave 5 betreft het omzetten van 0,4 naar de breuk $\frac{2}{5}$ en opgave 6 betreft het aangeven van de plaats van $3\frac{1}{3}$ op een getallenlijn van 3 tot 4 die in 6 stukken verdeeld is. Bij deze opgave 7 moet de leerling weten hoeveel jongens 20% is, als bekend is dat 50% van het totaal gelijk is aan 2500 jongens. Een deel van de informatie bij deze opgave moest uit een taartdiagram gehaald worden. Opgave 8 betreft het ordenen van klein naar groot van breuken als $\frac{1}{3}$, $\frac{3}{4}$, $\frac{1}{2}$ en $\frac{2}{5}$. Bij opgave 9 moet de leerling bepalen hoeveel korting je ongeveer krijgt als je € 14,98 moet betalen en de oorspronkelijke prijs € 24,98 was. In de laatste voorbeeldopgave (10) is gegeven dat er met een vulmachine 7985 kuipjes gevuld worden in 5 minuten en de vraag die de leerlingen moeten beantwoorden is hoe lang het vullen van 32650 kuipjes ongeveer duurt.

Naast de afbeelding van de opgaven in figuur 5.5 is ook weer de beschrijving te geven in termen van wat de leerlingen van de diverse niveaus kunnen.

De percentiel-10-leerling beheerst alleen opgave 1 (bijna) goed: de kans op het goed beantwoorden is bijna 0,80. Het is de enige opgave in de EB die leerlingen op dat niveau beheersen. Van de overige 19 opgaven zijn er 15 te moeilijk: de kans op het goed beantwoorden is kleiner dan een half. In figuur 5.5 zijn dat opgaven 3 tot en met 10. Het zijn de elementaire opgaven die een dergelijke leerling beheerst.

De percentiel-25-leerling beheerst slechts één opgave meer goed in de EB dan een P10-leerling, en dat is opgave 2. Het aantal opgaven dat voor deze leerling te moeilijk is echter wel aardig wat minder: er is slechts een opgave die iets te moeilijk is, te weten opgave 7. Er zijn 6 opgaven waarop deze leerlingen ongeveer een kans van een half hebben om ze goed te beantwoorden. Opgave 5 en 8 zijn daar voorbeelden van. Opgaven 3 en 4 zijn voorbeelden van uitdagende opgaven voor deze leerlingen.

De percentiel-50-leerling beheerst 6 van de 19 opgaven van de EB (zeer) goed en 2 opgaven bijna goed. Opgaven 5 en 6 zijn opgaven die net iets minder goed beheerst worden door een gemiddelde leerling, maar waarvan de beheersing ook behoorlijk is (een kans van rond de 0,75 om deze opgaven goed te beantwoorden).

De percentiel-75-leerling beheerst 15 opgaven (zeer) goed, en nog eens 3 opgaven bijna goed. Opgaven 8 en 9 zitten rond de 0,80 voor deze leerlingen. De andere voorbeeldopgaven zijn gemakkelijk voor deze leerling. De enige iets moeilijker opgave voor deze leerlingen uit de EB is opgave 10.

De percentiel-90-leerling beheerst alle 19 opgaven (zeer) goed, ook opgave 10.

De vaardigheidsverdelingen van verschillende groepen leerlingen bij Breuken, procenten en verhoudingen (tabel 5.5.1) lijken veel op die bij de vaardigheid Getallen en bewerkingen gevonden zijn. De grootste verschillen zijn gevonden bij de variabele thuistaal. De gemiddelden bij anderstaligen liggen bij de vaardigheid Breuken, procenten en verhoudingen dicht tegen het gemiddelde 250 dan bij Getallen en bewerkingen het geval is.

Ook wat betreft de gecorrigeerde effectschattingen (tabel 5.5.2) zijn de verschillen bij de vaardigheid Breuken, procenten en verhoudingen vergelijkbaar met die bij Getallen en bewerkingen. Het verschil tussen jongen en meisjes is iets groter en er is geen enkel effect meer bij regio, maar veel meer echte verschillen zijn er niet tussen deze twee vaardigheden.

Tabel 5.5.1 Verdelingen van de leerlingen op de vaardigheid Rekenen/Wiskunde – Breuken, procenten en verhoudingen

Niveau	Percentiel 10-leerling (90% v.d. leerlingen is vaardiger)	186
	Percentiel 25-leerling (75% v.d. leerlingen is vaardiger)	216
	Percentiel 50-leerling (50% v.d. leerlingen is vaardiger)	250
	Percentiel 75-leerling (25% v.d. leerlingen is vaardiger)	284
	Percentiel 90-leerling (10% v.d. leerlingen is vaardiger)	314

Variabele		Gemiddelde	SD	(Percentiel)
Geslacht	Jongens	261	49	(59)
	Meisjes	239	49	(41)
Leertijd	Regulier	257	48	(56)
	Vertraagd	222	48	(29)
Stratum	Stratum 1	255	49	(54)
	Stratum 2	245	50	(46)
	Stratum 3	233	51	(37)
Formatiegewicht	1.00	256	49	(54)
	1.25	226	49	(31)
	1.90	227	50	(32)
Toets	EB	252	49	(51)
	NT	157	50	(3)
Doorstroom naar het VO	Vmbo-BB	174	37	(7)
	Vmbo-KB	208	37	(20)
	Vmbo-GT	234	37	(37)
	Havo	269	37	(65)
	Vwo	314	36	(90)
Regio	Noord	244	50	(45)
	Oost	249	50	(49)
	West	250	50	(50)
	Zuid	253	50	(53)
Urbanisatiegraad	Zeer sterk	247	51	(48)
	Sterk	250	50	(50)
	Matig	252	50	(51)
	Weinig	251	50	(51)
	Niet	248	50	(48)
Thuis taal	Alleen Nederlands (NLs)	252	50	(51)
	NLs en een andere taal	242	51	(44)
	Alleen een andere taal	239	51	(41)
	> Turks (eventueel met NLs)	240	50	(42)
	> Arabisch (eventueel met NLs)	235	50	(38)
	> Surinaams/Ant. (eventueel met NLs)	219	51	(26)
	> West-Europees (eventueel met NLs)	246	51	(47)

Tabel 5.5.2 Effectgrootte op de vaardigheid Rekenen/Wiskunde – Breuken, procenten en verhoudingen

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	-0,52	matig
Leertijd	Vertraagd – Regulier	-0,72	matig
Stratum	Stratum2 – Stratum1	-0,11	geen effect
	Stratum3 – Stratum2	-0,04	geen effect
	Stratum3 – Stratum1	-0,16	geen effect
Formatiegewicht	F125 – F100	-0,50	klein
	F190 – F125	0,13	geen effect
	F190 – F100	-0,36	klein
Toets	NT – EB	-1,75	groot
Doorstroom	KB – BB	0,83	groot
	GT – KB	0,73	matig
	havo – GT	1,00	groot
	vwo – havo	1,36	groot
Regio	Oost – Noord	0,10	geen effect
	West – Noord	0,16	geen effect
	Zuid – Noord	0,16	geen effect
	West – Oost	0,06	geen effect
	Zuid – Oost	0,06	geen effect
	Zuid – West	0,00	geen effect
Urbanisatiegraad	sterk – zeer sterk	-0,07	geen effect
	matig – sterk	0,00	geen effect
	weinig – matig	-0,04	geen effect
	niet – weinig	-0,08	geen effect
Thuis taal	NLs+Ander – Alleen NLs	-0,02	geen effect
	Alleen ander – Alleen NLs	0,07	geen effect
	Alleen ander – NLs+Ander	0,09	geen effect
	Turks – Alleen NLs	0,04	geen effect
	Arabisch – Alleen NLs	0,00	geen effect
	Surinaams/Antilliaans – Alleen NLs	-0,47	klein
	West-Europees – Alleen NLs	-0,03	geen effect

5.6 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 het 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 toepassingsituaties.

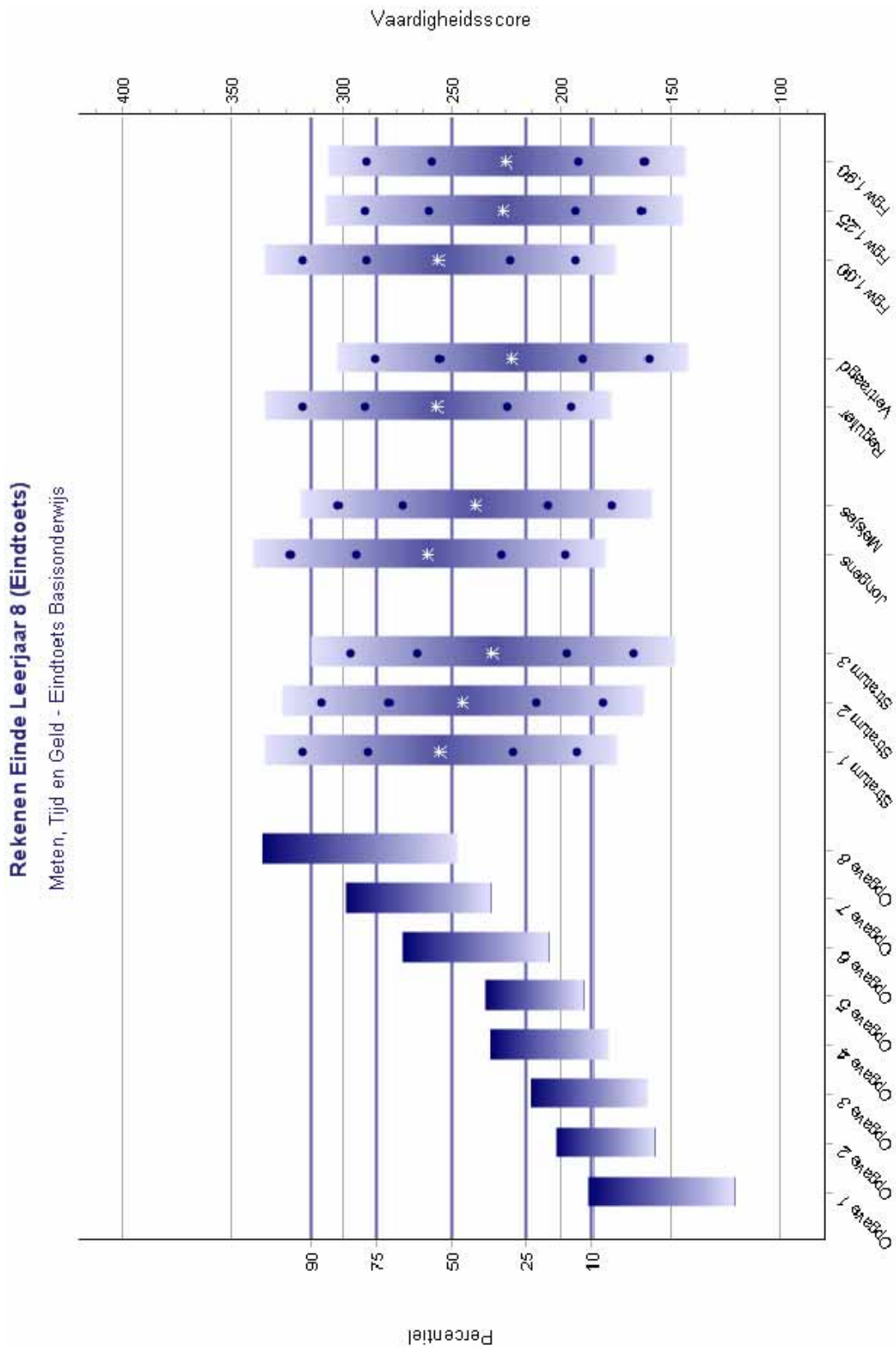
Geld. Hierbij gaat het om toepassingsgericht rekenen met geld waarbij specifieke handelingen met munten en bankbiljetten uitgevoerd moeten worden.

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 waarmee deze vaardigheid is gemeten bevatte ondanks de breedte van het onderwerp met 37 opgaven de minste opgaven. Het had wel de meeste ankeropgaven (8) die allen ook op de schaal pasten. Ook bij deze schaal was er één EB-opgave die niet op de schaal paste en ook niet in de analyses is opgenomen. Dit had ook weer weinig effect op de betrouwbaarheid en de conclusies. Alle 15 opgaven van de NT pasten wel op de schaal. Net als bij de overige schalen waren deze NT-opgaven weer gemakkelijk. Alle 15 opgaven waren voor een P10-leerling met een kans van 0,50 goed te beantwoorden. Het percentage opgaven dat door de P10-leerling goed beheerst wordt lag echter wat lager: 5 van de 15 opgaven konden met een kans van 0,80 of meer gemaakt worden. Dat gold voor de P25-leerling voor 10 van de 15 opgaven. Een gemiddelde leerling (P50) had nog geen goede beheersing op 20% van de NT-opgaven, maar een P75-leerling beheerst alle opgaven goed. Hiermee was het een relatief moeilijke NT-toets. Wat dit echter inhoudelijk betekent, kan pas bekeken worden als naar de inhoud van de opgaven gekeken wordt. Dat gebeurt weer door EB-opgaven op de referentiepercentielen te beschouwen. Doordat het een kleine schaal is, betreft dit nu 8 in plaats van 10 voorbeeldopgaven.

Figuur 5.6 De vaardigheid Rekenen/Wiskunde – Meten, meetkunde, tijd en geld



De eerste voorbeeldopgave in figuur 5.6 betreft het herkennen van het feit dat het, als het 20:25 uur is, nog 3 uur en 35 minuten duurt voordat het middernacht is. In opgave 2 moet de leerling aangeven hoeveel bakjes met 250 gram kunnen worden gevuld als er in totaal 5 kg beschikbaar is. Opgave 3 vraagt hoeveel glazen van 20 cl gevuld kunnen worden als er in totaal 10 liter is, opgave 4 vraagt aan te geven dat 250 ml evenveel is als $\frac{1}{4}$ liter, en opgave 5 vraagt de inhoud te bepalen van een bakje van 5 dm bij 1 dm bij 3,5 dm in kubieke decimeters. In voorbeeldopgave 6 wordt een veld van 20 bij 10 meter geheel opgedeeld in 20 kleinere velden van gelijke grootte, waarbij de vraag is wat de oppervlakte is van een dergelijk veldje. Opgave 7 was een geldopgave. De leerling moet aangeven hoeveel je bij een briefje van 50 euro moet leggen om 5 euro terug te krijgen als je € 45,15 moet betalen. In de laatste voorbeeldopgave (8) moeten de leerlingen bepalen hoeveel tegels van 50 cm bij 50 cm nodig zijn voor de vloer van een rechthoekige kamer van 3 meter bij 2,50 meter. Wat de leerlingen kunnen wordt weer weergegeven per niveau.

De percentiel-10-leerling beheerst slechts een opgave van de EB (bijna) goed: opgave 1. Van de 14 EB opgaven waren er vier duidelijk te moeilijk zoals opgave 6, 7 en 8, en twee iets te moeilijk, waaronder opgave 5. De overige opgaven, zoals 2, 3 en 4 zijn uitdagend voor deze leerlingen.

De percentiel-25-leerling beheerst opgave 1 tot en met 3 goed. Er is ook een item in de EB dat een leerling op percentiel 25 niveau bijna goed beheerst. Twee van de opgaven, waaronder opgave 7 en 8 in Figuur 5.6 zijn te moeilijk voor een dergelijke leerling. De overige opgaven zijn als uitdagend te beoordelen, waarbij opgave 4 aan de makkelijke kant is en opgave 6 aan de moeilijke kant.

De percentiel-50-leerling beheerst acht EB opgaven (zeer) goed. Daarvan zijn opgave 1 tot en met 5 voorbeelden. Geen van de overige zes opgaven is te moeilijk, al is opgave 8, als enige in de EB, aan de moeilijke kant voor de gemiddelde leerling.

De percentiel-75-leerling beheerst negen opgaven (zeer) goed en de andere opgaven van ruim voldoende tot redelijk. Geen van de opgaven was voor deze leerlingen moeilijk te noemen.

De percentiel-90-leerling beheerst bijna al de 14 opgaven (zeer) goed. Er zijn twee opgaven waarbij dergelijke leerlingen een kans net iets lager dan 0,80 hebben om die opgaven goed te maken, waarbij opgave 8 de moeilijkste was.

Deze schaal was wat betreft de moeilijkheid vergelijkbaar met de andere twee rekenschalen. Ook de verdeling van de leerlingen verschilde niet veel van de resultaten die bij die schalen gevonden waren. De verdelingen van de verschillende groepen bij Meten, meetkunde, tijd en geld staan in tabel 5.6.1. De verdelingen van de variabelen geslacht, leertijd, stratum en formatiegewicht zijn vrijwel gelijk aan die bij Breuken, procenten en verhoudingen. De verdelingen van de diverse thuistaalgroepen lijkt meer op die bij Getallen en bewerkingen. Dat houdt in dat de gemiddelde vaardigheden van anderstalige groepen iets verder van het gemiddelde afliggen dan bij Breuken, procenten en verhoudingen.

De gecorrigeerde effecten bij Meten, meetkunde, tijd en geld (tabel 5.6.2) lijken het meest op de effecten die ook gevonden werden bij Breuken, procenten en verhoudingen. Het verschil tussen jongens en meisjes is hier zelfs nog iets groter. Ook groter is het effect van vwo tegenover havo. Van alle rekenschalen is het effect op dit onderdeel het grootst. Dat geldt ook voor het effect tussen vmbo-BB en vwo. Ook bij deze vaardigheid waren de effecten voor regio en urbanisatiegraad afwezig. Taaleffecten waren ook afwezig, waarbij Surinaams/Antilliaans andermaal weer de enige uitzondering was.

Tabel 5.6 Verdelingen van de leerlingen op de vaardigheid Rekenen/Wiskunde – Meten, meetkunde, tijd en geld

Niveau	Percentiel 10-leerling (90% v.d. leerlingen is vaardiger)	186
	Percentiel 25-leerling (75% v.d. leerlingen is vaardiger)	216
	Percentiel 50-leerling (50% v.d. leerlingen is vaardiger)	250
	Percentiel 75-leerling (25% v.d. leerlingen is vaardiger)	284
	Percentiel 90-leerling (10% v.d. leerlingen is vaardiger)	314

Variabele		Gemiddelde	SD	(Percentiel)
Geslacht	Jongens	261	49	(59)
	Meisjes	239	49	(42)
Leertijd	Regulier	257	48	(55)
	Vertraagd	223	49	(29)
Stratum	Stratum 1	255	49	(54)
	Stratum 2	245	50	(46)
	Stratum 3	232	50	(36)
Formatiegewicht	1.00	256	49	(55)
	1.25	227	49	(32)
	1.90	225	50	(31)
Toets	EB	252	49	(51)
	NT	159	50	(3)
Doorstroom naar het VO	Vmbo-BB	178	36	(8)
	Vmbo-KB	208	36	(20)
	Vmbo-GT	231	35	(35)
	Havo	267	35	(63)
	Vwo	321	35	(92)
Regio	Noord	245	50	(46)
	Oost	249	50	(49)
	West	249	50	(49)
	Zuid	254	50	(53)
Urbanisatiegraad	Zeer sterk	242	51	(44)
	Sterk	250	50	(50)
	Matig	252	50	(51)
	Weinig	252	49	(52)
	Niet	250	49	(50)
Thuis taal	Alleen Nederlands (NLs)	252	49	(52)
	NLs en een andere taal	236	51	(39)
	Alleen een andere taal	232	50	(36)
	> Turks (eventueel met NLs)	230	50	(34)
	> Arabisch (eventueel met NLs)	226	50	(31)
	> Surinaams/Ant. (eventueel met NLs)	211	51	(22)
	> West-Europees (eventueel met NLs)	243	51	(45)

Tabel 5.6.2 Effectgrootte op de vaardigheid Rekenen/Wiskunde – Meten, meetkunde, tijd en geld

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	-0,50	klein
Leertijd	Vertraagd – Regulier	-0,69	matig
Stratum	Stratum2 – Stratum1	-0,12	geen effect
	Stratum3 – Stratum2	-0,06	geen effect
	Stratum3 – Stratum1	-0,18	geen effect
Formatiegewicht	F125 – F100	-0,49	klein
	F190 – F125	0,09	geen effect
	F190 – F100	-0,40	klein
Toets	NT – EB	-1,72	groot
Doorstroom	KB – BB	0,81	groot
	GT – KB	0,61	matig
	havo – GT	1,02	groot
	vwo – havo	1,57	groot
Regio	Oost – Noord	0,12	geen effect
	West – Noord	0,17	geen effect
	Zuid – Noord	0,23	klein
	West – Oost	0,05	geen effect
	Zuid – Oost	0,10	geen effect
	Zuid – West	0,05	geen effect
Urbanisatiegraad	sterk – zeer sterk	-0,01	geen effect
	matig – sterk	-0,02	geen effect
	weinig – matig	-0,02	geen effect
	niet – weinig	-0,07	geen effect
Thuis taal	NLs+Ander – Alleen NLs	-0,03	geen effect
	Alleen ander – Alleen NLs	0,04	geen effect
	Alleen ander – NLs+Ander	0,07	geen effect
	Turks – Alleen NLs	0,02	geen effect
	Arabisch – Alleen NLs	-0,09	geen effect
	Surinaams/Antilliaans – Alleen NLs	-0,53	matig
	West-Europees – Alleen NLs	-0,05	geen effect

6 Resultaten van leerlingen in jaargroep 4

De presentatie van de resultaten van de leerlingen in jaargroep 4 komt in grote lijnen overeen met hoe de resultaten voor jaargroep 8 weergegeven worden. Voor het overzicht van de opgaven, verdelingen van de te onderscheiden groepen binnen jaargroep 4 en de geschatte effecten wordt worden dezelfde type figuren en tabellen gegeven. De verslaglegging van de zeven verschillende vaardigheden die gemeten zijn in jaargroep 4 is minder uitgebreid dan die van jaargroep 8 en zal zich meer beperken tot het presenteren van deze figuren en tabellen. Naast een minder uitgebreide uitwerking van de resultaten zijn er nog wat meer kleine verschillen tussen de analyses in beide groepen. Deze verschillen zijn behandeld in hoofdstuk 4 (paragraaf 4.4 Presentatie van de resultaten bij leerjaar 4).

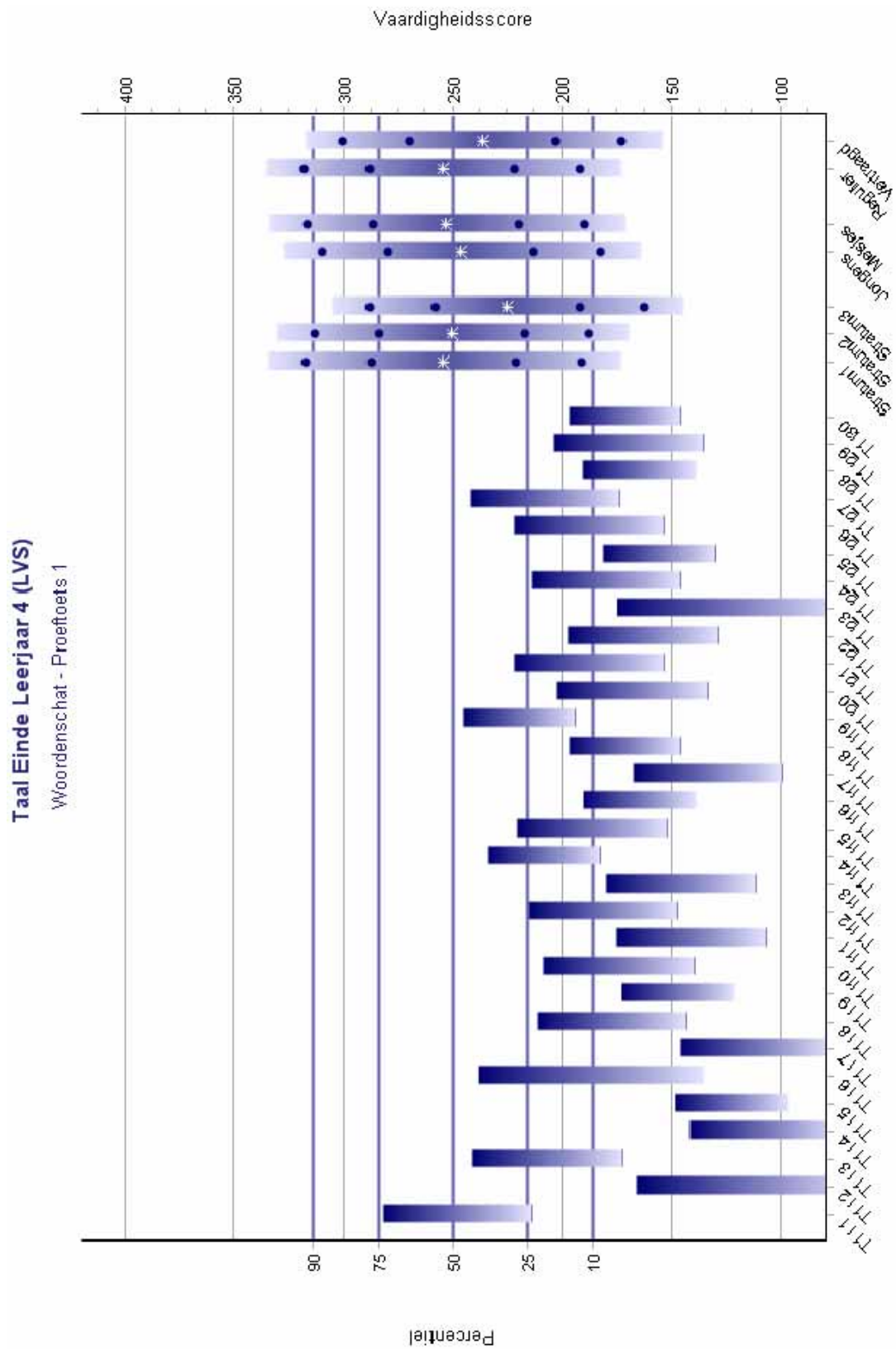
Een belangrijke opmerking uit dat hoofdstuk willen we graag herhalen bij de presentatie van de resultaten in jaargroep 4. Dat betreft het feit dat de schaal van de vaardigheden voor leerjaar 4 net zoals bij leerjaar 8, op een schaal gezet is waarbij het gemiddelde van de schaal gelijk gesteld is aan 250, en de standaarddeviatie gelijk aan 50. Dat maakt de gevonden schaalwaarden gemakkelijk interpreteerbaar. Daarbij moet wel een waarschuwing worden gegeven. De schaalwaarden voor de leerlingen in jaargroep 4 zijn niet rechtstreeks vergelijkbaar met de schaalwaarden zoals gevonden in leerjaar 8. Het moet duidelijk zijn dat dit gemiddelde van 250 niet exact dezelfde inhoudelijke vaardigheid betreft als in jaargroep 8. Een vergelijking tussen jaargroep 4 en jaargroep 8 is wel mogelijk door de gezuiverde effecten van de achtergrondvariabelen met elkaar te vergelijken op de verschillende schalen in beide onderzoekspopulaties. Die vergelijking wordt gegeven in hoofdstuk 7.

6.1 Nederlandse taal: Woordenschat

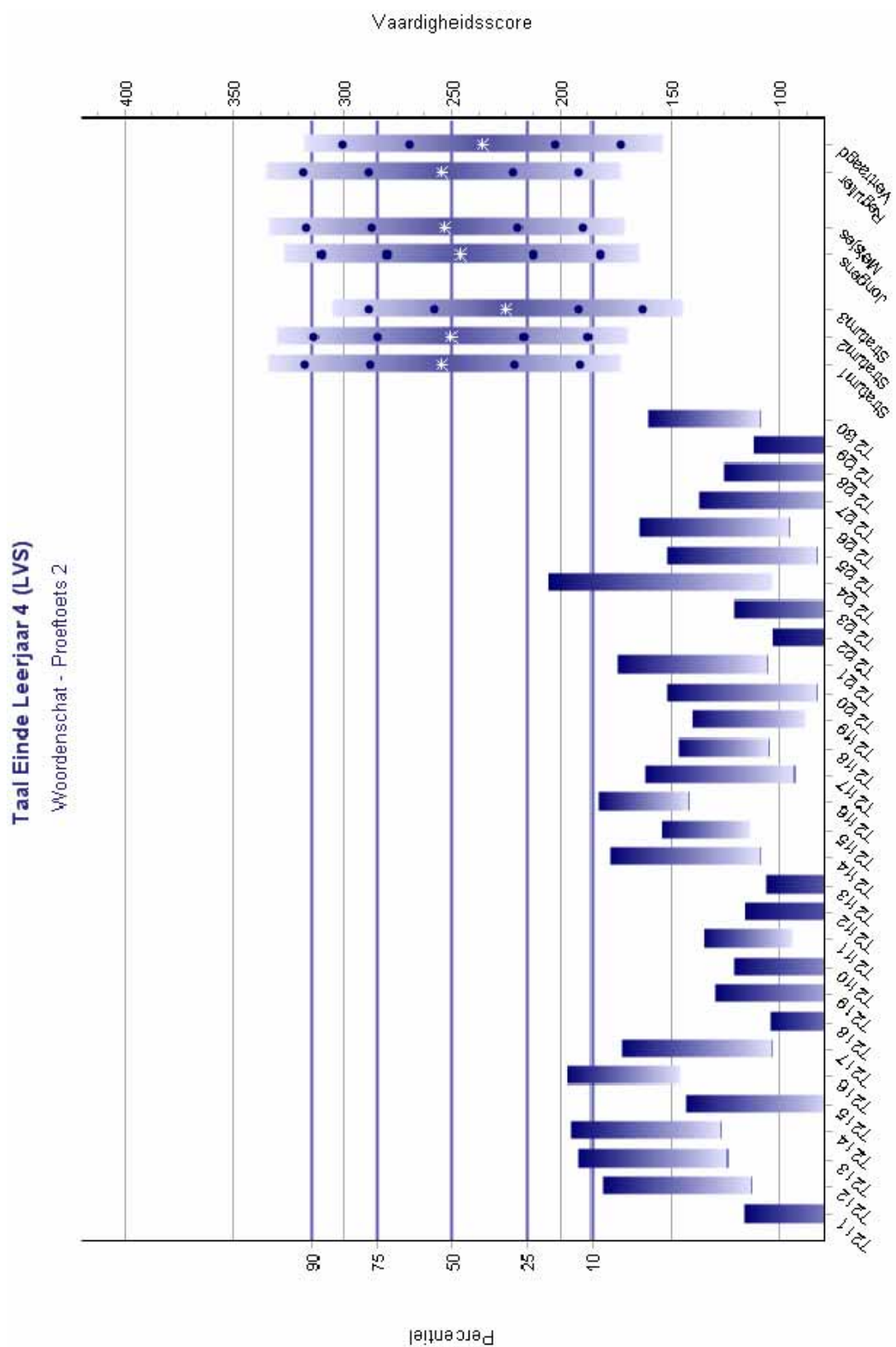
In tegenstelling tot bij leerjaar 8 was het formatiegewicht van de leerlingen uit jaargroep 4 die aan de proeftoetsen voor de vaardigheid Woordenschat meededen niet bekend. Variabelen die wel in het onderzoek meegenomen konden worden waren stratum, geslacht en leertijd. Deze zijn ook gegeven in figuur 6.1 a tot en met e. Het is opvallend dat veel van de opgaven aan de onderkant van de schaal liggen en de meeste proeftoetsen aan de gemakkelijke kant zijn. Zie hiervoor ook tabel 6.1.3.

Naast de variabelen stratum, geslacht en leertijd zijn ook regio en urbanisatiegraad bekend. De verdelingen van de verschillende groepen over die variabelen zijn ook gegeven in tabel 6.1.1. De effecten van de verschillende variabelen zijn gegeven in tabel 6.1.2. De wijze waarop de figuur en de tabellen geïnterpreteerd kunnen worden, is vergelijkbaar met de interpretatie voor leerjaar 8. De verdeling van de moeilijkheid van de opgaven per percentielniveau en per toets is gegeven in tabel 6.1.3. Daarin staat ook de verdeling van de moeilijkheden van de alle 150 opgaven waarmee de vaardigheid Woordenschat is gemeten.

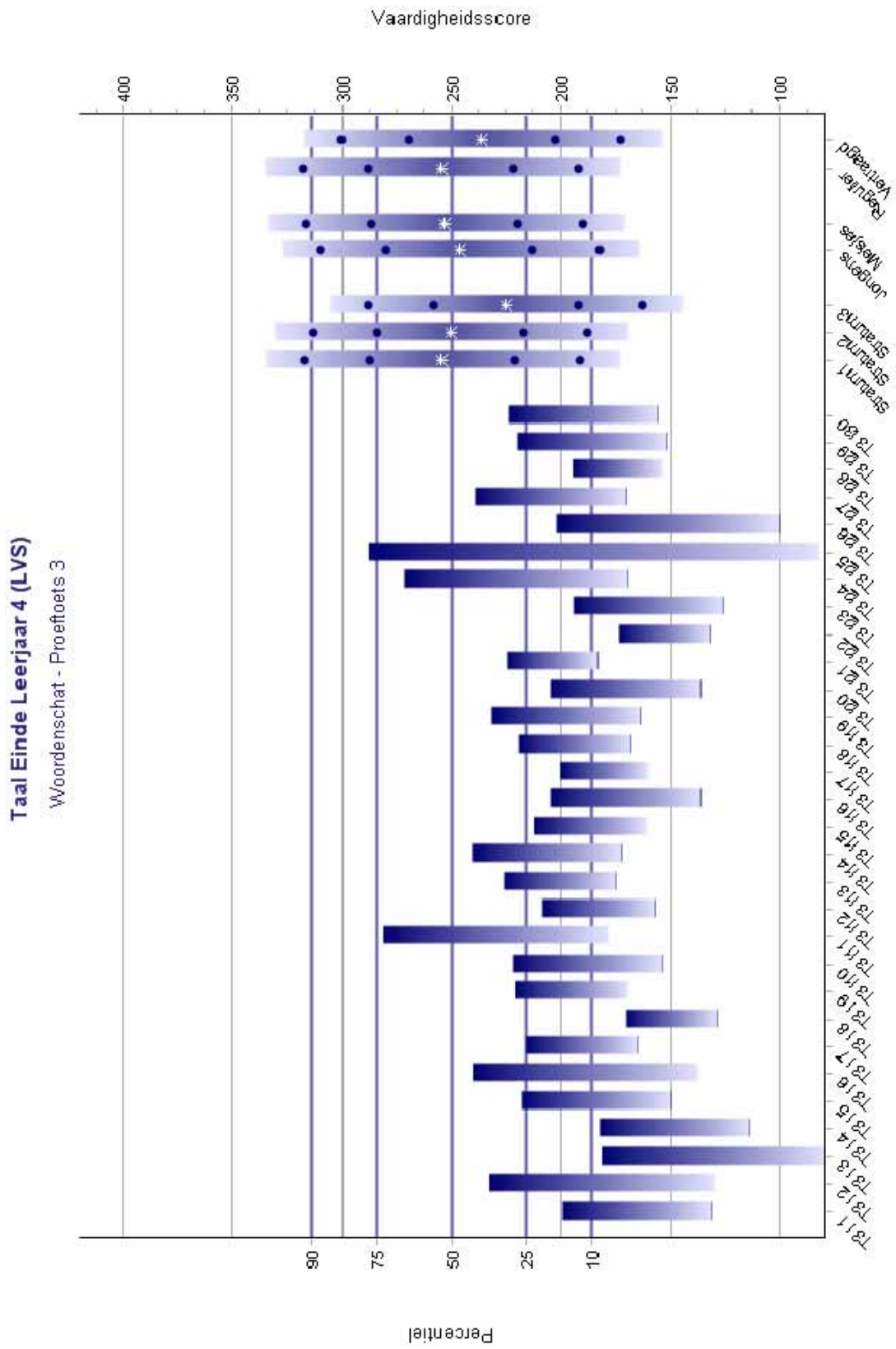
Figuur 6.1.a Verdeling van de opgaven van de eerste proeftoets Woordenschat voor leerjaar 4, geordend naar de positie binnen deze proeftoets



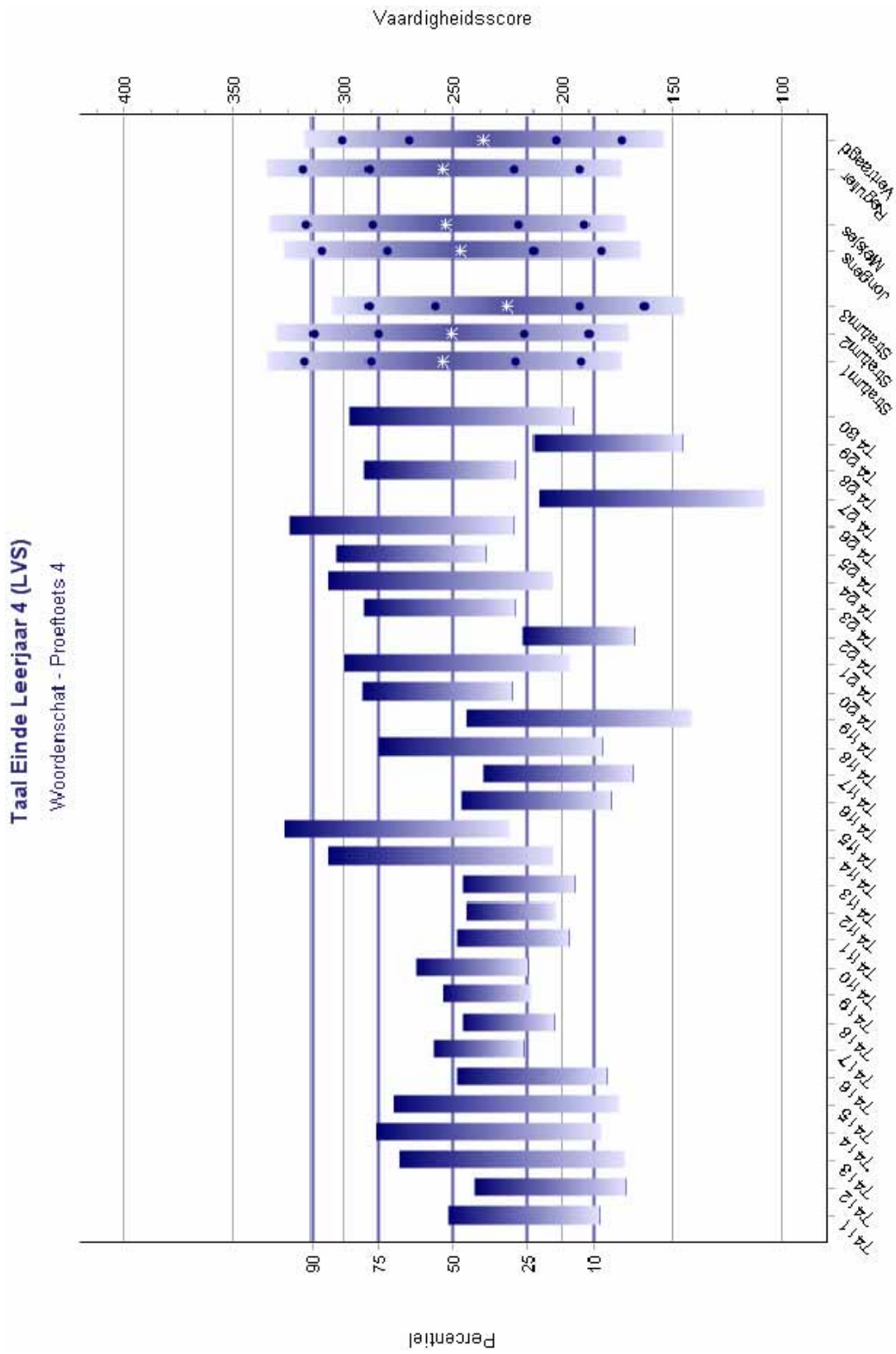
Figuur 6.1.b Verdeling van de opgaven van de tweede proeftoets Woordenschat voor leerjaar 4, geordend naar de positie binnen deze proeftoets



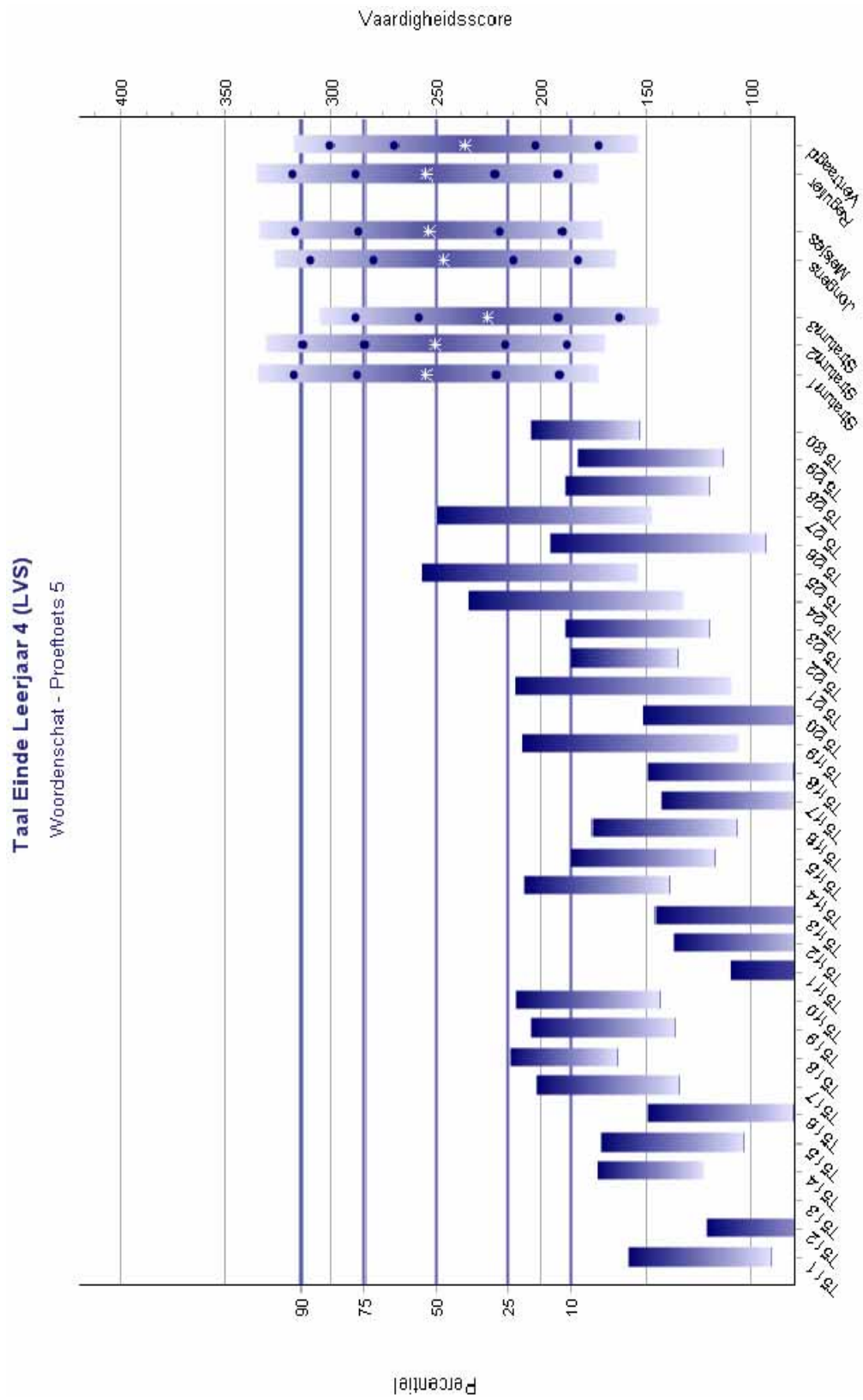
Figuur 6.1.c Verdeling van de opgaven van de derde proeftoets Woordenschat voor leerjaar 4, geordend naar de positie binnen deze proeftoets



Figuur 6.1.d Verdeling van de opgaven van de vierde proeftoets Woordenschat voor leerjaar 4, geordend naar de positie binnen deze proeftoets



Figuur 6.1.e Verdeling van de opgaven van de vijfde proeftoets Woordenschat voor leerjaar 4, geordend naar de positie binnen deze proeftoets



Tabel 6.1 Verdelingen van de leerlingen op de vaardigheid Woordenschat – leerjaar 4

Niveau	Percentiel 10-leerling (90% v.d. LJ4 lln. is vaardiger)	186
	Percentiel 25-leerling (75% v.d. LJ4 lln. is vaardiger)	216
	Percentiel 50-leerling (50% v.d. LJ4 lln. is vaardiger)	250
	Percentiel 75-leerling (25% v.d. LJ4 lln. is vaardiger)	284
	Percentiel 90-leerling (10% v.d. LJ4 lln. is vaardiger)	314

Variabele		Gemiddelde	SD	(LJ 4 Percentiel)
Geslacht	Jongens	246	50	(47)
	Meisjes	253	50	(53)
Leertijd	Regulier	255	49	(54)
	Vertraagd	236	50	(39)
Stratum	Stratum 1	254	49	(54)
	Stratum 2	251	49	(51)
	Stratum 3	225	49	(31)
Regio	Noord	244	49	(46)
	Oost	255	50	(54)
	West	250	50	(50)
	Zuid	248	51	(48)
Urbanisatiegraad	Zeer sterk	239	52	(41)
	Sterk	236	50	(39)
	Matig	247	49	(48)
	Weinig	254	49	(54)
	Niet	260	50	(58)

Tabel 6.1.2 Effectgrootte op de vaardigheid Nederlandse taal – Woordenschat in leerjaar 4

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	0,13	niet significant
Leertijd	Vertraagd – Regulier	-0,39 **	klein effect
Stratum	Stratum2 – Stratum1	-0,01	niet significant
	Stratum3 – Stratum2	-0,53 **	matig effect
	Stratum3 – Stratum1	-0,55 ***	matig effect
Regio	Oost – Noord	0,27	niet significant
	West – Noord	0,07	niet significant
	Zuid – Noord	0,32	niet significant
	West – Oost	-0,20	niet significant
	Zuid – Oost	0,05	niet significant
	Zuid – West	0,25	niet significant
Urbanisatiegraad	sterk – zeer sterk	-0,24	niet significant
	matig – sterk	-0,03	niet significant
	weinig – matig	-0,02	niet significant
	niet – weinig	0,04	niet significant

Overschrijdingskansen: * : ,05; ** : ,01; *** : ,000001

In tegenstelling tot leerjaar 8, waar jongens hoger op Woordenschat scoren, hebben in leerjaar 4 meisjes een vaardigheid iets boven het gemiddelde. Bij de effectschattingen is het effect voor geslacht echter niet significant en zo klein dat het als afwezig beschouwd kan worden. De enige significante effecten betreffen hier leertijd en stratum, waarbij er een matig effect is van stratum 1 ten opzichte van zowel stratum 2 als stratum 3. Hierbij moet wel opgemerkt worden dat hier niet, in tegenstelling tot de andere vaardigheden, gecorrigeerd is voor formatiegewicht. Het is aannemelijk dat deze stratumeffecten (duidelijk) kleiner zouden zijn als formatiegewicht wel bekend zou zijn.

Tabel 6.1.3 Verdeling van de opgaven over de toetsen en percentielgroepen: aantallen opgaven van de toetsen Woordenschat in leerjaar 4

	W1		W2		W3		W4		W5		WS	
	P <	P >	P <	P >	P <	P >	P <	P >	P <	P >	P <	P >
	0,5	0,8	0,5	0,8	0,5	0,8	0,5	0,8	0,5	0,8	0,5	0,8
P10	2	10	0	26	0	4	17	0	0	14	19	54
P25	0	21	0	30	0	13	7	2	0	27	7	93
P50	0	29	0	30	0	27	0	12	0	28	0	126
P75	0	30	0	30	0	29	0	18	0	30	0	137
P90	0	30	0	30	0	30	0	28	0	30	0	148
Totaal	30	30	30	30	30	30	30	30	30	30	150	150

Uit tabel 6.1.3 is af te leiden dat een leerling op P10-niveau (P10) bij 19 van de 150 opgaven een kans kleiner dan een half heeft ($P < 0,5$) om die opgaven goed te maken. Het aantal opgaven dat voor een dergelijke leerling in leerjaar 4 als moeilijk wordt beschouwd is dus gelijk aan 19. Bij 54 van de 150 opgaven

hebben leerlingen op dat niveau een goede beheersing, aangezien de kans op succes bij 54 opgaven groter is dan 0,80 ($P > 0,8$). Voor leerlingen op P25-niveau is het aantal moeilijke opgaven gelijk aan 7 en het aantal beheerste is opgaven 93. Vanaf het niveau P50 is er geen enkele opgave meer als moeilijk te classificeren. Voor een leerling op P90 zijn er nog twee opgaven die niet goed beheerst worden. Deze verdelingen zijn ook terug te vinden in de figuren 6.1 a tot en met e.

6.2 Nederlandse taal: Spelling

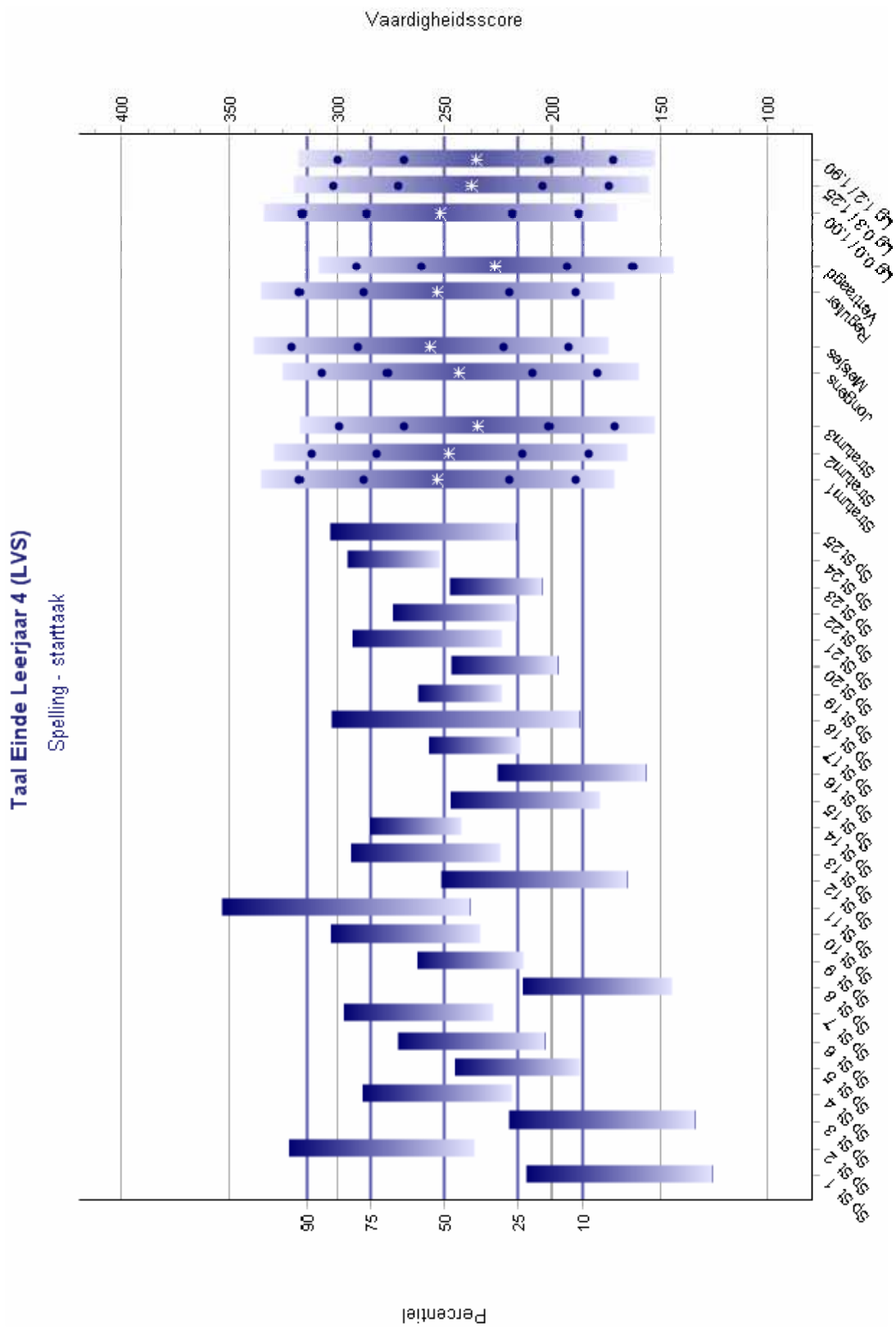
Bij de vaardigheid Spelling waren meer achtergrondgegevens beschikbaar dan bij Woordenschat. Formatiegewicht en thuistaal zijn bij de analyses van deze vaardigheid in jaargroep 4 wel beschikbaar. Bij een deel van de leerlingen zijn de oude gewichten bekend, bij een ander deel de nieuwe. In de presentatie van de resultaten zijn de formatiegewichten teruggebracht naar drie groepen: '0.0 of 1.00', '0.3 of 1.25' en '1.2 of 1.90'. De verschillen in vaardigheid binnen de groep '0.0 of 1.00' zijn verwaarloosbaar, maar bij de andere twee groepen zijn de verschillen iets groter. Hierover meer bij de beschrijving van tabel 6.2.2. In figuur 6.2.a worden de opgaven afgebeeld van de starttoets Spelling, in figuur 6.2.b de opgaven van de gemakkelijke Vervoltoets 1 en in figuur 6.2 c die van de moeilijke Vervoltoets 2. Het verschil in moeilijkheid van de toets is duidelijk te zien in de figuren en komt ook naar voren in tabel 6.2.1, waar de verdeling gegeven is van de moeilijkheid van de alle 75 opgaven waarmee de vaardigheid Spelling gemeten is in jaargroep 4. De verdelingen van de verschillende groepen zijn gegeven in tabel 6.2.2. De verdelingen van geslacht, leertijd, stratum en formatiegewicht zijn ook afgebeeld in de figuren. De effecten van de variabelen staan in tabel 6.2.3.

Er moet opgemerkt worden dat de definitie van de formatiegewichten, door de combinatie van oude en nieuwe gewichten die gebruikt is, niet eenduidig is. De verschillen tussen het oude en het nieuwe gewicht zijn bij vergelijking tussen de gewichten 0.0 en 1.00 zeer klein. Maar bij de gewichten 0.3 en 1.25 en bij 1.2 en 1.90 zijn wel kleine effecten gevonden. In deze steekproef waren de leerlingen met het nieuwe gewicht 0.30 iets vaardiger dan die met gewicht oude gewicht 1.25. Bij 1.2 en 1.90 werd juist gevonden dat de leerlingen met het oude gewicht (1.90) iets beter presteerden dan die met nieuw gewicht 1.2. De eigenschappen van de leerling zijn natuurlijk anders bij de oude en de nieuwe gewichten, en vandaar uit zijn die verschillen hier ook wel te verklaren. Het aantal leerlingen was echter te beperkt om daarover zeer sterke uitspraken te doen. Deze resultaten worden zodoende ook meer als indicatief beschouwd.

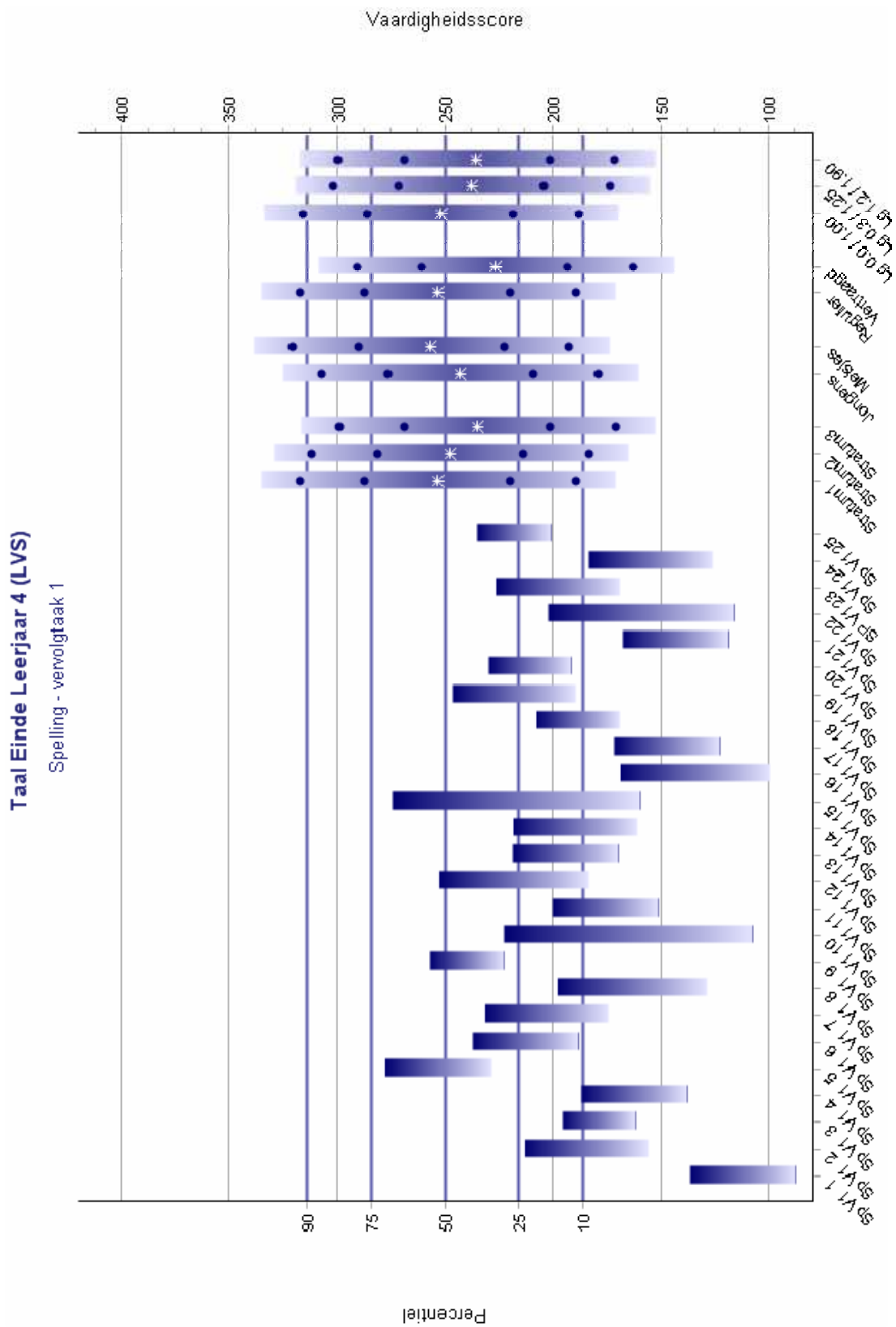
Tabel 6.2.1 Verdeling van het aantal opgaven voor de vaardigheid Spelling over de Starttoets (St), Vervoltoets 1 (V1) en Vervoltoets 2 (V2) en over de percentiële niveaus in leerjaar 4

	SP St		SP V1		SP V2		SP	
	P <	P >	P <	P >	P <	P >	P <	P >
	0,5	0,8	0,5	0,8	0,5	0,8	0,5	0,8
P10	19	0	6	5	24	0	49	5
P25	12	2	2	12	21	0	35	14
P75	1	8	0	21	7	2	8	31
P50	0	14	0	25	1	13	1	52
P90	0	23	0	25	1	19	1	67
Totaal	25	25	25	25	25	25	75	75

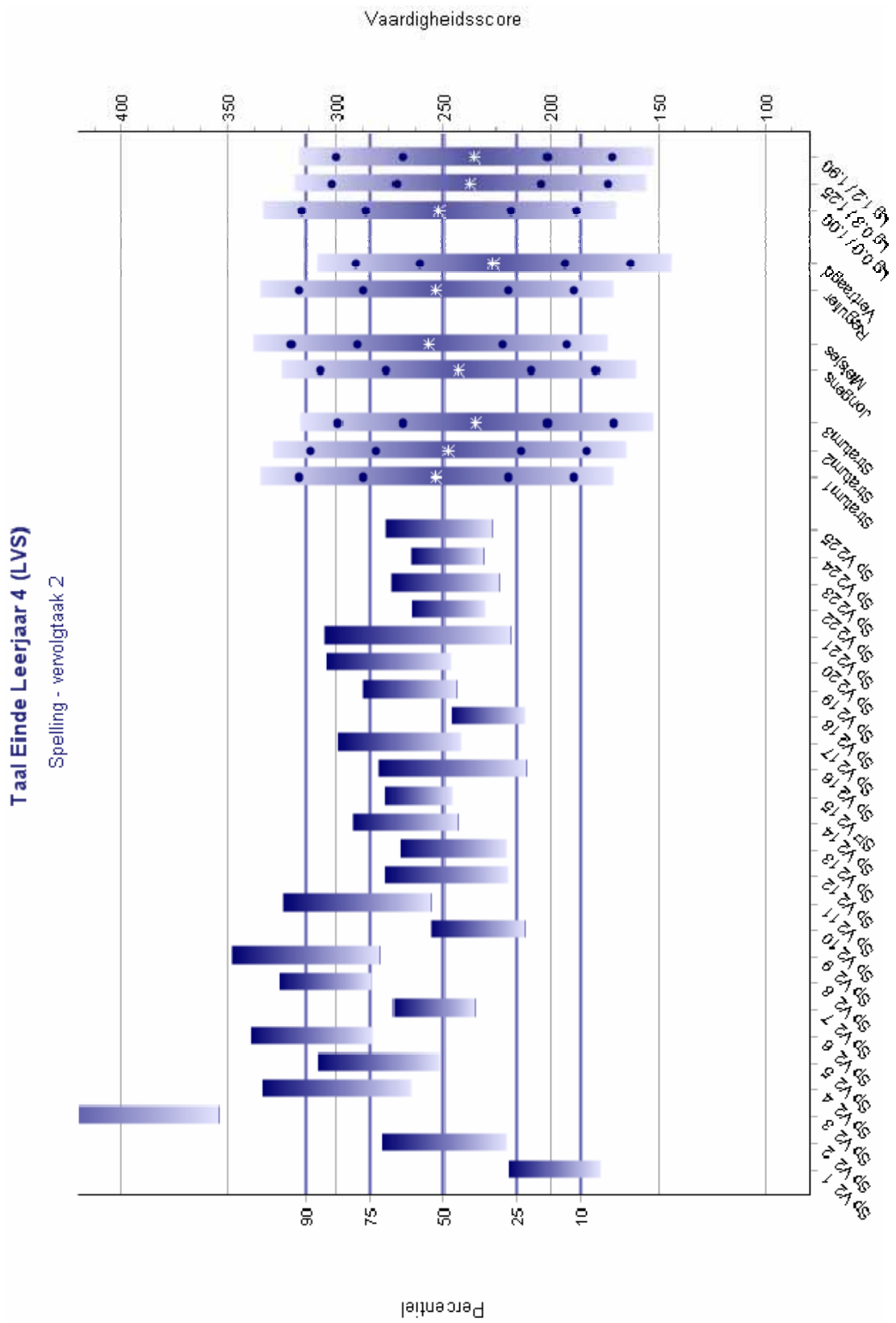
Figuur 6.2.a De opgaven van de starttoets Spelling E4 geordend naar de positie binnen deze toets, en de verdelingen van de groepen binnen leerjaar 4



Figuur 6.2.b De opgaven van Vervolgtoets 1 voor Spelling E4 geordend naar de positie binnen deze toets, en de verdelingen van de groepen binnen leerjaar 4.



Figuur 6.2.c De opgaven van Vervoltoets 2 voor Spelling E4 geordend naar de positie binnen deze toets, en de verdelingen van de groepen binnen leerjaar 4.



Tabel 6.2.2 Verdeling van de groepen in leerjaar 4 op de vaardigheid Spelling

Variabele		Gemiddelde	SD.	Percentielniveau
Geslacht	Jongens	243	50	(44)
	Meisjes	257	49	(55)
Leertijd	Regulier	254	49	(53)
	Vertraagd	227	49	(32)
Stratum	Stratum 1	254	50	(53)
	Stratum 2	248	50	(48)
	Stratum 3	235	50	(38)
Formatiegewicht	0.0 / 1.00	252	50	(52)
	0.3 / 1.25	238	50	(40)
	1.2 / 1.90	236	50	(39)
Regio	Noord	237	49	(40)
	Oost	258	49	(56)
	West	248	50	(48)
	Zuid	254	50	(53)
Urbanisatiegraad	Zeer sterk	239	50	(41)
	Sterk	253	50	(52)
	Matig	249	50	(49)
	Weinig	255	49	(54)
	Niet	250	50	(50)
Thuis taal	Alleen Nederlands (NLs)	254	49	(53)
	NLs en een andere taal	233	50	(37)
	Alleen een andere taal	236	50	(39)
	Streektaal	236	50	(39)

Tabel 6.2.3 Effectgrootte op de vaardigheid Spelling in leerjaar 4

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	0,25 ***	klein effect
Leertijd	Vertraagd – Regulier	-0,48 ***	klein effect
Stratum	Stratum2 – Stratum1	-0,10	niet significant
	Stratum3 – Stratum2	-0,18 ***	geen effect
	Stratum3 – Stratum1	-0,27 ***	klein effect
Formatiegewicht	.3/1.25 – 0.0/1.00	-0,20 ***	geen effect
	1.2/1.90 – 0.3/1.25	0,11	niet significant
	1.2/1.90 – 0.0/1.00	-0,09	niet significant
Regio	Oost – Noord	0,39 ***	klein effect
	West – Noord	0,22 ***	klein effect
	Zuid – Noord	0,36 ***	klein effect
	West – Oost	-0,16 ***	geen effect
	Zuid – Oost	-0,03	niet significant
	Zuid – West	0,13 ***	geen effect
Urbanisatiegraad	sterk – zeer sterk	0,21 ***	klein effect
	matig – sterk	-0,10	niet significant
	weinig – matig	0,08	niet significant
	niet – weinig	-0,07	niet significant
Thuis taal	NLs+Ander – Alleen NLs	-0,29 ***	klein effect
	Alleen ander – Alleen NLs	-0,21	niet significant
	Alleen ander – NLs+Ander	0,09	niet significant
	Streektaal – Alleen NLs	-0,28 ***	klein effect

Overschrijdingskansen: * : ,05; ** : ,01; *** : ,000001

6.3 Nederlandse taal: Begrijpend lezen

De presentatie van de resultaten voor Begrijpend lezen is vergelijkbaar met die van de resultaten voor Spelling. Ook hier zijn oude en de nieuwe leerlinggewichten samengevoegd tot nieuwe categorieën. De opzet van de toetsen Begrijpend lezen is eveneens vergelijkbaar met die van de toetsen Spelling. De leerlingen krijgen eerst een starttoets (figuur 6.3.a), waarna zij op basis van de resultaten op die toets een gemakkelijke (Vervolg 1, figuur 6.3.b) dan wel een moeilijke vervoltoets (Vervolg 2, figuur 6.3.c) krijgen. De verdeling van de opgaven in de startmodule (St), en de gemakkelijke (V1) en moeilijke (V2) vervolgmodes staat in tabel 6.3.3. De verdelingen van de verschillende groepen zijn gegeven in tabel 6.3.2, de effecten staan in tabel 6.3.3.

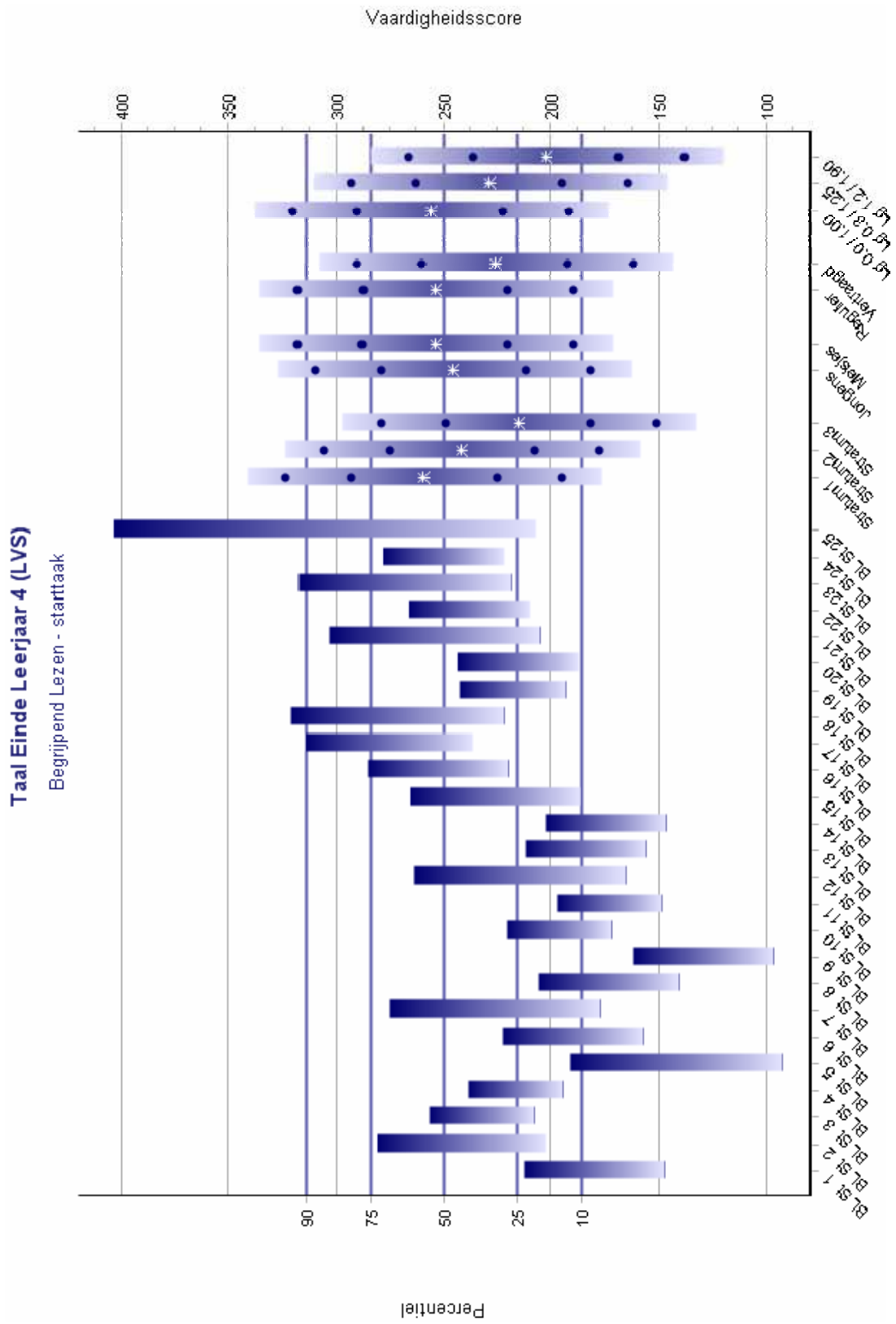
Bij de vaardigheid Begrijpend lezen zijn er oude en nieuwe leerlinggewichten samengevoegd, op dezelfde wijze als is gebeurd bij Spelling. De verschillen binnen die categorieën zijn ook vergelijkbaar met wat bij Spelling gevonden is: tussen leerlingen met gewicht 0.0 en 1.00 is er amper verschil in vaardigheid, leerlingen met het nieuwe gewicht 0.3 presteren over het algemeen iets beter dan leerlingen met oud gewicht 1.25 en leerlingen met het oude gewicht 1.90 scoren weer iets beter dan leerlingen met het nieuwe

gewicht 1.2. Ook geldt hier dezelfde terughoudendheid als bij Spelling met betrekking tot de stelligheid van uitspraken rond dit resultaat. Het geeft wel een mogelijke indicatie van verschillen tussen de gewichten, maar voor sterkere uitspraken zou meer onderzoek noodzakelijk zijn.

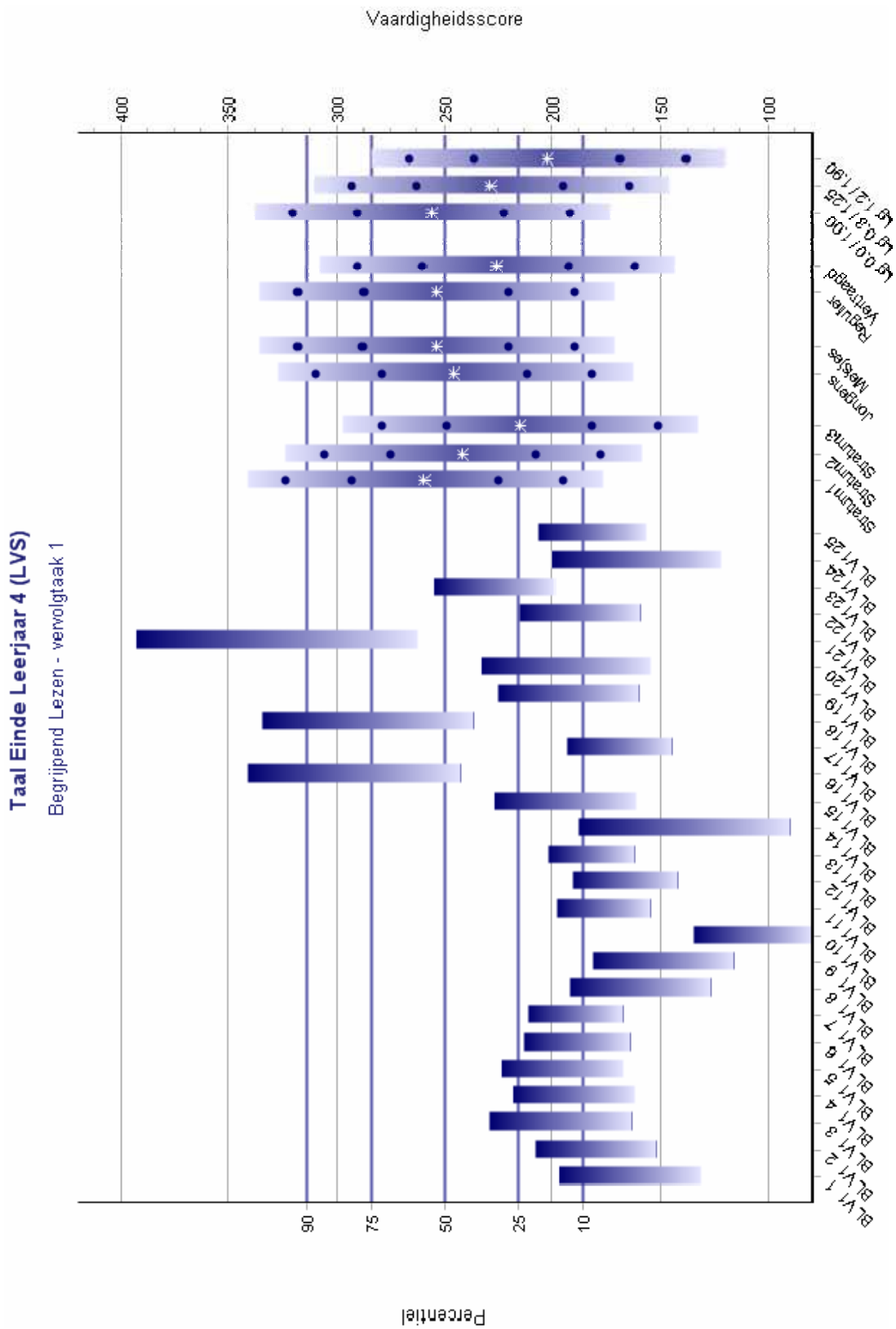
Tabel 6.3.1 Verdeling van het aantal opgaven voor de vaardigheid Begrijpend lezen over de Starttoets (St), Vervoltoets 1 (V1) en Vervoltoets 2 (V2) en over de percentielniveaus in leerjaar 4

	BL St		BL V1		BL V2		BL	
	P <	P >	P <	P >	P <	P >	P <	P >
	0,5	0,8	0,5	0,8	0,5	0,8	0,5	0,8
P10	14	1	4	2	24	0	42	3
P25	5	7	3	15	14	0	22	22
P75	0	12	1	21	6	3	7	36
P50	0	19	0	22	3	13	3	54
P90	0	21	0	22	2	19	2	62
Totaal	25	25	25	25	25	25	75	75

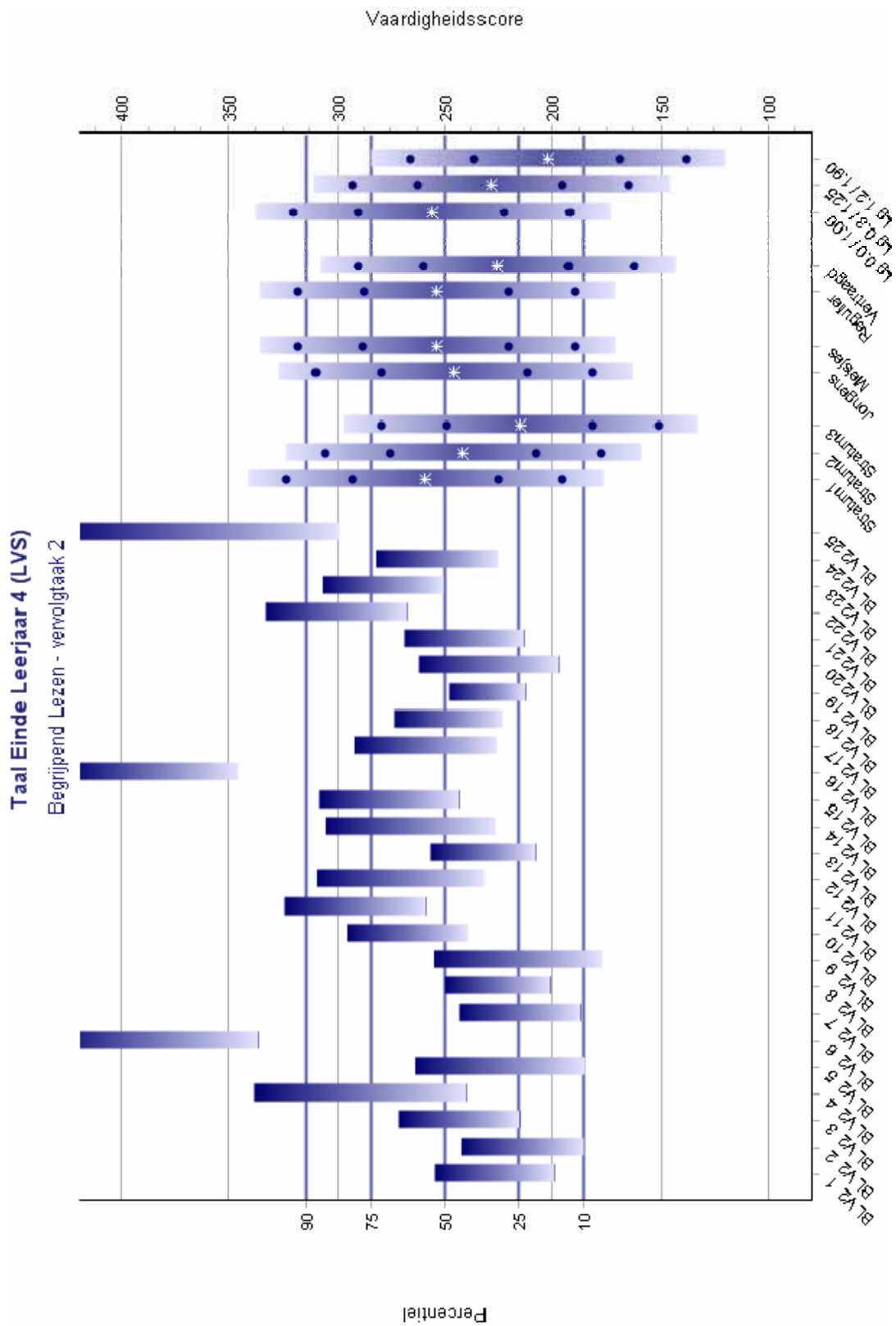
Figuur 6.3.a De opgaven van de starttoets Begrijpend lezen E4 geordend naar de positie binnen deze toets, en de verdelingen van de groepen binnen leerjaar 4



Figuur 6.3.b: De opgaven van Vervoltoets 1 voor Begrijpend lezen E4 geordend naar de positie binnen deze toets, en de verdelingen van de groepen binnen leerjaar 4.



Figuur 6.3.c De opgaven van Vervoltoets 2 voor Begrijpend lezen E4 geordend naar de positie binnen deze toets, en de verdelingen van de groepen binnen leerjaar 4.



Tabel 6.3.2 Verdeling van de groepen in leerjaar 4 op de vaardigheid Begrijpend lezen

Variabele		Gemiddelde	SD.	(Percentielniveau)
Geslacht	Jongens	246	50	(47)
	Meisjes	254	50	(53)
Leertijd	Regulier	254	49	(53)
	Vertraagd	226	50	(32)
Stratum	Stratum 1	260	47	(58)
	Stratum 2	242	48	(43)
	Stratum 3	216	50	(25)
Formatiegewicht	0.0 / 1.00	256	47	(55)
	0.3 / 1.25	229	48	(34)
	1.2 / 1.90	203	48	(17)
Regio	Noord	237	48	(39)
	Oost	263	48	(60)
	West	247	50	(48)
	Zuid	250	50	(50)
Urbanisatiegraad	Zeer sterk	229	52	(34)
	Sterk	252	49	(52)
	Matig	256	50	(55)
	Weinig	257	47	(55)
	Niet	250	48	(50)
Thuis taal	Alleen Nederlands (NLs)	257	47	(56)
	NLs en een andere taal	199	49	(15)
	Alleen een andere taal	228	49	(33)
	Streektaal	238	49	(41)

Tabel 6.3.3 Effectgrootte op de vaardigheid Begrijpend lezen in leerjaar 4

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	0,16 ***	geen effect
Leertijd	Vertraagd – Regulier	-0,44 ***	klein effect
Stratum	Stratum2 – Stratum1	-0,30 ***	klein effect
	Stratum3 – Stratum2	-0,25 ***	klein effect
	Stratum3 – Stratum1	-0,55 ***	matig effect
Formatiegewicht	.3/1.25 – 0.0/1.00	-0,39 ***	klein effect
	1.2/1.90 – 0.3/1.25	-0,35 ***	klein effect
	1.2/1.90 – 0.0/1.00	-0,74 ***	matig effect
Regio	Oost – Noord	0,48 ***	klein effect
	West – Noord	0,27 ***	klein effect
	Zuid – Noord	0,36 ***	klein effect
	West – Oost	-0,21 ***	klein effect
	Zuid – Oost	-0,12	niet significant
	Zuid – West	0,09	niet significant
Urbanisatiegraad	sterk – zeer sterk	0,23 ***	klein effect
	matig – sterk	0,01	niet significant
	weinig – matig	-0,09	niet significant
	niet – weinig	-0,10	niet significant
Thuis taal	NLs+Ander – Alleen NLs	-0,36 ***	klein effect
	Alleen ander – Alleen NLs	-0,62 ***	matig effect
	Alleen ander – NLs+Ander	-0,25	niet significant
	Streektaal – Alleen NLs	-0,23 ***	klein effect

Overschrijdingskansen: * : ,05; ** : ,01; *** : ,000001

6.4 Rekenen

De vier verschillende rekenvaardigheden zijn met verschillende opgaven uit een toets gemeten. Deze toets bestaat uit twee delen van ieder 27 opgaven. De opgaven die verschillende vaardigheden meten wisselen elkaar af in deze twee toetsdelen. De classificatie van de moeilijkheid van de opgaven over deze vaardigheden wordt gegeven in tabel 6.4.1. De opgaven zijn ook te vinden in figuur 6.4 a tot en met d. Deze figuren worden per vaardigheid gegeven. In ieder figuur zijn ook de vaardigheidsverdelingen gegeven van de verschillende groepen voor de variabelen geslacht, leertijd, stratum en formatiegewichtgroep. Bij deze formatiegewichtgroepen zijn oude en nieuwe leerlinggewichten op dezelfde wijze samengevoegd als bij de vaardigheden voor Nederlandse taal is gebeurd.

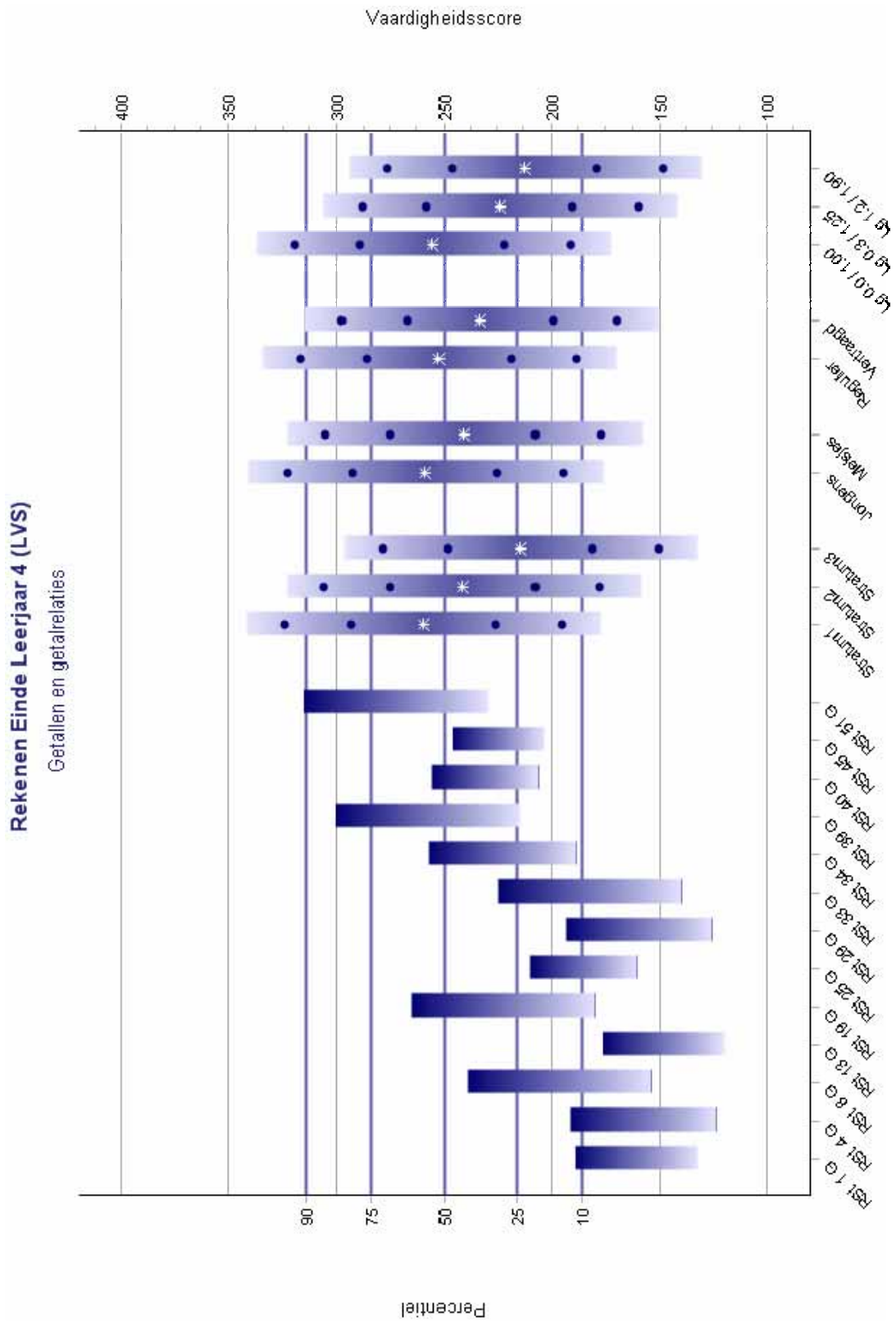
Tabel 6.4.1 *Verdeling van het aantal opgaven voor de vaardigheden Getallen en getalrelaties (G), Optellen/Aftrekken (O/A), Vermenigvuldigen/Delen (V/D) en Meten, tijd en geld (MTG) over de toetsen en percentielniveaus in leerjaar 4*

	RE G		RE O/A		RE V/D		RE MTG	
	P <	P >	P <	P >	P <	P >	P <	P >
	0,5	0,8	0,5	0,8	0,5	0,8	0,5	0,8
P10	5	1	8	0	11	0	10	0
P25	1	5	5	0	7	1	7	2
P75	0	8	1	5	0	3	0	7
P50	0	11	0	9	0	11	0	14
P90	0	12	0	12	0	14	0	15
Totaal	13	13	12	12	14	14	15	15

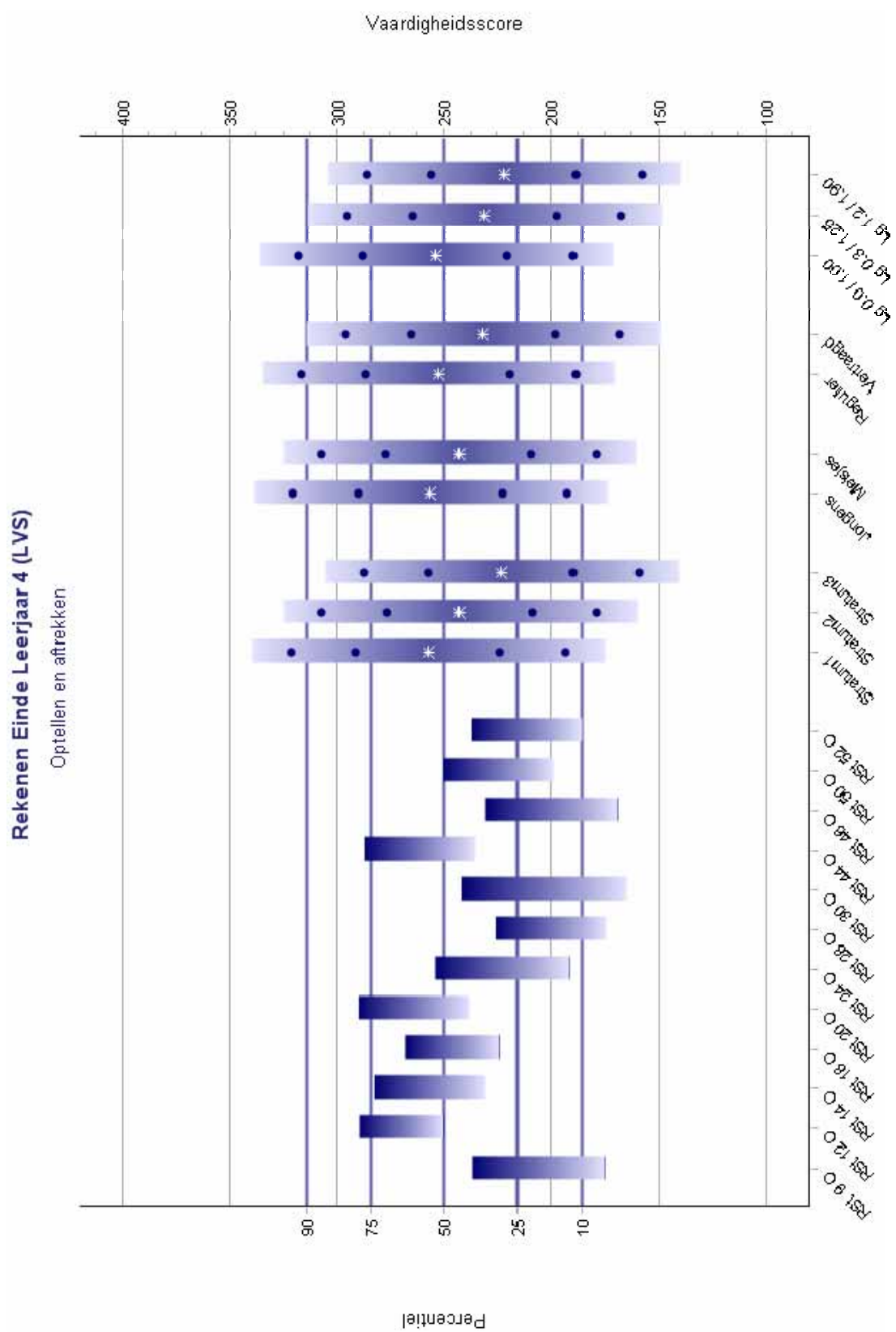
De vaardigheidsverdelingen van de verschillende groepen in de figuren 6.4 a tot en met d lijken behoorlijk op elkaar, maar zijn niet identiek aan elkaar. De vaardigheidsniveaus voor de vier verschillende rekenvaardigheden zijn naast elkaar gegeven in tabel 6.4.2. Deze tabel lijkt op de andere tabellen in dit onderzoek waarin verdelingen gegeven zijn voor de verschillende groepen. In deze tabel ontbreken alleen de standaarddeviaties en percentielniveaus. In deze tabel staan niet alleen de gegevens voor de groepen die in de figuren staan maar ook die voor de groepen van leerlingen die verschillen in thuistaal, regio en urbanisatiegraad van de plaats waar de school zich bevindt.

De gecorrigeerde effecten van alle variabelen zijn gegeven voor de vier rekenvaardigheden. Voor de vaardigheid Getallen en getalrelaties staan de effecten in tabel 6.4.3, voor Optellen en aftrekken staan de effecten in tabel 6.4.4, voor Vermenigvuldigen en delen in tabel 6.4.5 en voor Meten, tijd en geld in tabel 6.4.6.

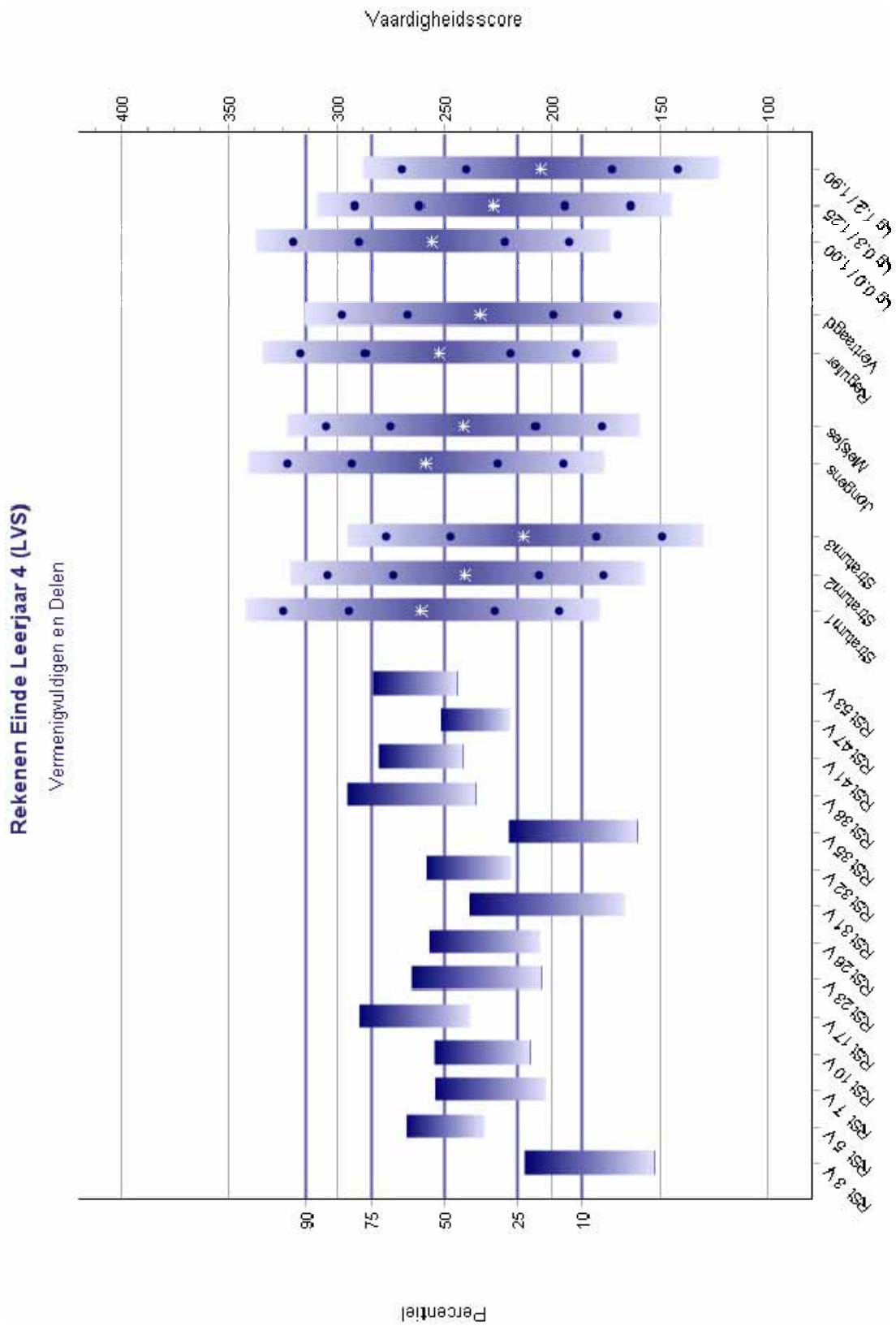
Figuur 6.4.a De opgaven voor Getallen en getalrelaties van de toets Rekenen E4 geordend naar de positie binnen deze toets, en de verdelingen van de groepen binnen leerjaar 4



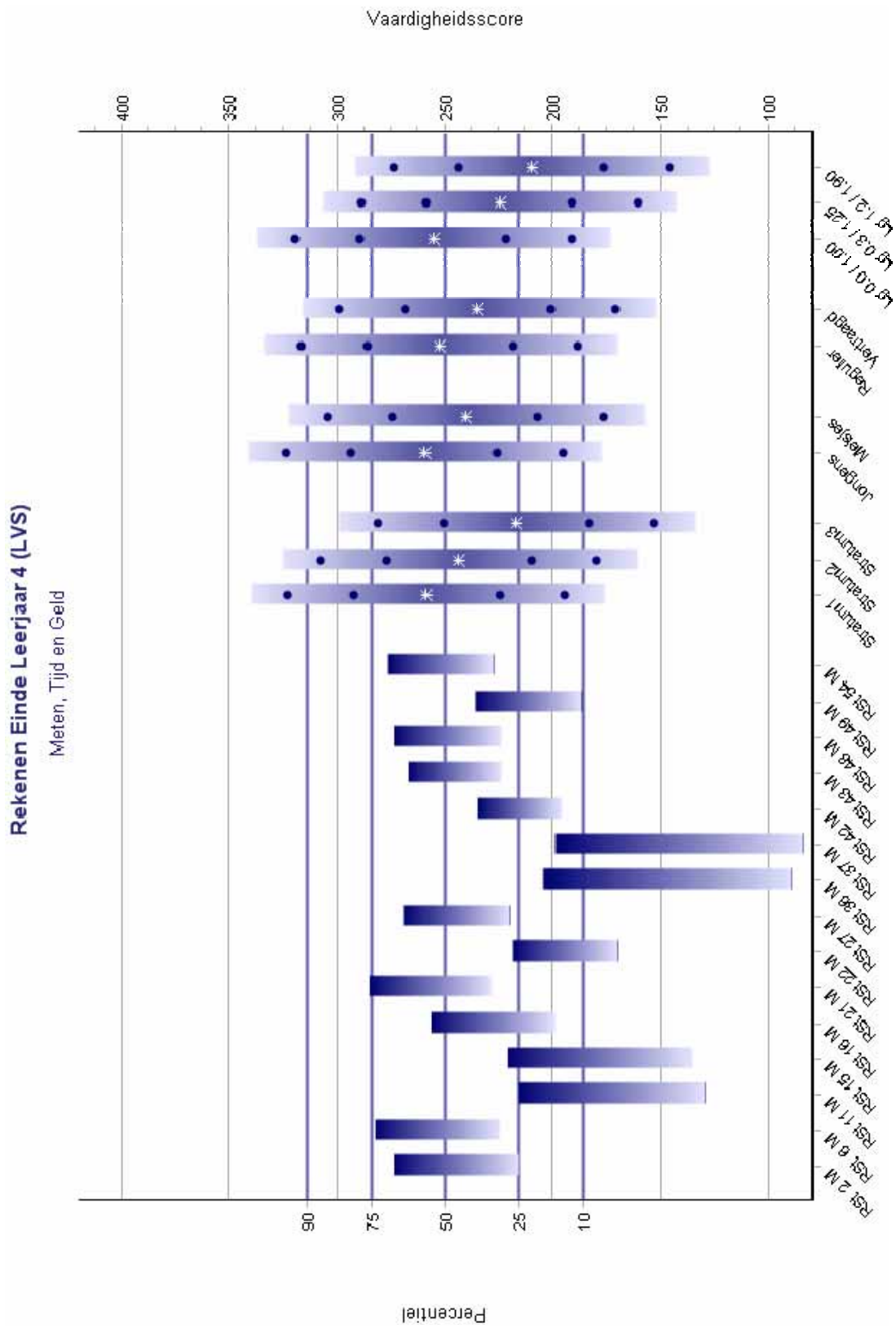
Figuur 6.4.b De opgaven voor Optellen en aftrekken van de toets Rekenen E4 geordend naar de positie binnen deze toets, en de verdelingen van de groepen binnen leerjaar 4



Figuur 6.4.c De opgaven voor Vermenigvuldigen en delen van de toets Rekenen E4 geordend naar de positie binnen deze toets, en de verdelingen van de groepen binnen leerjaar 4



Figuur 6.4.d De opgaven voor Meten, tijd en geld van de toets Rekenen E4 geordend naar de positie binnen deze toets, en de verdelingen van de groepen binnen leerjaar 4



Tabel 6.4.2.a Verdeling van de groepen in leerjaar 4 op de rekenschalen Getallen en getalrelaties (G), Optellen/Aftrekken (O/A), Vermenigvuldigen/Delen (V/D) en Meten, tijd en geld (MTG)

Variabele		Gemiddelden				Standaarddeviaties			
		G	O/A	V/D	MTG	G/R	O/A	V/D	MTG
Geslacht	Jongens	259	257	259	260	49	50	49	49
	Meisjes	241	243	241	240	49	50	49	49
Leertijd	Regulier	253	253	253	253	49	49	49	49
	Vertraagd	233	232	234	235	51	50	51	51
Stratum	Stratum 1	260	258	261	259	47	48	47	48
	Stratum 2	242	243	240	244	48	49	48	49
	Stratum 3	215	224	213	217	48	49	49	49
Formatiegewicht	0.0 / 1.00	256	254	256	256	48	49	48	48
	0.3 / 1.25	224	231	228	225	49	49	48	48
	1.2 / 1.90	213	222	206	210	49	49	48	48
Regio	Noord	245	241	247	245	48	49	48	48
	Oost	263	258	262	263	48	49	48	48
	West	247	249	248	246	50	50	50	50
	Zuid	247	249	245	248	50	50	50	50
Urbanisatiegraad	Zeer sterk	232	237	233	236	52	51	53	53
	Sterk	251	250	249	251	49	50	49	49
	Matig	253	250	255	255	50	50	51	51
	Weinig	258	258	258	253	48	49	47	48
	Niet	250	250	251	250	48	49	48	48
Thuis taal	Alleen Nederlands (NLs)	257	255	256	257	48	49	48	48
	Streektaal	236	237	241	238	49	49	50	49
	NLs en een andere taal	223	226	223	224	49	49	50	49
	Alleen een andere taal	212	227	212	213	48	49	49	49

Tabel 6.4.2.b Percentielniveaus van de gemiddelden in leerjaar 4 op de rekenschalen Getallen en getalrelaties (G), Optellen/Aftrekken (O/A), Vermenigvuldigen/Delen (V/D) en Meten, tijd en geld (MTG)

Variabele		G	O/A	V/D	MTG
Geslacht	Jongens	57	55	57	58
	Meisjes	43	45	43	42
Leertijd	Regulier	52	52	52	52
	Vertraagd	37	36	37	38
Stratum	Stratum 1	58	56	58	57
	Stratum 2	43	44	42	45
	Stratum 3	24	30	23	26
Formatiegewicht	0.0 / 1.00	54	53	55	55
	0.3 / 1.25	30	36	33	31
	1.2 / 1.90	23	29	19	21
Regio	Noord	46	43	48	46
	Oost	60	56	60	61
	West	48	49	48	47
	Zuid	47	49	46	48
Urbanisatiegraad	Zeer sterk	36	40	36	39
	Sterk	50	50	49	51
	Matig	52	50	54	54
	Weinig	56	57	56	53
	Niet	50	50	51	50
Thuis taal	Alleen Nederlands (NLs)	55	54	55	55
	NLs en een andere taal	29	31	30	30
	Alleen een andere taal	22	32	22	23
	Streektaal	39	39	43	41

Tabel 6.4.3 Effectgrootte op de vaardigheid Rekenen: Getallen en getalrelaties in leerjaar 4

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	-0,41 ***	klein effect
Leertijd	Vertraagd – Regulier	-0,33 ***	klein effect
Stratum	Stratum2 – Stratum1	-0,38 ***	klein effect
	Stratum3 – Stratum2	-0,33 ***	klein effect
	Stratum3 – Stratum1	-0,71 ***	matig effect
Formatiegewicht	.3/1.25 – 0.0/1.00	-0,34 ***	klein effect
	1.2/1.90 – 0.3/1.25	-0,26 ***	klein effect
	1.2/1.90 – 0.0/1.00	-0,59 ***	matig effect
Regio	Oost – Noord	0,26 ***	klein effect
	West – Noord	0,10	niet significant
	Zuid – Noord	0,06	niet significant
	West – Oost	-0,16 ***	geen effect
	Zuid – Oost	-0,19 ***	geen effect
	Zuid – West	-0,04	niet significant
Urbanisatiegraad	sterk – zeer sterk	0,03	niet significant
	matig – sterk	0,09	niet significant
	weinig – matig	-0,11	niet significant
	niet – weinig	-0,09	niet significant
Thuis taal	NLs+Ander – Alleen NLs	-0,39 ***	klein effect
	Alleen ander – Alleen NLs	-0,24	niet significant
	Alleen ander – NLs+Ander	0,16	niet significant
	Streektaal – Alleen NLs	-0,18	niet significant

Overschrijdingskansen: * : ,05; ** : ,01; *** : ,000001

Tabel 6.4.4 Effectgrootte op de vaardigheid Rekenen: Optellen en aftrekken in leerjaar 4

Variabele	Contrast	Effectgrootte	Effect kwalificatie
Geslacht	Meisjes – Jongens	-0,43 ***	klein effect
Leertijd	Vertraagd – Regulier	-0,31 ***	klein effect
Stratum	Stratum2 – Stratum1	-0,25 ***	klein effect
	Stratum3 – Stratum2	-0,31 ***	klein effect
	Stratum3 – Stratum1	-0,56 ***	matig effect
Formatiegewicht	.3/1.25 – 0.0/1.00	-0,44 ***	klein effect
	1.2/1.90 – 0.3/1.25	-0,15	niet significant
	1.2/1.90 – 0.0/1.00	-0,59 ***	matig effect
Regio	Oost – Noord	0,36 ***	klein effect
	West – Noord	0,12	niet significant
	Zuid – Noord	0,17	niet significant
	West – Oost	-0,24 ***	klein effect
	Zuid – Oost	-0,19 ***	geen effect
	Zuid – West	0,05	niet significant
Urbanisatiegraad	sterk – zeer sterk	0,03	niet significant
	matig – sterk	0,05	niet significant
	weinig – matig	-0,18 ***	geen effect
	niet – weinig	-0,03	niet significant
Thuis taal	NLs+Ander – Alleen NLs	-0,45 ***	klein effect
	Alleen ander – Alleen NLs	-0,38 ***	klein effect
	Alleen ander – NLs+Ander	0,07	niet significant
	Streektaal – Alleen NLs	-0,27 ***	klein effect

Significantie: * : alpha =,05; ** : alpha ,01; *** : alpha ,0001

Tabel 6.4.5 Effectgrootte op de vaardigheid Rekenen: Vermenigvuldigen en delen in leerjaar 4

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	-0,41 ***	klein effect
Leertijd	Vertraagd – Regulier	-0,35 ***	klein effect
Stratum	Stratum2 – Stratum1	-0,34 ***	klein effect
	Stratum3 – Stratum2	-0,38 ***	klein effect
	Stratum3 – Stratum1	-0,71 ***	matig effect
Formatiegewicht	.3/1.25 – 0.0/1.00	-0,40 ***	klein effect
	1.2/1.90 – 0.3/1.25	-0,03	niet significant
	1.2/1.90 – 0.0/1.00	-0,42 ***	klein effect
Regio	Oost – Noord	0,32 ***	klein effect
	West – Noord	0,13	niet significant
	Zuid – Noord	0,14	niet significant
	West – Oost	-0,19 ***	geen effect
	Zuid – Oost	-0,18 ***	geen effect
	Zuid – West	0,01	niet significant
Urbanisatiegraad	sterk – zeer sterk	0,09	niet significant
	matig – sterk	0,01	niet significant
	weinig – matig	-0,03	niet significant
	niet – weinig	-0,12	niet significant
Thuis taal	NLs+Ander – Alleen NLs	-0,48 ***	klein effect
	Alleen ander – Alleen NLs	-0,40 ***	klein effect
	Alleen ander – NLs+Ander	0,07	niet significant
	Streektaal – Alleen NLs	-0,34 ***	klein effect

Significantie: * : alpha =,05; ** : alpha ,01; *** : alpha ,0001

Tabel 6.4.6 Effectgrootte op de vaardigheid Meten, tijd en geld in leerjaar 4

Variabele	Contrast	Effectgrootte	Effectkwalificatie
Geslacht	Meisjes – Jongens	-0,30 ***	klein effect
Leertijd	Vertraagd – Regulier	-0,39 ***	klein effect
Stratum	Stratum2 – Stratum1	-0,26 ***	klein effect
	Stratum3 – Stratum2	-0,27 ***	klein effect
	Stratum3 – Stratum1	-0,53 ***	matig effect
Formatiegewicht	.3/1.25 – 0.0/1.00	-0,25 ***	klein effect
	1.2/1.90 – 0.3/1.25	-0,01	niet significant
	1.2/1.90 – 0.0/1.00	-0,26 ***	klein effect
Regio	Oost – Noord	0,31 ***	klein effect
	West – Noord	0,22 ***	klein effect
	Zuid – Noord	0,25 ***	klein effect
	West – Oost	-0,08	niet significant
	Zuid – Oost	-0,06	niet significant
	Zuid – West	0,03	niet significant
Urbanisatiegraad	sterk – zeer sterk	0,07	niet significant
	matig – sterk	-0,04	niet significant
	weinig – matig	0,06	niet significant
	niet – weinig	-0,14	niet significant
Thuis taal	NLs+Ander – Alleen NLs	-0,42 ***	klein effect
	Alleen ander – Alleen NLs	-0,14	niet significant
	Alleen ander – NLs+Ander	0,28 ***	klein effect

Significantie: * : alpha =,05; ** : alpha ,01; *** : alpha ,0001

Voorlopige resultaten waarbij gekeken is naar de verschillen tussen de oude en de nieuwe leerlinggewichten geven het volgende beeld bij Rekenen in leerjaar 4. De verschillen tussen 0.0 en 1.00 zijn gering. Het effect is het hoogst bij Optellen en aftrekken, en kan met een waarde van 0,12 (0.0 scoort hoger 1.00) als afwezig beschouwd worden. Bij de gewichten 0.3 en 1.25 worden iets grotere verschillen gevonden. Bij Vermenigvuldigen en delen doen de leerlingen met het nieuwe gewicht het iets beter dan die met een oud gewicht. Bij de vaardigheden Getallen en getalrelaties wordt daar juist het omgekeerde gevonden. Bij de gewichten 1.2 en 1.90 wordt ook gevonden dat de leerlingen met het oude gewicht het op alle schalen iets beter doen. Dat resultaat was te verwachten, gezien de veranderde definitie van de gewichten. Wel moet nogmaals opgemerkt worden dat dit voorlopige resultaten zijn, die gebaseerd zijn op kleine aantallen leerlingen; de verschillen zijn dan ook niet significant en over het algemeen kleiner dan 0,2. Ook over de oorzaak van de (zeer kleine) verschillen is geen harde uitspraak te doen. Naast verschillen in de leerlingen zouden de verschillen ook kunnen liggen in de scholen, die al dan niet de nieuwe gewichten gebruiken.

7 Discussie

7.1 Samenvatting van het onderzoek: wat is er onderzocht?

In dit verslag beschrijven wij het onderzoek naar de taal- en rekenvaardigheid van leerlingen in jaargroep 8 en in jaargroep 4 van het basisonderwijs. De onderzochte taalvaardigheden waren *Begrijpend lezen*, *Spelling* en *Woordenschat*. De rekenvaardigheden verschilden voor jaargroep 8 en jaargroep 4. In jaargroep 8 zijn drie vaardigheden gemeten: *Getallen en bewerkingen*, *Breuken, procenten en verhoudingen* en *Meten, meetkunde, tijd en geld*. In jaargroep 4 zijn vier rekenvaardigheden onderzocht: *Getallen en getalrelaties*, *Optellen en aftrekken*, *Vermenigvuldigen en delen*, en *Meten, tijd en geld*.

De vaardigheden zijn beschreven door middel van de inhoud van de opgaven. Deze opgaven zijn gerelateerd aan het niveau van verschillende groepen leerlingen. Er zijn hierbij op twee verschillende wijzen groepen gedefinieerd. De eerste wijze is door verschillende groepen te maken op basis van hun vaardigheid, en de tweede is door groepen te onderscheiden die van elkaar verschillen op basis van achtergrondvariabelen.

Bij de eerste wijze worden per vaardigheid vijf verschillende percentiepunten beschreven. Dat zijn de percentiepunten 10, 25, 50, 75 en 90 in de gehele populatie. Door middel van die percentiepunten kan inzicht gegeven worden in de vaardigheid van de (veel) minder vaardige, gemiddelde en (veel) vaardiger leerlingen. Merk overigens op dat als een leerling bij de ene vaardigheid behoort tot een bepaalde vaardigheidsgroep, dit niet hoeft te betekenen dat de leerling bij een andere vaardigheid in dezelfde vaardigheidsgroep zit. Verschillende leerlingen hebben verschillende talenten.

Bij de tweede manier van indelen in groepen zijn leerlingen met dezelfde achtergrondkenmerken samengenomen. Een vraag die op basis van een dergelijke indeling kan worden beantwoord is bijvoorbeeld wat de gemiddelde vaardigheid van jongens is, of die van meisjes. De groepen worden niet alleen onderscheiden door geslacht, maar ook door leertijd, leerlinggewicht (formatiegewicht), schoolgewicht (stratum), regio en urbanisatiegraad van de plaats van de school, en de thuistaal. Voor het onderzoek in jaargroep 8 zijn ook andere variabelen meegenomen in de analyse: de toets die de leerlingen gemaakt hebben en het ingeschatte niveau van het vervolgonderwijs in VO.

Niet alleen is er gekeken naar de gemiddelden van de verschillende groepen, maar ook naar de gecorrigeerde verschillen tussen deze groepen. Met behulp van effectschattingen van de vaardigheden is bijvoorbeeld gekeken of er verschillen zijn tussen scholen in het westen en het oosten van Nederland, als gecorrigeerd wordt voor de variabelen geslacht, leertijd, leerlinggewicht en stratum. We kijken hierbij naar de grootte van het effect en of het effect significant groter dan 0 is. Dat is onderzocht voor alle variabelen waar ook gemiddelden voor zijn berekend.

7.2 Samenvatting van de resultaten: jaargroep 8 vergeleken met jaargroep 4

Beschrijving van wat kinderen kunnen

Een belangrijk deel van de resultaten betreft het beschrijven van wat leerlingen wel en niet kunnen. De meeste aandacht gaat uit naar het beschrijven van de vaardigheden in jaargroep 8. Dit is een belangrijk moment in de schoolcarrière, aangezien dat het snijpunt is tussen het basisonderwijs en het voortgezet onderwijs. Wat kunnen en kennen leerlingen nu na acht jaar basisonderwijs en met welke reken- en taalvaardigheid gaan zij het voortgezet onderwijs in?

De beschrijving van de vaardigheden op dit meetmoment gebeurt met behulp van opgaven uit de Eindtoets Basisonderwijs (EB) 2008. Het is een toets waarvoor de meeste leerlingen goed gemotiveerd zijn, waardoor de prestaties waarschijnlijk optimaal zijn. Sommige leerlingen in jaargroep 8 hebben de Niveautoets (NT) gemaakt. Voor de beschrijving van de vaardigheid van de leerlingen in jaargroep 4 wordt gebruikgemaakt van de toetsen van het Leerling- en onderwijsvolgsysteem (LOVS) en proeftoetsen die voor het LOVS zijn uitgevoerd.

De opgaven worden zowel bij jaargroep 8 als bij jaargroep 4 afgebeeld op vaardigheidsschalen, waarbij de opgaven direct vergeleken kunnen worden met de vaardigheidsniveaus van de minder vaardige, gemiddeld vaardige en de meer vaardige leerlingen zoals die met behulp van de percentielgroepen beschreven zijn. De vaardigheidsniveaus van de jaargroepen 8 en 4 zijn echter niet rechtstreeks met elkaar te vergelijken in dit onderzoek. Dat komt omdat de toetsen geen overlap hebben. De vergelijking kan wel gemaakt worden door middel van de opgaven en hun verdeling over de percentielpunten. Er kan bijvoorbeeld bekeken worden wat een gemiddelde leerling in jaargroep 8 kan en wat kan een gemiddelde leerling in jaargroep 4 kan. Hiermee is de vooruitgang van de leerlingen inhoudelijk te zien.

Effect van achtergrondvariabelen bij taalvaardigheden

Wat wel onderzocht kan worden in de vergelijking van jaargroep 8 met jaargroep 4 is het verschil tussen de effecten van de achtergrondvariabelen op de vaardigheden. Binnen de taalvaardigheden zijn er wat verschillen over de vaardigheden heen, maar de meeste effecten zijn binnen een leerjaar vergelijkbaar. Een opvallend resultaat was dat in jaargroep 8 jongens beter presteerden op de vaardigheid Woordenschat dan meisjes. In jaargroep 4 was dat nog omgekeerd, zij het dat het in jaargroep 4 gevonden effect zo klein was dat het niet significant van nul verschilde. Bij Spelling en Begrijpend lezen zijn in beide groepen de meisjes vaardiger dan de jongens. De effecten zijn echter (zeer) klein.

Bij de achtergrondvariabele leertijd valt op dat de effecten groter worden van leerjaar 4 naar leerjaar 8. De afstand in vaardigheid tussen de leerlingen die een vertraagd traject doorlopen ten opzichte van de leerlingen die een regulier (of vervoegd) traject doorlopen is groter geworden. Een mogelijke oorzaak zou bijvoorbeeld kunnen zijn dat al eerder vertraagde leerlingen minder snel in vaardigheid groeien dan de reguliere leerlingen. Een andere mogelijke oorzaak zou kunnen zijn dat leerlingen die later in hun schoolloopbaan vertraagd raken, ook eerder minder vaardig waren. Wanneer zij in de vier jaar die volgen blijven zitten, dan wordt het verschil tussen de twee groepen nog groter. Voor het achterhalen van de oorzaak is echter meer onderzoek nodig.

In jaargroep 8 worden geen effecten gevonden voor de variabelen stratum en thuistaal bij Spelling en Begrijpend lezen. Er is wel een effect voor leerlinggewicht, maar er is in jaargroep 8 geen bijkomend effect van de thuistaal. Dat betekent dat, als gecorrigeerd wordt voor leerlinggewicht, geslacht en leertijd, het er in jaargroep 8 voor deze vaardigheden niet toe doet of iemand thuis een andere taal naast of in plaats van het Nederlands spreekt. Voor een aantal specifieke talen wordt bij Begrijpend lezen nog wel kleine verschillen gevonden, maar bij Spelling zien we alleen een klein effect bij leerlingen die thuis Surinaams of Antilliaans spreken. Bij Woordenschat worden nog wel effecten voor thuistaal gevonden: anderstaligen presteren slechter op deze vaardigheid, ook als er gecorrigeerd is voor leerlinggewicht. Daarbij maakt het niet uit of zij daarnaast thuis ook nog Nederlands spreken. In groep 4 zijn op alle schalen waarbij thuistaal bekend was dergelijke effecten te vinden.

Het is opvallend dat aan het eind van de basisschool de effecten van thuistaal vrijwel verdwenen zijn voor de meeste vaardigheden, behalve bij Woordenschat. Mogelijk is Woordenschat een minder schoolse vaardigheid, en gebeurt een groot deel van de ontwikkeling daarvan thuis. Om deze hypothese te toetsen is meer onderzoek nodig.

Bij de variabele formatiegewicht verdwijnt het verschil tussen de 1.00-leerlingen en de 1.90-leerlingen niet, maar het verschil wordt wel kleiner. Iets vergelijkbaars zien we bij het verschil tussen de leerlingen met leerlinggewicht 1.25 en 1.00. Bij formatiegewicht is de vergelijking echter wat lastiger, aangezien in jaargroep 4 ook de nieuwe leerlinggewichten gebruikt zijn. Of het verschil komt door een veranderde definitie van de leerlinggewichten of in een daadwerkelijk verschil tussen leerlingen in leerjaar 4 en leerjaar 8 is daarmee niet geheel duidelijk.

In leerjaar 4 zien we net als in leerjaar 8 ook weinig verschillen tussen regio's, al komt in leerjaar 4 en bij een aantal vaardigheden in leerjaar 8 regio Noord als iets minder vaardig naar voren. In jaargroep 4 is gevonden dat de leerlingen in het oosten beter presteren dan in het noorden, maar een goede verklaring valt daar niet voor te geven. Bij Woordenschat wordt dat verschil niet gevonden en ook in jaargroep 8 doen leerlingen in het oosten van het land het niet beter. Vooral nog kan aangenomen worden dat dit geen

systematisch verschil is in jaargroep 4, maar om dat te weten te komen zal dit gevolgd moeten worden. Uiteraard geldt dit laatste in principe ook voor de andere vergelijkingen die hierboven gemaakt worden. Ook wat betreft de verschillen in urbanisatiegraad zijn er weinig verschillen tussen leerjaar 4 en leerjaar 8. Het kleine effect van een sterke naar een zeer sterke urbanisatiegraad verdwijnt wel in jaargroep 8 bij de taalvaardigheden. In jaargroep 8 worden juist relatief betere resultaten gevonden, als gecorrigeerd is voor formatiegewicht, stratum, leertijd en geslacht.

Effect van achtergrondvariabelen bij rekenvaardigheden

Bij Rekenen lijken de resultaten van de vaardigheden binnen de leerjaren behoorlijk op elkaar. Als gekeken wordt naar de resultaten voor de schalen binnen leerjaar 8 zijn er weinig verschillen, en ook de verschillen binnen jaargroep 4 over de schalen zijn niet groot. Er zijn wel een aantal verschillen aan te wijzen tussen jaargroep 8 en jaargroep 4. Een aantal van die verschillen is vergelijkbaar met wat bij de taalschalen gevonden is.

Zowel in jaargroep 8 als jaargroep 4 is gevonden dat de rekenvaardigheden over het algemeen beter beheerst worden door jongens dan door meisjes. De verschillen zijn iets groter geworden van jaargroep 4 naar jaargroep 8.

Bij de variabele leertijd worden de verschillen ook groter van jaargroep 4 naar jaargroep 8. Een toename in verschil tussen deze twee groepen was ook gevonden bij de taalvaardigheden. Ook bij de rekenvaardigheden was het duidelijk dat de effecten van schoolgewicht en van thuistaal afnemen van jaargroep 4 naar jaargroep 8. In jaargroep 8 zijn de meeste effecten binnen deze variabele zelfs afwezig. Wat betreft de formatiegewichten is opvallend dat in jaargroep 4 de leerlingen met gewicht 1.25 (of 0.3) nog beter presteerden dan leerlingen met gewicht 1.90 (of 1.2), terwijl dat in jaargroep 8 omgekeerd is. Dit verschil tussen de leerjaren kan echter ook veroorzaakt zijn door een veranderde definitie van het leerlinggewicht. Ook bij leerjaar 4 geven de voorlopige resultaten namelijk aan dat de leerlingen met nieuw gewicht 1.2 iets slechter presteren dan leerlingen met oud gewicht 1.90. De verschillen tussen de regio's en de opvolgende urbanisatiegraden zijn ook bij rekenen vrijwel afwezig in jaargroep 8. In jaargroep 4 is bij regio gevonden dat de leerlingen in het oosten beter presteren dan in het noorden, maar een goede verklaring valt daar niet voor te geven.

7.3 Tot slot

De resultaten zoals die gepresenteerd zijn geven een stand van zaken weer. Hoewel deze resultaten wellicht kunnen dienen om ideeën te vormen over mogelijke oorzaken en verbanden, zijn deze uiteraard met dit onderzoek niet aan te tonen. Daar zal meer onderzoek voor gedaan moeten worden.

Ook is verder onderzoek noodzakelijk voor het interpreteren van de prestaties van de leerlingen. We hebben beschreven wat leerlingen van verschillende niveaus kunnen, maar daarmee is nog niet de vraag beantwoord of het al dan niet *voldoende* is wat de leerlingen beheersen. Om daar een antwoord op te geven is standaardonderzoek nodig. Daarmee kan het beoogde fundamentele niveau en het beoogde streefniveau beschreven worden. Die resultaten kunnen in een later stadium vergeleken worden met de resultaten uit dit onderzoek om te zien waar bij Nederlandse leerlingen de sterke en zwakke punten liggen: bij welke vaardigheden en welke groepen. Wellicht is het ook mogelijk om in een later stadium deze jaarlijkse peiling te integreren met de periodieke peiling. Die periodieke peiling gaat veel dieper op de schalen in. Er wordt veel meer in gedifferentieerd, maar door overlap in opgaven zou kruisvalidering plaats kunnen vinden doordat resultaten van verschillende schalen naar elkaar toe vertaald kunnen worden.

Ander toekomstig onderzoek betreft de vergelijking met de resultaten in het komende jaar. Daarbij worden de gegevens van de volgende Eindtoets vergeleken worden met die van de EB 2008. Bij de vergelijking van de resultaten over de jaren spelen de ankeropgaven een belangrijke rol. Dat zijn de opgaven die in het speciale EB-Anker toetsboekje zaten. Door die overlap in opgaven is een vergelijking mogelijk. Hierbij merken we op dat er voor Spelling er geen ankeropgaven gebruikt zijn. De vergelijking met het volgende jaar zal dus via een andere schaal plaatsvinden. We moeten ons wel realiseren dat te verwachten

jaareffecten klein zullen zijn. We hebben gelukkig wel de beschikking over grote hoeveelheden gegevens, zodat, wanneer er relevante verschillen zullen zijn, die ook opgemerkt kunnen worden.

De vergelijking voor leerjaar 4 over de jaren heen zal via een steekproef moeten gebeuren. De LOVS-toetsen voor Begrijpend lezen, Spelling en de rekenschalen zullen niet veranderen waardoor een rechtstreekse vergelijking mogelijk zal zijn. Bij Woordenschat is een nieuwe toets ontwikkeld, maar alle opgaven daarvan zaten in de proeftoets die in dit onderzoek geanalyseerd is.

De effecten van de gewijzigde leerlinggewichten die nu al een rol speelden, zullen in de toekomst nadere aandacht krijgen. Zoals nu al blijkt wordt een directe vergelijking tussen groepen met verschillende typen gewichten lastig. Dit punt zal speciale aandacht krijgen. Voor komend jaar is dat binnen jaargroep 8 nog geen probleem, maar in schooljaar 2009-2010 zullen alle leerlingen nieuwe gewichten hebben.

Onderzoek dat minder ver in de toekomst ligt, betreft het Technisch lezen. Het onderzoek naar die vaardigheid is op zoveel verschillende punten anders dan het onderzoek dat hier nu beschreven is, dat het niet in dit rapport past. Het type toetsen, de wijze van afname, het type opgaven, het type gegevens, het soort analyses en in totaal daarmee ook de presentatie van de gegevens verschillen. Wat betreft het type analyse kan zelfs opgemerkt worden dat afgelopen maanden nieuwe modellen uitgewerkt zijn en nieuwe programmatuur is ontwikkeld om deze opgaven beter te kunnen analyseren.

Literatuur

- 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.
- 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.
- 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. Arnhem, Cito.

Sijstra 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*. PPOON-reeks 19. Arnhem: Citogroep.

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

Nieuwe Oeverstraat 50
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